26:010:680
Current Topics in Accounting Research

Dr. Peter R. Gillett
Associate Professor
Department of Accounting and Information Systems
Rutgers Business School–Newark and New Brunswick
OVERVIEW

- Rule-Based Expert Systems
- Belief Functions
- Assignments for Week 6
- Homework 4
Assignments for Week 6

Assignments for Week 7


Representing Uncertainty

- Conditional probability
  \[ P(d \mid s) = \frac{P(d \& s)}{P(s)} \]

- Bayes’ Rule
  \[ P(d \mid s) = \frac{P(s \mid d) \cdot P(d)}{P(s)} \]
Representing Uncertainty

- Criticisms of relevance and applicability of *objective* probabilities (based on long-run frequencies)

- Consideration of *subjective* probabilities
  * Bayesian updating important here
  * Subjective probabilities must exhibit
    • Coherence: avoid certain loss
    • Total Evidence: use all the evidence
    • Conditionalization: update using Bayes’ Rule
  * In practice bounded rationality makes this difficult
Representing Uncertainty

- The more general form of Bayes’ rule

\[
P(d | s_1 & \ldots & s_k) = \frac{P(s_1 & \ldots & s_k | d) \cdot P(d)}{P(s_1 & \ldots & s_k)}
\]

requires computation of \((mn)^k + m + n^k\) probabilities (for \(m\) diseases and \(n\) symptoms)

- Tractability requires independence assumptions
Probability theory thus leaves us with a trade-off

* assume data are independent
  • fewer numbers
  • simpler calculations
  • sacrifice accuracy

* track dependencies
  • pay computational price
Representing Uncertainty

- Kahneman & Tversky etc.
  * Humans are poor Bayesian reasoners
  * Discount prior odds
  * Recency effects
  * Over-confident in judgments
  * Poor understanding of sampling theory

- N.B. Constructive probabilities
Certainty Factors

- Designed originally for use in MYCIN
- CF: \{propositions\} $\rightarrow [-1, +1]$
  - $CF(X) = 1$  X is certainly true
  - $CF(X) = -1$  X is certainly false
  - $CF(X) = 0$  X is entirely unknown

- Generally:
  $CF(\text{action}) = CF(\text{rule}) \times CF(\text{Premise})$
Certainty Factors

- As applied in MYCIN
  - IF patient has symptoms $s_1 \& \ldots \& s_k$ and background conditions $t_1 \& \ldots \& t_m$
  - THEN conclude patient has disease $d_i$ with certainty $\tau$

- Background knowledge constrains application of the rules
- Buchanan & Shortcliffe argue that rigorous application of Bayes’ rule would not be more accurate because conditional probabilities are subjective
- They intend CFs and their associated manipulations as approximations of probabilistic reasoning
Certainty Factors

- Computation of certainty factors is modular (Pearl)
  * i.e., we don’t need to consider information not contained in the rule
  * conditional probabilities are not modular in this sense
  * thus, when A is true, we cannot conclude \( P(B) = \tau \) from \( P(B|A) = \tau \) unless A is all that we know
  * otherwise, if we acquire additional knowledge \( E \), we may need to consider \( P(B|A,E) \)
Certainty Factors

- In order to combine support provided by two different rules, Shortcliffe & Buchanan looked for a method that was
  * commutative
    - independent of order of firing
  * asymptotic
    - certainty arises only from an absolute proof

- Note also the argument in S & B (1975) that imperfect evidence in favor of a hypothesis is not to be construed as evidence against it
Certainty Factors

- This is expressed rather more formally:
  \[ C[h,e] \neq 1 - C[\neg h,e] \]
  confirmation is not 1 - disconfirmation

- This is an idea we will re-visit e.g. when we consider Dempster-Shafer Belief Functions and their potential application in auditing
Certainty Factors

- Measure of Belief
  * the measure of increased belief in the hypothesis $h$, based on the evidence $e$, is $x$
  $$MB[h,e] = x$$

- Measure of Disbelief
  * the measure of increased disbelief in the hypothesis $h$, based on the evidence $e$, is $y$
  $$MD[h,e] = y$$
Certainty Factors

