Chapter 3 (Part 1) of "The Book of Why"

From Evidence to Causes—Reverend Bayes meets Mr. Holmes

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

June 6, 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 LaTeX Beamer—Metropolis theme

A patient presents with chest pain to clinical provider.

The patient has a history of coronary artery disease.

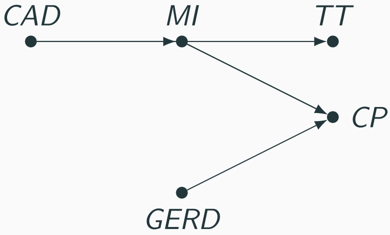


Figure 1: A patient with a history of coronary artery disease (CAD) presents to a provider complaining of prolonged chest pain (CP). The provider's differential diagnosis (hypotheses) are myocardial infarction (MI) and gastroesophageal reflux disease (GERD). The provider sends a blood specimen for a Troponin Test (TT) to "rule out" a MI.

Sherlock Holmes: deduction vs. induction

Deduction vs. induction

"It's elementary, my dear Watson."

So spoke Sherlock Holmes ... Holmes performed not just **deduction**, which works *from a hypothesis to a conclusion*.[1] His great skill was **induction**, which works in the opposite direction, *from evidence to hypothesis*.[2]

Multi-cause reasoning (an extension of causal reasoning)

"When you have eliminated the impossible, whatever remains, however improbable, must be the truth." Having *induced* several hypothesis, Holmes eliminated them one by one in order to *deduce* (by elimination) the correct one.

[1] Causal reasoning

[2] Evidential reasoning

Synonyms

$$\text{Cause} \bullet \rightarrow \bullet \text{Effect}$$

$$\text{Hypothesis} \bullet \rightarrow \bullet \text{Evidence}$$

Graph	Conditional prob.	Probability	Synonym	Reasoning type
H $\rightarrow$ E	$P(E \mid H)$	Forward prob.	deduction	causal reasoning[3]
H $\rightarrow$ E	$P(H \mid E)$	"Inverse" prob.	induction	evidential reasoning[4]

[3] Also called "predictive" reasoning

[4] Also 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	$P(E \mid H)$	Likelihood (TP [sensitivity], FP [1-specificity])
Evidential reasoning	3	$P(E)$	Marginal probability of E
... leads to	4	$P(H \mid E)$	Posterior probability (conditional probability)

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

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

Bayes' theorem for causal (Hypothesis) $\rightarrow$ (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])

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

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

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

Public health examples: Cause (hypothesis) $\bullet \to \bullet$ Effect (evidence)

Exposure $\bullet \to \bullet$ Disease

Disease $\bullet \to \bullet$ Test result

Smoking $\bullet \to \bullet$ Lung cancer

Jury trial: Guilty $\bullet \to \bullet$ Evidence

Citrus fruit consumption $\bullet \to \bullet$ Scurvy

Green house gases $\bullet \to \bullet$ Global warming

Global warming $\bullet \to \bullet$ Extreme weather events

Implicit (nonconscious) bias $\bullet \to \bullet$ Discrimination

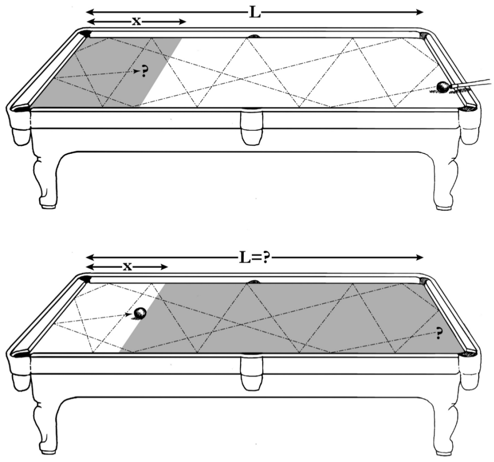
Reverend Thomas Bayes' pool table example

L = length of pool table
x = feet from left end of table

$$L \text{ (cause)} \rightarrow x \text{ (effect)}$$

Forward probability
$$P(x \mid L)$$

"Inverse" probability
$$P(L \mid x)$$



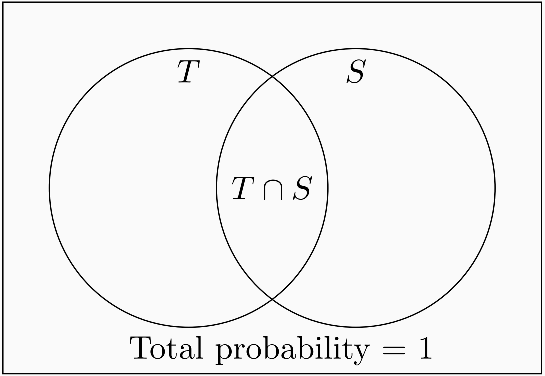
Tea-Scones example (default: assume probabilistic dependence)

