Bayes Theorem

- **The notion of conditional probability : P(H\E)**

Let :

$P(H_i \backslash E)$ = the probability that hypothesis H_i is true given evidence E
$P(E \backslash H_i)$ = the probability that we will observe evidence E given that hypothesis i is true
$P(H_i)$ = the *a priori* probability that hypothesis i is true in the absence of any specific evidence. These probabilities are called prior probabilities or *prlors*.
k = the number of possible hypotheses

Bayes' theorem then states that

$$P(H_i \backslash E) = \frac{P(E \mid H_i) \cdot P(H_i)}{\sum_{n=1}^{k} P(E \mid H_n) \cdot P(H_n)}$$

- **Bayes' theorem then states that**

$$P(H_i \backslash E) = \frac{P(E \mid H_i) \cdot P(H_i)}{\sum_{n=1}^{k} P(E \mid H_n) \cdot P(H_n)}$$

- **Further, if we add a new piece of evidence, e, then**

$$P(H \backslash E, e) = P(H \mid E) \cdot \frac{P(e \mid E, H)}{P(e \mid E)}$$

Adding Certainty Factors to Rules

- **Example of a Mycin Rule :**

```
(1)  the stain of the organism is gram-positive, and
(2)  the morphology of the organism is coccus, and
(3)  the growth conformation of the organism is clumps,
     then there is suggestive evidence (0.7) that
     the identity of the organism is staphylococcus.
```

This is the form in which the rules are stated to the user. They are actually represented internally in an easy-to-manipulate LISP list structure. The rule we just saw would be represented internally as

```
PREMISE:   ($AND   (SAME CNTXT GRAM GRAMPOS)
                   (SAME CNTXT MORPH COCCUS)
                   (SAME CNTXT CONFORM CLUMPS))
ACTION:            (CONCLUDE CNTXT IDENT STAPHYLOCOCCUS TALLY 0.7)
```

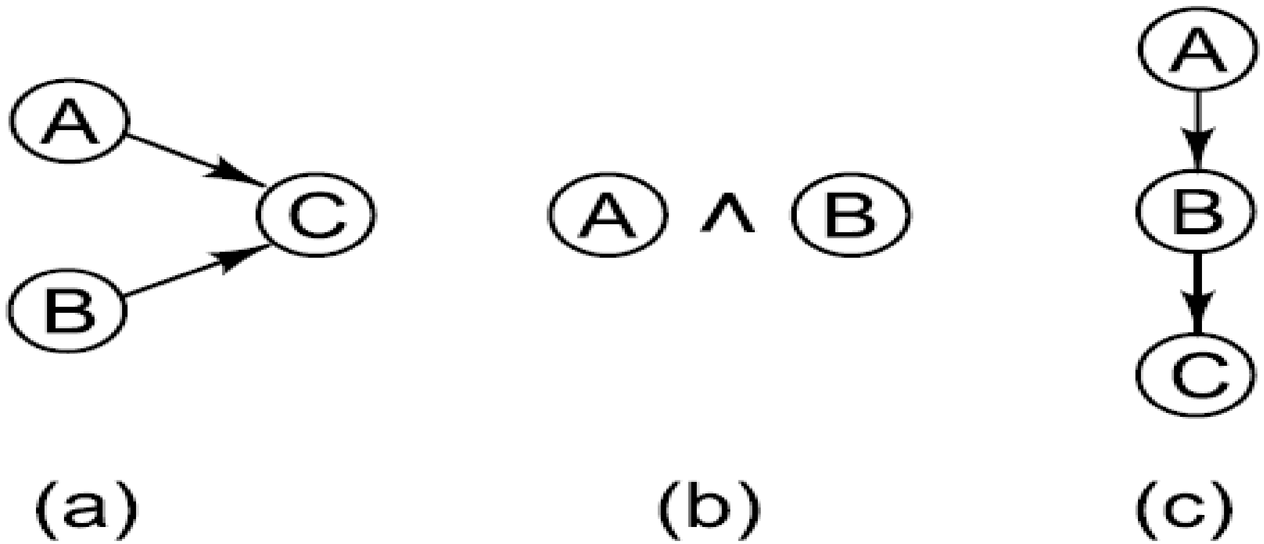
Measures of Belief

- $MB[h, e]$—a measure (between 0 and 1) of belief in hypothesis h given the evidence e. MB measures the extent to which the evidence supports the hypothesis. It is zero if the evidence fails to support the hypothesis.

