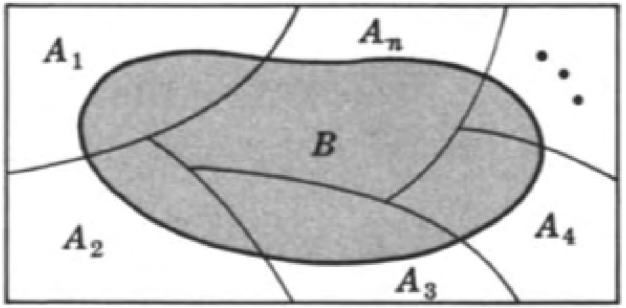
Bayes Theorem

Suppose the events $A_1, A_2, \ldots, A_n$ form a partition of a sample space S; that is, the events A_i are mutually exclusive and their union is S. Now let B be any other event. Then

$$B = S \cap B = (A_1 \cup A_2 \cup \cdots \cup A_n) \cap B$$
$$= (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B)$$



B is shaded.

where the $A_i \cap B$ are also mutually exclusive. Accordingly,

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)$$

Thus by the multiplication theorem,

$$P(B) = P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \cdots + P(A_n)P(B \mid A_n) \tag{1}$$

On the other hand, for any i, the conditional probability of A_i given B is defined by

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)}$$

In this equation we use (1) to replace $P(B)$ and use $P(A_i \cap B) = P(A_i)P(B \mid A_i)$ to replace $P(A_i \cap B)$, thus obtaining

Bayes' Theorem : Suppose $A_1, A_2, \ldots, A_n$ is a partition of S and B is any event. Then for any i,

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \cdots + P(A_n)P(B \mid A_n)}$$

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**

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

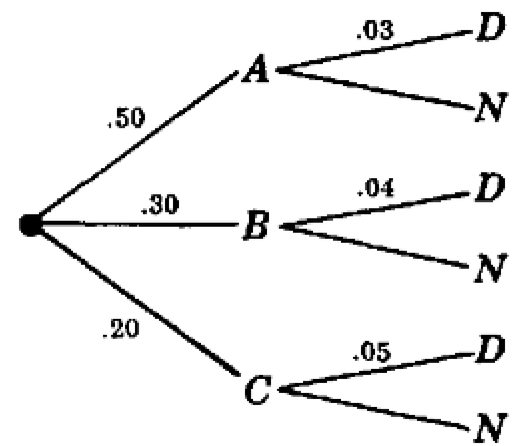
Three machines A, B and C produce respectively 50%, 30% and 20% of the total number of items of a factory. The percentages of defective output of these machines are 3%, 4% and 5%. If an item is selected at random, find the probability that the item is defective.

Three machines A, B and C produce respectively 50%, 30% and 20% of the total number of items of a factory. The percentages of defective output of these machines are 3%, 4% and 5%. If an item is selected at random, find the probability that the item is defective.

Let X be the event that an item is defective.

Then

$$\begin{aligned} P(X) &= P(A)\,P(X \mid A) + P(B)\,P(X \mid B) \\ &\quad + P(C)\,P(X \mid C) \\ &= (.50)(.03) + (.30)(.04) + (.20)(.05) \\ &= .037 \end{aligned}$$



Consider the factory in the preceding example. Suppose an item is selected at random and is found to be defective. Find the probability that the item was produced by machine A; that is, find $P(A \mid X)$.

Consider the factory in the preceding example. Suppose an item is selected at random and is found to be defective. Find the probability that the item was produced by machine A; that is, find $P(A \mid X)$.

By Bayes' theorem,

$$P(A \mid X) = \frac{P(A)\,P(X \mid A)}{P(A)\,P(X \mid A) + P(B)\,P(X \mid B) + P(C)\,P(X \mid C)}$$

$$= \frac{(.50)(.03)}{(.50)(.03) + (.30)(.04) + (.20)(.05)} = \frac{15}{37}$$

In other words, we divide the probability of the required path by the probability of the reduced sample space, i.e. those paths which lead to a defective item.

