Uncertainty: Wrap up & Decision Theory: Intro

Alan Mackworth

UBC CS 322 – Decision Theory 1

March 25, 2013

Textbook §6.4.1 & §9.2

Announcements (1)

- Teaching Evaluations are online
  - You should have received a message about them
  - Secure, confidential, mobile access
- Your feedback is important!
  - Allows us to assess and improve the course material
  - I use it to assess and improve my teaching methods
  - The department as a whole uses it to shape the curriculum
  - Teaching evaluation results are important for instructors
    - Appointment, reappointment, tenure, promotion and merit, salary
  - UBC takes them very seriously (now)
  - Evaluations close at 11:59PM on April 9, 2013.
    - Before exam, but instructors can't see results until *after* we submit grades
  - Please do it!
- Take a few minutes and visit https://eval.olt.ubc.ca/science

Announcements (2)

- Assignment 4 due Wednesday, April 3rd,1pm
- Final exam
  - Thursday, April 18th, 8:30am – 11am in PHRM 1101
  - Same general format as midterm (~60% short questions)
    - *List of short questions is now on Connect*
  - *Practice final is now available in Connect*
  - More emphasis on material after midterm
  - How to study?
    - Practice exercises, assignments, short questions, lecture notes, text, problems in text, learning goals …
    - Use TA and my office hours (extra office hours TBA if needed)
    - Review sessions: last class plus more TBA if needed
    - *Submit topics you want reviewed in response to message on Connect*

Hints for Assignment 4

- Question 4 (Bayesian networks)
  - "correctly represent the situation described above" means "do not make any independence assumptions that aren't true"
  - "(Hint: remember that Bayes nets do not necessarily encode causality."
  - Another hint:
    - Step 1: identify the causal network
    - Step 2: for each network, check if it entails (conditional or marginal) independencies the causal network does not entail. If so, it's incorrect
  - Failing to entail some (or all) independencies does not make a network incorrect (only computationally suboptimal)

Lecture Overview

- Variable elimination: recap and some more details
- Variable elimination: pruning irrelevant variables
- Summary of Reasoning under Uncertainty
- Decision Theory
  - Intro
  - Time-permitting: Single-Stage Decision Problems

Recap: Factors and Operations on them

- A factor is a function from a tuple of random variables to the real numbers R
- Operation 1: assigning a variable in a factor
  - E.g., assign X=t

Factor of Y,X,Z

X	Y	Z	$f_1(X,Y,Z)$
t	t	t	0.1
t	t	f	0.9
t	f	t	0.2
t	f	f	0.8
~~f~~	~~t~~	~~t~~	~~0.4~~
~~f~~	~~t~~	~~f~~	~~0.6~~
~~f~~	~~f~~	~~t~~	~~0.3~~
~~f~~	~~f~~	~~f~~	~~0.7~~

$$f_1(X,Y,Z)_{X=t} = f_2(Y,Z)$$

Y	Z	$f_2(Y,Z)$
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

Factor of Y,Z

Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers R
- Operation 1: assigning a variable in a factor
  - $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor

$$\sum_B f_3(A,B,C) = f_4(A,C)$$

B	A	C	$f_3(A,B,C)$
t	t	t	0.03
t	t	f	0.07
f	t	t	0.54
f	t	f	0.36
t	f	t	0.06
t	f	f	0.14
f	f	t	0.48
f	f	f	0.32

A	C	$f_4(A,C)$
t	t	0.57
t	f	0.43
f	t	0.54
f	f	0.46

Recap: Operation 3: multiplying factors

$f_5(A,B) \times f_6(B,C) = f_7(A,B,C)$, i.e

$f_5(A=a,B=b) \times f_6(B=b,C=c) = f_7(A=a,B=b,C=c)$

A	B	$f_5(A,B)$
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

A	C	$f_6(A,C)$
t	t	0.3
t	f	0.7
f	t	0.6
f	f	0.4

A	B	C	$f_7(A,B,C)$
t	t	t	0.03
t	t	f	0.1x0.7
t	f	t	0.9x0.6
t	f	f	…
f	t	t	
f	t	f	
f	f	t	
f	f	f	

Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers R
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A{=}a,B{=}b,C{=}c) = f_5(A{=}a,B{=}b) \times f_6(B{=}b,C{=}c)$

- If we assign variable A=a in factor $f_7(A,B)$, what is the correct form for the resulting factor?

f(A) | f(B) | f(A,B) | A number

Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers R
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

- If we assign variable A=a in factor $f_7(A,B)$, what is the correct form for the resulting factor?
  - f(B).
    When we assign variable A we remove it from the factor's domain

Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers R
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

- If we marginalize variable A out from factor $f_7(A,B)$, what is the correct form for the resulting factor?

f(A) | f(B) | f(A,B) | A number

Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers R
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

- If we assign variable A=a in factor $f_7(A,B)$, what is the correct form for the resulting factor?
  - f(B).
    When we marginalize out variable A we remove it from the factor's domain

Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers R
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

- If we multiply factors $f_4(X,Y)$ and $f_6(Z,Y)$, what is the correct form for the resulting factor?

f(X,Z) f(X) f(X,Y) f(X,Y,Z)

Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers R
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

- If we multiply factors $f_4(X,Y)$ and $f_6(Z,Y)$, what is the correct form for the resulting factor?
  - $f(X,Y,Z)$
  - When multiplying factors, the resulting factor's domain is the union of the multiplicands' domains

Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers R
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

- What is the correct form for $\sum_B f_5(A,B) \times f_6(B,C)$
  - As usual, product before sum: $\sum_B (\; f_5(A,B) \times f_6(B,C) \;)$

f(A,B,C)   f(B)   f(A,C)   f(B,C)

Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers R
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

- What is the correct form for $\sum_B f_5(A,B) \times f_6(B,C)$
  - As usual, product before sum: $\sum_B (f_5(A,B) \times f_6(B,C))$
  - Result of multiplication: f(A,B,C). Then marginalize out B: f'(A,C)

Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers R
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$
- Operation 4: normalize the factor
  - Divide each entry by the sum of the entries. The result will sum to 1.

