The Book of Why notes

Tomás J. Aragón

Version April 16, 2019

For our Population Health Data Science (PHDS)[1] book club at the San Francisco Department of Public Health we are reading Judea Pearl's *The Book of Why* (TBoW) (http://bayes.cs.ucla.edu/WHY/) [1].

These are my notes for selected (hopefully all the) chapters. The notes are not comprehensive, and sometimes present an outline or expansion of concepts covered in the book. I've inserted some personal insights that I hope you find informative.

0 Mind over Data

Some definitions

1. causal inference
2. strong artificial intelligence
3. big data, machine learning, deep learning

Limitations of mathematics for causal inference

Consider that a barometer reading (B) is a function of atmospheric pressure (P) and a constanct (k).

$$B = kP$$
$$P = B/k$$
$$k = B/P$$

Questions

1. What are limitations of mathematical equations for causality?
2. Why did it take so long for humans to develop a science of causality?
3. Do you agree with Pearl that "data are profoundly dumb"? Why or why not?

Calculus of causation (p. 7)

Judea Pearl calls the mathematical framework that led to this transformation Structural Causal Models (SCM). The SCM deploys three parts

1. Graphical models,
2. Structural equations, and
3. Counterfactual and interventional logic

What is the difference between *seeing* and *doing*, and why does it matter?

What is a counterfactual?

The "Inference Engine" (Figure 1 on p. 12 of TBoW).

1. knowledge (mostly implicit)

[1] Population health is a systems approach to studying and improving the health of populations through collective action and learning. PHDS is the art and science of transforming data into information and actionable knowledge that informs, influences or optimizes decisions and actions that protect and improve health.

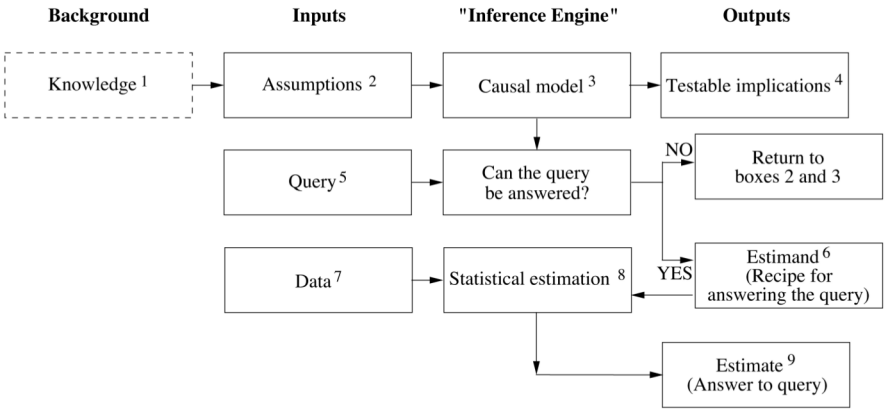


Figure 1. The Inference Engine

2. assumptions (explicit)
3. causal models (i.e., SCM)
4. testable implications
5. queries; e.g., $P(L \mid do(D))$
6. estimand; e.g., $P(L \mid D, Z) \times P(Z)$
7. data
8. statistical estimation
9. estimate

Here is an alternative depiction of the inference engine [2].

