Fundamentals of AI and KR - Module 3

4. Exact inference

Paolo Torroni

Fall 2023

Notice

Credits

The present slides are largely an adaptation of existing material, including:

- slides from Russel & Norvig
- slides by Daphne Koller on Probabilistic Graphical Models
- slides by Fabrizio Riguzzi on Data Mining and Analytics

I am especially grateful to these authors.

Downloading and ~~sharing~~

Table of Contents

Exact inference

Inference tasks

query variable

- Simple queries: compute posterior marginal $\mathbf{P}(X_i|\mathbf{E}=\mathbf{e})$
  e.g., $P(NoGas|Gauge=empty, Lights=on, Starts=false)$
- Conjunctive queries: $\mathbf{P}(X_i, X_j|\mathbf{E}=\mathbf{e}) = \mathbf{P}(X_i|\mathbf{E}=\mathbf{e})\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})$
- 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:

$\mathbf{P}(B|j,m)$

$$= \alpha \, \underbrace{P(B,j,m)}_{} = \alpha \sum_a \sum_e P(B,j,m,a,e)$$

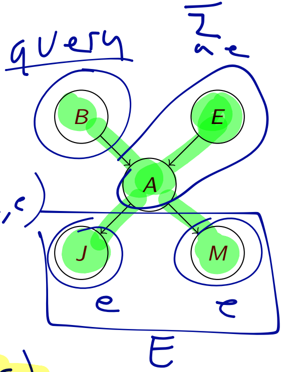
$$= \alpha \sum_a \sum_e P(B)\,P(a|B,e)\,P(e)\,P(j|a)\,P(m|a)$$

a,e

$\boxed{a,\neg e} \rightarrow P(B)P(a|B,\neg e)P(\neg e)P(j|a)P(m|a)$

$\neg a,e$

$\neg a,\neg e$



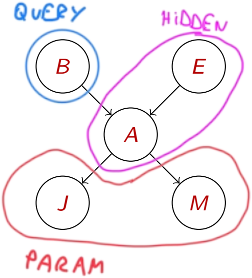
Inference by enumeration

Slightly intelligent way to sum out variables from the joint without actually constructing its explicit representation

Simple query on the burglary network:

$\mathbf{P}(B|j,m)$
$= \mathbf{P}(B,j,m)/P(j,m)$
$= \alpha\mathbf{P}(B,j,m)$
$= \alpha \sum_e \sum_a \mathbf{P}(B,e,a,j,m)$

hidden variables
e, a

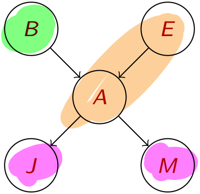


Inference by enumeration

Slightly intelligent way to sum out variables from the joint without actually constructing its explicit representation

Simple query on the burglary network:

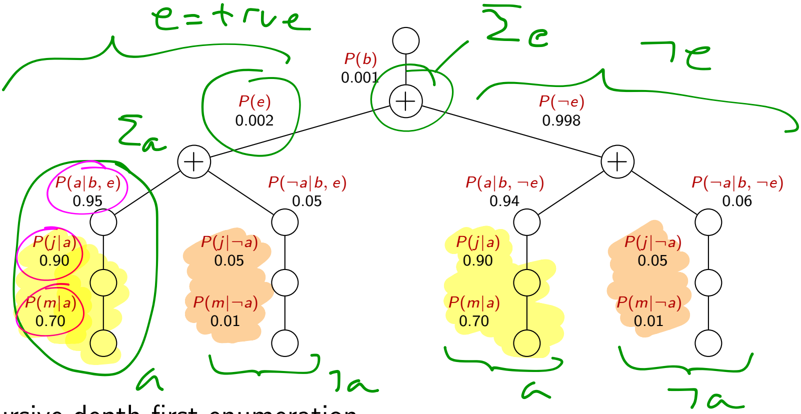
$\mathbf{P}(B|j,m)$
$= \mathbf{P}(B,j,m)/P(j,m)$
$= \alpha\mathbf{P}(B,j,m)$
$= \alpha \sum_e \sum_a \mathbf{P}(B,e,a,j,m)$



Rewrite full joint entries using product of CPT entries:

$\mathbf{P}(B|j,m) = \alpha \sum_e \sum_a \mathbf{P}(B)P(e)\mathbf{P}(a|B,e)P(j|a)P(m|a)$
$= \alpha\mathbf{P}(B) \sum_e P(e) \sum_a \mathbf{P}(a|B,e)P(j|a)P(m|a)$

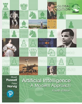
Evaluation tree



Recursive depth-first enumeration

$O(n)$ space (no need to construct full joint), $O(d^n)$ time

Inefficient: repeated computation (e.g., see $P(j|a)P(m|a)$)



