Bayesian networks
Chapter 14.4–5

Outline

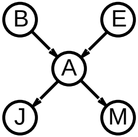
- Exact inference by enumeration
- Exact inference by variable elimination
- Approximate inference by stochastic simulation
- Approximate inference by Markov chain Monte Carlo

Inference tasks

- Simple queries: compute posterior marginal $P(X_i|\mathbf{E}=e)$
  e.g., $P(NoGas|Gauge = empty, Lights = on, Starts = false)$
- Conjunctive queries:
  $P(X_i, X_j|\mathbf{E}=e) = P(X_i|\mathbf{E}=e)P(X_j|X_i, \mathbf{E}=e)$
- Optimal decisions: decision networks include utility information; probabilistic inference required for $P(outcome|action, evidence)$
- Value of information: which evidence to seek next?
- Sensitivity analysis: which probability values are most critical?
- Explanation: why do I need a new starter motor?

Inference by enumeration

- Slightly intelligent way to sum out variables from the joint without actually constructing its explicit representation
- Simple query on the burglary network:



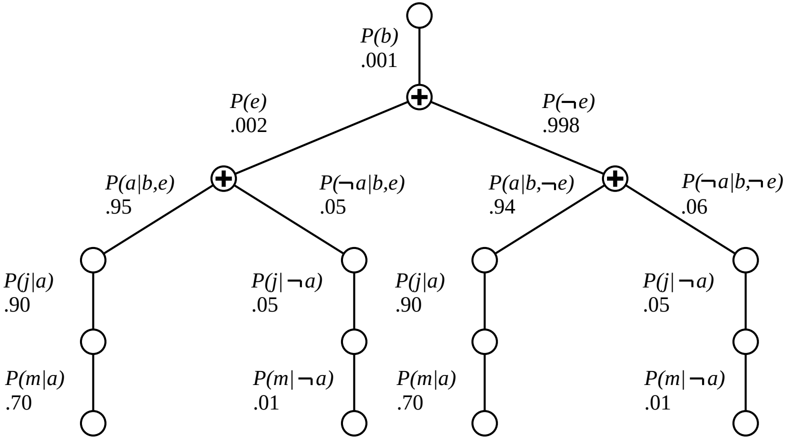
$$
\begin{aligned}
&P(B|j,m) = \\
&= P(B,j,m)/P(j,m) = \\
&= \alpha P(B,j,m) \\
&= \alpha \sum_e \sum_a P(B,e,a,j,m)
\end{aligned}
$$

- Rewrite full joint entries using product of CPT entries:

$$
\begin{aligned}
&P(B|j,m) = \alpha \sum_e \sum_a P(B)P(e)P(a|B,e)P(j|a)P(m|a) \\
&= \alpha P(B) \sum_e P(e) \sum_a P(a|B,e)P(j|a)P(m|a)
\end{aligned}
$$

- Recursive depth-first enumeration: $O(n)$ space, $O(d^n)$ time

Evaluation tree



Enumeration is inefficient: repeated computation e.g., computes $P(j|a)P(m|a)$ for each value of e

Inference by variable elimination

Variable elimination: carry out summations right-to-left, storing intermediate results (**factors**) to avoid re-computation

$$
\begin{aligned}
P(B|j,m) &= \alpha \underbrace{P(B)}_{B} \textstyle\sum_e \underbrace{P(e)}_{E} \textstyle\sum_a \underbrace{P(a|B,e)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M} \\
&= \alpha P(B) \textstyle\sum_e P(e) \textstyle\sum_a P(a|B,e) P(j|a) f_M(a) \\
&= \alpha P(B) \textstyle\sum_e P(e) \textstyle\sum_a P(a|B,e) f_J(a) f_M(a) \\
&= \alpha P(B) \textstyle\sum_e P(e) \textstyle\sum_a f_A(a,b,e) f_J(a) f_M(a) \\
&= \alpha P(B) \textstyle\sum_e P(e) f_{\bar{A}JM}(b,e) (sumout\,\mathrm{A}) \\
&= \alpha P(B) f_{\bar{E}\bar{A}JM}(b) (sumout\,\mathrm{E}) \\
&= \alpha f_B(b) \times f_{\bar{E}\bar{A}JM}(b).
\end{aligned}
$$

