

Chapter 3 (Part 2) of “The Book of Why”

From Evidence to Causes—Reverend Bayes meets Mr. Holmes



Tomás Aragón and James Duren (Version C; Last compiled August 4, 2019)

June 20, 2019, San Francisco, CA

Health Officer, City & County of San Francisco

Director, Population Health Division (PHD)

San Francisco Department of Public Health

<https://taragonmd.github.io/> (GitHub page)

PDF slides produced in Rmarkdown $\text{\LaTeX}$ Beamer—Metropolis theme

- 1 Review of Part 1: Bayes' theorem
- 2 Bayesian networks: What causes say about data
- 3 Where is my bag? From Aachen to Zanzibar
- 4 Bayesian networks in the real world
- 5 From Bayesian networks to causal diagrams

1 Review of Part 1: Bayes' theorem

Cause • $\rightarrow$ • Effect
Hypothesis • $\rightarrow$ • Evidence

Graph	Conditional prob.	Probability	Synonym	Reasoning type
$H \rightarrow E$	$P(E H)$	Forward prob.	deduction	causal reasoning ¹
$H \rightarrow E$	$P(H E)$	"Inverse" prob.	induction	evidential reasoning ²

¹Also called "predictive" reasoning

²Also called "diagnostic" reasoning

Bayes' theorem for causal Hypothesis • → • Evidence

Starting from left to right, and then from right to left:

Reasoning	N	Probability	Description
Causal reasoning	1	$P(H)$	Prior probability (margin probability of H)
... leads to	2	$P(E H)$	Likelihood (TP [sensitivity], FP [1-specificity])
Evidential reasoning	3	$P(E)$	Marginal probability of E
... leads to	4	$P(H E)$	Posterior probability (conditional probability)

For **Bayes' theorem** just substitute probability expressions from table above:

$$(4) = \frac{(1)(2)}{(3)}$$

Bayes' theorem for causal (Hypothesis) → (Evidence)

1. **Prior probability** is the marginal probability of causal Hypothesis

2. **Causal reasoning** is evaluating the **likelihood** (true positive [sensitivity], and false positive [1-specificity])

$$P(H | E) = \frac{P(H)P(E | H)}{P(E)}$$

4. **Evidential reasoning** is evaluating the **posterior probability**

3. Marginal probability of Evidence summed over all possibilities;
Likelihood Ratio is $(2) \div (3)$

2 Bayesian networks: What causes say about data

Bayesian networks are nodes (variables) with probabilistic dependence

Here is a **non-causal** Bayesian network. Smelling smoke increases the *credibility* (belief) of a fire nearby, but smoke does not cause fire.



Here is a **causal** Bayesian network. Fire *causes* smoke. Smoke is *evidence* of a fire (cause). Causal BNs have causal and evidential implications.



Directed acyclic graphs (DAGs) are **causal Bayesian networks**. The mammography example was a causal Bayesian network.



Bayesian networks generalize Bayes' theorem for complex causal graphs

Junctions: Core DAG patterns for three nodes and two edges

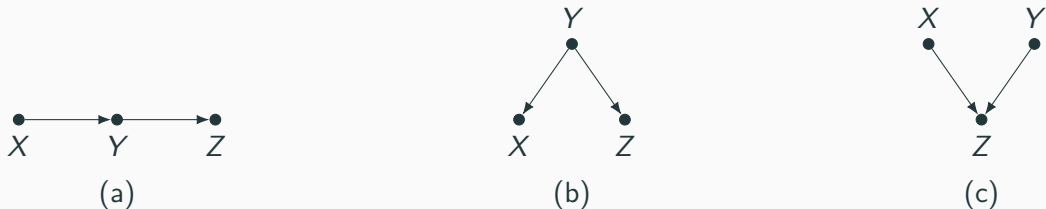


Figure 1: Core DAG patterns for three nodes and two edges: (a) chain (sequential cause; mediation), (b) fork (common cause; confounding), and (c) collider (common effect, multi-cause reasoning, explaining away).

Recap: A patient presents with chest pain to clinical provider.

The patient has a history of coronary artery disease.