- **Formal definitions in terms of probability**

\[
MB[h,e] = \begin{cases} 
1 & \text{if } P(h) = 1 \\
\frac{\max\left[P(h), P(h|e)\right] - P(h)}{\max[1,0] - P(h)} & \text{otherwise}
\end{cases}
\]

\[
MD[h,e] = \begin{cases} 
1 & \text{if } P(h) = 0 \\
\frac{\min\left[P(h), P(h|e)\right] - P(h)}{\min[1,0] - P(h)} & \text{otherwise}
\end{cases}
\]

\[
\]
Certainty Factors

Characteristics
\[ 0 \leq \text{MB}[h,e] \leq 1, \quad 0 \leq \text{MD}[h,e] \leq 1, \quad -1 \leq \text{CF}[h,e] \leq 1 \]

If \( P[h|e] = 1 \)
\[ \text{MB}[h,e] = 1, \quad \text{MD}[h,e] = 0, \quad \text{CF}[h,e] = 1 \]

If \( P[\neg h|e] = 1 \)
\[ \text{MB}[h,e] = 0, \quad \text{MD}[h,e] = 1, \quad \text{CF}[h,e] = -1 \]

\( \text{MB}[h,e] = 0 \) if \( h \) is not confirmed by \( e \)
\( \text{MD}[h,e] = 0 \) if \( h \) is not disconfirmed by \( e \)
\( \text{CF}[h,e] = 0 \) if \( h \) is neither confirmed nor disconfirmed by \( e \)
Certainty Factors

- CF as defined here has the desired property
  * confirmation is not 1 - disconfirmation

- In fact
  * confirmation + disconfirmation = 0

- CF judgments must be elicited carefully from experts to ensure that they respect the constraints implied by these formal definitions
Certainty Factors

- Defining criteria
  - **Limits**
    \[ \text{MB}[h,e+] \rightarrow 1, \quad \text{MD}[h,e-] \rightarrow 1, \]
    \[ \text{CF}[h,e-] \leq \text{CF}[h,e- & e+] \leq \text{CF}[h,e+] \]
  - **Absolutes**
    \[ \text{MB}[h,e+] = 1 \Rightarrow \text{MD}[h,e-] = 0 \]
    \[ \text{MD}[h,e-] = 1 \Rightarrow \text{MB}[h,e+] = 0 \]
    \[ \text{MB}[h,e-] = \text{MD}[h,e-] \text{ is undefined} \]
Certainty Factors

- Defining criteria
  * Commutativity
    \[ MB[h, s_1 \& s_2] = MB[h, s_2 \& s_1] \]
    \[ MD[h, s_1 \& s_2] = MD[h, s_2 \& s_1] \]
    \[ CF[h, s_1 \& s_2] = CF[h, s_2 \& s_1] \]
  * Missing information
    \[ MB[h, s_1 \& s?] = MB[h, s_1] \]
    \[ MD[h, s_1 \& s?] = MD[h, s_1] \]
    \[ CF[h, s_1 \& s?] = CF[h, s_1] \]
Certainty Factors

- Combining functions
  * Incrementally acquired evidence

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

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

- Combining functions
  * Conjunctions of hypotheses
    \[ MB[h_1 \& h_2, e] = \min(MB[h_1, e], MB[h_2, e]) \]
    \[ MD[h_1 \& h_2, e] = \max(MD[h_1, e], MD[h_2, e]) \]
  * Disjunctions of hypotheses
    \[ MB[h_1 \lor h_2, e] = \max(MB[h_1, e], MB[h_2, e]) \]
    \[ MD[h_1 \lor h_2, e] = \min(MD[h_1, e], MD[h_2, e]) \]
Certainty Factors

Strength of evidence

* Suppose evidence $s_1$ is not known with certainty, but a CF based on prior evidence $e$ is known. If $MB'$ and $MD'$ are the degrees of belief and disbelief when $s_1$ is known with certainty, then the actual degrees of belief and disbelief are given by

$$MB[h, s_1] = MB'[h, s_1] \cdot \max(0, CF[h, s_1])$$

$$MD[h, s_1] = MD'[h, s_1] \cdot \max(0, CF[h, s_1])$$
Certainty Factors

- Note that in S & B (1975) MYCIN computes and maintains MBs and MDs separately, only computing CFs at the end, although CFs are then used to generate recommendations.