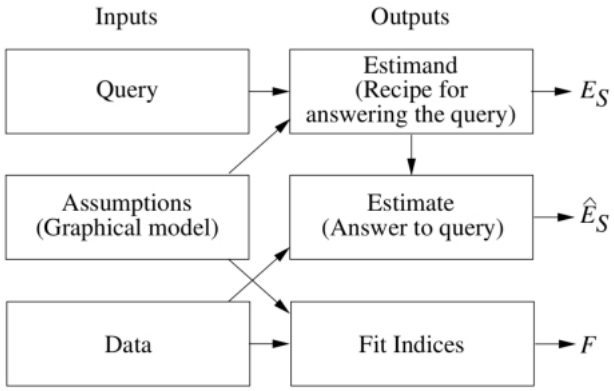


Figure 2. Alternative depiction of Inference Engine

Chapters ahead

1. The Ladder of Causation is observation (seeing), intervention (doing), and counterfactual (imagining)
2. "how the discipline of statistics inflicted causal blindness on itself"
3. "how I became a convert to causality through my work in AI and ... Bayesian networks."
4. "the major contribution of statistics to causal inference: the randomized controlled trial (RCT)"
5. Causation in the absence of a RCT: case study of smoking and lung cancer
6. Paradoxes that stumped the brightest minds for many years

7. Rung 2—interventions
8. Rung 3—counterfactuals
9. Mediation: The Search for a Mechanism
10. Big Data and strong AI

REFLECTION: Book club as an intervention

A book club is an intervention. Book club participation could be

a. mandatory
b. incentivized
c. voluntary

I (PHD director) *chose* to make the book club *voluntary*.

Staff *chose* (self-select) to participate or not.

> **Insight:** Every *decision* has a *causal assumption* (causal inference) and a *prediction* (probabilistic reasoning). If my long-term goal is to spread causal thinking, what short term causal models did I likely consider that led to making the book club voluntary? For staff, what causal models did they consider in self-selecting to participate in the book club? If months later I show good data that those who attended the book club improved their work performance compared to those who did not attend, what causal inferences can I make or not make? Explain your reasoniing.

1 Ladder of Causation

Table 1 (adapted from [2]) corresponds to the Ladder of Causation on p. 28 of TBoW.

Table 1. The Causal Hierarchy. Questions at level i can only be answered if information from level i or higher is available.

i	Level	Activity	Questions
1	Association $P(y \mid x)$	Seeing	What if I *see* ...? How would seeing X change my belief in Y? Are X and Y correlated (associated)? Does X predict Y? What are the predictors of Y?
2	Intervention $P(y \mid do(x), y)$	Doing Intervening	What if I *do* ...? How would doing X change Y? How can I change Y?
3	Counterfactuals $P(y_x \mid x', y')$	Imagining Retrospection	What if I *had done* ...? Was it X that caused Y? What if I had acted differently? What caused Y? What are the causes of Y?

> **Insight:** Between *seeing* and *doing* is *choosing*. Choice can be based on *intuition* ("gut", nonconscious) and/or *deliberation* (conscious, reflection, discussion). Choice always involves a *causal assumption* (causal inference) and *prediction* (probabilistic reasoning). Together, causal inference, probabilistic reasoning, and

decision-making are the core competencies of **population health thinking** (PHT). PHT supports PHDS.

Figure 3 is the causal diagram for the firing squad example (TBoW, p. 39).

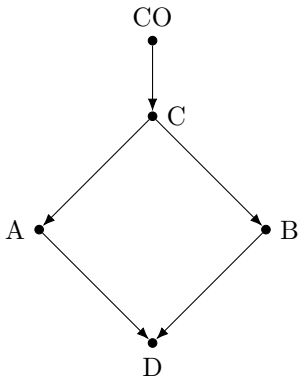


Figure 3. Causal diagram for the firing squad example in TBoW, p. 39 (Figure 1.4). A and B represent (the actions of) Soldiers A and B. CO = Court Order and C = Captain.