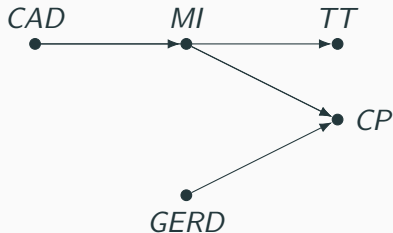
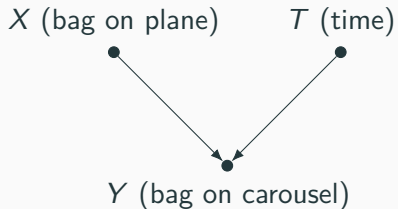


Figure 2: A patient with a history of coronary artery disease (CAD) presents to a provider complaining of prolonged chest pain (CP). The provider's differential diagnosis (hypotheses) are myocardial infarction (MI) and gastroesophageal reflux disease (GERD). The provider sends a blood specimen for a Troponin Test (TT) to "rule out" a MI. The pattern $CAD \rightarrow MI \rightarrow TT$ is a **chain** (sequential cause); $TT \leftarrow MI \rightarrow CP$ is a **fork** (common cause or confounder); and $MI \rightarrow CP \leftarrow GERD$ is a **collider** (common effect). Providers reason like Sherlock Holmes.

3 Where is my bag? From Aachen to Zanzibar

Bayesian network for airport/bag example (TBoW, p. 118)

Question: Is this a causal BN or a noncausal BN? Why or why not?



Joint probability

$$P(X, Y, T) = P(X)P(T)P(Y | X, T)$$

Marginal probability

$$P(X) = \text{Probability bag on plane} = 0.5$$

Conditional probability table (CPT)

$$P(Y | X, T) = \text{Probability bag on carousel given } X \text{ and elapsed time } T$$

Bayesian network for airport/bag example (TBoW, p. 120)

Conditional probability table: X = bag on plane, Y = bag on carousel, T = elapsed time

```
## , , X = False
```

```
##
```

```
##          T
```

```
## Y          0 1 2 3 4 5 6 7 8 9 10
```

```
## False 1 1 1 1 1 1 1 1 1 1 1
```

```
## True  0 0 0 0 0 0 0 0 0 0 0
```

```
##
```

```
## , , X = True
```

```
##
```

```
##          T
```

```
## Y          0  1  2  3  4  5  6  7  8  9 10
```

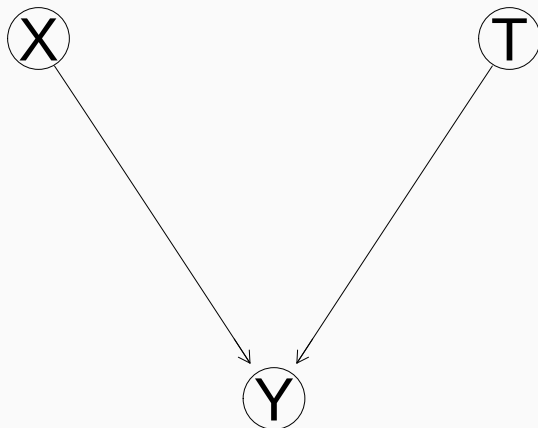
```
## False 1 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0
```

```
## True  0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
```

Bayesian network for airport/bag example (TBoW, p. 120)

Conditional probability table: X = bag on plane, Y = bag on carousel, T = elapsed time

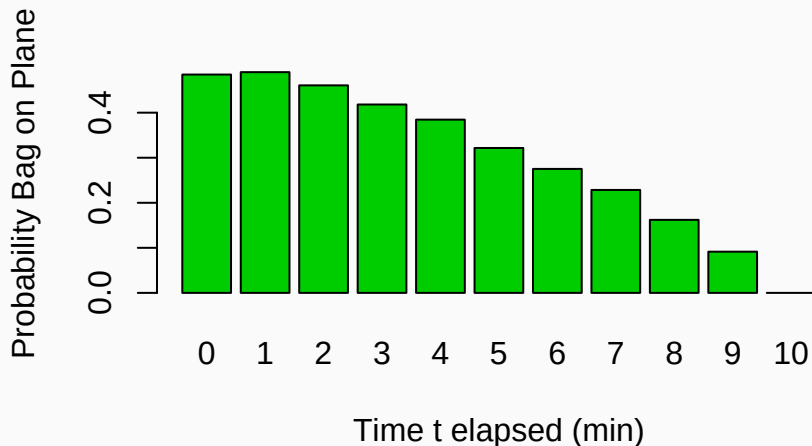
```
graphviz.plot(dag)
```



Bayesian network for airport/bag example (p. 121): Evidential reasoning in R

If bag has not appeared in t minutes, what is the probability that bag was on the plane?

```
cpquery(bn, event = (X=='True'), evidence = (Y=='False') & (T==T.lv[i]))
```



“Bayesian networks are almost like a living organic tissue . . .”

