Bayesian Causal Induction

Pedro A. Ortega
Max-Planck Institute for Biological Cybernetics
Spemannstraße 38, 72070 Tübingen
`peortega@dcc.uchile.cl`

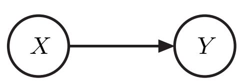
Abstract

Discovering causal relationships is a hard task, often hindered by the need for intervention, and often requiring large amounts of data to resolve statistical uncertainty. However, humans quickly arrive at useful causal relationships. One possible reason is that humans use strong prior knowledge; and rather than encoding hard causal relationships, they encode beliefs over causal structures, allowing for sound generalization from the observations they obtain from directly acting in the world.

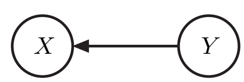
In this work we propose a Bayesian approach to causal induction which allows modeling beliefs over multiple causal hypotheses and predicting the behavior of the world under causal interventions. We then illustrate how this method extracts causal information from data containing interventions and observations.

1 Introduction

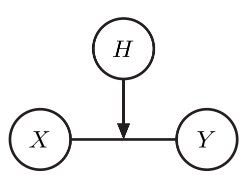
A fundamental problem of statistical causality is the problem of *causal induction*[1]; namely, the generalization from particular instances to abstract causal laws [Hume, 1739-1740]. For instance, how do we determine from experience whether X causes Y or viceversa? That is, which of the two causal hypotheses



or



over X and Y is correct when we assume that they model identical joint distributions? The main difficulty of this problem is that the hypothesis, say H, plays the role of a confounder that determines the causal direction. In other words, a more accurate graphical representation would be the model:



which cannot be analyzed using the framework of graphical models alone. In this work, we have overcome this difficulty by mapping the alternative causal hypotheses and their causal dependencies into probability trees [Shafer, 1996]. Furthermore, we define interventions [Pearl, 2009] on probability trees so as to predict the statistical behavior after manipulation. We then show that such a

[1] Causal induction is also know as *causal discovery*.

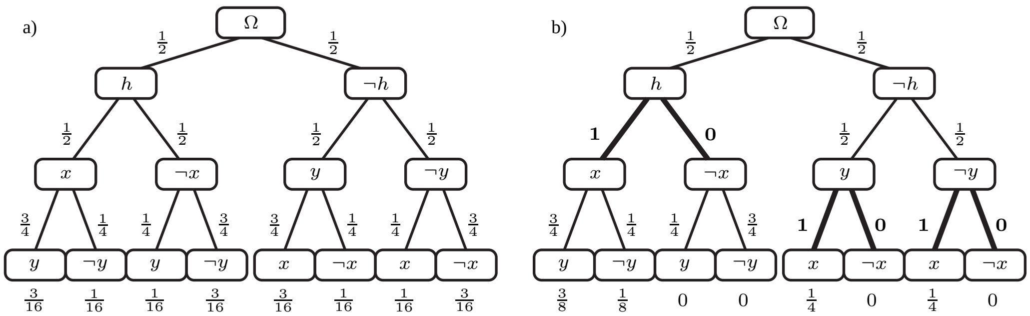


Figure 1: a) The probability tree representing the statistics of the device with two lights. b) The probability tree resulting from (a) after setting $X = x$.

formalization leads to a fully probabilistic method for causal induction that is intuitively appealing and actually trivial in hindsight.

2 Causal Induction in Probability Trees

Imagine we are given a device with two lights whose states, say X and Y, obey a hidden mechanism that correlates them positively. Moreover, the box has a switch that allows us controlling the state of the first light: we can either leave it undisturbed, or we can intercept the mechanism by turning the light on or off as we please. We encode the "on" and "off" states of the first light as $X = x$ and $X = \neg x$ respectively. Analogously, $Y = y$ and $Y = \neg y$ denote the "on" and "off" states of the second light. We ponder the explanatory power of two competing hypotheses: either the state of the first light influences the state of the second ($H = h$) or viceversa ($H = \neg h$).

Furthermore, assume that the probabilities governing the realization of H, X and Y are as detailed in Figure 1a. In this tree, each (internal) node is interpreted as a causal mechanism; hence a path from the root node to one of the leaves corresponds to a particular sequential realization of causal mechanisms[2]. The logic underlying the structure of this tree is self-explanatory:

1. A node causally precedes its descendants. For instance, the root node corresponding to the sure event Ω precedes all other nodes.
2. Each node resolves the node of one random variable. Thus, sibling nodes form a set of mutually exclusive and complete events given the parent node. For instance, given $H = h$ and $X = \neg x$, either $Y = y$ or $Y = \neg y$ will happen.
3. An event can span several nodes. For example, the event $Y = y$ corresponds to the collection of three nodes.
4. The order of events can vary across different branches. For instance, X precedes Y under $H = h$ but Y precedes X under $H = \neg h$. This allows modeling different causal hypotheses.

