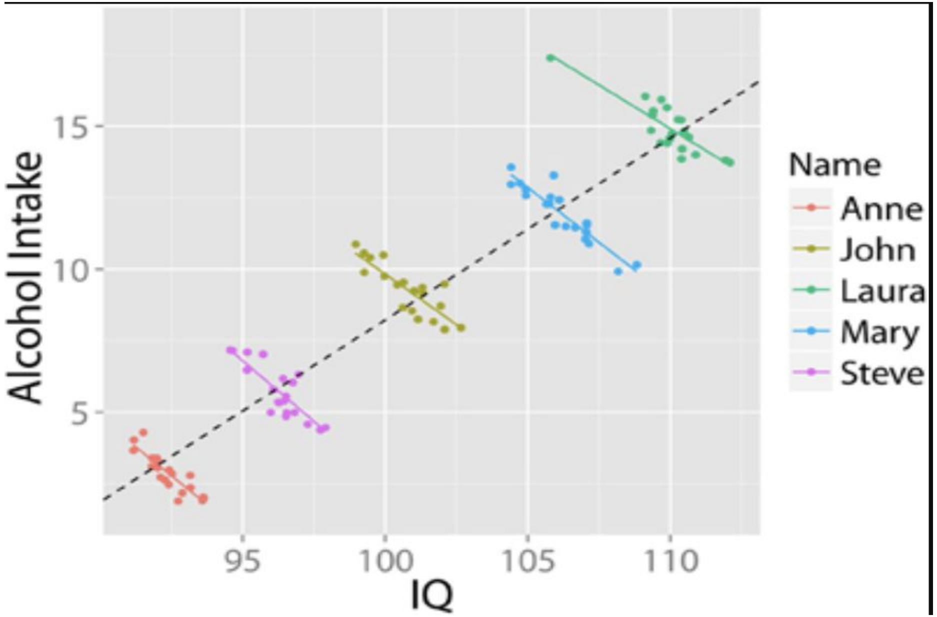
Do-Calculus

Outline

Simpson's paradox	
Confunder and deconfunder	Book of Why
Covariates Adjustment	PJS
Back Door Adjustment	Pearl/PJS
Front Door Adjustment	Pearl/PJS
Do-calculus	Pearl/Book of Why

Alcohol Study



#Not real data

Regression analysis find a positive relationship with IQ and Alcohol intake.

But for each person, the relationship is negative.

What caused this???

Lurking variable,confounding variable, confounder

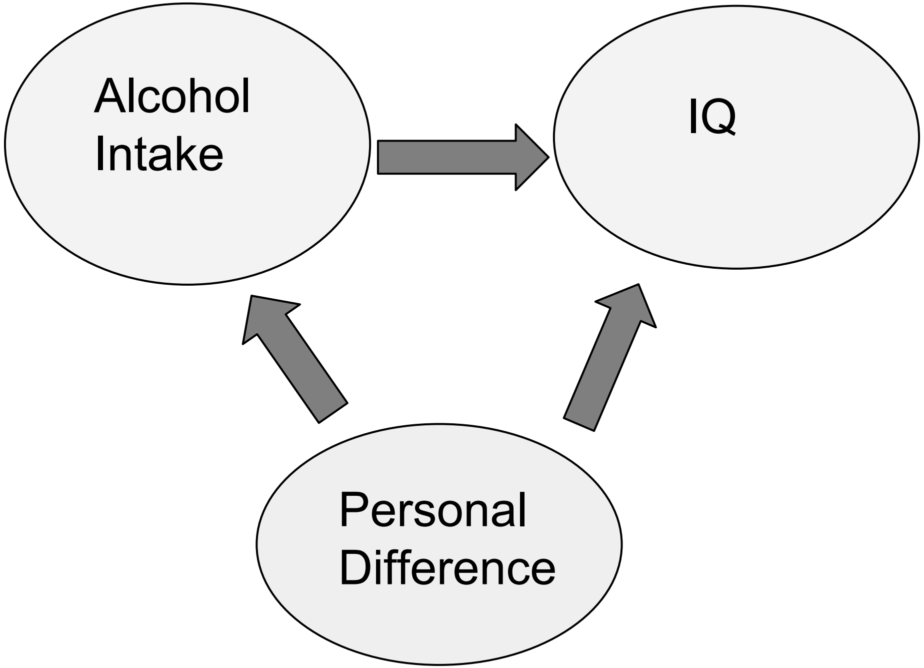
Solution: Adjustment

Stratification in statistics

Truncated factorization [Pearl, 1993]

