Decision Support Technology

6.871

Knowledge Based Systems

Howard Shrobe & Randall Davis

- Bayes Theorem
  - Intuition
  - Formalization
  - Use in Decision Making Programs
  - Problems with Application
- Decision Flow Diagrams
  - Basic Concepts of Decision Diagrams
  - Expected Monetary Value
  - Averaging Out and Folding Back
- Maximum Entropy and Choosing the Best Test

Bayes Theorem
What is it for?

- **A Framework for Interpreting Evidence in an Uncertain World.**
- **A Means for inferring *Causes* from observation of their *Effects***
- **Generally it is more appropriate to refer to *Hypothesis* and *Evidence*
  since causation is a very tricky subject, particular in statistics.**

An Example

Should There be Universal AIDS Testing?

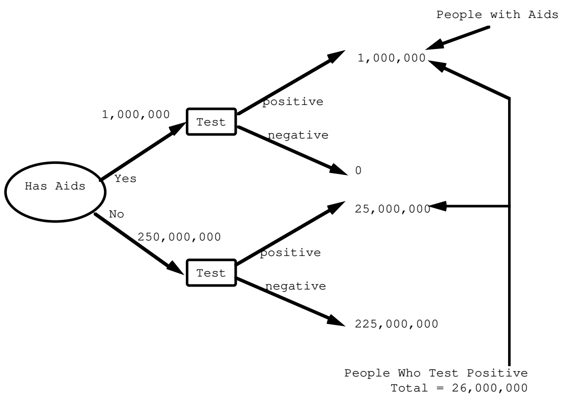
Somebody Claims to have a test for AIDS which is 90% accurate. He suggests that, because of the severity of the disease and the power of his test, we should test everybody for AIDS.

Suppose that you were tested and the results were positive. Should you be worried?

How likely is it that you have the disease?

Bayes Rule (or theorem) tells you how to answer these questions.

Understanding Bayes Theorem



Given that you tested positive, the probability that you have the disease is:

- There are about 1 million HIV positive individuals in the U.S.
- There are about 250 million people in the U.S.
- Assume that the test has no false negatives.
- Assume that the 90% accuracy means 1% false positive

Bayes Theorem Formalization

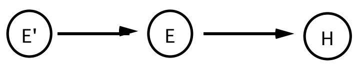
- **~H** — **The negation of H**
- **P(H|E)** — **The probability of H given that E has been seen. This is called the Posterior Probability of H.**
- **P(E|H)** — **The probability that E will be manifested given that H exists.**
- **P(H)** — **The *a priori* probabiliy that H exists.**
- **P(E)** — **The *a priori* probability of E manifesting, whether H exists or not.**

The Theorem

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

- All the quantities on the right hand side can, in principle, be determined *a priori.*
- If H is better thought of as part of a set of alternatives, then the denominator needs to sum over all of these which is the reason that the second form is a more useful statement.
- In practice, the numbers required by the theory can't really be gathered with accuracy. In medicine, for example, the best we have is often just subjective estimates by physicians.

Bayes Theorem: Using Uncertain Evidence



- **Suppose H can be inferred from E which is itself uncertain (perhaps since it is inferrred from E').**

  How do we handle such uncertain evidence (or chained inference)?

$$
\begin{aligned}
\mathbf{P(H|E')} &= \mathbf{P(H,E|E') + P(H,\sim E|E')} && \mathbf{(1)}\\
&= \mathbf{P(H|E,E') \cdot P(E|E')} \; + \\
&\quad\; \mathbf{P(H|\sim E,E') \cdot P(\sim E,E')} && \mathbf{(2)}\\
&= \mathbf{P(H|E) \cdot P(E|E') + P(H|\sim E) \cdot P(\sim E|E')} && \mathbf{(3)}
\end{aligned}
$$

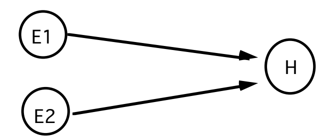
- P(H|E) is calculated using Bayes Theorem.
- P(E|E') is either a given degree of certainty in the evidence or the result of using a "certainty factor" for a rule to derate the evidence supporting the rule.
- This requires us to know P(E) and P(E|H) to apply Bayes Theorem.

(1) Follows just because you can always distribute any probability over a pair of negated propositions.

(2) Follows by use of Bayes Theorem (ignoring the same denominator on both sides)

(3) Follows because E' doesn't make us any more certain about H than would E alone (or ~E alone).

