Probabilistic Reasoning; Network-based reasoning

COMPSCI 276, Spring 2011

Set 1: Introduction and Background

Rina Dechter

(Reading: Pearl chapter 1-2, Darwiche chapters 1,3)

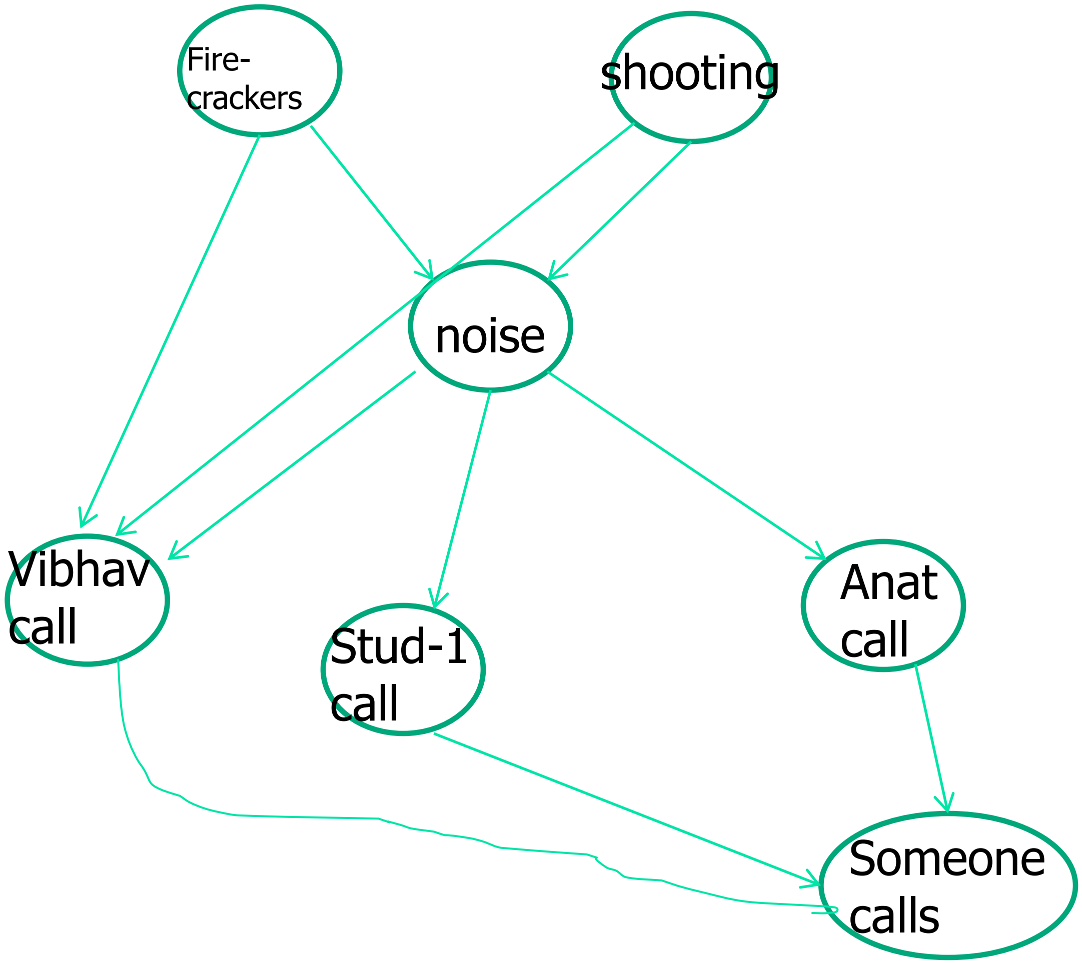
Class Description

- Instructor: Rina Dechter
- Days: Tuesday & Thursday
- Time: 11:00 - 12:20 pm
- Class page:
- http://www.ics.uci.edu/~dechter/courses/ics-275b/spring-11/

Example of common sense reasoning

- Explosive noise at UCI
- Parking in Cambridge
- The missing garage door
- Years to finish an undergrad degree in college

Shooting at UCI



Why uncertainty

- **Summary of exceptions**
  - Birds fly, smoke means fire (cannot enumerate all exceptions.
- **Why is it difficult?**
  - Exception combines in intricate ways
  - e.g., we cannot tell from formulas how exceptions to rules interact:

A→C
B→C

A and B -→ C

The problem

All men are mortal	T
All penguins are birds	T
…	
Socrates is a man	
Men are kind	p1
Birds fly	p2
T looks like a penguin	
Turn key –> car starts	P_n

True propositions

Uncertain propositions

Q: Does T fly? Logic?....but how we handle exceptions

P(Q)? Probability: astronomical

Managing Uncertainty

- Knowledge obtained from people is almost always loaded with uncertainty
- Most rules have exceptions which one cannot afford to enumerate
- Antecedent conditions are ambiguously defined or hard to satisfy precisely
- First-generation expert systems combined uncertainties according to simple and uniform principle
- Lead to unpredictable and counterintuitive results
- Early days: logicist, new-calculist, neo-probabilist

Extensional vs Intensional Approaches

- **Extensional** (e.g., Mycin, Shortliffe, 1976) certainty factors attached to rules and combine in different ways.