- This differs from “simplified” explanation in textbooks; e.g., in Durkin Chapter 12.
Certainty Factors

- In accordance with the limiting properties, multiple items of confirming evidence will result in MB --> 1 (say, 0.99)
- Suppose, however, we have a single item of disconfirming evidence with MD = 0.8
- Then CF = MB - MD = 0.19, i.e., many sources of confirmation have been almost completely offset by a single disconfirming item
Certainty Factors

To de-sensitize this effect, the definition of CF was subsequently modified to

$$\text{CF}[h,e] = \frac{\text{MB}[h,e] - \text{MD}[h,e]}{1 - \min[\text{MB}[h,e], \text{MD}[h,e]]}$$
Certainty Factors

- If we are only interested in updating CFs without retaining MBs and MDs, we can perform incremental updating using

\[
CF_{\text{COMBINE}} = \begin{cases} 
CF_1 + CF_2 \cdot (1 - CF_1) & \text{if both } > 0 \\
CF_1 + CF_2 \cdot (1 + CF_1) & \text{if both } < 0 \\
\frac{CF_1 + CF_2}{1 - \min(|CF_1|, |CF_2|)} & \text{otherwise}
\end{cases}
\]
Certainty Factors

- CFs may be used
  * to direct a best-first search
  * to control search explicitly
  * to prune the search
    - e.g., to drop goals when their CFs fall within the range [-0.2, +0.2]
  * to rank order hypotheses
Certainty Factors

- Durkin recommends
  * Obtain CFs from expert’s use of qualified terms
  * Don’t elicit CFs directly
  * Avoid deep inference chains (because approximate departs increasingly from probabilistic values)
  * Avoid many rules with the same hypothesis
  * Avoid rules with many premises - split into multiple rules
Adam (1976) criticized certainty factors

* CF associated with a hypothesis by MYCIN does not correspond to a simple probability model based on Bayes’ rule
  - did S & B (1975) claim that it did?

* Degrees of belief from different evidence cannot always be chosen independently
  - e.g., absolute diagnostic indicators

* min and max are not always appropriate for conjunctions
  - e.g., mutually exclusive alternatives
Certainty Factors

* CF ranking may reverse probability ranking

• Suppose

\[ P(h_1) = 0.8 \quad P(h_2) = 0.2 \]

\[ P(h_1 | e) = 0.9 \quad P(h_2 | e) = 0.8 \]

• Note

\[ P(h_1 | e) = 0.9 > P(h_2 | e) = 0.8 \]

• But

\[ CF(h_1, e) = \frac{P(h_1 | e) - P(h_1)}{1 - P(h_1)} = \frac{0.9 - 0.8}{0.2} = 0.5 \]

\[ CF(h_2, e) = \frac{P(h_2 | e) - P(h_2)}{1 - P(h_2)} = \frac{0.8 - 0.2}{0.8} = 0.75 \]

• Hence

\[ CF(h_1, e) < CF(h_2, e) \]
Certainty Factors

* Transitivity across chains of reasoning is not generally valid
* CFs are defined from MBs and MDs in terms of *increases* or *decreases* in belief, but elicited for MYCIN as *absolute values*
Certainty Factors

- Heckerman (1986)
  * Provides an example to show that the S & B (1975) definition of CFs, in conjunction with the rules for combining (incremental updating), lead to non-commutativity.
  * His conclusion from this is that we should take desirable properties of CFs as axiomatic, retain the combination rules, and seek an alternative formulation of CFs in probabilistic terms.
### Certainty Factors

- **Heckerman (1986)**
  - Axiomatizes the “desiderata” for certainty factors using a somewhat modified (simplified) notation, but formally conditioning on prior evidence,
  - Exhibits an example of non-commutativity
  - States a formal requirement for a probabilistic interpretation of CFs
  - Gives the odds-likelihood form of Bayes’ Theorem

\[
O(h|e,e_p) = \frac{P(e|h,e_p)}{P(e|\neg h,e_p)} \cdot O(h|e_p) = \lambda(h,e,e_p) \cdot O(h|e_p)
\]
Certainty Factors

- **Heckerman (1986)**
  * Defines conditional independence of $e$ and $e_p$ given $H$ and $\neg H$
  * Shows that $\lambda$ is a candidate for a probabilistic interpretation of CFs except that it ranges from 0 to $\infty$
  * Shows that any monotonic increasing transformation of the likelihood ratio satisfying
    \[
    F\left(\frac{1}{x}\right) = -F(x) \quad \text{and} \quad F(\infty) = 1
    \]
    is a probabilistic interpretation for CFs (and conversely)
Certainty Factors

