

# Learning how to act: making good decisions with machine learning

Finnian Lattimore

October 30, 2015

## 1 Introduction

Learning the outcome of an action is central to many real world problems. Does de-worming children in poor countries improve health and educational outcomes [worm-wars]? Would increasing the minimum wage benefit lead to higher unemployment? Will offering this customer a discount improve my revenue? These questions are difficult as they require more than identifying a pattern in data. There are two, very different, major approaches to this problem with the machine learning community: reinforcement learning and causal inference.

Reinforcement learning addresses the problem of learning from explicitly taking actions. There is typically some state or environment, an agent chooses an action from those available in the current state, the state evolves stochastically as a function of the selected action and the agent then receives some feedback or reward that is a function of the new state. This setting differs from the standard classification problem in that the agent must learn from feedback on the selected action, rather than being presented with the correct action for a given state. A common modelling assumption is to assume that the state evolves only as a function of the previous state and the action chosen, and given these, is independent of the previous history of states and actions. This is known as a Markov decision process or MDP. A particularly well studied and understood model are MDPs with only a single state. In this case, there are a set of actions, each associated with a fixed but unknown reward distribution and at each time step our agent selects an action and receives corresponding feedback. This is known as the multi-armed bandit problem.

Causal inference makes use of assumptions to allow the outcome of actions to be predicted from observational data. The key to causal inference is a framework that can model how actions change the state of the world. This framework then allows us to map information collected in one setting to another.

Both approaches can be seen as extensions to the concept of randomised controlled trials. Bandit algorithms deal with the sequential nature of the decision making process. Causal inference with the problem that full randomisation is not always feasible, affordable or ethical. The similarities between the problems that these techniques have been developed to address raises the question of if there are problems best addressed by a combination of these approaches and how they can be combined. The goal of my thesis is to explore these questions. In the next sections I briefly review the key literature in causal inference and bandits. I then present a general approach to how causal models might be incorporated into bandit settings and conclude by demonstrating an algorithm that leverages causal assumptions to improve performance in a specific bandit setting.

## 2 Causal Inference

### 2.1 Models of causality and intervention

Predicting the outcome of actions without explicitly taking them requires assumptions about how the actions will change the system. A very powerful and general model that underlies much of the recent work in causality is the causal bayesian network. A causal Bayesian network, or directed acyclic graph (DAG), is a Bayesian network in which a link  $V_i \rightarrow V_j$ , is defined to mean  $V_i$  directly causes  $V_j$ . This means that if we intervene and change the value of  $V_i$ , we expect  $V_j$  to change, but if we intervene to change  $V_j$ ,  $V_i$  will not change. More generally, if  $G$  is a causal network for a distribution  $P$  defined over variables  $V_1 \dots V_N$ , then the distribution after an intervention where we set  $X \subset V$  to  $x$ , denoted  $do(X = x)$  is obtained by simply dropping the terms for each of the variables in  $X$  from the factorization given by the network. This is referred to as the truncated product formula [11].

If the network contains latent (unobservable) variables we will not be able to calculate all the terms in the truncated product formula. However, it may still be possible to determine the post-interventional distribution of specific variables of interest. A general causal query  $P(Y|do(X = x))$  is identifiable if it can be shown to be equivalent to an expression containing only pre-interventional quantities. This means that asymptotically we obtain an unbiased estimate for the distribution after an intervention based on purely observational data.

The Do Calculus is a set of three rules that allow transformations of interventional terms to non-interventional ones, given a causal graph []. They are derived directly from d-separation properties of graphical models and the definition of intervention in causal DAGs. These rules are complete. A query is identifiable if and only if it can be transformed to contain only non-interventional terms via the do-calculus and there is an equivalent algorithm that can take any causal graph and query and determine identifiability []. I have implemented this algorithm as a simple javascript application to demonstrate XXXX. If a query is not identifiable, it may still be possible to get bounds for causal effects, for example using instrumental variables [] or by making additional assumptions.

There are two other key frameworks that arise in causal inference. Counterfactuals and structural equation models. Counterfactuals are statements about imagined or alternate realities, are prevalent in everyday language and may play a role in the development of causal reasoning in humans [21]. Causal effects are differences in counterfactual variables; what is the difference between what would happen if we did one thing versus what would happen if we did something else [15, 16, 14, 17, 18].