Enumeration algorithm

function ENUMERATION-ASK(X,**e**,bn) **returns** a distribution over X
  inputs: X, the query variable
    e, observed values for variables **E**
    bn, a Bayesian network with variables $\text{VARS}[bn] = \{X\} \cup \mathbf{E} \cup \mathbf{Y}$
  $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
  for each value x_i of X **do**
    extend **e** with $X = x_i$
    $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL(VARS[bn],**e**)
  return NORMALIZE($\mathbf{Q}(X)$)

function ENUMERATE-ALL($vars$,**e**) **returns** a real number
  if EMPTY?($vars$) **then** **return** 1.0
  $Y \leftarrow$ FIRST($vars$)
  if Y has value y in **e** **then**
    return $P(y|Parents(Y)) \times$ ENUMERATE-ALL(REST($vars$),**e**)
  else return $\sum_y P(y|Parents(Y)) \times$ ENUMERATE-ALL(REST($vars$),$\mathbf{e} \cup \{Y = y\}$)

Inference by variable elimination

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

$$\mathbf{P}(B|j,m) = \alpha \underbrace{\mathbf{P}(B)}_{B} \sum_e \underbrace{P(e)}_{E} \boxed{\sum_a} \underbrace{\mathbf{P}(a|B,e)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M}$$

$f_1(a,B,e)\ f_2(\cancel{j},a)\ f_3(\cancel{m},a)$

b ¬b

f_4

	b	¬b
a e	.95×.9×.7	
a ¬e		
¬a ¬e		
¬a e		.71×.05×.01

$\mathbf{P}(B)$
0.001

$\mathbf{P}(E)$
.002

B	E	$\mathbf{P}(A\|B,E)$
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	$\mathbf{P}(J\|A)$
T	.90
F	.05

f_2

A	$\mathbf{P}(M\|A)$
T	.70
F	.01

f_3

REDO THE EXERCISE

$$P(B \mid j, m) = \alpha \, P(B) \sum_{e} P(e) \sum_{a} P(a \mid B, e) \cdot P(j \mid a) \cdot P(m \mid a) =$$

$$= f_1(a, B, e) \cdot f_2(j, a) \cdot f_3(m, a)$$ NOT VARIABLES

$\downarrow$ MORE CONVENIENT FORMULA

P(M|A)

P(J|A)

P(A|B,E)

	b	$\neg b$
a, e	$0.95 \times 0.9 \times 0.7$	
$a, \neg e$		
$\neg a, e$		$(1-0.29) \times 0.05 \times 0.01$
$\neg a, \neg e$		

$$P(\neg A \mid \neg B, E) = 1 - P(A \mid \neg B, E)$$

Inference by variable elimination

X	Y

$\Rightarrow P(B|j,m) = \left\langle \frac{X}{X+Y}, \frac{Y}{X+Y} \right\rangle$

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

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

P(B)
0.001

P(E)
.002

B	E	**P**($A \mid B, E$)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	**P**($J \mid A$)
T	.90
F	.05

A	**P**($M \mid A$)
T	.70
F	.01

e	¬e
.001	.998

e

¬e

e

¬e

Variable elimination: Basic operations

A	$\mathbf{P}(J \mid A)$
T	.90
F	.05

A	$\mathbf{P}(M \mid A)$
T	.70
F	.01

B	E	$\mathbf{P}(A \mid B, E)$
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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: Basic operations

A	$\mathbf{P}(J\|A)$
T	.90
F	.05

A	$\mathbf{P}(M\|A)$
T	.70
F	.01

B	E	$\mathbf{P}(A\|B,E)$
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

Summing out a variable from a product of factors:

1. move any constant factors outside the summation
2. 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

Variable elimination algorithm

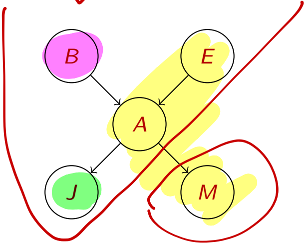
function ELIMINATION-ASK(X,**e**,bn) **returns** a distribution over X
  inputs: X, the query variable
    e, evidence specified as an event
    bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  factors ← [ ]
  for each *var* **in** ORDER(VARS[*bn*]) **do**
    factors ← [MAKE-FACTOR(*var*,**e**)|*factors*]
    if *var* is a hidden variable **then**
      factors ← SUM-OUT(*var*,*factors*)
  return NORMALIZE(POINTWISE-PRODUCT(*factors*))

ORDER(VARS[*bn*]) → every ordering yields a valid algorithm
Try to minimize size of next factor to be constructed