Variable elimination: Basic operations

Summing out a variable from a product of factors:

- move any constant factors outside the summation
- add up submatrices in pointwise product of remaining factors

$$\sum_x f_1 \times \cdots \times f_k = f_1 \times \cdots \times f_i \sum_x f_{i+1} \times \cdots \times f_k = f_1 \times \cdots \times f_i \times f_{\bar{X}}$$

assuming $f_1, \ldots, f_i$ do not depend on X.

Pointwise product of factors f_1 and f_2:

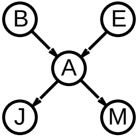
$$f_1(x_1, \ldots, x_j, y_1, \ldots, y_k) \times f_2(y_1, \ldots, y_k, z_1, \ldots, z_l)$$

$$= f(x_1, \ldots, x_j, y_1, \ldots, y_k, z_1, \ldots, z_l)$$

E.g., $f_1(a, b) \times f_2(b, c) = f(a, b, c)$

Irrelevant variables

Consider the query $P(JohnCalls|Burglary = true)$



$$P(J|b) = \alpha P(b) \sum_e P(e) \sum_a P(a|b,e) P(J|a) \sum_m P(m|a)$$

Sum over m is identically 1; M is **irrelevant** to the query.

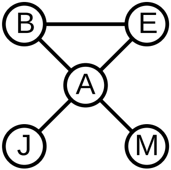
Thm 1:

Y is irrelevant to $P(X|\ldots)$, unless $Y \in Ancestors(\{X\} \cup \mathbf{E})$

Here, $X = JohnCalls$, $\mathbf{E} = \{Burglary\}$,
$Ancestors(\{X\} \cup \mathbf{E}) = \{Alarm, Earthquake\}$
$\implies MaryCalls$ is irrelevant
(Compare this to backward chaining from the query in Horn clause KBs)

Irrelevant variables (continued)

- **Definition:** moral graph of Bayes net: marry all parents and drop arrows
- **Definition:** **A** is m-separated from **B** by **C** $\iff$ separated by **C** in the moral graph
- **Theorem 2:** Y is irrelevant if m-separated from X by



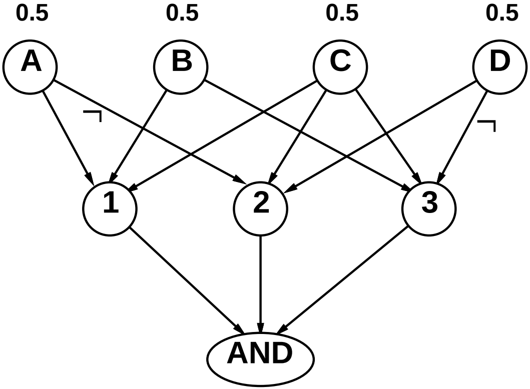
- For $P(\mathit{JohnCalls}|\mathit{Alarm} = \mathit{true})$, both *Burglary* and *Earthquake* are irrelevant

Complexity of exact inference –1

- **Singly connected networks (polytrees)**
  - any two nodes are connected by at most one (undirected) path
  - time and space cost of variable elimination are $O(d^k n)$
- **Multiply connected networks**
  - can reduce 3SAT to exact inference $\implies$ NP-hard
  - equivalent to **counting** 3SAT models $\implies$ #P-complete ("number-P hard"), i.e. strictly harder that NP-complete.