- Heckerman (1986)
  * Gives specific examples of such transformations
  * Observes that evidence combined using the S & B combination functions is required to be conditionally independent given both the hypothesis and its negation
  * Argues by example that the latter condition often fails in practice
  * Introduces axioms for sequential combination (corresponding to strength of evidence in S & B)
Certainty Factors

- Heckerman (1986)
  * Shows that these new axioms do not further constrain probabilistic interpretations of CFs
  * Demonstrates that although CFs have been applied to non-tree inference networks, updating is valid only in tree structures (rarely applicable in complex practical situations)
Certainty Factors

- Rules may conveniently be organized as an inference net, e.g.,
Certainty Factors

- **Rules:**
  - * R1: A v B --> C \(\text{CF} = 0.8\)
  - * R2: D --> E \(\text{CF} = 0.7\)
  - * R3: C & E --> F \(\text{CF} = 0.9\)

- **Facts**
  - * A \(\text{CF} = 0.4\)
  - * B \(\text{CF} = 0.6\)
  - * D \(\text{CF} = 0.9\)
  - * C,E \(\text{CF} = 0\)
  - * F \(\text{CF} = 0.2\)
Certainty Factors

- \( CF(A \lor B) = \max(0.4, 0.6) = 0.6 \)
- \( CF(R1') = 0.8 \times 0.6 = 0.48 \)
- \( CF(C|A \lor B) = 0 + 0.48 \times (1 - 0) = 0.48 \)
- \( CF(R2') = 0.7 \times 0.9 = 0.63 \)
- \( CF(E|D) = 0 + 0.63 \times (1 - 0) = 0.63 \)
- \( CF (C \land E) = \min (0.48, 0.63) = 0.48 \)
- \( CF(R3') = 0.9 \times 0.48 = 0.432 \)
- \( CF(F|C \land E) = 0.2 + 0.432 \times (1 - 0.2) = 0.5456 \)
We have already seen

\[ O(h|e) = \frac{P(e|h)}{P(e|\neg h)} \cdot O(h) = \lambda(h,e) \cdot O(h) \]

Now, defining the Likelihood of Sufficiency by

\[ LS = \frac{P(e|h)}{P(e|\neg h)} \text{ we can write } O(h|e) = LS \cdot O(h) \]
Similarly, if we define the Likelihood of Necessity by

$$\text{LN} = \frac{P(\neg e|h)}{P(\neg e|\neg h)}$$

we can write

$$O(h|\neg e) = \text{LN} \cdot O(h)$$

This enables us to develop rules of the form:

**IF** \( e \) **THEN** \( h \) (\text{LS, LN})

with both factors provided by an expert.
Mathematically, we have the constraints

\[ \begin{align*}
    \text{LS} > 1 & \implies \text{LN} < 1 \\
    \text{LS} < 1 & \implies \text{LN} > 1 \\
    \text{LS} = 1 & \implies \text{LN} = 1
\end{align*} \]

but real-world problems may contradict this

More generally, if we are uncertain of \( e \) itself, and it depends on observed evidence \( e' \), we can make adjustments.
The probability of $h$ given our belief $e'$ is

$$P(h|e') = P(h|e) \cdot P(e'|e) + P(h|\neg e) \cdot P(\neg e|e')$$

from which the following derive

$$P(e'|e) = P(e) \Rightarrow P(h|e') = P(h)$$

$e$ true $\Rightarrow P(e|e') = 1$ and $P(h|e') = P(h|e)$

$e$ false $\Rightarrow P(\neg e|e') = 1$ and $P(h|e') = P(h|\neg e)$

which in turn define a linear relationship between $P(h|e')$ and $P(e|e')$
Real-world situations may result in experts providing values that contradict these assumptions, and some adjustment therefore needs to be made.

Duda et al. proposed an *ad hoc* assumption to relate $P(h|e')$ and $P(e|e')$ following a piecewise linear function.