INDEPENDENCE

An event B is said to be *independent* of an event A if the probability that B occurs is not influenced by whether A has or has not occurred. In other words, if the probability of B equals the conditional probility of B given A: $P(B) = P(B \mid A)$. Now substituting $P(B)$ for $P(B \mid A)$ in the multiplications theorem $P(A \cap B) = P(A)\,P(B \mid A)$, we obtain

$$P(A \cap B) = P(A)\,P(B)$$

We use the above equation as our formal definition of independence.

Example

In Orange County, 51% of the adults are males. (It doesn't take too much advanced mathematics to deduce that the other 49% are females.) One adult is randomly selected for a survey involving credit card usage.

Find the prior probability that the selected person is a male.

Let's use the following notation:

M = male F = female (or not male)

we know that 51% of the adults in Orange County are males, so the probability of randomly selecting an adult and getting a male is given by *P*(M) = 0.51.

Part B

It is later learned that the selected survey subject was smoking a cigar. Also, 9.5% of males smoke cigars, whereas 1.7% of females smoke cigars (based on data from the Substance Abuse and Mental Health Services Administration). Use this additional information to find the probability that the selected subject is a male.

Based on the additional given information, we have the following:

P(M) = 0.51 because 51% of the adults are males

P(F) = 0.49 because 49% of the adults are females (not males)

P(C|M) = 0.095 because 9.5% of the males smoke cigars (That is, the probability of getting someone who smokes cigars, given that the person is a male, is 0.095.)

P(C|F) = 0.017. because 1.7% of the females smoke cigars (That is, the probability of getting someone who smokes cigars, given that the person is a female, is 0.017.)

$$\begin{aligned}
P(M \mid C) &= \frac{P(M)\cdot P(C|M)}{[P(M)\cdot P(C|M)]+[P(\overline{M})\cdot P(C|\overline{M})]} \\
&= \frac{0.51\cdot 0.095}{[0.51\cdot 0.095]+[0.49\cdot 0.017]} \\
&= 0.85329341 \\
&= 0.853 \text{ (rounded)}
\end{aligned}$$

Example

An aircraft emergency locator transmitter (ELT) is a device designed to transmit a signal in the case of a crash. The Altigauge Manufacturing Company makes 80% of the ELTs, the Bryant Company makes 15% of them, and the Chartair Company makes the other 5%. The ELTs made by Altigauge have a 4% rate of defects, the Bryant ELTs have a 6% rate of defects, and the Chartair ELTs have a 9% rate of defects (which helps to explain why Chartair has the lowest market share).

a. If an ELT is randomly selected from the general population of all ELTs, find the probability that it was made by the Altigauge Manufacturing Company.

Solution

We use the following notation:

A = ELT manufactured by Altigauge
B = ELT manufactured by Bryant
C = ELT manufactured by Chartair

D = ELT is defective
$\overline{D}$ = ELT is not defective (or it is good)

If an ELT is randomly selected from the general population of all ELTs, the probability that it was made by Altigauge is 0.8 (because Altigauge manufactures 80% of them).

b. If a randomly selected ELT is then tested and is found to be defective, find the probability that it was made by the Altigauge Manufacturing Company.

$P(\mathrm{A}) = 0.80$ because Altigauge makes 80% of the ELTs
$P(\mathrm{B}) = 0.15$ because Bryant makes 15% of the ELTs
$P(\mathrm{C}) = 0.05$ because Chartair makes 5% of the ELTs

$P(\mathrm{D}|\mathrm{A}) = 0.04$ because 4% of the Altigauge ELTs are defective
$P(\mathrm{D}|\mathrm{B}) = 0.06$ because 6% of the Bryant ELTs are defective
$P(\mathrm{D}|\mathrm{C}) = 0.09$ because 9% of the Chartair ELTs are defective

Here is Bayes' theorem extended to include three events corresponding to the selection of ELTs from the three manufacturers (A, B, C):

$$P(A|D) = \frac{P(A)\cdot P(D|A)}{[P(A)\cdot P(D|A)]+[P(B)\cdot P(D|B)]+[P(C)\cdot P(D|C)]}$$

$$= \frac{0.80\cdot 0.04}{[0.80\cdot 0.04]+[0.15\cdot 0.06]+[0.05\cdot 0.09]}$$

$= 0.703$ (rounded)

At a certain university, 4% of men are over 6 feet tall and 1% of women are over 6 feet tall. The total student population is divided in the ratio 3:2 in favour of women. If a student is selected at random from among all those over six feet tall, what is the probability that the student is a woman?