Although not presented in the book, these are the *structural equations* for Figure 3 where 0 = false and 1 = true.

$$f_{CO} : CO = \{1 \text{ or } 0 \tag{1}$$

$$f_C : C = \begin{cases} 1, & \text{if } CO = 1. \\ 0, & \text{if } CO = 0. \end{cases} \tag{2}$$

$$f_A : A = \begin{cases} 1, & \text{if } C = 1. \\ 0, & \text{if } C = 0. \end{cases} \tag{3}$$

$$f_B : B = \begin{cases} 1, & \text{if } C = 1. \\ 0, & \text{if } C = 0. \end{cases} \tag{4}$$

$$f_D : D = \begin{cases} 1, & \text{if } A = 1 \text{ or } B = 1. \\ 0, & \text{if } A = 0 \text{ and } B = 0. \end{cases} \tag{5}$$

Interventional reasoning: If Soldier A decides to fire on his or her own volition, this is an *intervention* ($A = 1$). Now, the probability of death is $P(D \mid do(A = 1), B)$. In a causal graph an intervention is depicted by *removing* all arrows pointing into node A and setting value of A to true, and keeping the remaining nodes are unchanged (Figure 4).

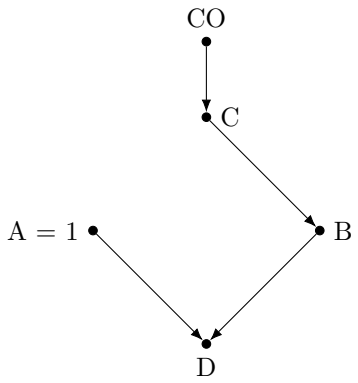


Figure 4. Interventional reasoning (from TBoW, p. 41). What *would happen if* Soldier A decides to fire? Soldier A decides to fire; arrow from C to A is deleted, and A is assigned the value true.

Counterfactual reasoning: If we observe a dead prisoner, we could ask: What *would have happened if* Soldier A had decided *not* to fire? This is a retrospective counterfactual intervention at node A ($A = 0$) knowing that the other nodes must have be set to true. Figure 5 depicts this counterfactual intervention. This is why it looks very similar to Figure 4 (i.e., no arrows pointing into to node A).

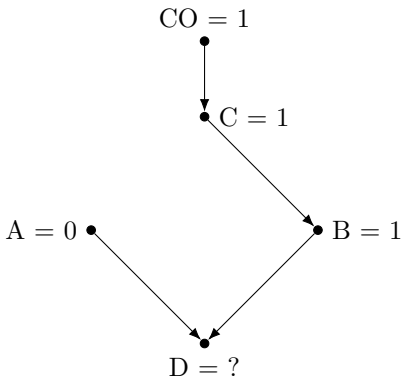


Figure 5. Counterfactual reasoning (from TBoW, p. 42). We observe that the prisoner is dead and ask: What *would have happened if* Soldier A had decided not to fire?

EXAMPLE: Smallpox vaccination (p. 44)

Assume that smallpox is an ongoing infectious disease threat. Suppose that out 1 million children, 99 percent are vaccinated against smallpox. Among the vaccinated, 9900 have a reaction, and 99 die. Among the 10,000 children not vaccinated, 200 get smallpox, and 40 die. The newspaper headlines read "More children died from smallpox vaccination than smallpox."

How do you explain the impact of the vaccination campaign?

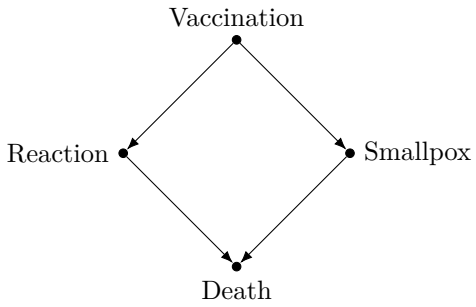


Figure 6. Causal diagram for the vaccination example (from TBoW, p. 45). Is vaccination beneficial or harmful?

Counterfactual reasoning is how we retrospectively **measure impact** for a completed intervention. This can only be accomplished with a causal model. Here is the information available for each variable (node).