A	$f_8(A)$
t	0.4
f	0.1

→

A	$f_9(A)$
t	$0.4/(0.4+0.1) = 0.8$
f	$0.1/(0.4+0.1) = 0.2$

Recap: General Inference in Bayesian Networks

Given

- A Bayesian Network BN, and
- Observations of a subset of its variables E: E=e
- A subset of its variables Y that is queried

Compute the conditional probability P(Y=y|E=e)

Definition of conditional probability

Marginalization over Y:
$P(E=e) = \Sigma_{y' \in dom(Y)} P(E=e, Y=y')$

$$P(Y=y \mid E=e) = \frac{P(Y=y, E=e)}{P(E=e)} = \frac{P(Y=y, E=e)}{\sum_{y' \in dom(Y)} P(Y=y', E=e)}$$

All we need to compute is the
joint probability of the query variable(s) and the evidence!

Recap: Key Idea of Variable Elimination

- To sum out a variable Z from a product $f_1 \times \ldots \times f_k$ of factors:
  - Partition the factors into
    - those that don't contain Z say $f_1 \times \ldots \times f_i$
    - those that contain Z say $f_{i+1} \times \ldots \times f_k$
- We know:

$$\sum_Z f_1 \times \ldots \times f_k = f_1 \times \ldots \times f_i \times \left(\sum_Z f_{i+1} \times \ldots \times f_k \right)$$

New factor! Let's call it f'

- We thus have $\sum_Z f_1 \times \ldots \times f_k = f_1 \times \ldots \times f_i \times f'$
- Store f' explicitly, and discard $f_{i+1} \ldots f_k$
- Now we've summed out Z

Recap: Variable Elimination (VE) in BNs

- The joint probability distribution of a Bayesian network is $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|pa(X_i))$
  - We make a factor f_i for each conditional probability table $P(X_i|pa(X_i))$
  - So we have $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} f_i$

- The variable elimination algorithm computes $P(Y| E_1=e_1, \ldots, E_j=e_j)$ as follows:
  - Assign $E_1=e_1, \ldots, E_j=e_j$
  - Sum out all non-query variables $Z_1, \ldots, Z_k$, one at a time
    - To sum out Z_i:
      - Multiply factors containing it Z_i
      - Then marginalize out Z_i from the product
    - The order in which we sum out variables is called our elimination ordering
  - Normalize the final factor $f(Y)$.
    - The resulting factor is exactly $P(Y| E_1=e_1, \ldots, E_i=e_i)$

VE_BN Algorithm

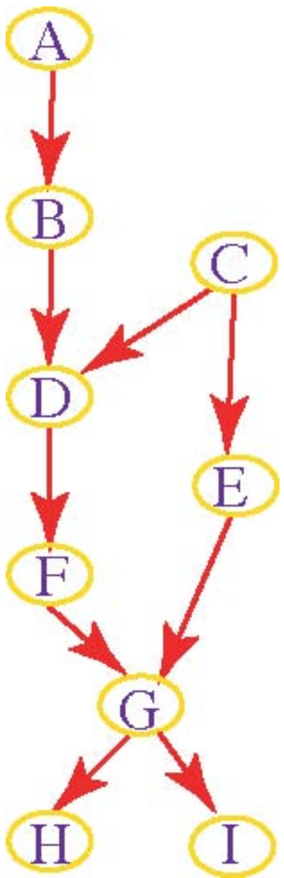
```
Procedure VE_BN(Vs,Ps,O,Q)
    Inputs
        Vs: set of variables
        Ps: set of factors representing the conditional probabilities
        O: set of observations of values on some of the variables
        Q: a query variable
    Output
        posterior distribution on Q
    Local
        Fs: a set of factors
    Fs ←Ps
    for each X∈Vs-{Q} using some elimination ordering do
        if (X is observed) then
            for each F ∈Fs that involves X do
                set X in F to its observed value in O
                project F onto remaining variables
        else
            Rs←{F ∈Fs: F  involves X}
            let T be the product of the factors in Rs
            N ←∑_X T
            Fs←Fs \ Rs∪{N}
    let T be the product of the factors in Rs
    N ←∑_Q T
    return T/N
```

Figure 6.8: Variable elimination for belief networks (P&M, Section 6.4.1, p. 254)

Recap: VE example: compute $P(G|H=h_1)$
Step 1: construct a factor for each cond. probability