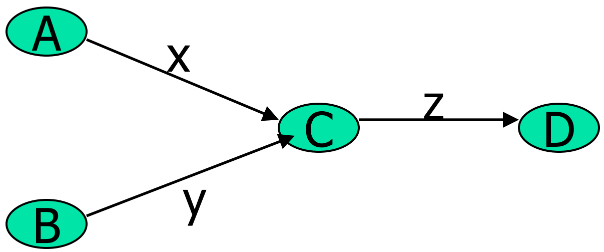
  $A \rightarrow B$: m

- **Intensional**, semantic-based, probabilities are attached to set of worlds.

  $P(A|B) = m$

Certainty combination in Mycin

If A then C (x)
If B then C (y)
If C then D (z)



1.Parallel Combination:
CF(C) = x+y-xy, if x,y>0
CF(C) = (x+y)/(1-min(x,y)), x,y have different sign
CF(C) = x+y+xy, if x,y<0
2. Series combination…
3.Conjunction, negation

Computational desire : locality, detachment, modularity

The limits of modularity

Deductive reasoning: modularity and detachment

$P \rightarrow Q$
P

Q

$P \rightarrow Q$
K and P

Q

$P \rightarrow Q$
$K \rightarrow P$
K

Q

Plausible Reasoning: violation of locality

Wet → rain
Wet
----
rain

wet → rain
Sprinkler and wet
----
rain?

Violation of detachment

Deductive reasoning

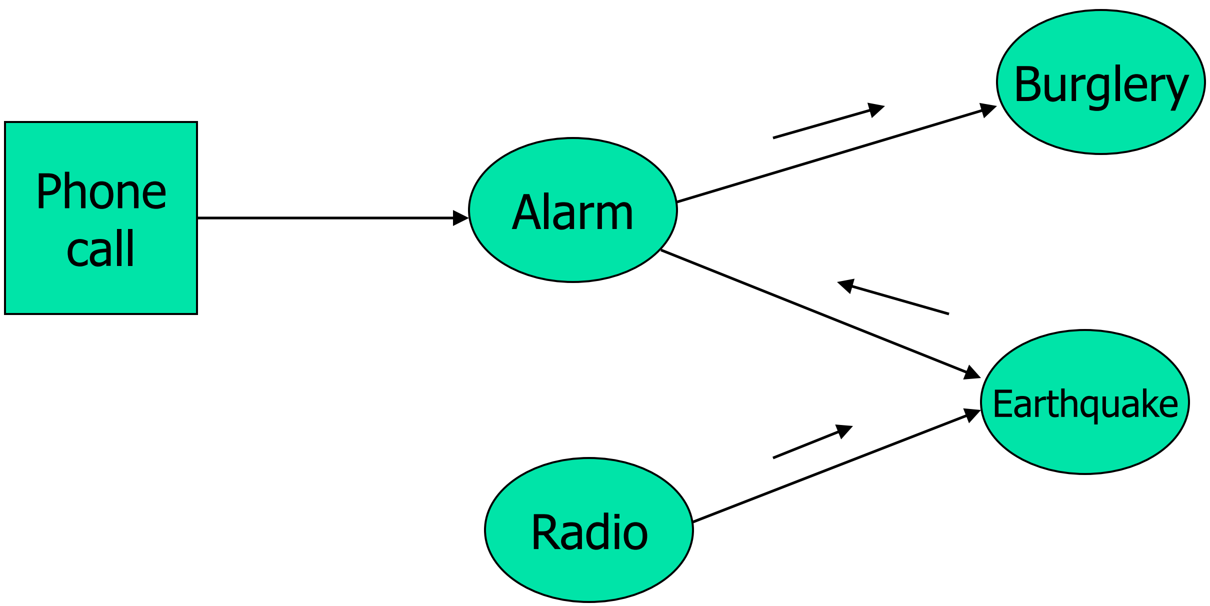
$P \rightarrow Q$
$K \rightarrow P$
K

Q

Plausible reasoning

Wet → rain
Sprinkler →wet
Sprinkler
----
rain?

