Inference in 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 $P(X_i|E = e)$
e.g., $P(\text{NoGas}|\text{Gauge} = \text{empty}, \text{Lights} = \text{on}, \text{Starts} = \text{false})$

Conjunctive queries: $P(X_i, X_j|E = e) = P(X_i|E = e)P(X_j|X_i, E = e)$

Optimal decisions: decision networks include utility information;
probabilistic inference required for $P(\text{outcome}|\text{action, evidence})$

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:
\[ P(B|j,m) = \frac{P(B,j,m)}{P(j,m)} = \alpha P(B,j,m) = \alpha \sum_e \sum_a P(B,e,a,j,m) \]

Rewrite full joint entries using product of CPT entries:
\[ P(B|j,m) = \alpha \sum_e \sum_a P(B)P(e)P(a|B,e)P(j|a)P(m|a) = \alpha P(B) \sum_e P(e) \sum_a P(a|B,e)P(j|a)P(m|a) \]

Recursive depth-first enumeration: \( O(n) \) space, \( O(d^n) \) time
**Enumeration algorithm**

\[
\text{function } \text{Enumeration-Ask}(X, e, bn) \text{ returns a distribution over } X \\
\text{inputs: } X, \text{ the query variable} \\
\hspace{0.5cm} e, \text{ observed values for variables } E \\
\hspace{0.5cm} bn, \text{ a Bayesian network with variables } \{X\} \cup E \cup Y \\
\hspace{0.5cm} Q(X) \leftarrow \text{a distribution over } X, \text{ initially empty} \\
\hspace{0.5cm} \text{for each value } x_i \text{ of } X \text{ do} \\
\hspace{1cm} \text{extend } e \text{ with value } x_i \text{ for } X \\
\hspace{1cm} Q(x_i) \leftarrow \text{Enumerate-All(Vars[bn], e)} \\
\text{return } \text{Normalize}(Q(X))
\]

\[
\text{function } \text{Enumerate-All}(vars, e) \text{ returns a real number} \\
\hspace{0.5cm} \text{if Empty?}(vars) \text{ then return } 1.0 \\
\hspace{0.5cm} Y \leftarrow \text{First}(vars) \\
\hspace{0.5cm} \text{if } Y \text{ has value } y \text{ in } e \\
\hspace{1cm} \text{then return } P(y | Pa(Y)) \times \text{Enumerate-All(Rest(vars), e)} \\
\hspace{1cm} \text{else return } \sum_y P(y | Pa(Y)) \times \text{Enumerate-All(Rest(vars), e}_y) \\
\hspace{1cm} \text{where } e_y \text{ is } e \text{ extended with } Y = y
\]
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 recomputation

\[ P(B|j, m) = \alpha P(B) \sum_e P(e) \sum_a P(a|B, e) P(j|a) P(m|a) \]
\[ = \alpha P(B) \sum_e P(e) \sum_a P(a|B, e) P(j|a) \sum_M f_M(a) \]
\[ = \alpha P(B) \sum_e P(e) \sum_a P(a|B, e) f_J(a) \sum_M f_M(a) \]
\[ = \alpha P(B) \sum_e P(e) \sum_a f_A(a, b, e) f_J(a) f_M(a) \]
\[ = \alpha P(B) \sum_e P(e) f_{\tilde{A}JM}(b, e) \text{ (sum out } A) \]
\[ = \alpha P(B) f_{\tilde{E}\tilde{A}JM}(b) \text{ (sum out } E) \]
\[ = \alpha f_B(b) \times f_{\tilde{E}\tilde{A}JM}(b) \]
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

\[ \sum_x f_1 \times \cdots \times f_k = f_1 \times \cdots \times f_i \sum_x f_{i+1} \times \cdots \times f_k = f_1 \times \cdots \times f_i \times f_X \]

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., \( f_1(a, b) \times f_2(b, c) = f(a, b, c) \)
Variable elimination algorithm

**function** `Elimination-Ask(X, e, bn)` **returns** a distribution over $X$

**inputs:** $X$, the query variable
  - $e$, evidence specified as an event
  - $bn$, a belief network specifying joint distribution $P(X_1, \ldots, X_n)$

1. `factors ← []; vars ← Reverse(VARS[bn])`
2. **for each** `var` **in** `vars` **do**
   - `factors ← [Make-Factor(var, e)]|factors`
   - **if** `var` is a hidden variable **then** `factors ← Sum-Out(var, factors)`
3. **return** `Normalize(Pointwise-Product(factors))`
Irrelevant variables

Consider the query \( P(\text{JohnCalls}|\text{Burglary} = \text{true}) \)

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

Sum over \( m \) is identically 1; \( M \) is irrelevant to the query

Thm 1: \( Y \) is irrelevant unless \( Y \in \text{Ancestors}([^X^] \cup E) \)

Here, \( X = \text{JohnCalls}, E = \{\text{Burglary}\}, \) and

\( \text{Ancestors}([^X^] \cup E) = \{\text{Alarm, Earthquake}\} \)

so \( \text{MaryCalls} \) is irrelevant

(Compare this to backward chaining from the query in Horn clause KBs)
Irrelevant variables contd.

Defn: **moral graph** of Bayes net: marry all parents and drop arrows

Defn: A is m-separated from B by C iff separated by C in the moral graph

Thm 2: Y is irrelevant if m-separated from X by E

For \( P(\text{JohnCalls}|\text{Alarm}=\text{true}) \), both 
*Burglary* and *Earthquake* are irrelevant
Complexity of exact inference

Singly connected networks (or 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 $\Rightarrow$ NP-hard
- equivalent to counting 3SAT models $\Rightarrow$ #P-complete
Summary

Exact inference by variable elimination:
- polytime on polytrees, NP-hard on general graphs
- space = time, very sensitive to topology

Approximate inference by LW, MCMC:
- LW does poorly when there is lots of (downstream) evidence
- LW, MCMC generally insensitive to topology
- Convergence can be very slow with probabilities close to 1 or 0
- Can handle arbitrary combinations of discrete and continuous variables