G-computation formula [Robins, 1986]

Manipulation theorem [Spirtes et al., 2000]



In this section we will make use of a somewhat trivial but very powerful invariance statement. Given an SCM $\mathfrak{C}$, and writing $pa(j) := \mathbf{PA}_j^{\mathcal{G}}$, we have

$$p^{\tilde{\mathfrak{C}}}(x_j \,|\, x_{pa(j)}) = p^{\mathfrak{C}}(x_j \,|\, x_{pa(j)}) \tag{6.7}$$

for any SCM $\tilde{\mathfrak{C}}$ that is constructed from $\mathfrak{C}$ by intervening on (some) X_k but not on X_j. Equation (6.7) shows that causal relationships are autonomous under interventions; this property is therefore sometimes called "autonomy." If we intervene on a variable, then the other mechanisms remain invariant (see the left box in Figure 2.2).

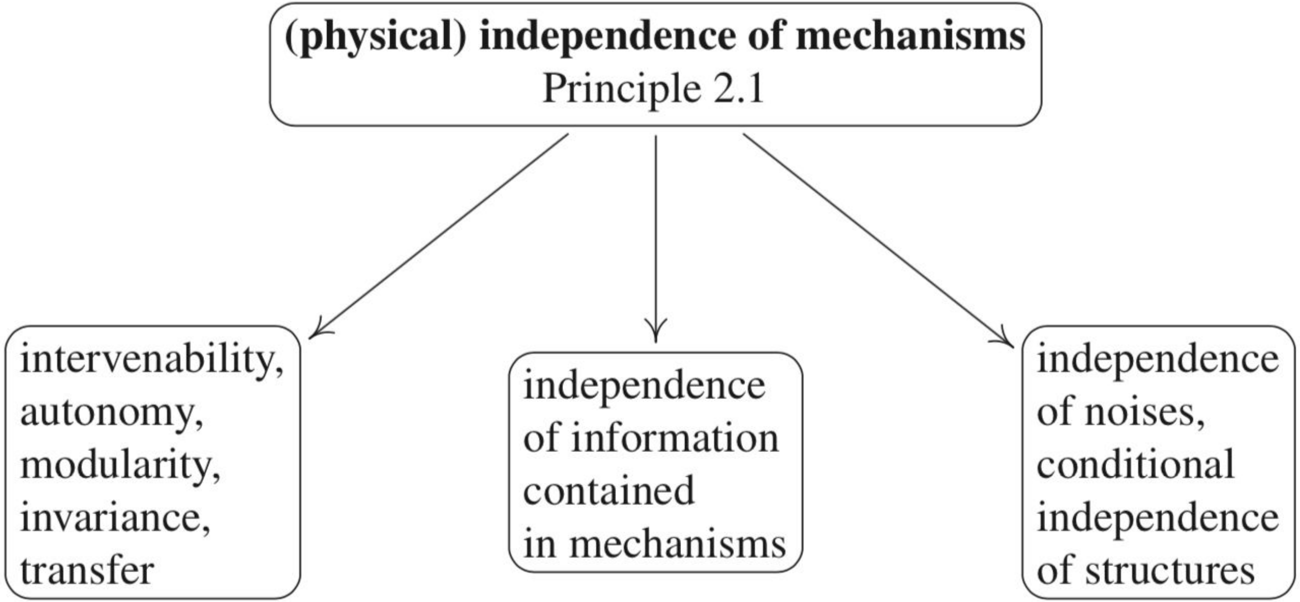
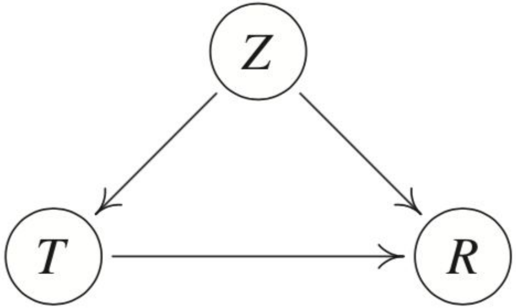