- $MD[h,e]$—a measure (between 0 and 1) of disbelief in hypothesis h given the evidence e. MD measures the extent to which the evidence supports the negation of the hypothesis. It is zero if the evidence supports the hypothesis.

$$CF[h, e] = MB[h, e] - MD[h, e]$$

Combining Uncertain Rules



Combining Uncertain Rules

Goals for combining rules :

- Since the order in which evidence is collected is arbitrary, the combining functions should be commutative and associative.
- Until certainty is reached, additional confirming evidence should increase MB (and similarly for disconfirming evidence and MD).
- If uncertain inferences are chained together, then the result should be less certain than either of the inferences alone.

Combining Two Pieces of Evidence

$$MB[h, s_1 \wedge s_2] = \begin{cases} 0 & \text{if } MD[h, s_1 \wedge s_2] = 1 \\ MB[h, s_1] + MB[h, s_2] \cdot (1 - MB[h, s_1]) & \text{otherwise} \end{cases}$$

$$MD[h, s_1 \wedge s_2] = \begin{cases} 0 & \text{if } MB[h, s_1 \wedge s_2] = 1 \\ MD[h, s_1] + MD[h, s_2] \cdot (1 - MD[h, s_1]) & \text{otherwise} \end{cases}$$

$$MB[h_1 \wedge h_2, e] = \min(MB[h_1, e], MB[h_2, e])$$

$$MB[h_1 \wedge h_2, e] = max(MB[h_1, e], MB[h_2, e])$$

$$MB[h, s] = MB'[h, s] \cdot \max(0, CF[s, e])$$

An Example of Combining Two Observations

$$MB[h, s_1] = 0.3$$

$$MD[h, s_1] = 0.0$$

$$CF[h, s_1] = 0.3$$

$$MB[h, s_2] = 0.2$$

$$\begin{aligned} MB[h, s_1 \wedge s_2] &= 0.3 + 0.2 \cdot 0.7 \\ &= 0.44 \end{aligned}$$

$$MD[h, s_1 \wedge s_2] = 0.0$$

$$CF[h, s_1 \wedge s_2] = 0.44$$

The Definition of Certainty Factors

- **Original definitions :**

$$MB[h, e] = \begin{cases} 1 & \text{if } P(h) = 1 \\ \dfrac{\max[P(h \mid e), P(h)] - P(h)}{1 - P(h)} & \text{otherwise} \end{cases}$$

- **Similarly, the MD is the proportionate decrease in belief in h as a result of e:**

$$MD[h, e] = \begin{cases} 1 & \text{if } P(h) = 0 \\ \dfrac{\min[P(h \mid e), P(h)] - P(h)}{-P(h)} & \text{otherwise} \end{cases}$$

- **But this definition is incompatible with Bayesian conditional probability. The following, slightly revised one is not :**

$$MB[h, e] = \begin{cases} 1 & \text{if } P(h) = 1 \\ \dfrac{\max[P(h \mid e), P(h)] - P(h)}{(1 - P(h) \cdot P(h \mid e)} & \text{otherwise} \end{cases}$$

What if the Observations are not Independent

- **Scenario (a) :**

Reconsider a rule with three antecedents and a *CF* of 0.7. Suppose that if there were three separate rules, each would have had a *CF* of 0.06. In other words, they are not independent. Then, using the combining rules, the total would be:

$$\begin{aligned} MB[h, s \wedge s_2] &= 0.6 + (0.6 \cdot 0.4) \\ &= 0.84 \\ MB[h, (s_1 \wedge s_2) \wedge s_3] &= 0.84 + (0.6 \cdot 0.16) \\ &= 0.936 \end{aligned}$$

- **This is very different than 0.7.**

What if the Observations are not Independent

- **Scenario (c) :**

  Events :

S: sprinkler was on last night
W: grass is wet
R: it rained last night

We can write MYCIN-style rules that describe predictive relationships among these three events:

```
If:  the sprinkler was on last night
then there is suggestive evidence (0.9) that
     the grass will be wet this morning
```

Taken alone, this rule may accurately describe the world. But now consider a second rule:

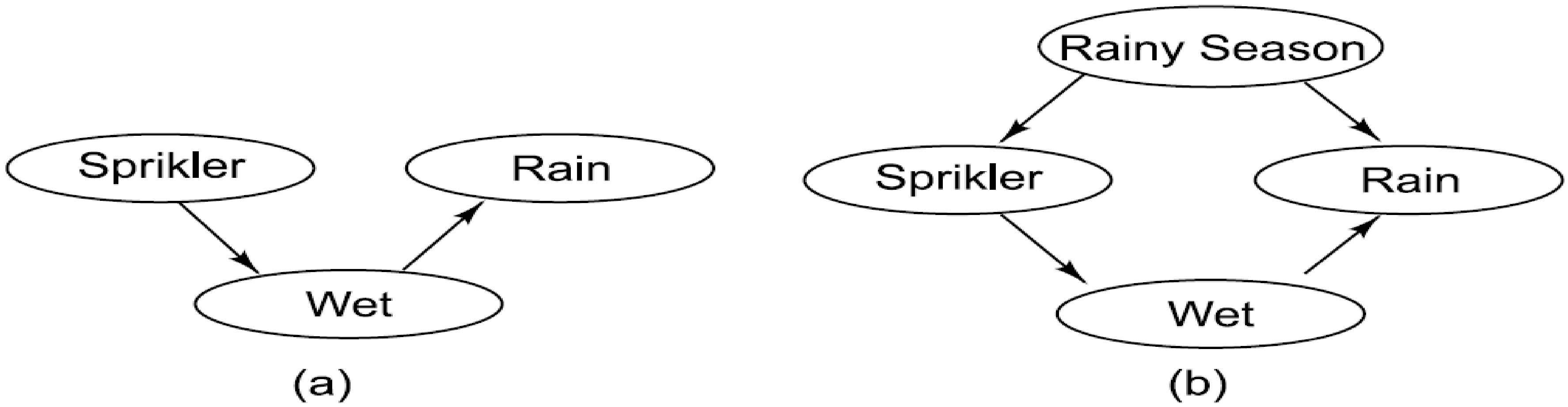
```
If:  the grass is wet this morning
then there is suggestive evidence (0.8) that
     it rained last night
```

Taken alone, this rule makes sense when rain is the most common source of water on the grass. But if the two rules are applied together, using MYCIN's rule for chaining, we get

$MB[W,S] = 0.8$ {sprinkler suggests wet}
$MB[R,W] = 0.8 \cdot 0.9 = 0.72$ {wet suggests rains}

- **So Sprinkler made us believe rain.**

Bayesian Networks : Representing Causality Uniformly



Conditional Probabilities for Bayesian Network

Attribute	Probability
$p(Wet \backslash Sprinkler,\ Rain)$	0.95
$P(Wet \backslash Sprinkler,\ \neg Rain)$	0.9
$p\{Wet \backslash \neg Sprinkler,\ Rain)$	0.8
$p(Wet \backslash \neg Sprinkler,\ \neg Rain)$	0.1
$p(Sprinkler \backslash RainySeason)$	0.0
$p(Sprinkler \backslash \neg RainySeason)$	1.0
$p(Rain\ \backslash RainySeason)$	0.9
$p(Rain\ \backslash\ \neg RainySeason)$	0.1
$p(RainySeason)$	0.5

Dempster -Shafer Theory

- **We consider the interval :**

$$[Belief,\ Plausibility]$$

- **Plausibility (Pl) is defined to be :**

$$Pl(s) = 1 - Bel(\neg s)$$

- **Let the frame of discernment be an Θ, austive, mutually exclusive set of hypothesis.**

- **Let m be a probability density function.**

- **We define the combination m_3 of m_1 and m_2 to be**

$$m_3(Z) = \frac{\sum_{X \cap Y = Z} m_1(X) \cdot m_2(Y)}{1 - \sum_{X \cap Y = \phi} m_1(X) \cdot m_2(Y)}$$

Dempster - Shafer Example

- **Let Θ be :**

  All : allergy

  Flu : flu

  Cold : cold

  Pneu : pneumonia
- **When we begin, with no information m is :**

  $\{\Theta\}$ (1.0)

- **Suppose m_1 corresponds to our belief after observing fever.**

  $\{Flu, Cold, Pneu\}$ (0.6)
  $\{\Theta\}$ (0.4)