For example, if we wanted to estimate the causal effect of a medical treatment, then we might let  $Y^1$  be a counterfactual random variable representing the (binary) potential outcome if treated. The distribution of  $Y^1$  is the distribution we would see of the outcome,  $Y$ , if everyone was treated. Similarly  $Y^0$  represents the potential outcome for the placebo. The causal effect of the drug is the difference between the probability of recovery, across the population, if everyone was treated, and the probability of recovery given placebo  $P(Y^1) - P(Y^0)$ . This quantity is estimable from observational data if we assume  $X \perp\!\!\!\perp Y^0$  and  $X \perp\!\!\!\perp Y^1$ . These assumptions are referred to as ignoreability assumptions [14]. They state that the treatment a each person receives is independent of whether they would recover if treated and if they would recover if not treated. Graphically, this is equivalent to the assumption that there is no variable that is a parent of both the treatment  $X$  and the outcome  $Y$ .

Structural equation models (SEMs) describe a deterministic world, where underlying mechanisms determine the output of any process for a given input. The mechanism (but not the

output) is assumed to be independent of what is fed into it. Linear structural equation models have a long history for causal estimation [22, 4]. Mathematically, each variable is a deterministic function of its direct causes and a noise term that captures unmeasured variables. The noise terms are required to be mutually independent. If there is the possibility that an unmeasured variable influences more than one variable of interest in a study, it must be modelled explicitly as a latent (unobserved) variable. Structural equation models can be represented visually as a network. Each variable is a node and arrows are drawn from causes to their effects. If the network for a structural equation model is acyclic then it implies a recursive factorization of the joint distribution over its variables. In other words, it is a causal bayesian network.

Remarkably for models developed relatively independently in fields with very different approaches and problems, the models we have discussed are almost equivalent. To determine if and how an interventional query can be non-parametrically identified, it is equivalent to specify assumptions graphically in terms of bayesian networks or as structural equation models or as conditional independence statements involving counterfactual variables (ignorability assumptions).

It is possible to pose causal queries in terms of counterfactuals that are not interventional and cannot be phrased in terms of the do-notation. The scientific and philosophical validity of such counterfactual queries remains under question [1, 2], however they are nonetheless widely posed in the form of attribution of causal effects to different pathways and mediation [12, 7, 20].

There are differences between the models we have considered when it comes to these non-interventional queries. Counterfactuals are not defined in causal Bayesian networks, as they only encode information on the interventional distribution over variables. Counterfactuals can be defined in terms of structural equation models [11] but there are subtle differences depending on the form of assumptions made. Structural equation models with independent errors allows the identification of quantities in mediation studies, which are not identifiable with the weak ignorability assumptions and cannot be tested experimentally [13].

In practice, differences in focus and approach across different fields eclipse these actual differences in the models. The work on causal graphical models [11, 19] focuses on non-parametric estimation in the population limit and rigorous theoretical foundations. The Neyman-Rubin framework builds on our understanding of randomized experiment and generalizes to quasi-experimental and observational settings, with a particular focus on non-random assignment to treatment. This research emphasises estimating average causal effects and provides practical methods for estimation, in particular, propensity scores; a method to control for multiple variables in high dimensional settings with finite data [14]. In economics, inferring causal effects from non-experimental data so as to support policy decisions is central to the field. Economists are often interested in broader measures of the distribution of causal effects than the mean and make extensive use of structural equation models, generally with strong parametric assumptions [5]. In addition, the parametric structural equation models favoured in economics can be extended to analyse cyclic (otherwise referred to as non-recursive) models.

## 2.2 Discovering causal structure

In the previous section we discussed when assumptions about the structure of the variables in a specific problem is sufficient to identify a causal effect. This approach relies on having enough prior knowledge or theory about the problem to allow you to, at least partially, specify the causal network. In this section, we consider the much harder problem of causal inference where you need to learn the network. Causal inference might seem impossible without specific assumptions about the structure of the variables involved but, amazingly, some aspects of causal structure can be determined from much more general assumptions.

### 2.2.1 Discovery based on conditional independence

One general approach is to look for clues about the structure of the network in the conditional independence relations in the distribution. Assume there is some acyclic causal network  $G$  that generated the distribution  $P(\mathbf{V})$  from which our data has been sampled. Our goal is to recover the network from this data.