Burglery Example



A→B
A more credible

B more credible

IF Alarm → Burglery
A more credible (after radio)
But B is less credible

Issue: Rule from effect to causes

Extensional vs Intensional

Extensional	Intensional
Uncertainty=truth value	Uncertainty = modality
Connectives combine certainty weight	Connectives combine set of worlds
Rules = Procedural license = summary of a problem solving history	Rules = constraints on the world = summary of world knowledge

What's in a rule?

A→B (m) C→B (n) **A and B→C (m+n-mn)**	P(B|A)= p A→B (p)
Semantic difficulties: Handling exceptions, Retracting conclusions Unidirectional references Incoherent updating	**Semantic clarity:** Syntax mirrors world knowledge Empirically testable parameters Bidirectional Inferences Coherent updating
Computational merit: Locality+detachment	**Computational difficulty:** Actions must wait verification of relevance

Probabilistic Modeling with Joint Distributions

Degrees of Belief

- Assign a degree of belief or probability in $[0,1]$ to each world ω and denote it by $\Pr(\omega)$.
- The belief in, or probability of, a sentence α:

$$\Pr(\alpha) \stackrel{\textit{def}}{=} \sum_{\omega \models \alpha} \Pr(\omega).$$

world	Earthquake	Burglary	Alarm	$\Pr(.)$
ω_1	true	true	true	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

Properties of Beliefs

- A bound on the belief in any sentence:

$$0 \leq \mathrm{Pr}(\alpha) \leq 1 \qquad \text{for any sentence } \alpha.$$

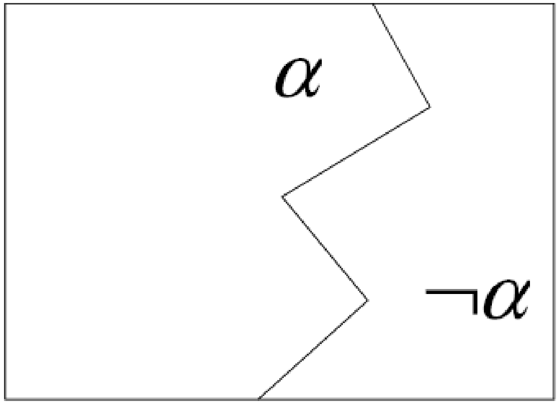
- A baseline for inconsistent sentences:

$$\mathrm{Pr}(\alpha) = 0 \qquad \text{when } \alpha \text{ is inconsistent.}$$

- A baseline for valid sentences:

$$\mathrm{Pr}(\alpha) = 1 \qquad \text{when } \alpha \text{ is valid.}$$

Properties of Beliefs



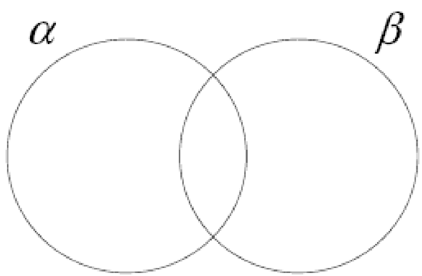
- The belief in a sentence given the belief in its negation:

$$\Pr(\alpha) + \Pr(\neg\alpha) = 1.$$

Example

$$\begin{aligned} \Pr(\text{Burglary}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_5) + \Pr(\omega_6) = .2 \\ \Pr(\neg\text{Burglary}) &= \Pr(\omega_3) + \Pr(\omega_4) + \Pr(\omega_7) + \Pr(\omega_8) = .8 \end{aligned}$$

Properties of Beliefs



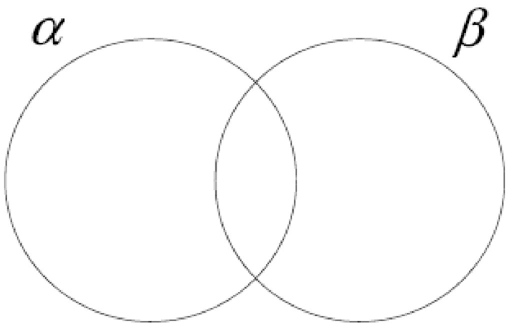
- The belief in a disjunction:

$$\Pr(\alpha \vee \beta) = \Pr(\alpha) + \Pr(\beta) - \Pr(\alpha \wedge \beta).$$

- Example:

$$\begin{aligned}
\Pr(\text{Earthquake}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4) = .1 \\
\Pr(\text{Burglary}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_5) + \Pr(\omega_6) = .2 \\
\Pr(\text{Earthquake} \wedge \text{Burglary}) &= \Pr(\omega_1) + \Pr(\omega_2) = .02 \\
\Pr(\text{Earthquake} \vee \text{Burglary}) &= .1 + .2 - .02 = .28
\end{aligned}$$

Properties of Beliefs



- The belief in a disjunction:

  $\Pr(\alpha \vee \beta) = \Pr(\alpha) + \Pr(\beta)$ when α and β are mutually exclusive.