- **Suppose m_2 corresponds to our belief after observing a runny nose.**

  $\{All, Flu, Cold\}$ (0.8)
  Θ (0.2)

Dempster - Shafer Example (Cont'd)

- Then we can combine m_1 and m_2 :

	$\{A, F, C\}$	(0.8)	Θ	(0.2)
$\{F, C, P\}$ (0.6)	$\{F, C\}$	(0.48)	$\{F, C, P\}$	(0.12)
Θ (0.4)	$\{A, F, C\}$	(0.32)	Θ	(0.08)

- So we produce a new, combined m_3 :

$$\begin{array}{ll} \{Flu,\ Cold\} & (0.48) \\ \{All,\ Flu, Cold\} & (0.32) \\ \{Flu, Cold, Pneu\} & (0.12) \\ \Theta & (0.08) \end{array}$$

- Suppose m_4 corresponds to our belief after that the problem goes away on trips :

$$\begin{array}{ll} \{All\} & (0.9) \\ \Theta & (0.1) \end{array}$$

Dempster - Shafer Example (Cont'd)

- Then we can combine m_1 and m_2 :

		$\{A\}$	(0.9)	Θ	(0.1)
$\{F, C\}$	(0.48)	ϕ	(0.432)	$\{F, C\}$	(0.048)
$\{A, F, C\}$	(0.32)	$\{A, F, C\}$	(0.288)	$\{A, F, C\}$	(0.032)
$\{F, C, P\}$	(0.12)	ϕ	(0.108)	$\{F, C, P\}$	(0.012)
Θ	(0.08)	$\{A\}$	(0.072)	Θ	(0.008)

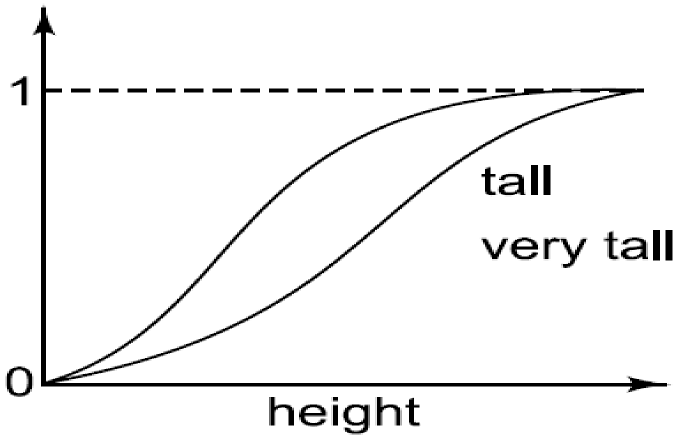
- Normalizing to get rid of the belief of 0.54 associated with ϕ gives m_5 :

$\{Flu, Cold\}$	(0.104)
$\{All, Flu, Cold\}$	(0.696)
$\{Flu, Cold, Pneu\}$	(0.026)
$\{All\}$	(0.157)
Θ	(0.017)

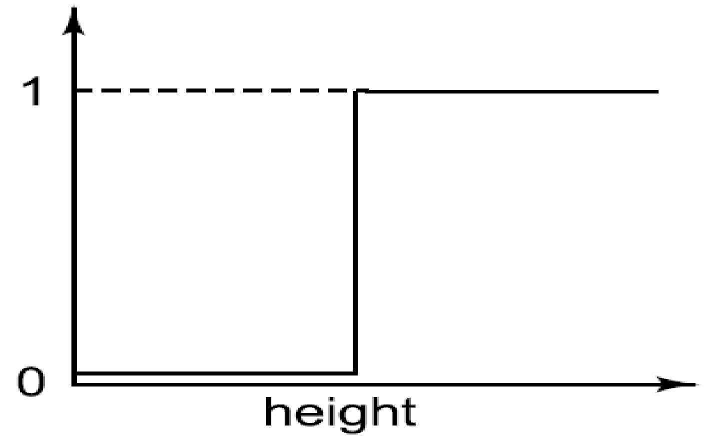
Fuzzy Logic

- Suppose we want to represent :
  - John is very tall.
  - Mary is slightly ill.
  - Sue and Linda are close friends.
  - Exceptions to the rule are nearly impossible.
  - Most Frenchmen are not very tall.

Fuzzy versus Conventional Set Membership



(a) Fuzzy Membership



(b) Conventional Membership