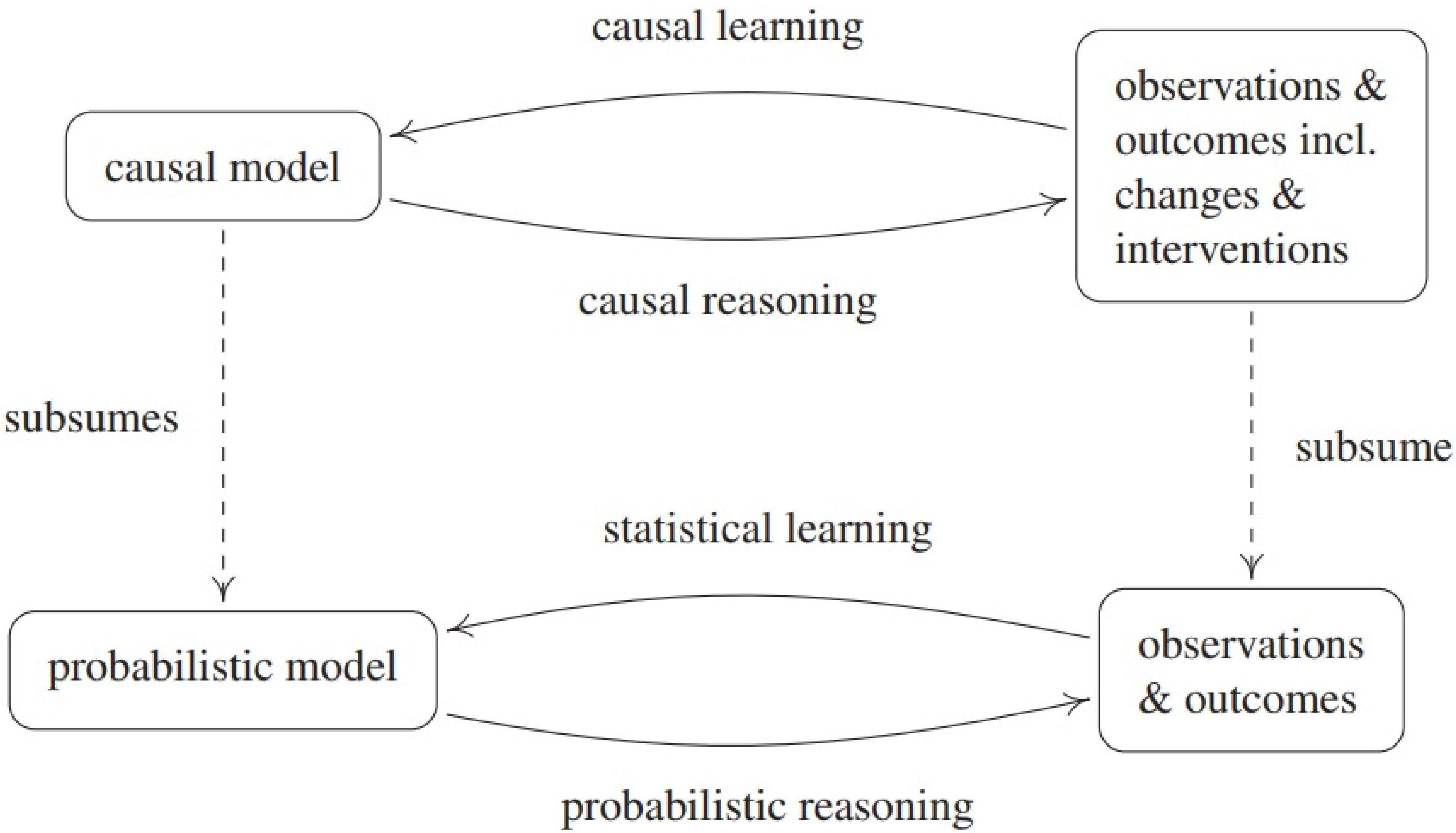
Causality

Dominik Janzing & Bernhard Schölkopf
Max Planck Institute for Intelligent Systems
Tübingen, Germany

http://ei.is.tuebingen.mpg.de

MAX-PLANCK-GESELLSCHAFT


causal learning
causal model
observations & outcomes incl. changes & interventions
causal reasoning
subsumes
subsume
statistical learning
probabilistic model
observations & outcomes
probabilistic reasoning

A Modeling Taxonomy

model	predict in IID setting	predict under changing distributions / interventions	answer counter-factual questions	obtain physical insight	automatically learn from data
mechanistic model ↓	Y	Y	Y	Y	?
structural causal model	Y	Y	Y	N	Y??
causal graphical model	Y	Y	N	N	Y?
statistical model	Y	N	N	N	Y