Bayes Conditioning

Alpha and beta are events

Closed form for Bayes conditioning:

$$\Pr(\alpha|\beta) = \frac{\Pr(\alpha \wedge \beta)}{\Pr(\beta)}.$$

Defined only when $\Pr(\beta) \neq 0$.

Degrees of Belief

world	Earthquake	Burglary	Alarm	$\Pr(.)$
ω_1	true	true	true	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

$$
\begin{aligned}
\Pr(\text{Earthquake}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4) = .1 \\
\Pr(\text{Burglary}) &= .2 \\
\Pr(\neg\text{Burglary}) &= .8 \\
\Pr(\text{Alarm}) &= .2442
\end{aligned}
$$

Belief Change

Burglary is independent of Earthquake

Conditioning on evidence Earthquake:

$$\begin{aligned}
\Pr(\text{Burglary}) &= .2 \\
\Pr(\text{Burglary}|\text{Earthquake}) &= .2 \\
\\
\Pr(\text{Alarm}) &= .2442 \\
\Pr(\text{Alarm}|\text{Earthquake}) &\approx .75 \uparrow
\end{aligned}$$

The belief in Burglary is not changed, but the belief in Alarm increases.

Belief Change

Earthquake is independent of burglary

Conditioning on evidence Burglary:

$$\begin{aligned}
\Pr(\text{Alarm}) &= .2442 \\
\Pr(\text{Alarm}|\text{Burglary}) &\approx .905 \uparrow \\
\\
\Pr(\text{Earthquake}) &= .1 \\
\Pr(\text{Earthquake}|\text{Burglary}) &= .1
\end{aligned}$$

The belief in Alarm increases in this case, but the belief in Earthquake stays the same.

Belief Change

The belief in Burglary increases when accepting the evidence Alarm. How would such a belief change further upon obtaining more evidence?

- Confirming that an Earthquake took place:

$$
\begin{aligned}
\Pr(\text{Burglary}|\text{Alarm}) &\approx .741 \\
\Pr(\text{Burglary}|\text{Alarm} \wedge \text{Earthquake}) &\approx .253 \downarrow
\end{aligned}
$$

  We now have an explanation of Alarm.

- Confirming that there was no Earthquake:

$$
\begin{aligned}
\Pr(\text{Burglary}|\text{Alarm}) &\approx .741 \\
\Pr(\text{Burglary}|\text{Alarm} \wedge \neg\text{Earthquake}) &\approx .957 \uparrow
\end{aligned}
$$

  New evidence will further establish burglary as an explanation.

Conditional Independence

$\Pr$ finds α conditionally independent of β given γ iff

$$\Pr(\alpha|\beta \wedge \gamma) = \Pr(\alpha|\gamma) \quad \text{or } \Pr(\beta \wedge \gamma) = 0.$$

Another definition

$$\Pr(\alpha \wedge \beta|\gamma) = \Pr(\alpha|\gamma)\Pr(\beta|\gamma) \quad \text{or } \Pr(\gamma) = 0.$$

Variable Independence

$\Pr$ finds **X** independent of **Y** given **Z**, denoted $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, means that $\Pr$ finds **x** independent of **y** given **z** for all instantiations **x**, **y** and **z**.

Example

$\mathbf{X} = \{A, B\}$, $\mathbf{Y} = \{C\}$ and $\mathbf{Z} = \{D, E\}$, where A, B, C, D and E are all propositional variables. The statement $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ is then a compact notation for a number of statements about independence:

$A \wedge B$ is independent of C given $D \wedge E$;

$A \wedge \neg B$ is independent of C given $D \wedge E$;

$\vdots$

$\neg A \wedge \neg B$ is independent of $\neg C$ given $\neg D \wedge \neg E$;

That is, $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ is a compact notation for $4 \times 2 \times 4 = 32$ independence statements of the above form.

Further Properties of Beliefs

Chain rule

$$\begin{aligned}
&\Pr(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n) \\
&\quad = \Pr(\alpha_1 | \alpha_2 \wedge \ldots \wedge \alpha_n)\Pr(\alpha_2 | \alpha_3 \wedge \ldots \wedge \alpha_n)\ldots\Pr(\alpha_n).
\end{aligned}$$

Case analysis (law of total probability)

$$\Pr(\alpha) = \sum_{i=1}^{n} \Pr(\alpha \wedge \beta_i),$$

where the events $\beta_1, \ldots, \beta_n$ are mutually exclusive and exhaustive.