Irrelevant variables

↳ ∉ ancestral tree
↳ d-separeted



Consider the query

$P(JohnCalls|Burglary = true)$

$P(J|b) = \alpha P(b) \sum_e P(e)$
$\sum_a P(a|b,e) P(J|a)$
$\cancel{\sum_m P(m|a)} = 1$

Sum over m is identically 1; $\boxed{M \text{ is } \textbf{irrelevant}}$ to the query

Theorem

Y is irrelevant unless $Y \in Ancestors(\{X\} \cup \mathbf{E})$

M IS NEITHER ANCESTOR OF J, NEITHER ANCESTOR OF B
QUERY VARIABLE
EVIDENCE VARIABLE

Here, $X = JohnCalls$, $\mathbf{E} = \{Burglary\}$, and
$Ancestors(\{X\} \cup \mathbf{E}) = \{Alarm, Earthquake\}$ so $MaryCalls$ is irrelevant

Irrelevant variables

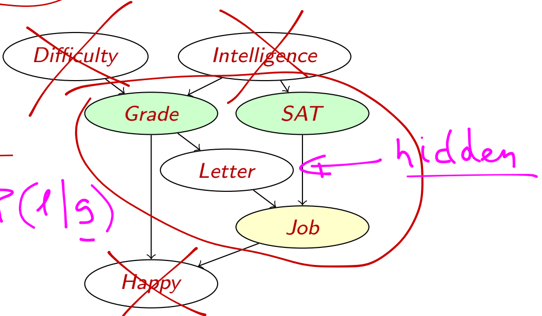
NO ACTIVE TRAIL BETWEEN ANY VALUES OF X AND Y, GIVEN E

Theorem

Y is irrelevant if d-separated from X by $\mathbf{E}$

$P(g)$? (SAT)

$\alpha \sum_{l} P(j|l,s)P(l|g)$



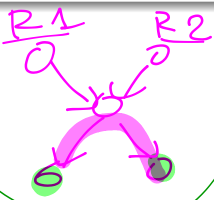
For $P(Job|Grade, SAT)$, not only *Happy* is irrelevant (outside of ancestral graph) but also *Difficulty* and *Intelligence* are (d-separated)

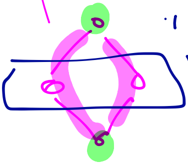
Complexity of exact inference

vs. approximate inference (sampling)

Theorem (Complexity results for variable elimination)

- *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:* NP-hard





'meganode'

Clustering algorithms

- Variable elimination OK for solving individual queries
- However, cost of estimating posterior probabilities for n variables in a polytree is $O(n^2)$
- **Clustering algorithms** (**join tree** algorithms) can do that in $O(n)$: used in commercial systems
- Idea is to join individual nodes to form cluster nodes (**meganodes**), obtaining a polytree
- Worst case still requires exponentially large CPTs
- Alternative approach: approximate methods

Exercise

Exam-style exercise: online shop

You are designing a new e-commerce system, to sell products online. You can display advertisements ("ads") of three types: book ads, toy ads, and holiday ads. However, your web site can only display one ad at a time. Your system should guess which type of ad has a higher chance of being clicked.

You don't have verified profile information on each customer. You can, however, guess at least some profile features, based on your domain knowledge and on your observations about which ads the customer does or doesn't click on ("clicking behaviour").

In particular, you have identified the following **profile features**:

- Young: is this a young customer (y, ¬y);
- Married: is the customer married (m, ¬m);
- Wealthy: is the customer wealthy (w, ¬w);
- Children: does the customer have children (c, ¬c).

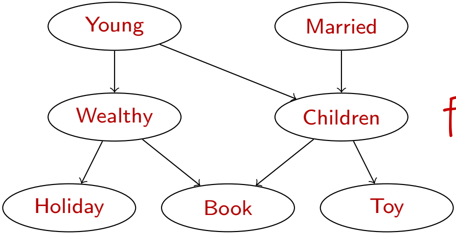
The customer's **clicking behaviour** is instead determined as follows:

- Holiday: the customer clicks on a holiday ad (h, ¬h);
- Book: the customer clicks on a book ad (b, ¬b);
- Toy: the customer clicks on a toy ad (t, ¬t).

☞ How would you model this domain using a Bayes net?