Let M={Student is Male}, F={Student is Female}.
Note that M and F partition the sample space of students.
Let T={Student is over 6 feet tall}.

We know that $\mathsf{P}(M) = 2/5$, $\mathsf{P}(F) = 3/5$, $\mathsf{P}(T|M) = 4/100$ and $\mathsf{P}(T|F) = 1/100$.

We require $\mathsf{P}(F|T)$. Using Bayes' theorem we have:

$$\begin{aligned}
\mathsf{P}(F|T) &= \frac{\mathsf{P}(T|F)\mathsf{P}(F)}{\mathsf{P}(T|F)\mathsf{P}(F) + \mathsf{P}(T|M)\mathsf{P}(M)} \\
&= \frac{\frac{1}{100} \times \frac{3}{5}}{\frac{1}{100} \times \frac{3}{5} + \frac{4}{100} \times \frac{2}{5}} \\
&= \frac{3}{11}
\end{aligned}$$

An engineering company advertises a job in three newspapers, A, B and C. It is known that these papers attract undergraduate engineering readerships in the proportions 2:3:1. The probabilities that an engineering undergraduate sees and replies to the job advertisement in these papers are 0.002, 0.001 and 0.005 respectively. Assume that the undergraduate sees only one job advertisement.

(a) If the engineering company receives only one reply to it advertisements, calculate the probability that the applicant has seen the job advertised in place A.
(i) A, (ii) B, (iii) C.

(b) If the company receives two replies, what is the probability that both applicants saw the job advertised in paper A?

$A = \{\text{Person is a reader of paper } A\}$,

$B = \{\text{Person is a reader of paper } B\}$,

$C = \{\text{Person is a reader of paper } C\}$,

$R = \{\text{Reader applies for the job}\}$.

We have the probabilities

(a)

$$
\begin{aligned}
\mathsf{P}(A) &= 1/3 \qquad \mathsf{P}(R|A) = 0.002 \\
\mathsf{P}(B) &= 1/2 \qquad \mathsf{P}(R|B) = 0.001 \\
\mathsf{P}(C) &= 1/6 \qquad \mathsf{P}(R|C) = 0.005 \\
\mathsf{P}(A|R) &= \frac{\mathsf{P}(R|A)\mathsf{P}(A)}{\mathsf{P}(R|A)\mathsf{P}(A) + \mathsf{P}(R|B)\mathsf{P}(B) + \mathsf{P}(R|C)\mathsf{P}(C)} = \frac{1}{3}
\end{aligned}
$$

Similarly

$$
\mathsf{P}(B|R) = \frac{1}{4} \qquad \text{and} \qquad \mathsf{P}(C|R) = \frac{5}{12}
$$

(b) Now, assuming that the replies and readerships are independent

$$
\begin{aligned}
\mathsf{P}(\text{Both applicants read paper} A) &= \mathsf{P}(A|R) \times \mathsf{P}(A|R) \\
&= \frac{1}{3} \times \frac{1}{3} \\
&= \frac{1}{9}
\end{aligned}
$$

It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

$$A = \text{event that an email is detected as spam,}$$
$$B = \text{event that an email is spam,}$$
$$B^c = \text{event that an email is not spam.}$$

We know $P(B) = P(B^c) = .5, \qquad P(A \mid B) = 0.99, \qquad P(A \mid B^c) = 0.05.$
Hence by the Bayes's formula we have

$$\begin{aligned} P(B^c \mid A) &= \frac{P(A \mid B^c)P(B^c)}{P(A \mid B)P(B) + P(A|B^c)P(B^c)} \\ &= \frac{0.05 \times 0.5}{0.05 \times 0.5 + 0.99 \times 0.5} \\ &= \frac{5}{104}. \end{aligned}$$

In a study, physicians were asked what the odds of cancer would be in a woman who was initially thought to have a 1% risk of cancer but who ended up with a positive mammogram result (a mammogram accurately classifies about 80% of cancerous tumors and 90% of benign tumors.) 95 out of a hundred physicians estimated the probability of cancer to be about 75%. Do you agree?