Customer	Tea	Scones
1	Yes	Yes
2	No	Yes
3	No	No
4	No	No
5	Yes	Yes
6	Yes	No
7	Yes	No
8	Yes	Yes
9	Yes	No
10	Yes	No
11	No	No
12	Yes	Yes

Tea	Scones		
	No	Yes	Total
No	3	1	4
Yes	4	4	8
Total	7	5	12

$$P(X, Y, T) = P(X)P(T)P(Y \mid X, T)$$

Tea-Scones example (default: assume probabilistic dependence)



$$P(\text{Tea}) = P(T)$$
$$P(\text{Scone}) = P(S)$$
$$P(\text{Tea} \cap \text{Scone}) = P(T, S)$$
$$= P(S, T)$$

$$P(S, T) = P(T)P(S \mid T) \qquad (1)$$
$$P(T, S) = P(S)P(T \mid S) \qquad (2)$$

Bayes' Theorem

$$P(T)P(S \mid T) == P(S)P(T \mid S)$$

Bayes' theorem

$$P(S \mid T) = \frac{P(S)P(T \mid S)}{P(T)}$$

$$P(\text{Hypothesis} \mid \text{Evidence}) = \frac{P(\text{Hypothesis})P(\text{Evidence} \mid \text{Hypothesis})}{P(\text{Evidence})}$$

Bayes' theorem

$$P(\text{Hypothesis} \mid \text{Evidence}) = \frac{P(\text{Hypothesis})P(\text{Evidence} \mid \text{Hypothesis})}{P(\text{Evidence})}$$

Excerpt from "The Book of Why" (p. 101)

"This is perhaps the most important role of Bayes' rule in statistics: we can estimate the conditional probability directly in one direction, for which our judgment is more reliable (*deduction*),[5] and use mathematics to derive the conditional probability in the other direction, for which our judgment is rather hazy (*induction*)."[6]

[5] causal reasoning

[6] evidential reasoning

Bayes' theorem

$$P(S \mid T) = \frac{P(S, T)}{P(T)} = \frac{P(S)P(T \mid S)}{P(T)}$$

$$P(\text{Hypothesis} \mid \text{Evidence}) = \frac{P(\text{Hypothesis})P(\text{Evidence} \mid \text{Hypothesis})}{P(\text{Evidence})}$$

Excerpt from "The Book of Why" (p. 103)

[Bayes' rule] acts, in fact, as a normative rule for updating beliefs (causal hypotheses) in response to evidence. . . . [T]he belief a person attributes to S (cause) after discovering T (evidence) is never lower than the degree of belief that person attributes to S AND T before discovering T. Also, it implies that the more surprising the evidence T—that is, the smaller $P(T)$ is—the more convinced one should become of its cause S.

Example: Mammogram for breast cancer screening

$$\text{Cause} \bullet \rightarrow \bullet \text{ Effect}$$
$$\text{Disease} \bullet \rightarrow \bullet \text{ Test}$$

$$\begin{aligned}
P(\text{Disease} \mid \text{Test}) &= \frac{P(\text{Disease}, \text{Test})}{P(\text{Test})} \\
&= \frac{P(\text{Disease})P(\text{Test} \mid \text{Disease})}{P(\text{Test})} \\
&= P(\text{Disease})\left(\frac{P(\text{Test} \mid \text{Disease})}{P(\text{Test})}\right) \\
&= P(\text{Disease})(\textbf{Likelihood Ratio})
\end{aligned}$$

Bayes' theorem example: Mammogram for breast cancer screening

$$\text{Cause} \bullet \rightarrow \bullet \text{ Effect}$$

$$\text{Disease} \bullet \rightarrow \bullet \text{ Test}$$

$$
\begin{aligned}
P(\text{Disease} \mid \text{Test}) &= \frac{P(\text{Disease, Test})}{P(\text{Test})} \\
&= \frac{P(\text{Disease})P(\text{Test} \mid \text{Disease})}{P(\text{Test})} \\
&= \frac{P(D)P(T \mid D)}{P(D)P(T \mid D) + P(\bar{D})P(T \mid \bar{D})}
\end{aligned}
$$