This lead to PROSPECTOR.
PROSPECTOR

- PROSPECTOR uses two simple functions to avoid inconsistencies:
  
  \[ P(h|e') = P(h|\neg e) + \frac{P(e'|e)}{P(e)} \cdot (P(h) - P(h|\neg e)) \quad \text{for } 0 \leq P(e'|e) \leq P(e) \]

  \[ P(h|e') = \frac{P(h) - P(h|e) \cdot P(e)}{1 - P(e)} + P(e'|e) \cdot \frac{P(h|e) - P(h)}{1 - P(e)} \quad \text{for } P(e) \leq P(e'|e) \leq 1 \]

- PROSPECTOR is an expert system that assists geologists in mineral deposit exploration
A PROSPECTOR network is a set of nodes representing evidence or hypotheses and links connecting the nodes together with uncertain relationships represented by $LS$ or $LN$ values and prior probabilities for the nodes.

Probabilities are propagated upward to the topmost node.
Where multiple nodes affect a single hypothesis, conditional independence is assumed, and rules combine conjunctively or disjunctively

* **Conjunctive rules**
  - each $e_i$ is based on the partial evidence $e_i'$
  - PROSPECTOR assumes $P(e|e') = \min \{ P(e_i|e') \}$
  - the resulting value is combined using the linear function given above

* **Disjunctive rules**
  - as above, but using max instead of min
PROSPECTOR

- Updating odds

* Each time new evidence is provided, the odds are updated, assuming conditional independence

\[
O(h|e_1', e_2', ..., e_n') = \prod_{i=1}^{i=n} \text{LS}_i \cdot O(h) \quad \text{where} \quad \text{LS}_i' = \frac{P(e_i|h)}{P(e_i|\neg h)}
\]

\[
O(h|\neg e_1', \neg e_2', ..., \neg e_n') = \prod_{i=1}^{i=n} \text{LN}_i \cdot O(h) \quad \text{where} \quad \text{LN}_i' = \frac{P(\neg e_i|h)}{P(\neg e_i|\neg h)}
\]
Beliefs were elicited from users of PROSPECTOR using certainty measures, which were subsequently converted to conditional probabilities using the same piecewise linear approach outlined earlier.
Using probabilities directly is a powerful but challenging technique

* Probabilities must be known
* Probabilities must be updated
* Total probability must equal unity
* Conditional independence is required
PROSPECTOR

- PROSPECTOR incorporates many simplifying assumptions, but it is still a demanding system.
- A large number of probabilities are still typically required to be provided:
  * difficult to obtain
  * computationally expensive
- Need to restart when new hypotheses are added: there is no incremental updating.
- Such a system is called *intensional* or *global* - by contrast, MYCIN is *extensional* and has a *modular* structure.
Prospector

- Other concerns about the updating methods
  - Rednault et al. (1981)
    - If A and B are intersections of the evidence $e_1 \ldots e_m$, then they are independent
  - Hussain (1972) sought to show
    - for exhaustive and mutually exclusive hypotheses $h_1 \ldots h_n$ and $e_1 \ldots e_m$ conditionally independent, no updating is possible
  - Glymour (1985)
    - gave a counter-example to disprove this
  - Johnson (1986)
    - showed that multiple updating of any hypothesis is impossible; i.e., there is at most one piece of evidence for which posteriors not the same as the prior
Belief Functions

- The standard text for definitions, etc. is, of course:

Belief Functions

A belief function on a frame $\Theta$ is a function $\text{Bel}: 2^\Theta \rightarrow [0, 1]$ such that:

1. $\text{Bel}(\emptyset) = 0$
2. $\text{Bel}(\Theta) = 1$
3. $\text{Bel}(A_1 \cup \ldots \cup A_n) \geq \sum_i \text{Bel}(A_i) - \sum_{i<j} \text{Bel}(A_i \cap A_j) + \ldots + (-1)^{n+1} \text{Bel}(A_1 \cap \ldots \cap A_n)$

Plausibility is defined by $\text{Pl}(A) = 1 - \text{Bel}(\sim A)$
Belief Functions

Basic probability assignments are functions $m : 2^\Theta \rightarrow [0, 1]$ such that:

1. $m(\emptyset) = 0$
2. $\sum_{A \subseteq \Theta} m(A) = 1$

Then we may define $\text{Bel}(A) = \sum_{B \subseteq A} m(B)$
Belief Functions

Example:

* Consider a frame with three possible outcomes \( \{a, b, c\} \)