$$+ = \text{ mammogram result is positive,}$$

$$B = \text{tumor is benign,}$$

$$M = \text{tumor is malignant.}$$

Note that $B^c = M$. We are given $P(M) = .01$, so $P(B) = 1 - P(M) = .99$. We are also given the conditional probabilities $P(+ \mid M) = .80$ and $P(- \mid B) = .90$, where the event $-$ is the complement of $+$, thus $P(+ \mid B) = .10$

Bayes' formula in this case is

$$\begin{aligned} P(M \mid +) &= \frac{P(+ \mid M)P(M)}{(P(+ \mid M)P(M) + P(+ \mid B)P(B))} \\ &= \frac{0.80 \times 0.01}{(0.80 \times 0.01 + 0.10 \times 0.99)} \\ &\simeq 0.075 \end{aligned}$$

So the chance would be 7.5%.

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)
```

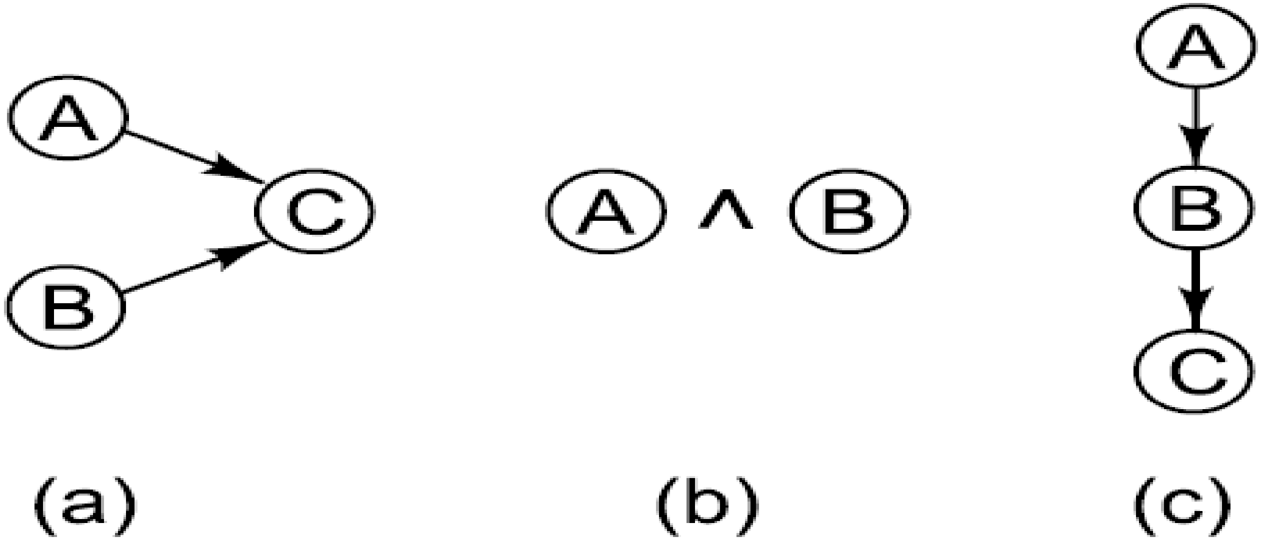
Measures of Belief

- $MB[h, e]$—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:

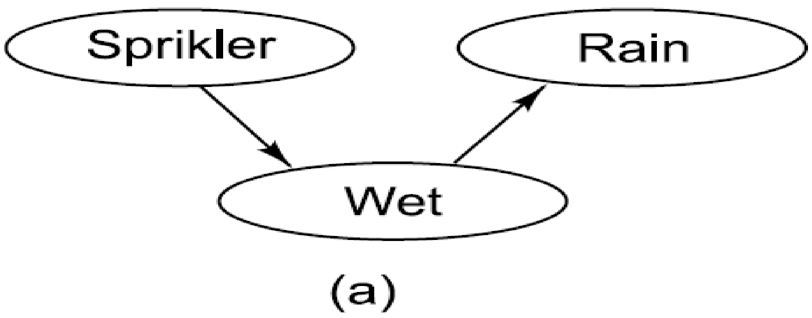
```
If:  the grass is wet this morning
then there is suggestive evidence (0.8) that
     it rained last night
```

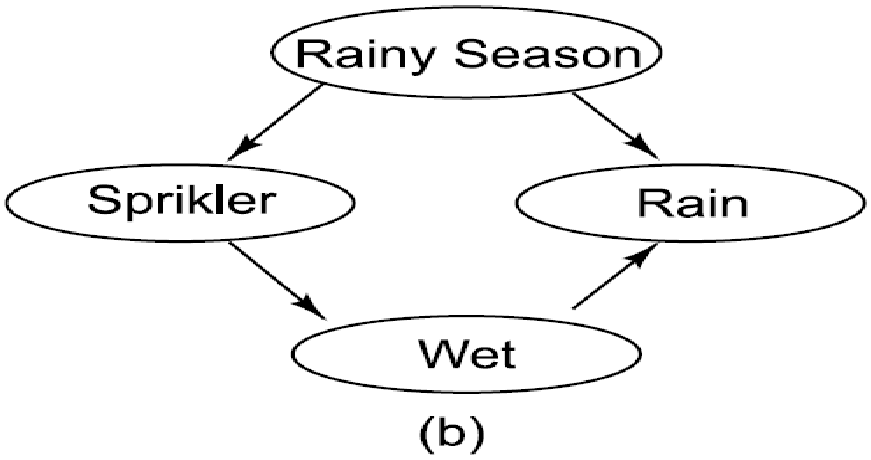