Further Properties of Beliefs

Another version of case analysis

$$\Pr(\alpha) = \sum_{i=1}^{n} \Pr(\alpha|\beta_i)\Pr(\beta_i),$$

where the events $\beta_1, \ldots, \beta_n$ are mutually exclusive and exhaustive.

Two simple and useful forms of case analysis are these:

$$\begin{aligned}
\Pr(\alpha) &= \Pr(\alpha \wedge \beta) + \Pr(\alpha \wedge \neg\beta) \\
\Pr(\alpha) &= \Pr(\alpha|\beta)\Pr(\beta) + \Pr(\alpha|\neg\beta)\Pr(\neg\beta).
\end{aligned}$$

The main value of case analysis is that, in many situations, computing our beliefs in the cases is easier than computing our beliefs in α. We shall see many examples of this phenomena in later chapters.

Further Properties of Beliefs

Bayes rule

$$\Pr(\alpha|\beta) = \frac{\Pr(\beta|\alpha)\Pr(\alpha)}{\Pr(\beta)}.$$

- Classical usage: α is perceived to be a cause of β.
- Example: α is a disease and β is a symptom–
- Assess our belief in the cause given the effect.
- Belief in an effect given its cause, $\Pr(\beta|\alpha)$, is usually more readily available than the belief in a cause given one of its effects, $\Pr(\alpha|\beta)$.

Difficulty: Complexity in model construction and inference

- In Alarm example:
  - 31 numbers needed,
  - Quite unnatural to assess: e.g.

$$P(B=y, E=y, A=y, J=y, M=y)$$

  - Computing $P(B{=}y|M{=}y)$ takes 29 additions.
- In general,
  - $P(X_1, X_2, \ldots, X_n)$ needs at least $2^n - 1$ numbers to specify the joint probability. Exponential model size.
  - Knowledge acquisition difficult (complex, unnatural),
  - Exponential storage and inference.

Chain Rule and Factorization

Overcome the problem of exponential size by exploiting conditional independence

- The chain rule of probabilities:

$$
\begin{aligned}
P(X_1, X_2) &= P(X_1)P(X_2|X_1) \\
P(X_1, X_2, X_3) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \\
&\cdots \\
P(X_1, X_2, \ldots, X_n) &= P(X_1)P(X_2|X_1)\ldots P(X_n|X_1, \ldots, X_{n-1}) \\
&= \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}).
\end{aligned}
$$

- No gains yet. The number of parameters required by the factors is: $2^{n-1} + 2^{n-1} + \ldots + 1 = 2^n - 1$.

Conditional Independence

- About $P(X_i|X_1, \ldots, X_{i-1})$:
  - Domain knowledge usually allows one to identify a subset $pa(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$ such that
    - Given $pa(X_i)$, X_i is independent of all variables in $\{X_1, \ldots, X_{i-1}\} \setminus pa(X_i)$, i.e.

$$P(X_i|X_1, \ldots, X_{i-1}) = P(X_i|pa(X_i))$$

- Then

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|pa(X_i))$$

- Joint distribution factorized.
- The number of parameters might have been substantially reduced.

Example

$P(B,E,A,J,M)=?$

Example continued

$$
\begin{aligned}
& P(B, E, A, J, M) \\
& \quad = \quad P(B)P(E|B)P(A|B, E)P(J|B, E, A)P(M|B, E, A, J) \\
& \quad = \quad P(B)P(E)P(A|B, E)P(J|A)P(M|A)(Factorization)
\end{aligned}
$$

- $pa(B) = \{\}$, $pa(E) = \{\}$,$pa(A) = \{B, E\}$, $pa(J) = \{A\}$,$pa(M) = \{A\}$.
- Conditional probabilities tables (CPT)

B	P(B)
Y	.01
N	.99

E	P(E)
Y	.02
N	.98

A	B	E	P(A|B, E)
Y	Y	Y	.95
N	Y	Y	.05
Y	Y	N	.94
N	Y	N	.06
Y	N	Y	.29
N	N	Y	.71
Y	N	N	.001
N	N	N	.999

M	A	P(M|A)
Y	Y	.9
N	Y	.1
Y	N	.05
N	N	.95

J	A	P(J|A)
Y	Y	.7
N	Y	.3
Y	N	.01
N	N	.99

Example continued

- Model size reduced from 31 to 1+1+4+2+2=10
- Model construction easier
  - Fewer parameters to assess.
  - Parameters more natural to assess:e.g.