$P(G,H) =$

$$\sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_7(H,G)\, f_8(I,G)$$

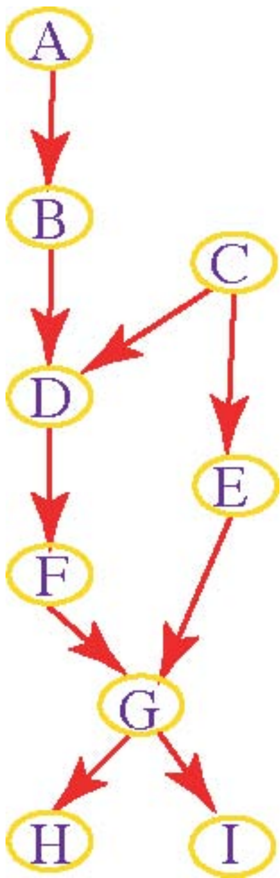


Recap: VE example: compute $P(G|H=h_1)$

Step 2: assign observed variables their observed value

$P(G,H) =$

$\sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_7(H,G)\, f_8(I,G)$



Assigning the variable $H=h_1$:

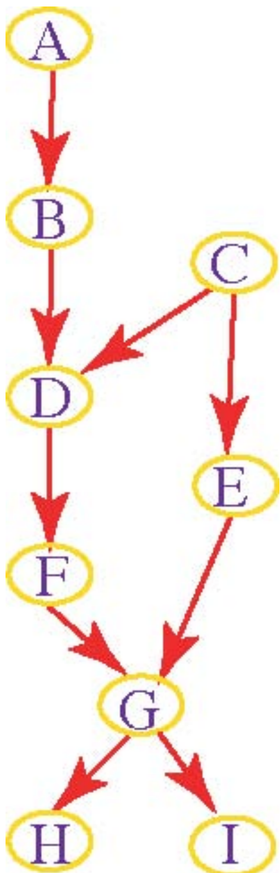
$f_9(G) = f_7(H,G)_{H=h1}$

$P(G,H=h_1)=\sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$

Recap: VE example: compute $P(G|H=h_1)$

Step 4: sum out non- query variables (one at a time)

$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$



Elimination ordering: A, C, E, I, B, D, F

Recap: VE example: compute $P(G|H=h_1)$

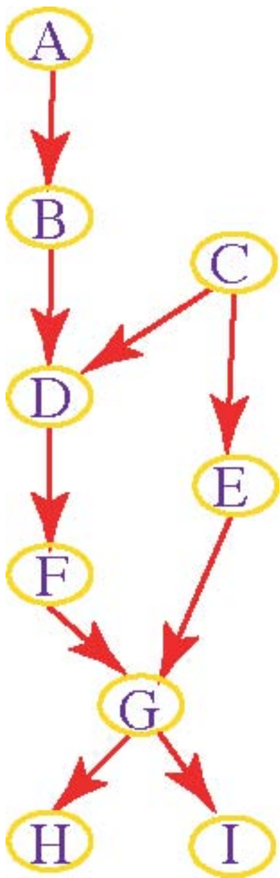
Step 4: sum out non- query variables (one at a time)

$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$

$= \sum_{B,C,D,E,F,I} f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)$

Summing out variable *A*:

$\sum_A f_0(A)\, f_1(B,A) = f_{10}(B)$



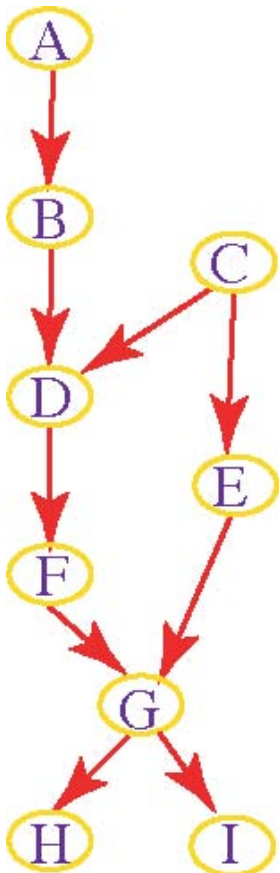
Elimination ordering: A, C, E, I, B, D, F

Recap: VE example: compute $P(G|H=h_1)$
Step 4: sum out non- query variables (one at a time)

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{11}(B,D,E)$$



Summing out variable *C:*

$\sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C) = f_{11}(B,D,E)$

Elimination ordering: A, C, E, I, B, D, F

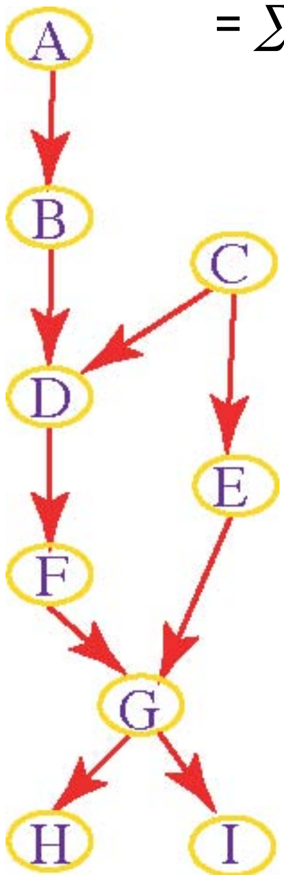
Recap: VE example: compute $P(G|H=h_1)$
Step 4: sum out non- query variables (one at a time)

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{11}(B,D,E)$$

$$= \sum_{B,D,F,I} f_5(F, D)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{12}(G,F,B,D)$$



Summing out variable *E:*

$$\sum_E f_6(G,F,E)\, f_{11}(B,D,E) = f_{12}(G,F,B,D)$$

Elimination ordering: A, C, E, I, B, D, F

Recap: VE example: compute $P(G|H=h_1)$

Step 4: sum out non- query variables (one at a time)