Exam-style exercise: online shop

Assume this is your network:

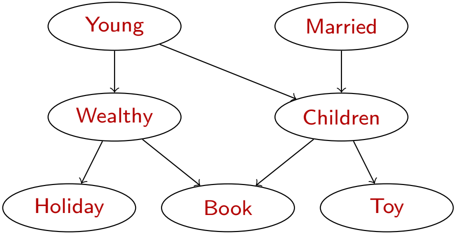


$P(W|B)$

- Q1: *You first display a sequence of ads and record which types of ads get clicked; then, you use that data to infer some of the customer's profile features.*
  - ☞ *Which reasoning pattern are you following?*
  - ☞ *Show a query that uses the customer's observed clicking behaviour as evidence.*

Exam-style exercise: online shop

Assume this is your network:



- Q2: *To fully define this Bayesian network, how many independent values are needed?*
- Q3: *Define a CPD for Children*

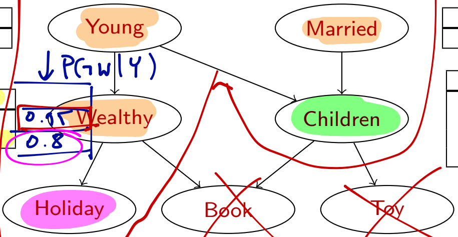
Q2 1 + 1 + 2 + ...

Exam-style exercise: online shop

Assume this is your network, with its parameters:

$\mathbf{P}(y)$
0.4

$\mathbf{P}(m)$
0.3



	$\mathbf{P}(w \mid Y)$
y	0.05
$\neg y$	0.2

	$\mathbf{P}(h \mid W)$
w	0.2
$\neg w$	0.05

		$\mathbf{P}(c \mid Y, M)$
y	m	0.1
y	$\neg m$	0.01
$\neg y$	m	0.5
$\neg y$	$\neg m$	0.45

$f_1(w, y)$

	y	$\neg y$
w	$.2 \times .05$	$.2 \times .2$
$\neg w$	$.05 \times .95$	$.05 \times .8$

f_2

$\downarrow P(\neg w \mid Y)$: 0.95, 0.8

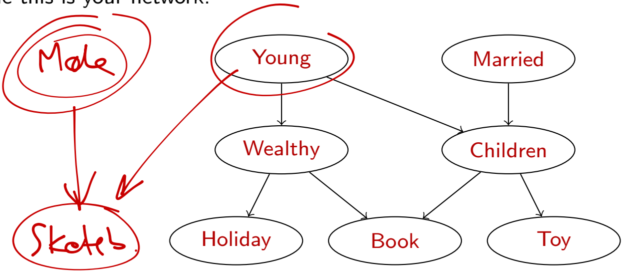
- Q4: *Calculate $P(c \mid h)$ using variable elimination.*

$$P(C|h) = \alpha \sum_y \sum_m \sum_w P(h|w)P(w|y)P(m)P(C|y,m)P(y)$$

$$\alpha \sum_y P(y) \sum_m P(m)P(C|y,m) \underbrace{\sum_w P(h|w)P(w|y)}_{f_1}$$

Exam-style exercise: online shop

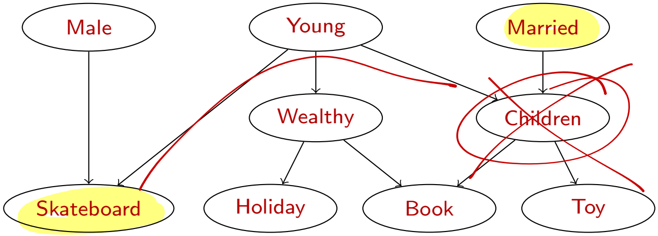
Assume this is your network:



- Q5: *You expand your business to sell Skateboards: an article that's especially popular with young male customers.*
  ☞ *How would you expand your Bayesian network to include also skateboard ads?*

Exam-style exercise: online shop

Assume this is your extended network:

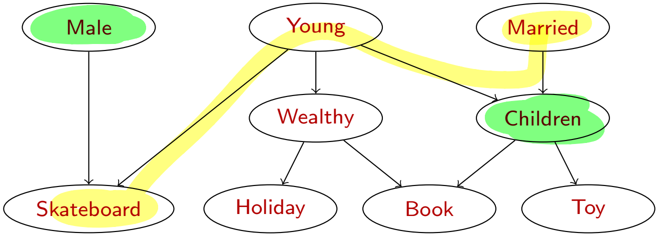


- Q6: *You display a skateboard ad, and observe that your customer doesn't click on it.*
  - ☞ *Can this observation alone help predict if your customer is married?* ← no

$$P \overset{?}{\models} M \perp S \iff P(M) \overset{?}{=} P(M|S)$$

Exam-style exercise: online shop

Assume this is your extended network:



- Q7: *The same customer joins a fidelity programme and fills in a form where she declares to be a female parent. She chooses not to disclose any other information about herself.*
  - ☞ *Given this new evidence, does her not clicking on a skateboard ad give you any information about her being married?*

yes

Suggested exercise from Russel & Norvig, 3^{rd} Ed.

- 14.21 (soccer teams), (a)-(d)

Questions?