Bayes' Nets

Inference by Enumeration

- **General case:**
  - **Evidence variables:** \( E_1, \ldots, E_k = e_1, \ldots, e_k \)
  - **Query variable:** \( Q \)
  - **Hidden variables:** \( H_1, \ldots, H_r \)

- We want: \( \Pr(Q|e_1, \ldots, e_k) \)
  - *Works fine with multiple query variables, too*

- **Step 1:** Select the entries consistent with the evidence

- **Step 2:** Sum out \( H \) to get joint of \( Q \) and evidence

- **Step 3:** Normalize.

\[
P(Q, e_1, \ldots, e_k) = \sum_{h_1, \ldots, h_r} P(Q, h_1, \ldots, h_r, e_1, \ldots, e_k)
\]

\[
\Pr(Q|e_1, \ldots, e_k) = \frac{1}{Z} \sum_{h_1, \ldots, h_r} P(Q, h_1, \ldots, h_r, e_1, \ldots, e_k)
\]

- Given unlimited time, inference in BNs is easy.

- \( P(B|+j, +m) \propto P(B, +j, +m) \)

Examples:

- Posterior probability
- Most likely explanation: Inference
  - Inference: calculating some useful quantity from a joint probability distribution

Inference by Enumeration in Bayes' Net

\[
P(B|+j, +m) = \sum_a P(B, e, a, +j, +m)
\]

\[
= \sum_a P(B) P(e) P(a|B, e) P(+j|a) P(+m|a)
\]

\[
= P(B) P(+e) P(+a|B, +e) P(+j|+a) P(+m|+a) + P(B) P(+e) P(-a|B, +e) P(+j|-a) P(+m|-a) + P(B) P(-e) P(+a|B, -e) P(+j|+a) P(+m|+a) + P(B) P(-e) P(-a|B, -e) P(+j|-a)
\]

\[
P(\text{antilock}|\text{observed variables}) = ?
\]
Factor Zoo

Factor Zoo I

Joint distribution: $P(X,Y)$
- Entries $P(x,y)$ for all $x, y$
- Sums to? 1

Selected joint: $P(x,Y)$
- Slice of joint distribution
- Entries $P(x,y)$: fixed $x$, all $y$
- Table $P(cold, W)$?
- Sums to? $P(x)$.

Factor Zoo II

Single conditional: $P(Y|x)$
- Entries $P(y|x)$ for fixed $x$, all $y$
- Sum of entries? 1

Family of conditionals:
- $P(X|Y)$
- Multiple conditionals
- Entries $P(x|y)$ for all $x, y$
- Sum? $|Y|

Factor Zoo III

Specified family: $P(y|X)$
- Entries $P(y|x)$ for fixed $y$, but for all $x$
- Sums to? who knows!

Factor Zoo Summary

In general, when we write $P(Y_1, \ldots, Y_N | X_1, \ldots, X_M)$
- It is a "factor": a multi-dimensional array
- Its entries are $P(y_1, \ldots, y_N | x_1, \ldots, x_M)$
- Any assigned (=lower-case) $X$ or $Y$ is a dimension missing (selected) from the array

Example: Traffic Domain

Random Variables
- $R$: Raining
- $T$: Traffic
- $L$: Late for class!

$P(R)$
- $+r$ 0.1
- $-r$ 0.9

$P(T|R)$
- $+r +t$ 0.8
- $+r -t$ 0.2
- $-r +t$ 0.1
- $-r -t$ 0.9

$P(L|T)$
- $+t +l$ 0.3
- $+t -l$ 0.7
- $-t +l$ 0.1
- $-t -l$ 0.9
Inference by Enumeration: Procedural Outline

Track objects called factors.
Initial factors are local CPTs (one per node)

Example: Multiple Joins

Operation 1: Join Factors
First basic operation: joining factors
Combining factors:
- Just like a database join
- Get all factors over the joining variable
- Build a new factor over the union of the variables involved

Example: Join on R

Operation 2: Eliminate
Second basic operation: marginalization
Take a factor and sum out a variable
- Shrinks a factor to a smaller one
- A projection operation

Example: Multiple Elimination
Inference by Enumeration:

\[ P(L) = \sum_r \sum_t P(r)P(t|r)P(L|t). \]

Join on \( r \):
\[ P(r)P(t|r) \]

Eliminate \( r \):
\[ \sum_r \]

Join on \( t \):
\[ \times P(L|t) \]

Eliminate \( t \):
\[ \sum_t \]

Variable Elimination:

\[ P(L) = \sum_r P(L|t) \sum_r P(r)P(t|r). \]

Join on \( r \):
\[ P(r)P(t|r) \]

Eliminate \( r \):
\[ \sum_r \]

Join on \( t \):
\[ \times P(L|t) \]

Eliminate \( t \):
\[ \sum_t \]
General Variable Elimination

Query: Q
Start with initial factors:
- Local CPTs (but instantiated by evidence)
While still hidden variables (not Q or evidence):
- Pick a hidden variable H
- Join all factors mentioning H
- Eliminate (sum out) H

Join all remaining factors and normalize

Example

B E A M
P(B, m|A, j, m) ∝ P(B, j, m)
Selected CPT:
P(B) P(E) P(A|B, E)
P(j|A) P(m|A)

Choose Hidden Variable A.
Factors: P(j, m|B, E), P(B).
Join: P(j, m, E|B). Sum: P(j, m|B)

Remaining Factors: P(B), P(j, m|B).