Variable node	Probability
Vaccination	$P(V = 1) = 0.99$
Reaction	$P(R = 1 \mid V = 1) = 0.01$
	$P(R = 1 \mid V = 0) = 0$
Smallpox	$P(S = 1 \mid V = 1) = 0$
	$P(S = 1 \mid V = 0) = 0.02$
Death	$P(D = 1 \mid R = 1) = 0.01$
	$P(D = 1 \mid S = 1) = 0.2$

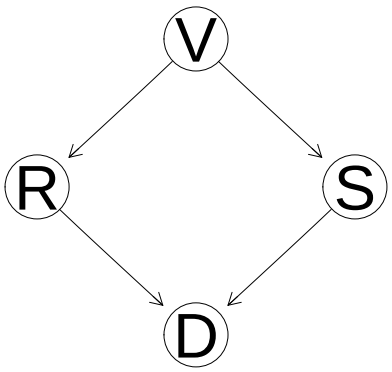
If there were no vaccination campaign—the counterfactual—then the number acquiring smallpox would be 20,000 $(1,000,000 \times 0.02)$, and the number of deaths would be 4,000 $(20,000 \times 0.2)$. In other words, the number of deaths prevented by the vaccination campaign was 3861 $(4000 - 99 - 40)$.

R implementation (optional)

The smallpox vaccination example above does not require a computer. However, if you can learn to use the computer for a simple problem, then you can solve complex problems with a little extra effort.

```
rm(list=ls())
library(bnlearn)
library(gRain)
# p.V <- 50/100 # neutral prior
p.V <- 99/100 # neutral prior
p.R.V <- 1/100
```

```
p.S.V <- 0
p.R.noV <- 0
p.S.noV <- 1/50
p.D.RnoS <- 1/100
p.D.SnoR <- 1/5
p.D.noRnoS <- 0
p.D.SR <- 1 - (1-p.D.RnoS)*(1-p.D.SnoR)
dag <- empty.graph(nodes = c('V', 'R', 'S', 'D'))
dag <- set.arc(dag, from = 'V', to = 'R')
dag <- set.arc(dag, from = 'V', to = 'S')
dag <- set.arc(dag, from = 'R', to = 'D')
dag <- set.arc(dag, from = 'S', to = 'D')
graphviz.plot(dag)
```




```
V.lv <- c('No', 'Yes')
R.lv <- c('No', 'Yes')
S.lv <- c('No', 'Yes')
D.lv <- c('No', 'Yes')
V.prob <- array(c(1-p.V, p.V), dim=2, dimnames=list(V = V.lv))
R.prob <- array(c(1-p.R.noV,p.R.noV,1-p.R.V,p.R.V), dim=c(2,2),
  dimnames=list(R = R.lv, V = V.lv))
S.prob <- array(c(1-p.S.noV,p.S.noV,1-p.S.V,p.S.V), dim=c(2,2),
  dimnames=list(S = S.lv, V = V.lv))
D.prob <- array(c(1-p.D.noRnoS,p.D.noRnoS,1-p.D.RnoS,p.D.RnoS,
  1-p.D.SnoR,p.D.SnoR,1-p.D.SR,p.D.SR),
  dim=c(2,2,2), dimnames=list(D = D.lv, R = R.lv, S = S.lv))
cpt <- list(V=V.prob, R=R.prob, S=S.prob, D=D.prob)
bn <- custom.fit(dag, cpt)
junction <- compile(as.grain(bn))
## Set Evidence to test hypotheses (one model)
## Alternative: change input parameters and rerun different models
j.V.yes <- setEvidence(junction, nodes = 'V', states = 'No')
querygrain(j.V.yes, nodes = 'S')$S*1000000
```

```
## S
##     No    Yes
## 980000  20000
```

```
querygrain(j.V.yes, nodes = 'D')$D*1000000
```

```
## D
##     No    Yes
```

```
## 996000     4000
```

References

1. Pearl J. The Book of Why: The new science of cause and effect. Basic Books; 2018.

2. Pearl J. The seven tools of causal inference, with reflections on machine learning. Communications of the ACM [Internet]. 2019 Mar;62(3):54–60. Available from: http://bit.ly/2OVXxFY