* Suppose we are given the following basic probability assignment:

\[
\begin{align*}
    m(\{a\}) &= .1; \\
    m(\{b\}) &= .1; \\
    m(\{c\}) &= .1; \\
    m(\{a, b\}) &= .1; \\
    m(\{a, c\}) &= .2; \\
    m(\{b, c\}) &= .3; \\
    m(\{a, b, c\}) &= .1
\end{align*}
\]
## Belief Functions

<table>
<thead>
<tr>
<th>Bpa</th>
<th>Bel</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>0</td>
</tr>
<tr>
<td>${a}$</td>
<td>0.1</td>
</tr>
<tr>
<td>${b}$</td>
<td>0.1</td>
</tr>
<tr>
<td>${c}$</td>
<td>0.1</td>
</tr>
<tr>
<td>${a,b}$</td>
<td>0.1</td>
</tr>
<tr>
<td>${a,c}$</td>
<td>0.2</td>
</tr>
<tr>
<td>${b,c}$</td>
<td>0.3</td>
</tr>
<tr>
<td>${a,b,c}$</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Current Topics in Accounting Research

February 23, 2011  Dr. Peter R. Gillett

58
## Belief Functions

<table>
<thead>
<tr>
<th>Bpa</th>
<th>Bel</th>
<th>Pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>∅</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>{a}</td>
<td>.1</td>
<td>.1</td>
</tr>
<tr>
<td>{b}</td>
<td>.1</td>
<td>.1</td>
</tr>
<tr>
<td>{c}</td>
<td>.1</td>
<td>.1</td>
</tr>
<tr>
<td>{a,b}</td>
<td>.1</td>
<td>.3</td>
</tr>
<tr>
<td>{a,c}</td>
<td>.2</td>
<td>.4</td>
</tr>
<tr>
<td>{b,c}</td>
<td>.3</td>
<td>.5</td>
</tr>
<tr>
<td>{a,b,c}</td>
<td>.1</td>
<td>1</td>
</tr>
</tbody>
</table>
Belief Functions

- Bpas may be recovered from Bel functions using

\[ m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \text{Bel}(B) \]
Belief Functions

- The *commonality* function is a function
  \[ Q : 2^\Theta \to [0, 1] \]
  defined by
  \[ Q(A) = \sum_{A \subseteq B} m(B) \]

- Bpas may be recovered from commonality functions using
  \[ m(A) = \sum_{A \subseteq B} (-1)^{|B - A|} Q(B) \]
# Belief Functions

<table>
<thead>
<tr>
<th>Bpa</th>
<th>Bel</th>
<th>Pl</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>∅</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>{a}</td>
<td>.1</td>
<td>.1</td>
<td>.5</td>
</tr>
<tr>
<td>{b}</td>
<td>.1</td>
<td>.1</td>
<td>.6</td>
</tr>
<tr>
<td>{c}</td>
<td>.1</td>
<td>.1</td>
<td>.7</td>
</tr>
<tr>
<td>{a,b}</td>
<td>.1</td>
<td>.3</td>
<td>.9</td>
</tr>
<tr>
<td>{a,c}</td>
<td>.2</td>
<td>.4</td>
<td>.9</td>
</tr>
<tr>
<td>{b,c}</td>
<td>.3</td>
<td>.5</td>
<td>.9</td>
</tr>
<tr>
<td>{a,b,c}</td>
<td>.1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Belief Functions

- Recall that the bpa function can be uniquely recovered from Pl, Bel or Q
- In fact, we can convert any one of the four representations uniquely into any of the others
- These conversions are examples of Möbius transforms
- There are Fast Möbius Transforms to do this efficiently (see Kennes)
Belief Functions

bpa \rightarrow Bel
Q \rightarrow Pl

bpa \leftarrow Bel
Q \leftarrow Pl
Belief Functions

- In expert systems based on belief functions:
  * user inputs are often in the form of bpas
  * propagation is most efficient implemented via commonalities
  * marginalization is most efficient implemented via Bel functions
  * output is often desired as Bel or Pl functions
Combining Belief Functions

- Dempster’s Rule
  * Consider two belief functions given by their bpas as follows:

  \[
  m_1(\{a\}) = .5; m_1(\{\sim a\}) = .3; m_1(\{a, \sim a\}) = .2;
  \]

  \[
  m_2(\{a\}) = .7; m_1(\{\sim a\}) = .2; m_1(\{a, \sim a\}) = .1
  \]
### Combining Belief Functions