Finish with B: Factors: P(j, m|B, E), P(B).
Join: P(j, m|B).

And Finally:
Normalize: P(B|j, m)

Evidence

Start with factors that select evidence
- Initial factors:
P(R) P(T|R) P(L|T)

+ t + l 0.1
- t - l 0.9

Then eliminate variables to answer query.

Evidence II

Result will be a selected joint of query and evidence
- E.g. for P(L + r), we would end up with:
  + e + l 0.026
  + e - l 0.074
To get our answer, just normalize this!

+ e + l 0.26
+ e - l 0.74
That's it!

Example

B
E
J M

P(B, j, m|A, E) ∝ P(B, j, m, A)
Selected CPT:
P(B) P(E) P(A|B, E)
P(j|A) P(m|A)

Choose Hidden Variable A.
Factors: P(j, m|B, E), P(B).
Join: P(j, m, E|B). Sum: P(j, m|B)

Remaining Factors: P(B), P(j, m|B).

Finish with B: Factors: P(j, m|B, E), P(B).
Join: P(j, m|B).

And Finally:
Normalize: P(B|j, m)

Factors:
P(B) P(E) P(A|B, E) P(j, A)

Sum ratio to get marginal

Evidence:
P(R) P(T|R) P(L|T)

+ e + l 0.8
- e + l 0.7
+ e - l 0.2
- e - l 0.1

Exploits:
uwy + uwz + uxy + uxz + vwy + vwz + vxy + vxz =
(u + v)(w + x)(y + z).

Same Example in Equations

Factors: P(B) P(E) P(A|B, E) P(j, A)

Sum ratio to get marginal

Evidence:
P(R) P(T|R) P(L|T)

+ e + l 0.8
- e + l 0.7
+ e - l 0.2
- e - l 0.1

Exploits:
uwy + uwz + uxy + uxz + vwy + vwz + vxy + vxz =
(u + v)(w + x)(y + z).

Another Variable Elimination Example

Query: P(X1|y, y2, y3)
Evidence:
P(Z) P(X1|Z) P(X2|Z) P(X3|Z) P(x1|Z) P(y1|j, m)

Eliminate X1, get P(j, z|y1) = Σz P(x1|z)P(y1|j, z).

Eliminate X2, get P(j, z|y2) = Σx2 P(x2|z)P(y2|j, x2).

Eliminate Z, get P(j, y1, y2, y3) = Σz P(x1|z)P(x2|z)P(y2|j, x2)P(z|y1, y2).

No more hidden variables. Join:
P(x1|y1, y2, y3) = P(y1|X1)P(y2|y1, y3)
Normalizing gives result.
Consider the following orderings:

- $Z, X_1, \ldots, X_{n-1}$
- $X_1, \ldots, X_{n-1}, Z$

What is the size of the maximum factor generated for each of the orderings?

- Eliminate $Z$? New factor with $X_1, \ldots, X_n$ of size $2^n$.
- Uh oh.
- Eliminate $X_1$? New factor with $Z$ and $y_1$ of size 2.
- Eliminate $X_2$? New factor with $Z$ and $y_2$ of size 2.
- \ldots Cool!

Answer: 2 versus $2^n$ (assuming binary variables.)

In general, the ordering can greatly affect efficiency.

---

Bayes’ Nets

**Representation**
- ✓

**D-separation**
- ✓

**Probabilistic Inference**
- ✓

- Enumeration (exact, exponential complexity).
- Variable elimination
  - (exact, worst-case exponential complexity, often better).
- Inference is NP-complete.
- Sampling (approximate) (Next up.)
- Learning Bayes’ Nets from Data. (Later.)

---

Variable Elimination Ordering

Query $P(X_i | y_1, \ldots, y_n)$.

Consider the following orderings:

- $Z, X_1, \ldots, X_{n-1}$ and $X_1, \ldots, X_{n-1}, Z$.

What is the size of the maximum factor generated for each of the orderings?

- Eliminate $Z$? New factor with $X_1, \ldots, X_n$ of size $2^n$.
- Uh oh.
- Eliminate $X_1$? New factor with $Z$ and $y_1$ of size 2.
- Eliminate $X_2$? New factor with $Z$ and $y_2$ of size 2.
- \ldots Cool!

Answer: 2 versus $2^n$ (assuming binary variables.)

In general: the ordering can greatly affect efficiency.

---

VE: Computational and Space Complexity

Computational and space complexity of variable elimination depends on largest factor.

Elimination ordering can greatly affect size of the largest factor.

- E.g., previous slide’s example $2^n$ vs. 2
- Does there always exist an ordering that only results in small factors?
  - No!

Worst Case Complexity?

CSP:

$((x_1 \lor x_2 \lor \neg x_3), (\neg x_1 \lor x_3 \lor \neg x_4), (x_2 \lor \neg x_2 \lor x_4), (\neg x_3 \lor \neg x_4 \lor \neg x_3), (x_3 \lor x_5 \lor x_7), (x_4 \lor x_5 \lor x_6), (\neg x_5 \lor x_6 \lor \neg x_7), (\neg x_5 \lor \neg x_6 \lor x_7)$

$P(X_i = 1) = 0.5, P(X_i = 0) = 0.5$

Exercise: Think about how the specifics would work out!