$$P(B = Y), P(E = Y), P(A = Y|B = Y, E = Y),$$

$$P(J = Y|A = Y), P(M = Y|A = Y)$$

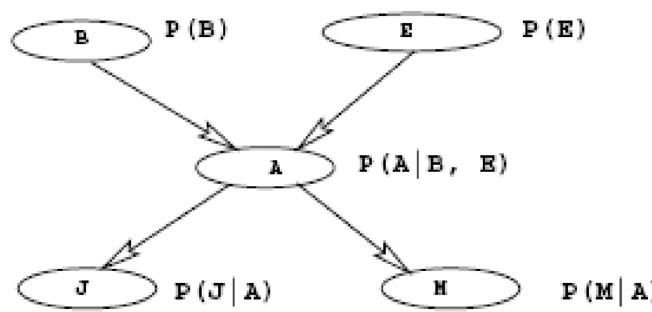
- Inference easier.Will see this later.

From Factorizations to Bayesian Networks

Graphically represent the conditional independency relationships:

- construct a directed graph by drawing an arc from X_j to X_i iff $X_j \in pa(X_i)$

$$pa(B) = \{\}, \; pa(E) = \{\}, \; pa(A) = \{B, E\}, \; pa(J) = \{A\}, \; pa(M) = \{A\}.$$



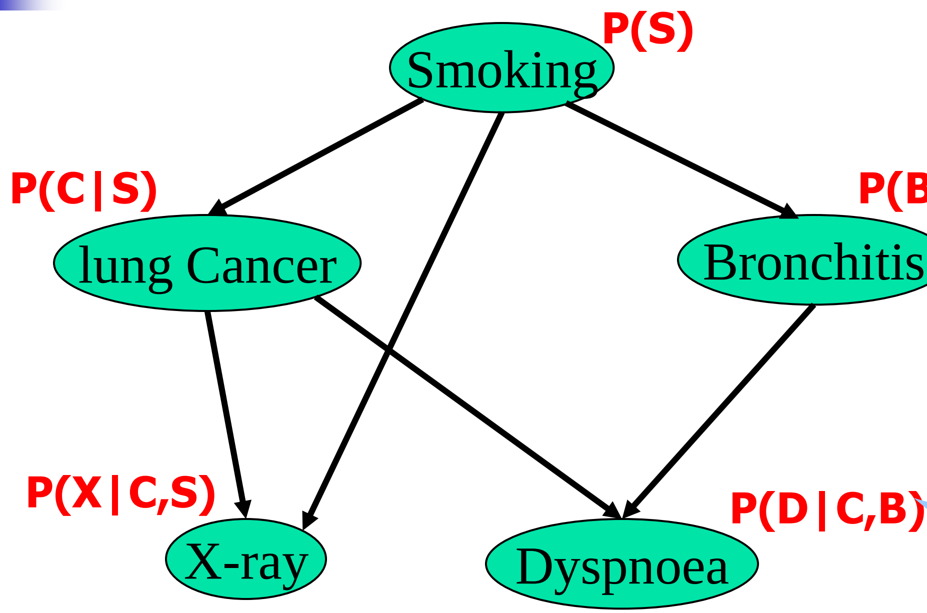
- Also attach the conditional probability (table) $P(X_i|pa(X_i))$ to node X_i.
- What results in is a Bayesian network. Also known as **belief network, probabilistic network**.

Formal Definition

A **Bayesian network** is:

- An directed acyclic graph (DAG), where
- Each node represents a random variable
- And is associated with the conditional probability of the node given its parents.

Bayesian Networks: Representation



$$BN = (G, \Theta)$$

CPD:

C	B	D=0	D=1
0	0	0.1	0.9
0	1	0.7	0.3
1	0	0.8	0.2
1	1	0.9	0.1

$$P(S, C, B, X, D) = P(S)\, P(C|S)\, P(B|S)\, P(X|C,S)\, P(D|C,B)$$

Conditional Independencies ➡ Efficient Representation

Soft Evidence

There are two types of evidence that one may encounter.

Hard evidence

is information to the effect that some event has occurred, which is also the type of evidence we have considered earlier.

Soft Evidence

There are two types of evidence that one may encounter.

Hard evidence

is information to the effect that some event has occurred, which is also the type of evidence we have considered earlier.

Soft evidence

is not conclusive: we may get an unreliable testimony that event β occurred, which may increase our belief in β, but not to the point where we would consider it certain.

Soft Evidence

There are two types of evidence that one may encounter.

Hard evidence

is information to the effect that some event has occurred, which is also the type of evidence we have considered earlier.

Soft evidence

is not conclusive: we may get an unreliable testimony that event β occurred, which may increase our belief in β, but not to the point where we would consider it certain.