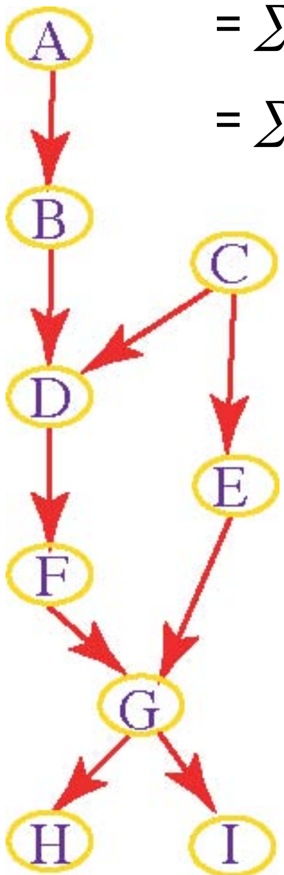
$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{11}(B,D,E)$$

$$= \sum_{B,D,F,I} f_5(F, D)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{12}(G,F,B,D)$$

$$= \sum_{B,D,F} f_5(F, D)\, f_9(G)\, f_{10}(B)\, f_{12}(G,F,B,D)\, f_{13}(G)$$



Elimination ordering: A, C, E, I, B, D, F

Recap: VE example: compute $P(G|H=h_1)$

Step 4: sum out non- query variables (one at a time)

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$$

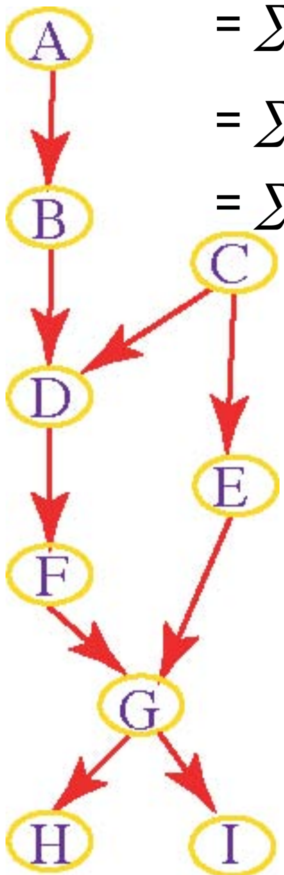
$$= \sum_{B,C,D,E,F,I} f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{11}(B,D,E)$$

$$= \sum_{B,D,F,I} f_5(F, D)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{12}(G,F,B,D)$$

$$= \sum_{B,D,F} f_5(F, D)\, f_9(G)\, f_{10}(B)\, f_{12}(G,F,B,D)\, f_{13}(G)$$

$$= \sum_{D,F} f_5(F, D)\, f_9(G)\, f_{11}(G,F)\, f_{12}(G)\, f_{14}(G,F,D)$$



Elimination ordering: A, C, E, I, B, D, F

Recap: VE example: compute $P(G|H=h_1)$
Step 4: sum out non- query variables (one at a time)

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\ f_1(B,A)\ f_2(C)\ f_3(D,B,C)\ f_4(E,C)\ f_5(F, D)\ f_6(G,F,E)\ f_9(G)\ f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C)\ f_3(D,B,C)\ f_4(E,C)\ f_5(F, D)\ f_6(G,F,E)\ f_9(G)\ f_8(I,G)\ f_{10}(B)$$

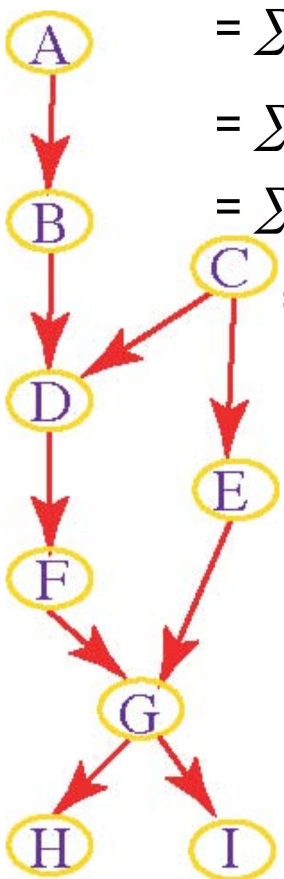
$$= \sum_{B,D,E,F,I} f_5(F, D)\ f_6(G,F,E)\ f_9(G)\ f_8(I,G)\ f_{10}(B)\ f_{11}(B,D,E)$$

$$= \sum_{B,D,F,I} f_5(F, D)\ f_9(G)\ f_8(I,G)\ f_{10}(B)\ f_{12}(G,F,B,D)$$

$$= \sum_{B,D,F} f_5(F, D)\ f_9(G)\ f_{10}(B)\ f_{12}(G,F,B,D)\ f_{13}(G)$$

$$= \sum_{D,F} f_5(F, D)\ f_9(G)\ f_{11}(G,F)\ f_{12}(G)\ f_{14}(G,F,D)$$

$$= \sum_{F} f_9(G)\ f_{11}(G,F)\ f_{12}(G)\ f_{15}(G,F)$$



Elimination ordering: A, C, E, I, B, D, F

Recap: VE example: compute $P(G|H=h_1)$
Step 4: sum out non- query variables (one at a time)