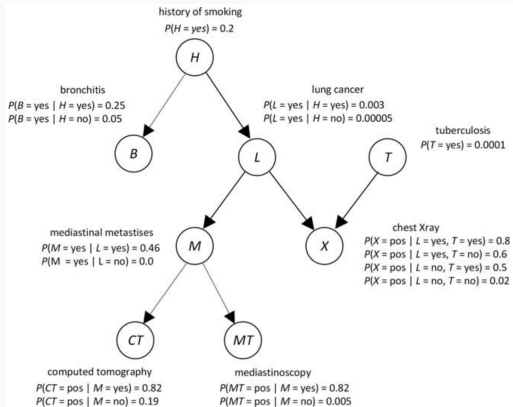
Judea Pearl, The Book of Why, p. 124–125

“Bayesian networks are almost like a living organic tissue, . . . I wanted Bayesian networks to operate like the neurons of a human brain; you touch one neuron, and the entire network responds by propagating the information to every other neuron in the system.

The transparency of Bayesian networks distinguishes them from most other approaches to machine learning, which tend to produce inscrutable “black boxes.” In a Bayesian network you can follow every step and understand how and why each piece of evidence changed the network’s beliefs.”

4 Bayesian networks in the real world

Bayesian network example: Relationships among respiratory disease variables

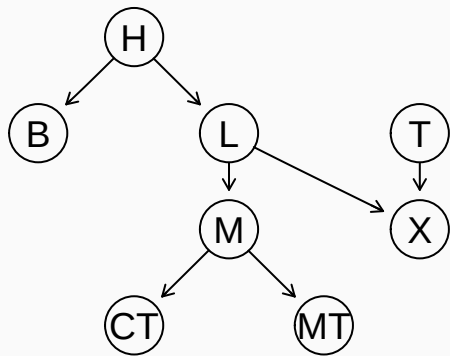


From Neapolitan:^a If a patient has a smoking history ($H = \text{yes}$), a positive chest X-ray ($X = \text{pos}$), and a positive computer tomogram ($CT = \text{pos}$), what is the probability of the patient having lung cancer ($L = \text{yes}$)?

$$P(L = \text{yes} \mid H = \text{yes}, X = \text{pos}, CT = \text{pos})$$

^a Neapolitan R, et al. A primer on Bayesian decision analysis with an application to a kidney transplant decision. Transplantation. 2016 Mar;100(3):489–96.

Bayesian network example: Relationships among respiratory disease variables



```
cpquery(bn2, event = (L == "Yes"), evidence = (H == "Yes") & (X=="Pos")  
& (CT=="Pos"), n = 10^6)
```

```
## [1] 0.2012012
```

5 From Bayesian networks to causal diagrams

From Bayesian networks to causal diagrams (TBoW, p. 128)

“... Bayesian networks hold the key that enables causal diagrams to interface with data. All the probabilistic properties of Bayesian networks (including the junctions³ ...) and the belief propagation algorithms that were developed for them remain valid in causal diagrams. They are in fact indispensable for understanding causal inference.

... A Bayesian network is literally nothing more than a compact representation of a huge probability table. The arrows mean only that the probabilities of child nodes are related to the values of parent nodes by a certain formula (the conditional probability tables) and that this relation is sufficient.”

³Chain (sequential cause, mediation), fork (common cause, confounding), collider (common effect, inter-causal reasoning, explaining away)

From Bayesian networks to causal diagrams (TBoW, p. 129)

“If, however, the same diagram has been constructed as a causal diagram, then both the thinking that goes into the construction and the interpretation of the final diagram change. In the construction phase, we need to examine each variable, say C , and ask ourselves which other variables it “listens” to before choosing its value. The chain structure $A \rightarrow B \rightarrow C$ means that B listens to A only, C listens to B only, and A listens to no one;”

From Bayesian networks to causal diagrams (TBoW, p. 129–130)

“First, it tells us that causal assumptions cannot be invented at our whim; they are subject to the scrutiny of data and can be falsified.