Example

Our neighbor who is known to have a hearing problem may call to tell us that they have heard the alarm trigger in our home. Such a call may not be used to categorically confirm the event Alarm, but can still increase our belief in alarm to some new level.

Soft Evidence

One of the key issues relating to soft evidence is how to specify its strength. There are two main methods for this.

The "All things considered" Method

Stating the new belief in β after the evidence has been accommodated. The new belief in β depends not only on the strength of evidence, but also on our initial beliefs that existed before the evidence was obtained.

Example

After receiving my neighbor's call, my belief in the alarm triggering stands now at .85.'

Soft Evidence: "All things considered" Method

- Soft evidence is a constraint $\mathrm{Pr}'(\beta) = q$, where Pr' denotes the new state of belief after accommodating the evidence.
- Pr' can be computed along the same principles we used for Bayes conditioning.
- $\mathrm{Pr}'(\beta) = q$ implies constraint $\mathrm{Pr}'(\neg\beta) = 1 - q$.
- If we insist on preserving the relative beliefs in worlds that satisfy β, and also on preserving the relative beliefs in worlds that satisfy $\neg\beta$, we get

$$\mathrm{Pr}'(\omega) \stackrel{def}{=} \begin{cases} \dfrac{q}{\mathrm{Pr}(\beta)}\mathrm{Pr}(\omega), & \text{if } \omega \models \beta \\ \dfrac{1-q}{\mathrm{Pr}(\neg\beta)}\mathrm{Pr}(\omega), & \text{if } \omega \models \neg\beta. \end{cases}$$

Jeffrey's rule

There is also a useful closed form known as Jeffrey's rule

$$\Pr'(\alpha) = q\Pr(\alpha|\beta) + (1-q)\Pr(\alpha|\neg\beta),$$

where $\Pr'$ is the new state of belief after accommodating the soft evidence $\Pr'(\beta) = q$.

Note that Bayes conditioning is a special case of Jeffrey's rule when $q = 1$, which is to be expected as they were both derived using the same principle.

Jeffrey's rule

When the evidence concerns a set of mutually exclusive and exhaustive events $\beta_1, \ldots, \beta_n$, with the new beliefs in these events being $q_1, \ldots, q_n$, respectively.

Generalization of Jeffrey's rule

$$\mathrm{Pr}'(\alpha) = \sum_{i=1}^{n} q_i \mathrm{Pr}(\alpha|\beta_i).$$

Jeffrey's rule

Example

Assume that we are given a piece of cloth C, where its color can be one of: green (c_g), blue (c_b), or violet (c_v). We want to know whether, in the next day, the cloth will be sold (s), or not sold ($\bar{s}$).

Our original state of belief is as follows:

worlds	S	C	$\Pr(.)$
ω_1	s	c_g	.12
ω_2	$\bar{s}$	c_g	.18
ω_3	s	c_b	.12
ω_4	$\bar{s}$	c_b	.18
ω_5	s	c_v	.32
ω_6	$\bar{s}$	c_v	.08

Our belief in the cloth being sold is $\Pr(s) = .56$. Our beliefs in the colors c_g, c_b, c_v are .3, .3, and .4, respectively.

Jeffrey's rule

Example

Assume that we now inspect the cloth by candlelight, and we conclude that our new beliefs in these colors should be .7, .25, and .05, respectively.

Using Jeffrey's rule:

$$\Pr'(s) = .7(\frac{.12}{.3}) + .25(\frac{.12}{.3}) + .05(\frac{.32}{.4}) = .42$$

The new state of belief according to Jeffrey's rule is:

worlds	S	C	$\Pr'(.)$
ω_1	s	c_g	$.28 = .12 \times .7/.3$
ω_2	$\bar{s}$	c_g	$.42 = .18 \times .7/.3$
ω_3	s	c_b	$.10 = .12 \times .25/.3$
ω_4	$\bar{s}$	c_b	$.15 = .18 \times .25/.3$
ω_5	s	c_v	$.04 = .32 \times .05/.4$
ω_6	$\bar{s}$	c_v	$.01 = .08 \times .05/.4$

The "Nothing else considered" Method

Soft evidence on event β is based on declaring the strength of this evidence, independently of currently held beliefs.

The odds of event β

$$O(\beta) \stackrel{def}{=} \frac{\Pr(\beta)}{\Pr(\neg\beta)}.$$

- An odds of 1 indicates that we believe β and $\neg\beta$ equally.
- An odds of 10 indicates that we believe β ten times more than we believe $\neg\beta$.

The "Nothing else considered" Method

We can specify soft evidence on event β by declaring the relative change it induces on the odds of β.