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{11}(B,D,E)$$

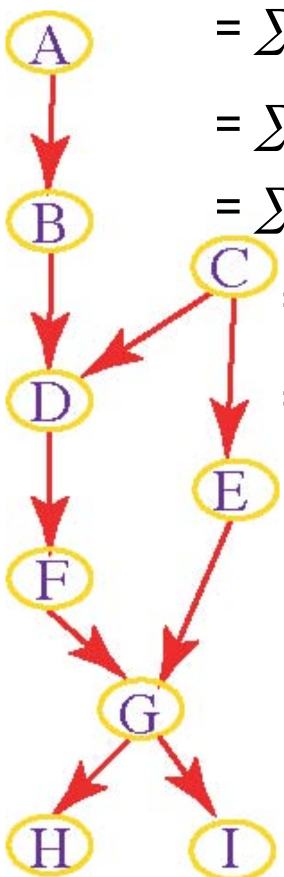
$$= \sum_{B,D,F,I} f_5(F, D)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{12}(G,F,B,D)$$

$$= \sum_{B,D,F} f_5(F, D)\, f_9(G)\, f_{10}(B)\, f_{12}(G,F,B,D)\, f_{13}(G)$$

$$= \sum_{D,F} f_5(F, D)\, f_9(G)\, f_{11}(G,F)\, f_{12}(G)\, f_{14}(G,F,D)$$

$$= \sum_{F} f_9(G)\, f_{11}(G,F)\, f_{12}(G)\, f_{15}(G,F)$$

$$= f_9(G)\, f_{12}(G)\, f_{16}(G)$$



Elimination ordering: A, C, E, I, B, D, F

Recap: VE example: compute $P(G|H=h_1)$

Step 5: multiply the remaining factors

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{11}(B,D,E)$$

$$= \sum_{B,D,F,I} f_5(F, D)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{12}(G,F,B,D)$$

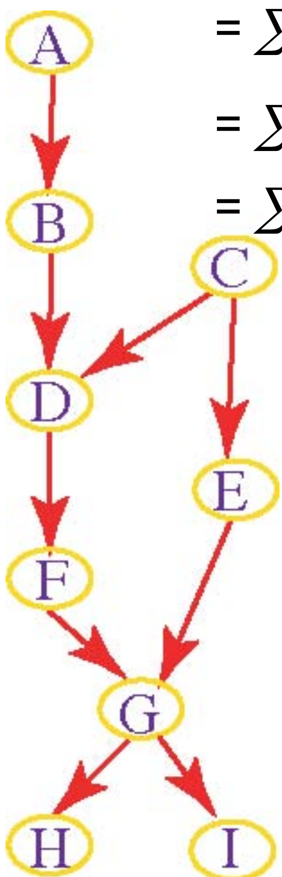
$$= \sum_{B,D,F} f_5(F, D)\, f_9(G)\, f_{10}(B)\, f_{12}(G,F,B,D)\, f_{13}(G)$$

$$= \sum_{D,F} f_5(F, D)\, f_9(G)\, f_{11}(G,F)\, f_{12}(G)\, f_{14}(G,F,D)$$

$$= \sum_{F} f_9(G)\, f_{11}(G,F)\, f_{12}(G)\, f_{15}(G,F)$$

$$= f_9(G)\, f_{12}(G)\, f_{16}(G)$$

$$= f_{17}(G)$$



Recap: VE example: compute $P(G|H=h_1)$
Step 6: normalize

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{11}(B,D,E)$$

$$= \sum_{B,D,F,I} f_5(F, D)\, f_9(G)\, f_8(I,G)\, f_{10}(B)\, f_{12}(G,F,B,D)$$

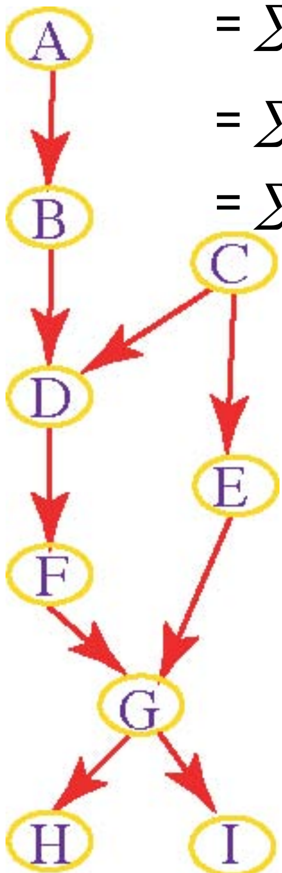
$$= \sum_{B,D,F} f_5(F, D)\, f_9(G)\, f_{10}(B)\, f_{12}(G,F,B,D)\, f_{13}(G)$$

$$= \sum_{D,F} f_5(F, D)\, f_9(G)\, f_{11}(G,F)\, f_{12}(G)\, f_{14}(G,F,D)$$

$$= \sum_{F} f_9(G)\, f_{11}(G,F)\, f_{12}(G)\, f_{15}(G,F)$$

$$= f_9(G)\, f_{12}(G)\, f_{16}(G)$$

$$= f_{17}(G)$$



$$P(G = g \mid H = h_1) = \frac{f_{17}(g)}{\sum_{g' \in dom(G)} f_{17}(g')}$$

AIspace: Belief and Decision Networks Applet

- http://aispace.org/bayes/ implements VE_BN.
- Try some of the sample problems.
- Provide some evidence (symptoms) and query some of the causes.
- Observe how the posterior probability of the causes changes from the prior as you add more evidence.
- Switch from Brief mode to Verbose mode to see the factors generated.
- Experiment with different elimination orderings.