Taken alone, this rule makes sense when rain is the most common source of water on the grass. But if the two rules are applied together, using MYCIN's rule for chaining, we get

$MB[W,S] = 0.8$ {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 :

  $$\{\Theta\} \qquad (1.0)$$

- Suppose m_1 corresponds to our belief after observing fever.

  $$\begin{array}{ll} \{Flu, Cold, Pneu\} & (0.6) \\ \{\Theta\} & (0.4) \end{array}$$

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

  $$\begin{array}{ll} \{All,\ Flu, Cold\} & (0.8) \\ \Theta & (0.2) \end{array}$$

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 :

$\{Flu, Cold\}$	(0.48)
$\{All, Flu, Cold\}$	(0.32)
$\{Flu, Cold, Pneu\}$	(0.12)
Θ	(0.08)

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

$\{All\}$	(0.9)
Θ	(0.1)

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)

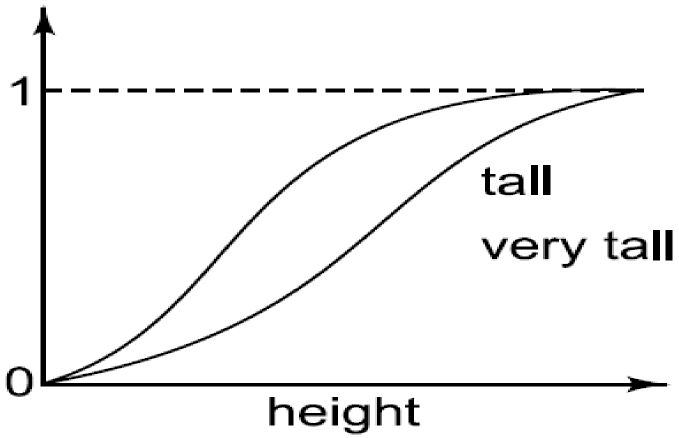
- Normalizing to get rid of the belief of 0.54 associated with ϕ gives m_5 :

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

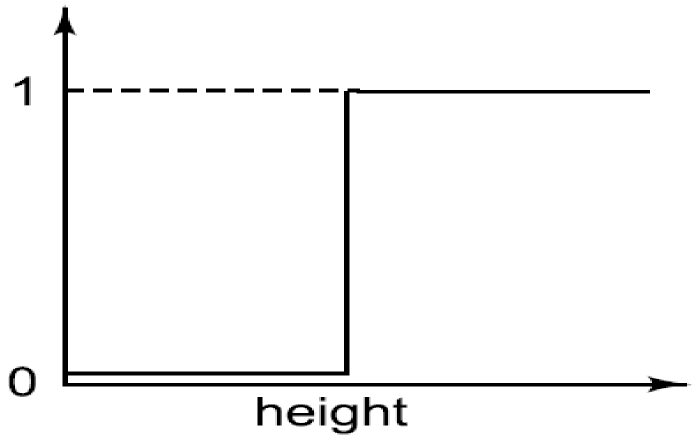
Fuzzy Logic

- **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