Figure 2.2: The principle of independent mechanisms and its implications for causal inference (Principle 2.1).

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 6.37 (Kidney stones, continued) Assume that the true underlying SCM allows for the graph



Here, Z is the size of the stone, T the treatment, and R the recovery (all binary). We see that the recovery is influenced by the treatment and the size of the stone. The treatment itself depends on the size, too. A large proportion of difficult cases was assigned to treatment A. Consider further the two SCMs $\mathfrak{C}_A$ and $\mathfrak{C}_B$ that we obtain after replacing the structural assignment for T with $T := A$ and $T := B$, respectively. Let us call the corresponding resulting probability distributions $P^{\mathfrak{C}_A}$ and $P^{\mathfrak{C}_B}$. Given that we are diagnosed with a kidney stone *without knowing its size*, we should base our choice of treatment on a comparison between

$$\mathbb{E}^{\mathfrak{C}_A} R = P^{\mathfrak{C}_A}(R=1) = P^{\mathfrak{C};do(T:=A)}(R=1)$$

$$\mathbb{E}^{\mathfrak{C}_B} R = P^{\mathfrak{C}_B}(R=1) = P^{\mathfrak{C};do(T:=B)}(R=1).$$

$$\begin{aligned}
P^{\mathfrak{C}_A}(R=1) &= \sum_{z=0}^{1} P^{\mathfrak{C}_A}(R=1, T=A, Z=z) \\
&= \sum_{z=0}^{1} P^{\mathfrak{C}_A}(R=1 \,|\, T=A, Z=z)\, P^{\mathfrak{C}_A}(T=A, Z=z) \\
&= \sum_{z=0}^{1} P^{\mathfrak{C}_A}(R=1 \,|\, T=A, Z=z)\, P^{\mathfrak{C}_A}(Z=z) \\
&\overset{(6.7)}{=} \sum_{z=0}^{1} P^{\mathfrak{C}}(R=1 \,|\, T=A, Z=z)\, P^{\mathfrak{C}}(Z=z).
\end{aligned} \tag{6.11}$$

$$P^{\mathfrak{C}_A}(R=1) \approx 0.93 \cdot \frac{357}{700} + 0.73 \cdot \frac{343}{700} = 0.832.$$

$$P^{\mathfrak{C}_B}(R=1) \approx 0.87 \cdot \frac{357}{700} + 0.69 \cdot \frac{343}{700} \approx 0.782,$$

$$P^{\mathfrak{C}_A}(R=1) - P^{\mathfrak{C}_B}(R=1) \approx 0.832 - 0.782 \tag{6.12}$$

is sometimes called the **average causal effect (ACE)** for binary treatments. It is important to realize that this is different from simple conditioning:

$$P^{\mathfrak{C}}(R=1 \,|\, T=A) - P^{\mathfrak{C}}(R=1 \,|\, T=B) = 0.78 - 0.83,$$

Do-Calculus

Theorem 3.4.1 (Rules of *do* Calculus)
Let G be the directed acyclic graph associated with a causal model as defined in (3.2), and let P(·) stand for the probability distribution induced by that model. For any disjoint subsets of variables X, Y, Z, and W, we have the following rules.

Rule 1 (*Insertion/deletion of observations*):

$$P(y \mid \hat{x}, z, w) = P(y \mid \hat{x}, w) \quad \text{if } (Y \perp\!\!\!\perp Z) \mid X, W)_{G_{\overline{X}}}. \tag{3.31}$$

When Z is irrelevant to Y(possible condition on some other value,i.e a mediator), then the distribution will not change if remove Z.