Since  $G$  is a bayesian network, if  $Z$  d-separates  $X$  and  $Y$  in  $G$  then  $(X \perp\!\!\!\perp Y|Z)$  in  $P$ . However, we want to work in the other direction, from conditional independence in the distribution to the structure of the network. This requires that we assume the reverse condition:  $(X \perp\!\!\!\perp Y|Z)$  in  $P$  must imply  $Z$  d-separates  $X$  and  $Y$  in  $G$ . This assumption, commonly referred to as **faithfulness**, says there are no additional independence relations that are satisfied in  $P$  but not in all distributions  $P'$  that are compatible with  $G$ . Stating that  $P$  is faithful to  $G$  is equivalent to  $G$  is a **perfect map** for  $P$ .

Faithfulness is an assumption. It does not always hold and we cannot verify it from the observational data we wish to use for causal inference. However, most distributions generated by a causal bayesian network will be faithful to that network. For faithfulness to be violated, different causal effects must exactly balance one-another out. For example, consider a simple binary variable model of chocolate consumption, income and obesity (figure). If the coefficients in the conditional probability tables are just right then the direct effect of chocolate on obesity will exactly balance the indirect effect through income and obesity will appear independent of chocolate consumption. However, this independence is not stable. It would disappear under a small perturbation to any of the parameters.

Given the faithfulness assumption, our causal discovery problem reduces to finding the set of bayesian networks that have exactly the dependency structure as we observe in  $P$ . A wide range of algorithms have been developed based on this key observation, see table 1. Constraint based methods such as the PC algorithm [], FCI algorithm [] and RFCI algorithm [], perform sequential conditional independence tests and eliminate inconsistent graphs. Search and score based methods, such as GES [], search over the space of graphs and score them according to how well they fit the independences given a complexity penalising prior. Constraint based methods are faster, particularly for sparse graphs, but can lack robustness as errors in early conditional independence tests can propagate. Search and score based methods are more robust for small samples sizes but difficult to scale to larger graphs. This has led to the development of hybrid approaches, such as the MMHC algorithm []. A key component of causal discovery is the ability to do high dimensional non-parametric conditional independence tests. Developments in kernalized conditional independence tests [],[24],[3] have made this possible.

If the end goal of causal discovery is to estimate causal effects, then it may not be necessary to learn the entire graph, only the subset of the graph surrounding target variables of interest. Such local causal discovery techniques can be scaled to problems with many more variables []. Once a set of causal graphs has been identified, causal effects of interest can be bounded by combining the results for the all the networks. This procedure is the IDA algorithm [10] and has been found to outperform standard regularization techniques at finding causal effects in a high-dimensional yeast gene expression data set [9]. An implementation is available in the R package [8]

### 2.2.2 Discovery with functional models

All of the algorithms we have considered so far return a Markov equivalence class. They cannot distinguish between two models that result in the same set of conditional independence relations. Consider the very simple case where we have only two variables and the only possible causal

Table 1: A comparison of key causal discovery algorithms

Alg.	Method	Scales (num.vars)	$\sim$ Vars	Latent
PC	Constraint based	Worst case exponential, polynomial for sparse graphs	5000	No
FCI	Constraint based	Worst case exponential, polynomial variant FCI+ for sparse graphs	30	Yes
RFCI	Constraint based	?	500	Yes
GES	Search & Score	Worst case exponential	50	No
MMHC	Hybrid	?	5000	No

structures are  $X \rightarrow Y$  or  $Y \rightarrow X$ . These models have the same dependency structure but in one case  $P(Y|do(X)) = P(Y|X)$  and in the other  $P(Y|do(X)) = P(Y)$ . No algorithm relying purely on conditional independence relations can separate these two cases.

One solution is to utilize structural equation models to specify additional assumptions. For example, if we assume that noise is additive, such that  $X \rightarrow Y \implies Y = f(X) + \epsilon$ , then this will only be invertible such that  $X = g(Y) + \epsilon'$  for specific pairs of functions  $f$  and noise distributions  $\epsilon$ . Thus in general we will be able to identify the causal direction [6]. This can be extended to post-non-linear additive noise,  $Y = h(f(X) + \epsilon)$ , [23]. These techniques can also be applied over more than two variables [1].

A more general approach is to leverage the assumption that the functions are independent of inputs [1]. This leads to the idea that  $P(X)$  and  $P(Y|X)$  are independent if  $X \rightarrow Y$  but not if  $Y \rightarrow X$ . [1] propose testing for this by applying both semi-supervised and standard techniques to estimate  $P(Y|X)$ . Semi-supervised methods, which utilize additional points from  $P(X)$  to learn  $P(Y|X)$  should only be able to outperform standard methods if  $Y \rightarrow X$ .