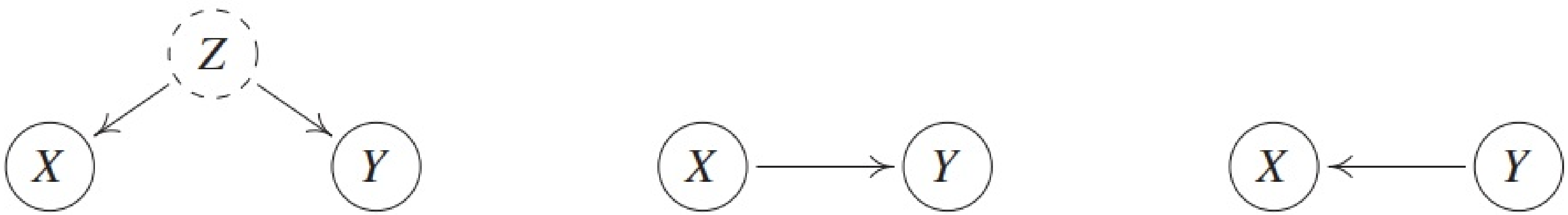
MAX-PLANCK-GESELLSCHAFT

“Correlation does not tell us anything about causality”

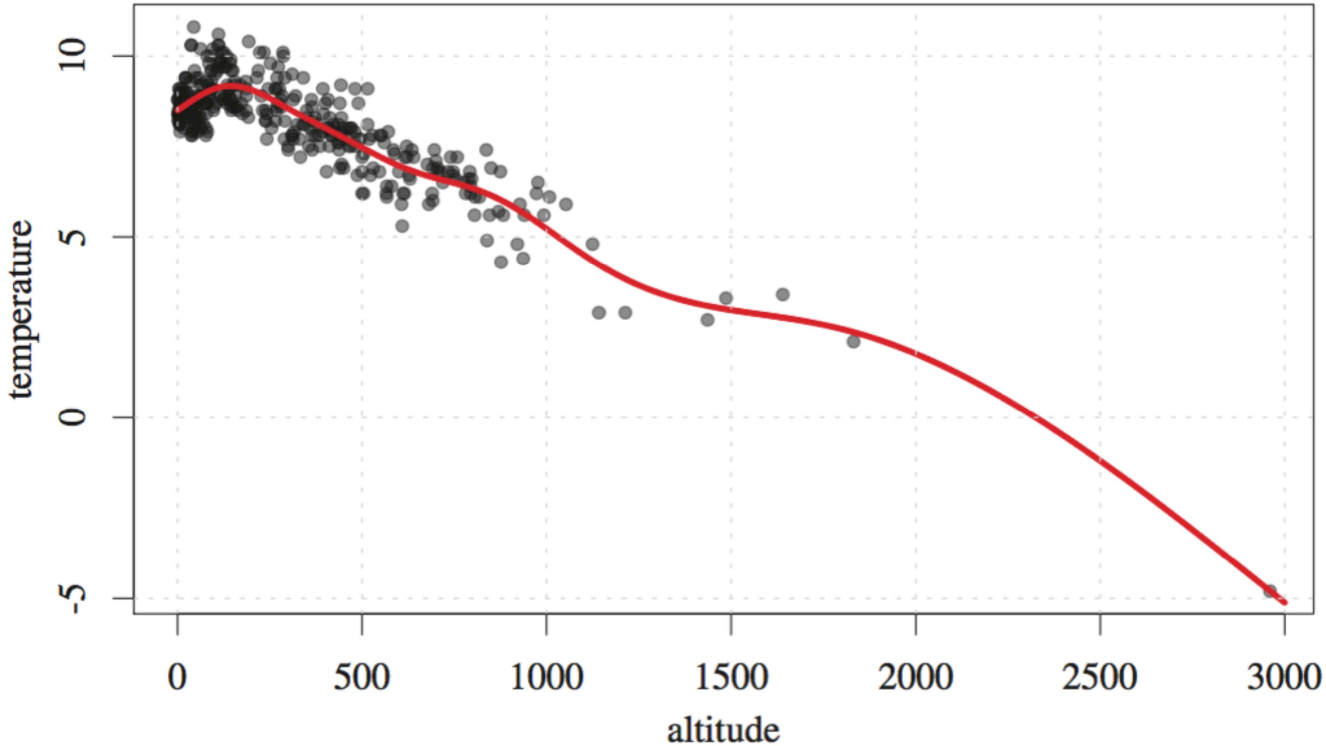
- Better to talk of dependence than correlation
- Most statisticians would agree that causality does tell us something about dependence
- But dependence does tell us something about causality too:

Principle 1.1 (Reichenbach's common cause principle) *If two random variables X and Y are statistically dependent ($X \not\perp\!\!\!\perp Y$), then there exists a third variable Z that* causally *influences both. (As a special case, Z may coincide with either X or Y.) Furthermore, this variable Z screens X and Y from each other in the sense that given Z, they become independent, $X \perp\!\!\!\perp Y \,|\, Z$.*

by permission of the University of Pittsburgh. All rights reserved.