Complexity of Variable Elimination (VE)

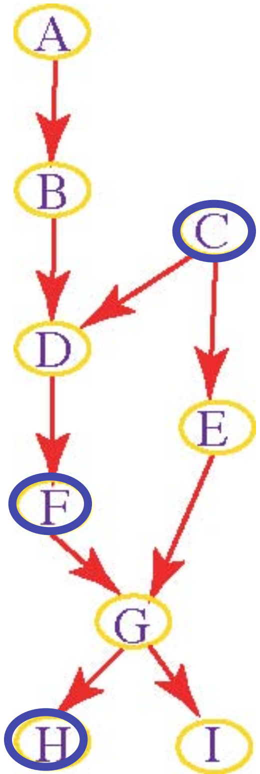
- A factor over n binary variables has to store 2^n numbers
  - The initial factors are typically quite small (variables typically only have few parents in Bayesian networks)
  - But variable elimination constructs larger factors by multiplying factors together
- The complexity of VE is exponential in the maximum number of variables in any factor during its execution
  - This number is called the treewidth of a graph (along an ordering)
  - Elimination ordering influences treewidth
- Finding the best ordering is NP-complete
  - I.e., the ordering that generates the minimum treewidth
  - Heuristics work well in practice (e.g. least connected variables first)
  - Even with best ordering, inference is sometimes infeasible
    - In those cases, we need approximate inference. See CS422 & CS540

Lecture Overview

- Variable elimination: recap and some more details
- Variable elimination: pruning irrelevant variables
- Summary of Reasoning under Uncertainty
- Decision Theory
  - Intro
  - Time-permitting: Single-Stage Decision Problems

VE and conditional independence

- So far, we haven't use conditional independence!
  - Before running VE, we can prune all variables Z that are conditionally independent of the query Y given evidence E: $Z \perp\!\!\!\perp Y \mid E$

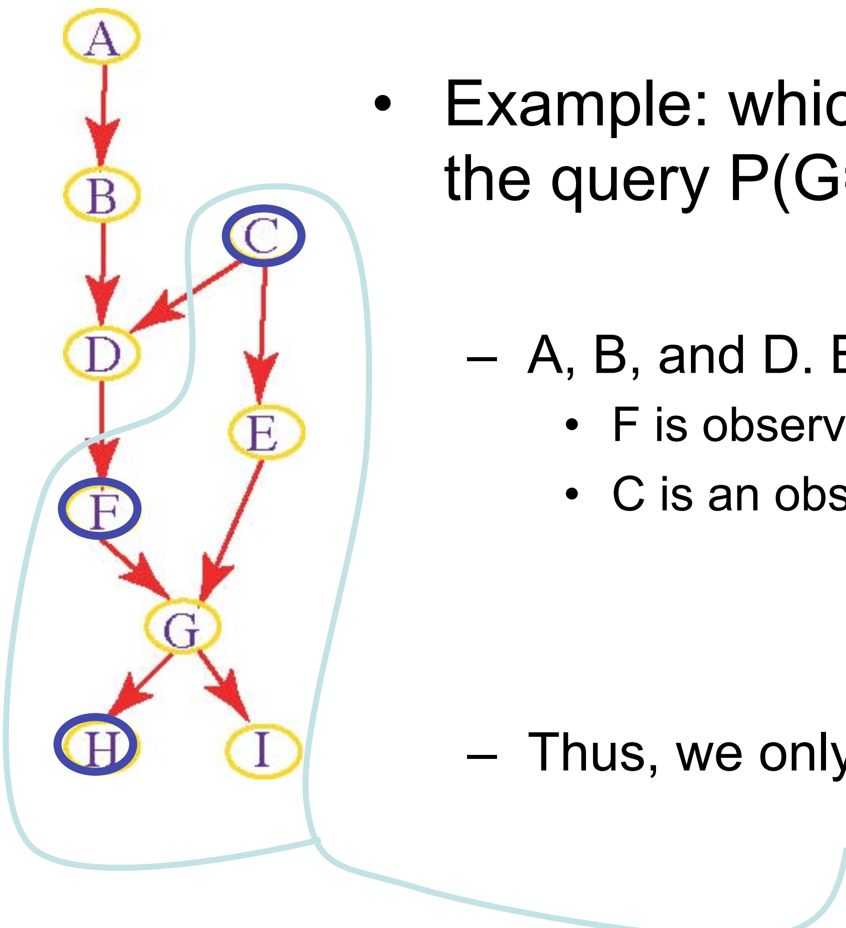


- Example: which variables can we prune for the query $P(G=g \mid C=c_1, F=f_1, H=h_1)$?