Finally, rather than explicitly developing an algorithm based on a specific asymmetry between cause and effect, [1] propose learning what causality looks like from data. They take as input a dataset where each row of data is itself a dataset in which either  $X \rightarrow Y$  or  $Y \rightarrow X$  and a corresponding label. Estimates of the distributions  $P(X)$ ,  $P(Y)$  and  $P()$  for each row are then mapped to features in some kernel space via mean kernel embeddings and finally a standard classification algorithm can be trained to learn label. New datasets where the direction of causality is unknown are then simply mapped to the kernel space and the causal direction is classified according to the trained classifier. In practice, the classifier is trained mostly on simulated data, as it is difficult to find a sufficient set of causal problems with only two variables where the direction of causality is known.

### 3 Multi-armed Bandits

Classical framework. Forms of feedback and structure. Key results graph feedback, etc.

stochastic vs adversarial.

## 4 Unifying the frameworks

stochastic most natural fit because graphical models are stochastic.

what kinds of structural assumptions lead to what?

### 4.1 Results for a specific problem

### 4.2 Open questions

## References

- [1] AP Dawid. Causal inference without counterfactuals. *J. Am. Stat. Assoc.*, 2000.
- [2] AP Dawid. Statistical Causality from a Decision-Theoretic Perspective. *arXiv Prepr. arXiv1405.2292*, 2014.
- [3] Arthur Gretton, K Fukumizu, CH Teo, and L Song. A kernel statistical test of independence. pages 1–8, 2008.
- [4] T Haavelmo. The statistical implications of a system of simultaneous equations. *Econom. J. Econom. Soc.*, 11(1):1–12, 1943.
- [5] James Heckman. Econometric causality. *Int. Stat. Rev.*, 2008.
- [6] Patrick Hoyer, Dominik Janzing, and Joris Mooij. Nonlinear causal discovery with additive noise models. In *NIPS*, 2009.
- [7] Kosuke Imai, Luke Keele, and Teppei Yamamoto. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Stat. Sci.*, 25(1):51–71, February 2010.
- [8] Markus Kalisch, Martin Mächler, Diego Colombo, Marloes H. Maathuis, and Peter Bühlmann. Causal inference using graphical models with the R package pcalg. *JSS*, VV(Ii), 2012.
- [9] Marloes H. Maathuis, D Colombo, Markus Kalisch, and Peter Bühlmann. Predicting causal effects in large-scale systems from observational data. *Nat. Methods*, 7(4):247–248, 2010.
- [10] Marloes H. Maathuis, Markus Kalisch, and Peter Bühlmann. Estimating high-dimensional intervention effects from observational data. *Ann. Stat.*, 37(6A):3133–3164, December 2009.
- [11] Judea Pearl. *Causality: models, reasoning and inference*. MIT Press, Cambridge, 2000.
- [12] Judea Pearl. Interpretation and Identification of Causal Mediation. *Psychol. Methods*, June 2014.
- [13] Thomas S. Richardson and James M. Robins. Single world intervention graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality. 2013.
- [14] PR Rosenbaum and DB Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.
- [15] DB Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.*, 1974.
- [16] DB Rubin. Bayesian inference for causal effects: The role of randomization. *Ann. Stat.*, 1978.

- [17] DB Rubin. Causal Inference Using Potential Outcomes. *J. Am. Stat. Assoc.*, 100(469):322–331, March 2005.
- [18] DB Rubin. For objective causal inference, design trumps analysis. *Ann. Appl. Stat.*, 2(3):808–840, September 2008.
- [19] Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*, volume 81. MIT press, 2000.
- [20] Tyler J VanderWeele and Sonia Hernández-Díaz. Is there a direct effect of pre-eclampsia on cerebral palsy not through preterm birth? *Paediatr. Perinat. Epidemiol.*, 25(2):111–5, March 2011.
- [21] Deena S Weisberg and Alison Gopnik. Pretense, counterfactuals, and Bayesian causal models: why what is not real really matters. *Cogn. Sci.*, 37(7):1368–81, 2013.
- [22] S Wright. Correlation and causation. *J. Agric. Res.*, 1921.
- [23] Kun Zhang and A Hyvärinen. Distinguishing causes from effects using nonlinear acyclic causal models. *NIPS 2008 Work. Causality*. URL [http://www ...](http://www...), 2008.
- [24] Kun Zhang, Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. Kernel-based conditional independence test and application in causal discovery. *pre-print*, 2012.