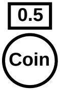
Complexity of exact inference – 2

1. A v B v C
2. C v D v ¬A
3. B v C v ¬D



Inference by stochastic simulation

- **Basic idea**
  1. Draw N samples from a sampling distribution



  2. Compute an approximate posterior probability $\hat{P}$
  3. Show this converges to the true probability P

- **Outline**
  1. Sampling from an empty network
  2. Rejection sampling: reject samples disagreeing with evidence
  3. Likelihood weighting: use evidence to weight samples
  4. Markov chain Monte Carlo (MCMC): sample from a stochastic process whose stationary distribution is the true posterior

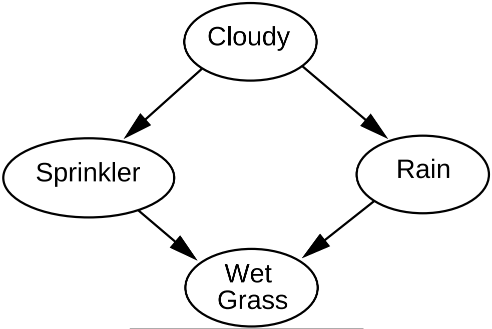
Sampling from an empty network

Prior-Sample(bn): an event sampled from bn

Input: **bn** - a belief network specifying joint distribution $P(X_1, \ldots, X_n)$
x is an event with n elements
For $i = 1$ to n do
    x_i: a random sample from $P(X_i | Parents(X_i))$
    given the values of $Parents(X_i)$ in x
Return

Example

P(C)
.50



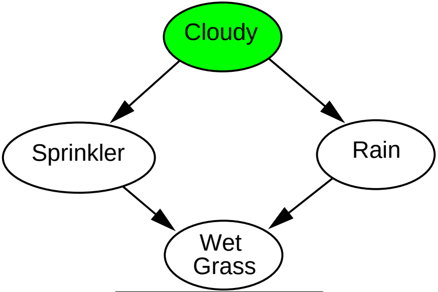
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90

Example

P(C)
.50



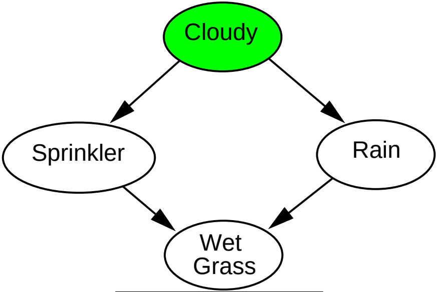
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example

P(C)
.50

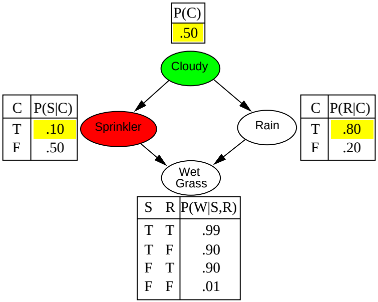


C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

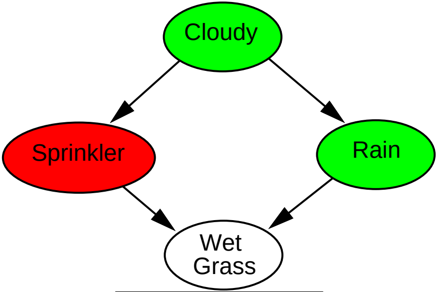
S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example


P(C)
.50
Cloudy
C P(S|C)
T .10
F .50
Sprinkler
Rain
C P(R|C)
T .80
F .20
Wet Grass
S R P(W|S,R)
T T .99
T F .90
F T .90
F F .01

Example

P(C)
.50



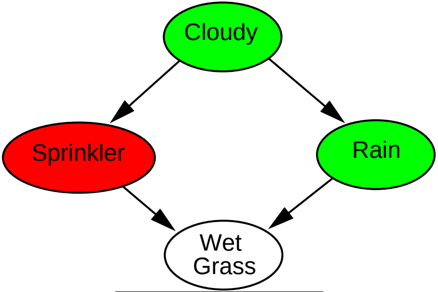
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example

P(C)
.50



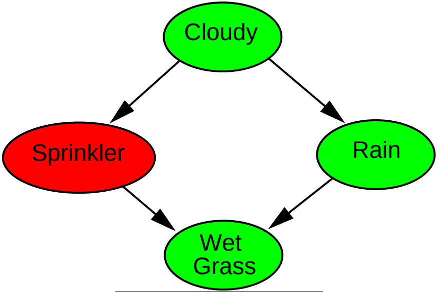
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example

P(C)
.50



C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01