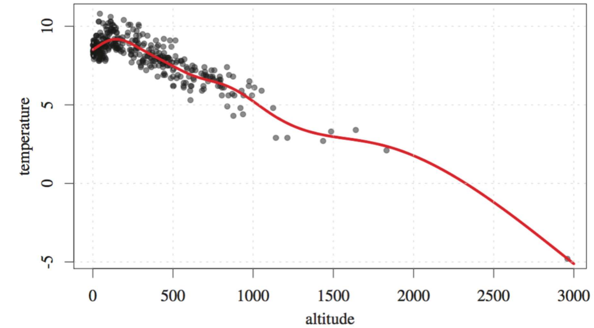


What is cause and what is effect?



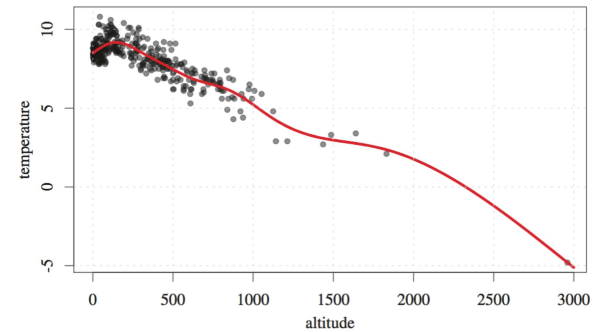
$$
\begin{aligned}
p(a,t) &= p(a|t)\ p(t) \qquad T \to A \\
&= p(t|a)\ p(a) \qquad A \to T
\end{aligned}
$$

Autonomous/invariant mechanisms



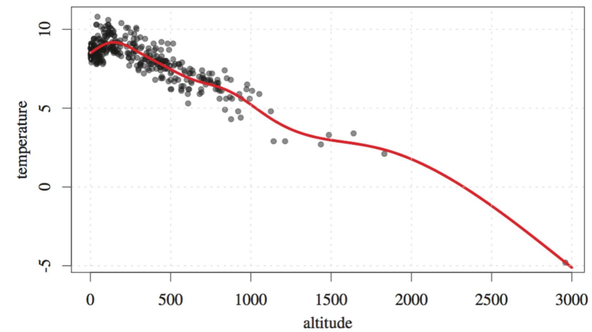
- intervention on a: raise the city, find that t changes
- hypothetical intervention on a: still expect that t changes, since we can think of a physical mechanism $p(t|a)$ that is independent of $p(a)$
- we expect that $p(t|a)$ is invariant across, say, different countries in a similar climate zone

Independence of cause & mechanism



- the conditional density $p(t|a)$ (viewed as a function of t and a) provides no information about the marginal density function $p(a)$
- this also applies if we only have a single density

Independence of noise terms



- view the distribution as entailed by a structural causal model (SCM)

$$A := N_A,$$
$$T := f_T(A, N_T),$$

where $N_T \perp\!\!\!\perp N_A$

- this allows identification of the causal graph under suitable restrictions on the functional form of f_T

Pearl's do-notation

- Motivation: goal of causality is to infer the effect of interventions

- distribution of Y given that X is set to x:

$$p(Y|do\,X = x) \;\; \text{or} \;\; p(Y|do\,x)$$

- don't confuse it with $P(Y|x)$

- can be computed from p and G

MAX-PLANCK-GESELLSCHAFT

Difference between seeing and doing

$$p(y|x)$$

probability that someone gets 100 years old given that we know that he/she drinks 10 cups of coffee per day

$$p(y|do\,x)$$

probability that some randomly chosen person gets 100 years old after he/she has been forced to drink 10 cups of coffee per day

MAX-PLANCK-GESELLSCHAFT

The Principle of Independent Mechanisms

Principle 2.1 (Independent mechanisms) *The causal generative process of a system's variables is composed of autonomous modules that do not inform or influence each other.*

In the probabilistic case, this means that the conditional distribution of each variable given its causes (i.e., its mechanism) does not inform or influence the other conditional distributions. In case we have only two variables, this reduces to an independence between the cause distribution and the mechanism producing the effect distribution.

Example 3.2 (Cause-effect interventions) Suppose that the distribution $P_{C,E}$ is entailed by an SCM $\mathfrak{C}$

$$\begin{aligned} C &:= N_C \\ E &:= 4 \cdot C + N_E, \end{aligned} \tag{3.3}$$

with $N_C, N_E \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, and graph $C \to E$. Then,

Computing $p(X_k|do\,x_i)$

summation over x_i yields

$$p(X_1,\dots,X_{i-1},X_{i+1},\dots,X_n|do\,x_i)=\prod_{j\neq i}p(X_j|PA_j(x_i))\,.$$

- distribution of X_j with $j \neq i$ is given by dropping $p(X_i|PA_i)$ and substituting x_i into PA_j to get $PA_j(x_i)$.
- obtain $p(X_k|do\,x_i)$ by marginalization