<table>
<thead>
<tr>
<th></th>
<th>( m_1 )</th>
<th>( m_2 )</th>
<th>( m_1 \otimes m_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( { a } )</td>
<td>0.5</td>
<td>0.7</td>
<td>( 0.7 \times 0.5 = 0.35 )</td>
</tr>
<tr>
<td>( { \sim a } )</td>
<td>0.3</td>
<td>0.2</td>
<td>( 0.7 \times 0.3 = 0.21 )</td>
</tr>
<tr>
<td>( { a, \sim a } )</td>
<td>0.2</td>
<td>0.2</td>
<td>( 0.7 \times 0.2 = 0.14 )</td>
</tr>
</tbody>
</table>

\[
\begin{array}{ccc}
\{ a \} & 0.7 & 0.7 \times 0.5 = 0.35 \\
\{ \sim a \} & 0.2 & 0.2 \times 0.5 = 0.10 \\
\{ a, \sim a \} & 0.1 & 0.1 \times 0.5 = 0.05 \\
\end{array}
\]

\[
m_1 \otimes m_2 \left( \{ a \} \right) = \frac{0.35 + 0.14 + 0.05}{1 - (0.21 + 0.10)} = 0.783
\]

\[
m_1 \otimes m_2 \left( \{ \sim a \} \right) = \frac{0.06 + 0.04 + 0.03}{1 - (0.21 + 0.10)} = 0.188
\]

\[
m_1 \otimes m_2 \left( \{ a, \sim a \} \right) = \frac{0.02}{1 - (0.21 + 0.10)} = 0.029
\]
Note, however, the following:

<table>
<thead>
<tr>
<th></th>
<th>m₁</th>
<th>Q₁</th>
<th>m₂</th>
<th>Q₂</th>
<th>Q₁xQ₂</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>{a}</td>
<td>.5</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.56</td>
<td>.54</td>
</tr>
<tr>
<td>{~a}</td>
<td>.3</td>
<td>.5</td>
<td>.2</td>
<td>.3</td>
<td>.15</td>
<td>.13</td>
</tr>
<tr>
<td>{a,~a}</td>
<td>.2</td>
<td>.2</td>
<td>.1</td>
<td>.1</td>
<td>.02</td>
<td>.02</td>
</tr>
</tbody>
</table>

After normalization, these are the same values as derived from Dempster’s Rule.
Combining Belief Functions

- In expert system applications, therefore, it is efficient to:
  - use Fast Möbius Transforms to convert bpas to commonalities
  - combine the commonalities by pointwise multiplication
  - (eventually) use Fast Möbius Transforms to convert the results back to bpas or other desired outputs
Types of Belief Functions

- If $A$ is a subset of the frame $\Theta$ of a belief function, then $A$ is a focal element if $m(A) > 0$
- The core of a belief function is the union of all its focal elements
- If, for some subset $A$, $m(A) = s$ and $m(\Theta) = 1 - s$ then $m$ is a simple support function
- Thus a simple support function has only one focal element other than the frame itself
Types of Belief Functions

- A belief function that is the combination of one or more simple support functions is called a *separable support function*.

- A belief function that results from marginalizing a separable support function may not itself be separable; it is called a *support function*; Shafer suggests these are fundamental for the representation of evidence.
Types of Belief Functions

- Simple support functions
  \[ \subset \]
- Separable support functions
  \[ \subset \]
- Support functions
  \[ \subset \]
- Belief functions

- A belief function whose focal elements are nested is called a consonant belief function
Types of Belief Functions

- A belief function that is not a support function is called a quasi support function.
- Quasi support functions arise as the limits of sequences of support functions.
- A belief function for which \( \text{Bel}(A \cup B) = \text{Bel}(A) + \text{Bel}(B) \) whenever \( A \cap B = \emptyset \) is called a Bayesian belief function.
- Equivalently, a Bayesian belief function is a belief function all of whose focal elements are singletons.
- Bayesian belief functions are quasi support functions (except when \( \text{Bel}(\{\theta\}) = 1 \) for some \( \theta \in \Theta \).)
Belief Functions in Expert Systems

- Belief functions can be propagated locally in Join Trees (Markov Trees) using the Shenoy-Shafer algorithm (coming soon to a class near you . . .)
- Belief functions can also be propagated locally in Junction Trees using the Aalborg architecture; this requires division (of commonalities) and intermediate results may not be interpretable
- In practice, it is most efficient to perform combination using commonalities and marginalization using Bels
Belief Functions in Expert Systems

