Overview

- Bayes rule: Significance
- Diagnostic vs. causal knowledge
- Calculating $P(T)$ given $P(T|D)$ and $P(D)$
- Ratios of conditional probabilities and causes of an phenomenon
- Example: object recognition
- Bayesian updating

Significance of Bayes’ Rule

- $P(T|D)$ may be easier to obtain: you can run the test on a small pool of known patients (say 100) at a hospital.
- $P(D|T)$ is much harder to obtain directly. Since the test makes 1 mistake out of 100 tests, if you run the test on 10,000 people, you’ll get 100 false-positives, and one genuine patient who tests positive (consider that $P(T) = 0.010098$). So, just to get about 100 people testing positive, you have to run the tests on 10,000 people.
- $P(D)$ serves as a prior in this case. In many cases, the prior represents subjective belief of the person calculating the probability in case $P(D)$ is not directly measurable.

Bayes’ Rule: Example Revisited

These are given:

\[
\begin{align*}
P(T|D) &= 0.99 \\
P(\neg T|\neg D) &= 0.99 \\
P(D) &= \frac{1}{10,000} = 0.0001
\end{align*}
\]

We want to calculate the probability that you have the disease given a positive test result:

\[
P(D|T) = \frac{P(T|D)P(D)}{P(T)}
\]

Diagnostic vs. Causal Knowledge

Consider these probabilities:

- $P(Symptom|Disease)$: causal knowledge
  - relatively fixed.

- $P(Disease)$: somewhat variable.

- $P(Disease|Symptom)$: diagnostic knowledge
  - fluctuates as $P(Disease)$ change.

$P(Disease|Symptom)$ directly measured can be no longer accurate when $P(Disease)$ changes (e.g. an epidemic outbreak), however the calculation based on Bayes’ rule can be much more robust.
Calculating $P(T)$ given $P(T|D)$ and $P(D)$

\[ P(T) = P(T \land D) + P(T \land \neg D) \]
\[ = P(T|D)P(D) + P(T|\neg D)P(\neg D) \]

- \{D\} $\cup$ \{\neg D\} completely account for the whole population, but
- \{T\} $\cup$ \{\neg T\} does not cover the whole population (because you did not test everyone!).

Calculating $P(T)$ given $P(T|D)$ and $P(D)$

Another way of deriving
\[ P(T) = P(T|D)P(D) + P(T|\neg D)P(\neg D) \]

From
\[ P(D|T) = \frac{P(T|D)P(D)}{P(T)} \]
\[ P(\neg D|T) = \frac{P(T|\neg D)P(\neg D)}{P(T)} \]

and from $P(D|T) + P(\neg D|T) = 1,$
\[ 1 = \frac{P(T|D)P(D)}{P(T)} + \frac{P(T|\neg D)P(\neg D)}{P(T)}, \text{thus} \]
\[ P(T) = P(T|D)P(D) + P(T|\neg D)P(\neg D) \]

Comparison of Conditional Probabilities

When $C_1$ or $C_2$ can cause phenomenon (or effect) $E$, to find out which is the more probable cause of phenomenon $E$, we do not need to explicitly calculate $P(E)$:

- $P(C_1|E) = \frac{P(E|C_1)P(C_1)}{P(E)}$
- $P(C_2|E) = \frac{P(E|C_2)P(C_2)}{P(E)}$

- From the above, we get:
\[ \frac{P(C_1|E)}{P(C_2|E)} = \frac{P(E|C_1)P(C_1)}{P(E|C_2)P(C_2)} = \frac{a}{b} \]
Example: The Problem of Object Recognition

Given an image projected on the retina, what is the more likely cause? the 2D hexagon? or a transparent 3D cube? This is basically a computer vision problem.

\[
\frac{P(\text{Hexagon}|\text{Image})}{P(\text{Cube}|\text{Image})} = \frac{P(\text{Image}|\text{Hexagon})P(\text{Hexagon})}{P(\text{Image}|\text{Cube})P(\text{Cube})} = \frac{a}{b}
\]

A probabilistic vision agent can make a decision based on such a ratio.

Example: Object Recognition (cont’d)

\[
\frac{P(\text{Hexagon}|\text{Image})}{P(\text{Cube}|\text{Image})} = \frac{P(\text{Image}|\text{Hexagon})P(\text{Hexagon})}{P(\text{Image}|\text{Cube})P(\text{Cube})} = \frac{a}{b}
\]

- Decision: if \(a/b > 1\), it is most likely that a hexagon generated the image. If \(a/b < 1\), it is most likely that a cube generated the image.

Example: Object Recognition (cont’d)

\[
\frac{P(\text{Hexagon}|\text{Image})}{P(\text{Cube}|\text{Image})} = \frac{P(\text{Image}|\text{Hexagon})P(\text{Hexagon})}{P(\text{Image}|\text{Cube})P(\text{Cube})} = \frac{a}{b}
\]

Combining Multiple Evidences

Suppose we have these conditional probabilities \(P(\text{Cavity}|\text{Toothache})\) and \(P(\text{Cavity}|\text{Catch})\). What if we want to know \(P(\text{Cavity}|\text{Toothache} \land \text{Catch})\)? These are the alternatives:

- Look up the joint probability table: not practical or even impossible in most cases
- We can calculate

\[
P(\text{Cav}|\text{Ache} \land \text{Catch}) = \frac{P(\text{Ache} \land \text{Catch}|\text{Cav})P(\text{Cav})}{P(\text{Ache} \land \text{Catch})}
\]

but, calculating the new conditional prob and the normalization factor is a pain.
Bayesian Updating

An Alternative: gradually work in the multiple evidences – Bayesian Updating

- Reformulate the Bayes' rule so that conditional probability of events given combined evidences (such as \( P(A | B \land C) \)) are not necessary.

- Use domain knowledge to replace the more complex conditional probabilities with known, simpler ones (utilize conditional independence).

Bayesian updating makes combing evidences efficient (more detail next time).

Key Points

- Why and when is Bayesian analysis useful?
- How to calculate priors from conditional distributions?
- How is subjective belief utilized in Bayesian analysis?

A Brief Summary

Topics covered so far:

- Search
- Logical inference
- Probabilistic inference

Topics to be covered:

- Learning
- Special topics

Next Time

- Conditional independence and efficient probabilistic inference.
- Probabilistic reasoning: chapter 15