Second, the graphical properties of the diagram dictate which causal models can be distinguished by data and which will forever remain indistinguishable, no matter how large the data. For example, we cannot distinguish the fork $A \leftarrow B \rightarrow C$ from the chain $A \rightarrow B \rightarrow C$ by data alone because, with C listening to B only, the two imply the same independence conditions.”

From Bayesian networks to causal diagrams (TBoW, p. 130)

“The **causal thinking** that goes into the construction of the causal network will pay off, of course, in the type of questions the network can answer. Whereas a Bayesian network can only tell us how likely one event is, given that we observed another (rung-one information), **causal diagrams** can answer **interventional** and **counterfactual** questions. For example, the causal fork $A \leftarrow B \rightarrow C$ tells us in no uncertain terms that wiggling A would have no effect on C , no matter how intense the wiggle. On the other hand, a Bayesian network is not equipped to handle a “wiggle,” or to tell the difference between seeing and doing, or indeed to distinguish a fork from a chain.”

From Bayesian networks to causal diagrams (TBoW, p. 130–131)

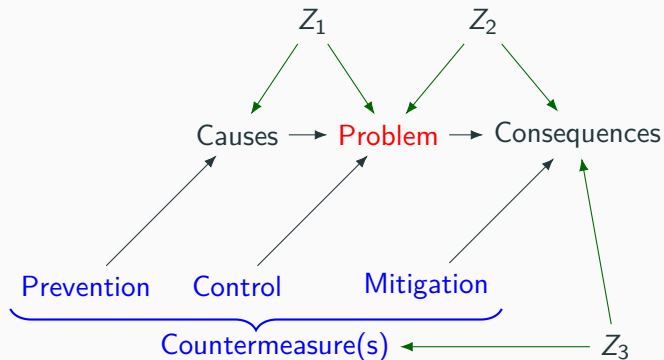
“... The relationships that were discovered between the graphical structure of the diagram and the data that it represents now permit us to emulate wiggling without physically doing so. Specifically, applying a smart sequence of conditioning operations enables us to predict the effect of actions or interventions without actually conducting an experiment. ... We can ask the diagram to emulate the experiment and tell us if any conditioning operation can reproduce the correlation that would prevail in the experiment.”

From Bayesian networks to causal diagrams (TBoW, p. 131)

“... We can now decide which set of variables we must measure in order to predict the effects of interventions from observational studies. We can also answer “Why?” questions. For example, someone may ask why wiggling A makes C vary. Is it really the direct effect of A , or is it the effect of a mediating variable B ? If both, can we assess what portion of the effect is mediated by B ?

To answer such mediation questions, we have to envision two simultaneous interventions: wiggling A and holding B constant (to be distinguished from conditioning on B).”

A3 Thinking (problem-solving) requires causal thinking/inference



Z_i = confounders

Summary

1. Bayes' theorem enable **causal and evidential reasoning**
2. Bayesian networks are **variables** (nodes) with probabilistic relationships.
3. A Bayesian network can be **noncausal** (smoke $\rightarrow$ fire)
4. Bayesian networks enable probabilistic reasoning (Rung 1)
5. A Bayesian network can be **causal** (fire $\rightarrow$ smoke): (directed acyclic graph)
6. DAG edges (arrows) are fixed, but BN probabilistic properties apply
7. Bayesian networks connect DAGs to data (via **statistical methods**)
8. DAGs have three core patterns (chain, fork, collider)
9. DAGs are used for **interventions** (Rung 2) and **counterfactuals** (Rung 3)
10. Bayesian networks can be generalized to conduct **decision analysis** by adding decision, deterministic (calculation), and value nodes.

References

1. Judea Pearl & Dan Mackenzie. The Book of Why: The New Science of Cause and Effect. Basic Books; 1 edition (May 15, 2018)
2. Marco Scutari & Jean-Baptiste Denis. Bayesian Networks: With Examples in R. Chapman and Hall/CRC; 1 edition (June 20, 2014) (`bnlearn` R package)
3. Norman Fenton & Martin Neil. Risk Assessment and Decision Analysis with Bayesian Networks. Chapman and Hall/CRC; 2 edition (September 12, 2018)
4. Neapolitan R, Jiang X, Ladner DP, Kaplan B. A Primer on Bayesian Decision Analysis With an Application to a Kidney Transplant Decision. Transplantation. 2016 Mar;100(3):489-96. doi: 10.1097/TP.0000000000001145. Review. PubMed PMID: 26900809; PubMed Central PMCID: PMC4818954.