- Xu and Kennes give efficient algorithms for carrying out belief function combination, for bit-array representations of subsets, and for Fast Möbius Transforms.
- The bit-array representation includes algorithms for testing subsets, forming intersections, unions, etc directly with the bit-arrays.
- Full details of the Fast Möbius Transform algorithms are given in Kennes.
Belief Functions in Expert Systems

- Efficient implementations are especially important for belief functions
  - $n$ binary variables generate a joint space with $2^n$ configurations in probability systems
  - $n$ binary variables generate a joint space with $2^{2^n}$ potential focal elements in belief function systems
Belief Functions in Expert Systems

- The Shafer & Srivastava paper we read for today sets out extensive arguments why belief functions might be considered superior to probabilities for certain applications, such as auditing.
- Among these reasons, the one that first attracted me to study belief functions when I was building an Expert System (ADAPT) is the argument that they better represent ignorance.
- In auditing, for example, accounts receivable, insufficient replies from customers might lead us to assess a probability of, say, only 70% that accounts receivable exist.
- Probability theory then forces us to assess a 30% probability that they do not exist, despite the fact that there is no evidence they do not exist - merely insufficient evidence that they do.
Belief Functions in Expert Systems

- Belief functions allow us to assign a 70% bpa to existence, and the balance to the whole frame, representing ignorance.
- In probability theory there would be no difference if some of the missing customers in fact wrote to deny the existence of the balance.
- Using belief functions, however, we could assign some part of the bpa to represent contrary evidence, and the remainder to ignorance.
  - perhaps \( m(\text{exist}) = 0.7; m(\sim \text{exist}) = 0.2; m(\text{exist,} \sim \text{exist}) = 0.1 \)
- Of course, in belief function terms, complete ignorance is represented by \( m(\text{exist,} \sim \text{exist}) = 1 \): it must be one of the outcomes, we don’t know which, or which is more likely.
- Probabilistically, ignorance is represented as \( P(\text{exist}) = P(\sim \text{exist}) = 0.5 \) and we have to assume the outcomes equally likely.
Bayesian and Belief-Function Formalisms

- Bayesian formalism
  - *Objective probabilities*
    - Repeated trials make no sense in auditing
  - *Subjective probabilities*
    - Additivity argument
      - based on two-sided betting rates
    - Problem of small worlds
    - Problem of non-existent preferences
    - Betting rate argument for conditioning
  - *Constructive probabilities*
Bayesian and Belief-Function Formalisms

- Belief-function formalism
  - Degrees of belief
  - Independence
  - Compatibility relations
  - Dempster’s Rule
  - Constructive interpretation

- Comparison of two formalisms
  - Bayes as a special case
  - Representation of ignorance
Belief Function Formulas for Audit Risk

- Audit risk model
  * Inherent Risk
  * Analytical Procedures
  * Internal Controls
  * Tests of Details
  * Evidential Networks
- Review of Belief-function approach
- p. 259 – audit risk is Plausibility
Belief Function Formulas for Audit Risk

- For the evidential network illustrated in the paper
  - Section IV defines m-values at each node
  - Section V gives explicit formulae for audit risk
  - A small example is calculated

- Note the acknowledgements on the first page – and the date of the paper
  - Yes, I have checked ALL these formulae and symbols!!!

- Note that the illustration is an “AND” tree
Aggregating Evidence in Auditing

- Reviews problems with probability models
- Discusses structure of audit risk
- Develops a propagation scheme
  * Draw a network
  * Express the impact of the evidence as belief functions
  * Construct a Markov tree (now more commonly called a Join Tree)
  * Propagate the belief functions in the Markov Tree
Aggregating Evidence in Auditing

- Illustrates the propagation scheme using an example
  * Evidence 4 prevents the original network from having a tree structure
  * The relational node is still an “AND” node
  * Note that there is also an issue regarding determining the m-values for Evidence 4 from the marginals, and the method adopted in this paper is not a conservative choice

- More complex illustrations follow, using software
Homework 4

- Homework 4 is a set of simple exercises using your new-found knowledge of belief functions.
- The important thing is not the answers, but your workings and whether you understand the manipulations you are carrying out – so by all means get help, but do the work yourself.
- Post your solutions as a WORD document on Blackboard as usual.