Bayes Theorem Using Multiple Pieces of Evidence



$$P(H|E1,E2) = \frac{P(E1,E2|H) \cdot P(H)}{P(E1,E2)}$$

- **If you assume "conditional independence" between the evidence then this takes on a nice multiplicative form.**
- **Conditional Independence is the notion that the various pieces of evidence are statistically independent of one another, given that the hypothesis obtains, i.e. the hypothesis "separates" the different pieces of evidence:**

  $\mathbf{P(E1,E2|H) = P(E1|H) \cdot P(E2|H)}$

  $\mathbf{P(\sim E1,\sim E2|H) = P(\sim E1|H) \cdot P(\sim E2|H)}$
- **Without conditional independence you need to build up a very large database of joint probabilities and joint conditional probabilities.**

The Problem Statement

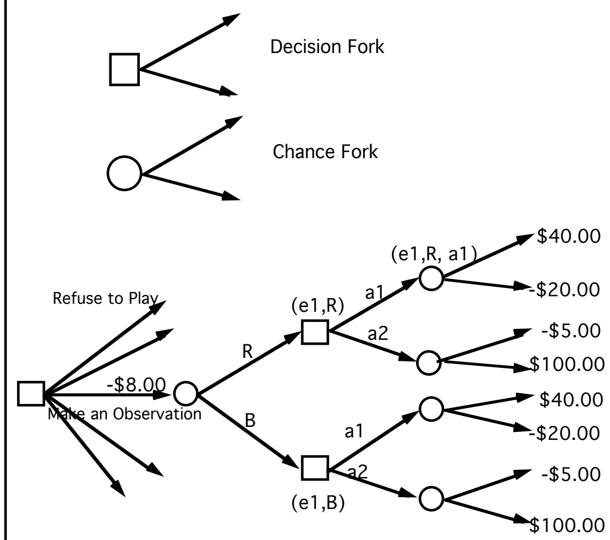
- There are two types of Urn U1 and U2.

  U1 contains 4 Red balls and 6 Black balls

  U2 contains nine Red balls and one Black ball
- An urn is selected at random and you are to guess which type it is.
- You have several courses of action:

Refuse to play	No Payoff no cost
Guess it is of type 1	$40 Payoff if right, $20 penalty if wrong
Guess it is of type 2	$100 Payoff if right, $5 penalty if wrong
Sample a ball	$8 payment for the right to sample

Decision Flow Diagrams

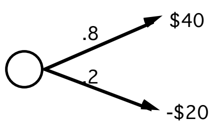


Expected Monetary Value

- Suppose there are several possible outcomes

  Each has a monetary payoff or penalty

  Each has a probability
- The Expected Monetary Value is the sum of the products of the monetary payoffs times their corresponding probabilites.



EMV = .8 · $40 + .2 · -$20 = $32 + (-$4) = $28

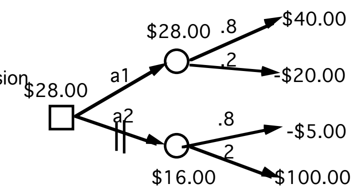
- EMV is a normative notion of what a person who has no other biases (risk aversion, e.g.) should be willing to accept in exchange for the situation facing him. In the picture above, you should be indifferent to the choice of taking $28 or playing the game.

- Most people have some extra biases and these can be incorporated in the form of a utility function applied to the calculated value.

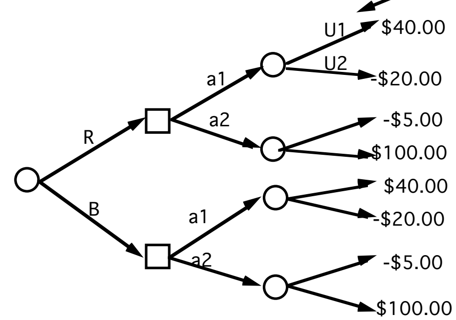
- A rational person should choose that course of action which has the highest EMV.

Averaging Out and Folding Back

EMV of Decision Node is Max Over all Branches



Bayes Theorem is Used to Calculate Probabilites At Chance Nodes following Decision Nodes That Provide Relevant Evidence.



P(U1) is .8 P(R|U1) = .4 P(R|U2) = .9 P(U2) = .2

P(U1|R) = P(R|U1) · P(U1) / P(R|U1) · P(U1) + P(R|U2) · P(U2)

= .4 · .8 / (.4 · .8 + .9 · .2) = .64

Maximum Entropy

.1 .2 .5 .2

Several Competing Hypotheses
Each with a Probability rating.

- **Entropy = Sum $-2 \cdot P(i) \cdot \text{Log}\ P(i)$ is a standard measure of Information.**

- **Suppose there are several tests you can make.**
- **Each test can change the probability of some (or all) of the hypotheses (using Bayes Theorem).**
- **Each outcome of the test has a probability.**
- **Which test should you make?**
- **For each outcome of a test calculate the change in entropy.**

  Weigh this by the probability of that outcome.

  Sum these to get an expected change of entropy for the test.
- **Chose that test which has the greatest expected change in entropy.**

  This is equivalent to chosing the test which is most likely to provide the most information.
- **Tests have different costs (sometimes quite drastic ones like life and death).**

  Normalize the benefits by the costs and then make choice.