A B D E

VE and conditional independence

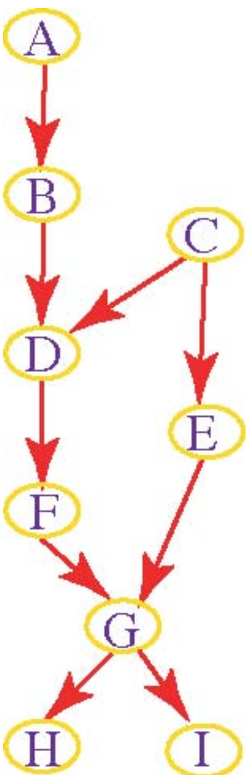
- So far, we haven't use conditional independence!
  - Before running VE, we can prune all variables Z that are conditionally independent of the query Y given evidence E: $Z \perp\!\!\!\perp Y \mid E$



- Example: which variables can we prune for the query $P(G=g \mid C=c_1, F=f_1, H=h_1)$?
  - A, B, and D. Both paths are blocked
    - F is observed node in chain structure
    - C is an observed common parent
  - Thus, we only need to consider this subnetwork

One last trick

- We can also prune unobserved leaf nodes
  - And we can do so recursively



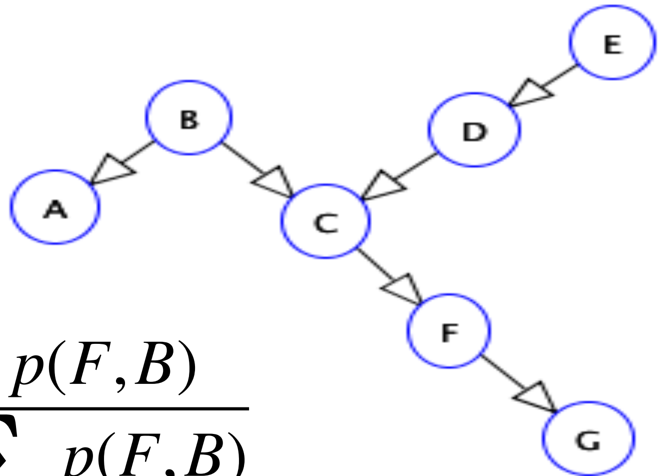
E.g., which nodes can we prune if the query is P(A)?

H   I   G   All nodes other than A

Recursively prune unobserved leaf nodes:

we can prune all nodes other than A !

