Inference in Bayesian networks

Chapter 14.4–5

Outline
- Exact inference by enumeration
- Approximate inference by stochastic simulation

Inference tasks
- Simple queries: compute posterior marginal
- Conjunctive queries:
- 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?

Complexity of exact inference
- Multiply connected networks:
  - Can reduce 3SAT to exact inference \( \Rightarrow \) NP-hard
  - Equivalent to counting 3SAT models \( \Rightarrow \) \#P-complete

Chapter 14.4–5
Inference by stochastic simulation

Basic idea:
1) Draw $N$ samples from a sampling distribution $S$
2) Compute an approximate posterior probability $\hat{P}$
3) Show this converges to the true probability $P$

Outline:
– Sampling from an empty network
– Rejection sampling: reject samples disagreeing with evidence
– Likelihood weighting: use evidence to weight samples

Chapter 14.4–5 7

Sampling from an empty network

function Prior-Sample($bn$) returns an event sampled from $bn$
inputs: $bn$, a belief network specifying joint distribution $P(X_1, \ldots, X_n)$

1. $x \leftarrow$ an event with $n$ elements
2. for $i = 1$ to $n$
   1. $x_i \leftarrow$ a random sample from $P(X_i|\text{parents}(X_i))$
   2. given the values of $\text{parents}(X_i)$
3. return $x$

Chapter 14.4–5 8

Example

Cloudy

| P(R|C) | C | T | F |
|-------|---|---|---|
| P(C)  | P(W|S,R) |
| .50   | .80   | .10 |
| .20   | .90   | .99 |

Chapter 14.4–5 9

Example

Cloudy

| P(R|C) | C | T | F |
|-------|---|---|---|
| P(C)  | P(W|S,R) |
| .50   | .80   | .10 |
| .20   | .90   | .99 |

Chapter 14.4–5 10

Example

Cloudy

| P(R|C) | C | T | F |
|-------|---|---|---|
| P(C)  | P(W|S,R) |
| .50   | .80   | .10 |
| .20   | .90   | .99 |

Chapter 14.4–5 11

Example

Cloudy

| P(R|C) | C | T | F |
|-------|---|---|---|
| P(C)  | P(W|S,R) |
| .50   | .80   | .10 |
| .20   | .90   | .99 |

Chapter 14.4–5 12
### Example

#### Cloudy

- **Probability**
  - $P(R|C) = 0.80$
  - $P(S|C) = 0.10$
  - $P(W|S,R) = 0.90$

#### Wet Grass

- **Prior Probability**
  - $P(C) = 0.50$
  - $P(S) = 0.01$

#### Analysis of Rejection Sampling

**Simple Rejection Sampling**

- **Estimation**
  - $\hat{P}(X|e) = \alpha \frac{N \text{Prior-Sample}(e)}{N \text{Prior-Sample}}$ (algorithm definition)
  - $\approx \frac{P(X|e)}{P(e)}$ (property of PriorSample)
  - $= P(X|e)$ (definition of conditional probability)

**Shorthand**

- $\hat{P}(X|e) \approx P(X|e)$

**Conclusion**

- Rejection sampling returns consistent posterior estimates.

**Problem**

- Hopelessly expensive if $P(e)$ is small.

$P(e)$ drops off exponentially with the number of evidence variables.

---

**Sampling from an empty network contd.**

**Probability that PriorSample generates a particular event**

$$P(S) = \begin{cases} 0.50 & \text{if } S = \text{true} \\ 0.01 & \text{if } S = \text{false} \end{cases}$$

**Example**

- **Rain and Sprinkler**
  - $P(R|C) = 0.80$
  - $P(S|C) = 0.10$
  - $P(W|S,R) = 0.90$

- **Wet Grass**
  - $P(C) = 0.50$
  - $P(S) = 0.01$
Likelihood weighting

Idea: fix evidence variables, sample only nonevidence variables, and weight each sample by the likelihood it accords the evidence function.

Likelihood-Weighting (X, e, bn, N) returns an estimate of P(X | e)

local variables:

W, a vector of weighted counts over X, initially zero

for j = 1 to N do
  x, w ← Weighted-Sample (bn)
  W[x] ← W[x] + w
where x is the value of X in x

return Normalize(W[X])

function Weighted-Sample (bn, e) returns an event and a weight

x ← an event with n elements; w ← 1
for i = 1 to n do
  if X_i has a value x_i in e then
    w ← w × P(X_i = x_i | parents(X_i))
  else
    x_i ← a random sample from P(X_i | parents(X_i))
return x, w
Likelihood weighting example

Cloudy
Rain
Sprinkler
Wet
Grass

\[ P(R|C) = \begin{pmatrix} C \ T \ F \\ 0.80 \ 0.20 \end{pmatrix} \]

\[ P(S|C) = \begin{pmatrix} S \ R \\ T \ T \ T \ F \ F \end{pmatrix} \]

\[ P(W|S,R) = \begin{pmatrix} 0.90 \ 0.90 \ 0.99 \end{pmatrix} \]

\[ W = 1 \times 0.1 \times 0.99 = 0.099 \]

Likelihood weighting analysis

Sampling probability for WeightedSample is

\[ S_{W S}(z,e) = \prod_{l=1}^{L} P(z_l | \text{parents}(Z_l)) \]

Note: pays attention to evidence in ancestors only

Weight for a given sample is

\[ w(z,e) = \prod_{m=1}^{M} P(e_m | \text{parents}(E_m)) \]

Weighted sampling probability is

\[ S_{W W}(z,e)w(z,e) = \prod_{l=1}^{L} P(z_l | \text{parents}(Z_l)) \prod_{m=1}^{M} P(e_m | \text{parents}(E_m)) = P(z,e) \text{(by standard global semantics of network)} \]

Hence likelihood weighting returns consistent estimates but performance still degrades with many evidence variables because a few samples have nearly all the total weight.

Summary

Exact inference by enumeration:
- NP-hard on general graphs

Approximate inference by L W:
- L W does poorly when there is lots of evidence
- L W, 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
- Useful for approximate inference with complex models
- Requires more samples to achieve accuracy