Bayes factor is the ratio

$$k = \frac{O'(\beta)}{O(\beta)},$$

where $O'(\beta)$ is the odds of β after accommodating the evidence, $\Pr'(\beta)/\Pr'(\neg\beta)$.

- A Bayes factor of 1 indicates a neutral evidence.
- A Bayes factor of 2 double the odds of β.
- As the Bayes factor tends to infinity, the soft evidence tends towards a hard evidence confirming β.
- As the factor tends to zero, the soft evidence tends towards a hard evidence refuting β.

Soft Evidence as a Noisy Sensor

Soft evidence on an event β can be emulated using a noisy sensor S having two states, with the strength of soft evidence captured by the false positive and negative rates of the sensor.

- The false positive rate of the sensor, f_p, is the belief that the sensor would give a positive reading even though the event β did not occur, $\Pr(S|\neg\beta)$.
- The false negative rate of the sensor, f_n, is the belief that the sensor would give a negative reading even though the event β did occur, $\Pr(\neg S|\beta)$.

Soft Evidence as a Noisy Sensor

Given a sensor with the above specifications that reads positive, we want to know the new odds of β given this positive sensor reading.

$$\begin{aligned}
\mathrm{O}'(\beta) &= \frac{\mathrm{Pr}'(\beta)}{\mathrm{Pr}'(\neg\beta)} \\
&= \frac{\mathrm{Pr}(\beta|S)}{\mathrm{Pr}(\neg\beta|S)}, \text{ emulating soft evidence by a positive sensor reading} \\
&= \frac{\mathrm{Pr}(S|\beta)\mathrm{Pr}(\beta)}{\mathrm{Pr}(S|\neg\beta)\mathrm{Pr}(\neg\beta)}, \text{ by Bayes Theorem} \\
&= \frac{1-f_n}{f_p}\frac{\mathrm{Pr}(\beta)}{\mathrm{Pr}(\neg\beta)} \\
&= \frac{1-f_n}{f_p}\mathrm{O}(\beta).
\end{aligned}$$

The relative change in the odds of β, the Bayes factor $\mathrm{O}'(\beta)/\mathrm{O}(\beta)$, is indeed a function of only the false positive and negative rates of the sensor (independent of initial beliefs).

Soft Evidence as a Noisy Sensor

A soft evidence with a Bayes factor of k^+ can be emulated by a positive sensor reading as long as

$$k^+ = \frac{1 - f_n}{f_p}.$$

The above equation shows that the specific false positive and negative rates are not as important as the above ratio.

Example

A positive reading from any of the following sensors will have the same impact on beliefs:

- Sensor 1: $f_p = 10\%$ and $f_n = 5\%$.
- Sensor 2: $f_p = 8\%$ and $f_n = 24\%$.
- Sensor 3: $f_p = 5\%$ and $f_n = 52.5\%$.

Soft Evidence as a Noisy Sensor

A negative sensor reading will not necessarily have the same impact for the different sensors.

Consider the Bayes factor corresponding to a negative reading:

$$
\begin{aligned}
\mathrm{O}'(\beta) &= \frac{\Pr'(\beta)}{\Pr'(\neg\beta)} \\
&= \frac{\Pr(\beta|\neg S)}{\Pr(\neg\beta|\neg S)}, \text{ emulating soft evidence by a negative sensor reading} \\
&= \frac{\Pr(\neg S|\beta)\Pr(\beta)}{\Pr(\neg S|\neg\beta)\Pr(\neg\beta)}, \text{ by Bayes Theorem} \\
&= \frac{f_n}{1-f_p}\frac{\Pr(\beta)}{\Pr(\neg\beta)} \\
&= \frac{f_n}{1-f_p}\mathrm{O}(\beta).
\end{aligned}
$$

A negative sensor reading corresponds to a soft evidence with a Bayes factor of

$$k^- = \frac{f_n}{1-f_p}.$$

Soft Evidence as a Noisy Sensor

Even though all of the sensors above have the same k^+, they do have different k^- values.

Example

$k^- \approx .056$ for Sensor 1; $k^- \approx .261$ for Sensor 2; and $k^- \approx .553$ for Sensor 3.

All negative sensor readings will decrease the odds of the corresponding hypothesis, but a negative reading from Sensor 1 is stronger than one from Sensor 2, which in turn is stronger than one from Sensor 3.

Soft Evidence as a Noisy Sensor

If $f_p + f_n < 1$, then $k^+ > 1$ and $k^- < 1$.

- A positive sensor reading will increase the odds of the corresponding event.
- A negative sensor reading will decrease those odds.
- The condition in the above equation is satisfied when the false positive and false negative rates are less than 50% each.
- The condition however can also be satisfied even if one of the rates is $\geq 50\%$.