While the probability tree represents our subjective model explaining the causal order in which the events are resolved, it does not necessarily correspond to the order in which the events are revealed to us. So for instance, under hypothesis $H = h$, the value of the variable Y might be revealed before X, even though X causally precedes Y, and the hypothesis H which precedes both X and Y is never observed.

2.1 Interventions

Suppose we observe that both lights are on. Have we learned anything about their causal dependency? A brief calculation shows that this is not the case because the posterior probabilities are

[2] Note that the set of paths is just the sample set of the probability space representing the experiment.

equal to the prior probabilities:

$$\mathsf{P}(h|x,y) = \frac{\mathsf{P}(y|h,x)\mathsf{P}(x|h)\mathsf{P}(h)}{\mathsf{P}(y|h,x)\mathsf{P}(x|h)\mathsf{P}(h) + \mathsf{P}(x|\neg h,y)\mathsf{P}(y|\neg h)\mathsf{P}(\neg h)} = \frac{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}}{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}+\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}} = \frac{1}{2} = \mathsf{P}(h),$$
$$\mathsf{P}(\neg h|x,y) = 1 - \mathsf{P}(h|x,y) = \frac{1}{2} = \mathsf{P}(\neg h).$$

This makes sense intuitively, because by just observing that the two lights are on, it is statistically impossible to tell which one caused the other. Note how the factorization of the likelihood $P(x,y|H)$ depends on whether $H = h$ or $H = \neg h$. How do we extract causal information then? To answer this question, we make use of two crucial insights of statistical causality:

1. To obtain new causal information from statistical data, old causal information needs to be supplied (paraphrased as "no causes in, no causes out" [Cartwright, 1994] or "to find out what happens if you kick the system, you have to kick the system");
2. Causal information is supplied by intervening the experiment, thereby altering its natural regime. In particular, controlling a variable is an intervention [Pearl, 2009].

Thus, we now repeat our experiment, but this time we control the first light by switching it on ($X = x$). We reflect this choice by changing all the mechanisms that resolve the random variable X, placing all the probability mass on the outcome $X = x$. The resulting probability tree is depicted in Figure 1b. Assume that we subsequently observe that the second light is on. Then, the posterior probabilities are

$$\mathsf{P}(h|\hat{x},y) = \frac{\mathsf{P}(y|h,\hat{x})\mathsf{P}(\hat{x}|h)\mathsf{P}(h)}{\mathsf{P}(y|h,\hat{x})\mathsf{P}(\hat{x}|h)\mathsf{P}(h) + \mathsf{P}(\hat{x}|\neg h,y)\mathsf{P}(y|\neg h)\mathsf{P}(\neg h)} = \frac{\frac{3}{4}\cdot 1\cdot\frac{1}{2}}{\frac{3}{4}\cdot 1\cdot\frac{1}{2}+1\cdot\frac{1}{2}\cdot\frac{1}{2}} = \frac{3}{5}$$
$$\text{and}\quad \mathsf{P}(\neg h|\hat{x},y) = 1 - \mathsf{P}(h|\hat{x},y) = \frac{2}{5},$$

where $\hat{x}$ is Pearl's notation to indicate a causal intervention of X. Looking at the result, we see that we have gathered evidence favoring the hypothesis "X influences Y".

This example is easily extended to sequential trials. With a slight abuse of notation, we model this with i.i.d. pairs of binary random variables X_n, Y_n taking on values in $\{x, \neg x\}$ and $\{y, \neg y\}$ respectively. If we repeatedly switch the first light on (that is, we set $\hat{X}_n = x$ for all n), then the posterior probabilities are:

$$\begin{aligned}\mathsf{P}(h|\hat{X}_{1\ldots N}, Y_{1\ldots N}) &= \frac{\mathsf{P}(h)\prod_{n=1}^{N}\mathsf{P}(Y_n|h,\hat{X}_n)\mathsf{P}(\hat{X}_n|h)}{\mathsf{P}(h)\prod_{n=1}^{N}\mathsf{P}(Y_n|h,\hat{X}_n)\mathsf{P}(\hat{X}_n|h) + \mathsf{P}(\neg h)\prod_{n=1}^{N}\mathsf{P}(\hat{X}_n|\neg h,Y_n)\mathsf{P}(Y_n|\neg h)} \\ &= \frac{(\frac{3}{4})^R\cdot(\frac{1}{4})^{N-R}}{(\frac{3}{4})^R\cdot(\frac{1}{4})^{N-R}+(\frac{1}{2})^N} = \frac{1}{1+\frac{2^N}{3^R}}\end{aligned}$$
$$\text{and}\quad \mathsf{P}(\neg h|\hat{X}_{1\ldots N}, Y_{1\ldots N}) = 1 - \frac{1}{1+\frac{2^N}{3^R}} = \frac{1}{1+\frac{3^R}{2^N}}.$$

where R is the number of times the second light switched on after we switched the first light on. To analyze the asymptotic evolution of the posterior probabilities, we note that due to the law of large numbers, the fraction $\frac{R}{N}$ will converge to $\frac{3}{4}$ if h is true, and to $\frac{1}{2}$ if $\neg h$ is true. Consequently, the posterior probability of h converges to

$$\mathsf{P}(h|\hat{X}_{1\ldots N}, Y_{1\ldots N}) \approx \frac{1}{1+\frac{2^N}{3^{\frac{3}{4}N}}} \approx \frac{1}{1+0.8774^N} \longrightarrow 1, \qquad \text{if } h \text{ is true,}$$
$$\mathsf{P}(h|\hat{X}_{1\ldots N}, Y_{1\ldots N}) \approx \frac{1}{1+\frac{2^N}{3^{\frac{1}{2}N}}} \approx \frac{1}{1+1.1547^N} \longrightarrow 0, \qquad \text{if } \neg h \text{ is true.}$$

This confirms the validity of our method for our example.

2.2 Analysis

Why does this work? The main difficulty in our example is that the hypothesis plays the role of a confounder that determines the causal order of X and Y. Probability trees provide a natural way to encode multiple competing hypotheses about the causal order of random variables, and it is easy to compute the effects of a causal intervention on them. This is vital because to discover the causal dependency, we have to statistically decouple X from its causal predecessors:

$$\mathsf{P}'(X|H=h) = \mathsf{P}'(X|H=\neg h, Y) = \mathsf{P}'(X),$$

where P' denotes the probabilities resulting after the intervention of X. We could have achieved the same result by replacing these mechanisms by any randomizing device, as long as X is chosen independently form its causal precedents. This is a well-known idea that underlies randomized controlled trials [Fisher, 1935].

3 Concluding Remarks

The problem of causal induction has been addressed relatively recently by the statistics and machine learning community, mainly under the context of graphical models [Pearl, 2009, Spirtes and Scheines, 2001, Silva, 2005, Dawid, 2007]. This has led to the development of many algorithms that propose a suitable causal graphical model explaining the data. Most, if not all, of these algorithms rely on independence assumptions, and hence naturally they proceed by exploiting the independence relations found in the data to construct a causal model.

This work outlines a general method for causal induction that is fully Bayesian in nature. It is based on the idea of representing multiple causal hypotheses about a system in a probability tree [Shafer, 1996], which has the advantage of containing sufficient causal and statistical information for predicting the behavior of a system under interventions [Pearl, 2009]. Furthermore, we have illustrated the need for interventions in order to extract causal information and we have analyzed the asymptotic evolution of beliefs in a concrete example. Of course, the causal relations that can be discovered are limited by the set of causal hypotheses and the interventions that we are allowed to apply to the system.

Acknowledgments

The author would like to thank Daniel A. Braun, Philip Dawid, David Balduzzi, Samory Kpotufe, Theofanis Karaletsos and Marcus Hutter for discussions of previous versions of this manuscript. This study was supported by the Emmy Noether Grant BR 4164/1-1, “Computational and Biological Principles of Sensorimotor Learning”.

References

N. Cartwright. *Nature’s Capacities and Their Measurements*. Oxford University Press, 1994.

A. P. Dawid. Fundamentals of statistical causality. Research Report 279, Department of Statistical Science, University College London, 2007.

R.A. Fisher. *The Design of Experiments*. New York: Hafner, 1935.

D. Hume. *A Treatise on Human Nature*. Penguine Classics (1975), 1739-1740.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK, 2009.

G. Shafer. *The Art of Causal Conjecture*. MIT Press, 1996.

R. Silva. *Automatic Discovery of Latent Variable Models*. PhD thesis, Machine Learning Department, Carnegie Mellon University, 2005.

P. Spirtes and R. Scheines. *Causation, Prediction, and Search, Second Edition*. MIT Press, 2001.