Bayes' theorem example: Mammogram for breast cancer screening

$$\text{Cause} \bullet \rightarrow \bullet \text{ Effect}$$
$$\text{Disease} \bullet \rightarrow \bullet \text{ Test}$$

$$\begin{aligned}
P(D+ \mid T+) &= \textbf{Positive Predictive Value} \\
&= \frac{P(D+)P(T+ \mid D+)}{P(D+)P(T+ \mid D+) + P(D-)P(T+ \mid D-)} \\
&= \frac{P(D+)(\text{True Positive})}{P(D+)(\text{True Positive}) + P(D-)(\text{False Positive})} \\
&= \frac{P(D+)(\text{Sensitivity})}{P(D+)(\text{Sensitivity}) + P(D-)(1 - \text{Specificity})}
\end{aligned}$$

Bayes' theorem example: Mammogram for breast cancer screening

$$\text{Disease} \bullet \rightarrow \bullet \text{ Test}$$

A 43 woman has a positive mammogram ($T+$). What is the probability of breast cancer given the positive test ($P(D+ \mid T+)$)? What do we know?

$P(D+) = 1/700$ for 43 year old women

$P(T+ \mid D+) = 0.73 =$ True Positive $=$ Sensitivity

$P(T+ \mid D-) = 0.12 =$ False Positive $=$ 1 - Specificity

$$\begin{aligned} P(D+ \mid T+) &= \frac{P(D+)(\text{TP})}{P(D+)(\text{TP}) + P(D-)(\text{FP})} \\ &= \frac{(1/700)(0.73)}{(1/700)(0.73) + (1 - 1/700)(0.12)} \approx 0.009 \end{aligned}$$

Bayes' theorem: Review sensitivity and specificity of a diagnostic test

Operating characteristics of a diagnostic test

Sensitivity $= P(T+ \mid D+) = \frac{TP}{TP+FN}$

Specificity $= P(T- \mid D-) = \frac{TN}{TN+FP}$

Remember "SnOut" and "SpIn"

SnOut: Use a very sensitive test (very low FN) to "rule out" a hypothesis. That is, when we have *confidence in a negative result* we can use the test to *rule out* a hypothesis.

SpIn: Use a very specific test (very low FP) to "rule in" a hypothesis. That is, when we have *confidence in a positive result* we can use the test to *rule in* a hypothesis.

Bayes' theorem

Bayes' rule is a distillation of the scientific method (TBoW, p. 108)

1. Formulate a causal hypothesis
2. Deduce a testable consequence of the hypothesis (causal reasoning)
3. Design an evaluation (study) and collect evidence
4. Update your belief in the causal hypothesis (evidential reasoning)

Bayes' theorem for causal (Disease) → (Test)

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])

$$P(D \mid T) = \frac{P(D)P(T \mid D)}{P(T)}$$

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

3. Marginal probability of Evidence summed over all possibilities; **Likelihood Ratio** is (2) ÷ (3)

Bayesian networks are nodes 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.

$$\text{Smell smoke} \bullet\!\longrightarrow\!\bullet\ \text{Fire nearby}$$

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

$$\text{Fire} \bullet\!\longrightarrow\!\bullet\ \text{Smoke}$$

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

$$\text{Disease} \bullet\!\longrightarrow\!\bullet\ \text{Test}$$

Bayesian networks generalize Bayes' theorem for complex causal graphs

Core DAG patterns for three nodes and two edges

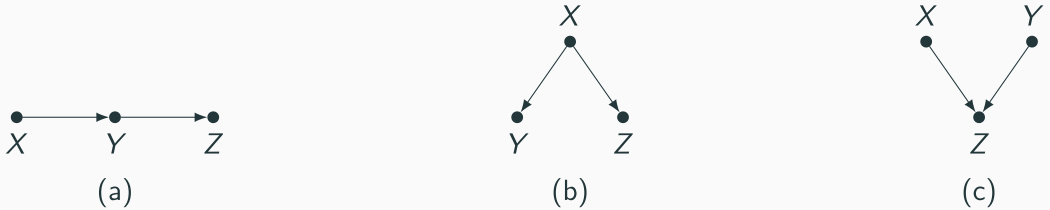


Figure 2: Core DAG patterns for three nodes and two edges: (a) chain (sequential cause), (b) fork (common cause), and (c) collider (common effect).

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

The patient has a history of coronary artery disease.

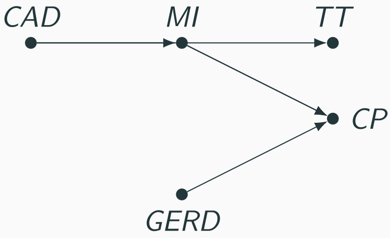


Figure 3: A patient with a history of coronary artery disease (CAD) presents to a provider complaining of prolonged chest pain (CP). The provider's differential diagnosis (hypotheses) are myocardial infarction (MI) and gastroesophageal reflux disease (GERD). The provider sends a blood specimen for a 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.