Rule 2 (*Action/observation exchange*):

$$P(y \mid \hat{x}, \hat{z}, w) = P(y \mid \hat{x}, z, w) \quad \text{if } (Y \perp\!\!\!\perp Z) \mid X, W)_{G_{\overline{X}\underline{Z}}}. \tag{3.32}$$

If all "Backdoors" are blocked, Do(x) = See(x)

$$P(Y \mid do(X), Z) = P(Y \mid X, Z)$$

Rule 3 (*Insertion/deletion of actions*):

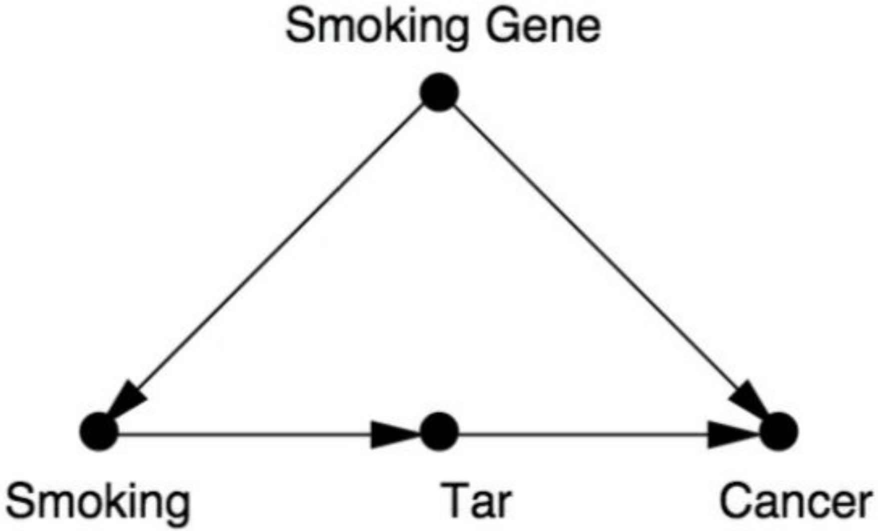
$$P(y \mid \hat{x}, \hat{z}, w) = P(y \mid \hat{x}, w) \text{ if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}, \overline{Z(W)}}}, \tag{3.33}$$

where Z(W) is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$.

If no path from X to Y, then we can remove Do(x) from P(Y|Do(x))

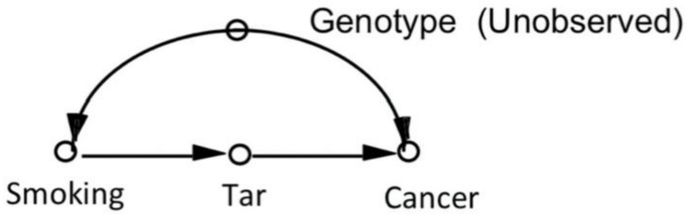
$$P(Y \mid do(X)) = P(Y)$$

Front-Door with Do-Calculus



DO-CALCULUS AT WORK

Query



$$P(c \mid do(s)) = \sum_t P(c \mid do(s), t) P(t \mid do(s))$$ Probability Axioms

$$= \sum_t P(c \mid do(s), do(t)) P(t \mid do(s))$$ Rule 2

$$= \sum_t P(c \mid do(s), do(t)) P(t \mid s)$$ Rule 2

$$= \sum_t P(c \mid do(t) P(t \mid s)$$ Rule 3

$$= \sum_{s'} \sum_t P(c \mid do(t), s') P(s' \mid do(t)) P(t \mid s)$$ Probability Axioms

$$= \sum_{s'} \sum_t P(c \mid t, s') P(s' \mid do(t)) P(t \mid s)$$ Rule 2

$$= \sum_{s'} \sum_t P(c \mid t, s') P(s') P(t \mid s)$$ Rule 3

Estimand

Figure 7.4. Derivation of the front-door adjustment formula from the rules of *do*-calculus.