Elimination of Irrelevant Variables in VE



$$p(F \mid B) = \frac{p(F,B)}{\sum_f p(F,B)}$$

$$p(F,B) = \sum_c \sum_d \sum_e p(B)p(F \mid C)p(C \mid B,D)p(D \mid E)p(E)$$

$$p(F,B) = \sum_c \sum_d p(B)p(F \mid C)p(C \mid B,D)f_1(D)$$

$$p(F,B) = \sum_c p(B)p(F \mid C)f_2(C,B)$$

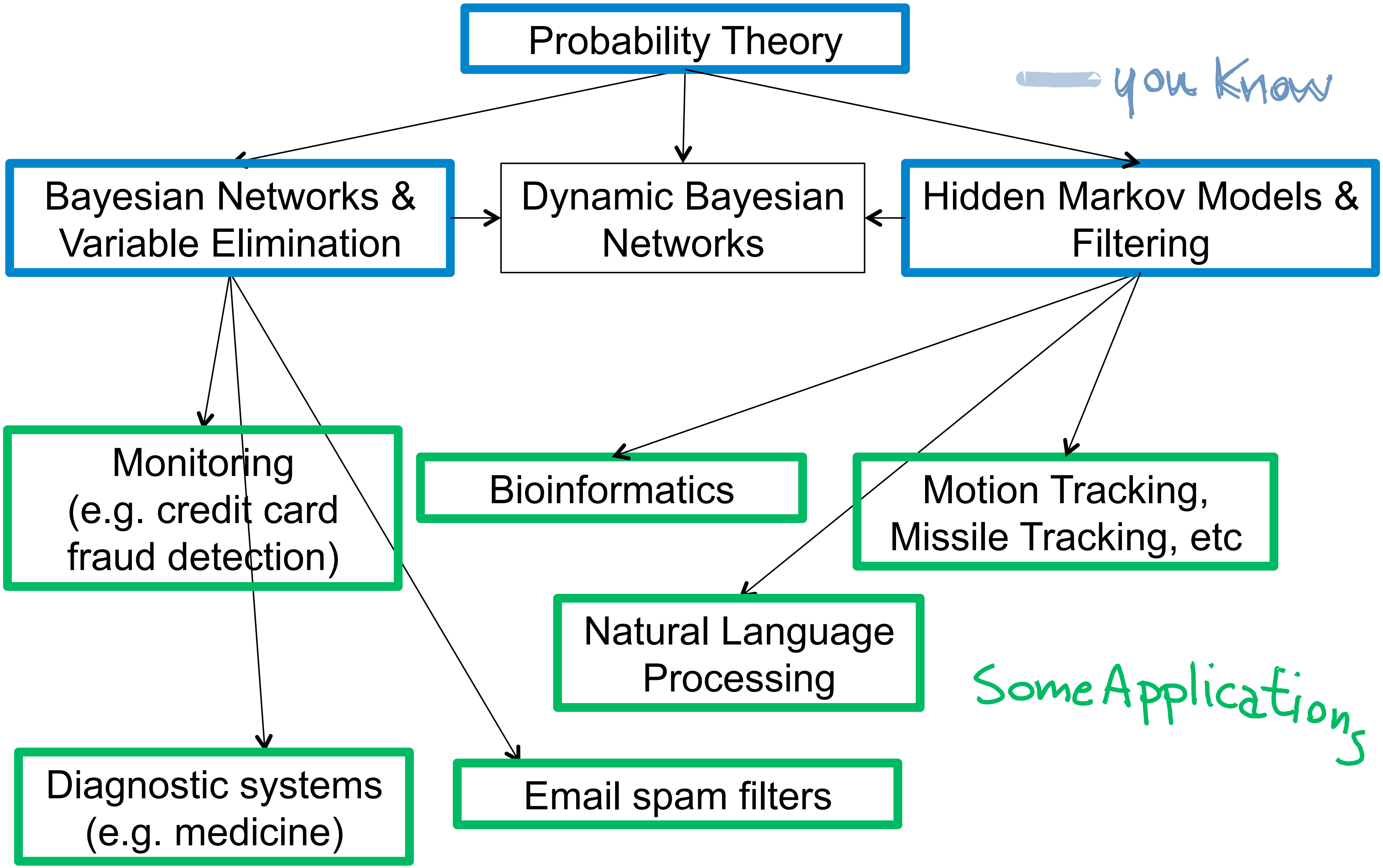
$$p(F,B) = p(B)f_3(B,F)$$

$$p(F,B) = f_4(B,F)$$

Lecture Overview

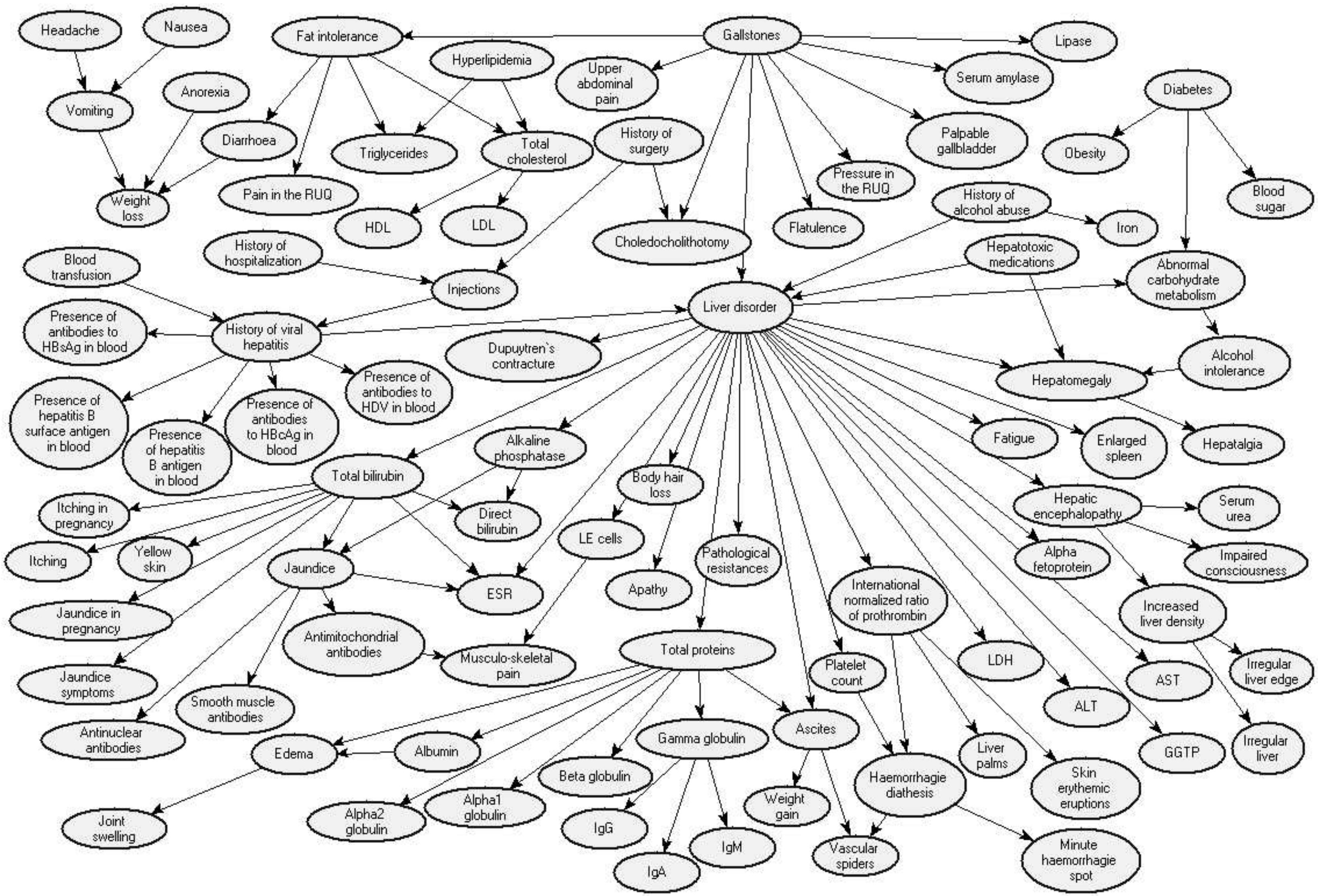
- Variable elimination: recap and some more details
- Variable elimination: pruning irrelevant variables
- Summary of Reasoning under Uncertainty
- Decision Theory
  - Intro
  - Time-permitting: Single-Stage Decision Problems

Big picture: Reasoning Under Uncertainty



One Realistic BN: Liver Diagnosis

Source: Onisko et al., 1999



~60 nodes, max 4 parents per node

Course Overview

Course Module: *Representation* — Reasoning Technique

Environment

Problem Type

Problem Type		Deterministic	Stochastic	Course Module
Static	Constraint Satisfaction	*Variables + Constraints*: Arc Consistency; Search		
Static	Logic	*Logics*: Search	*Bayesian Networks*: Variable Elimination	Uncertainty
Sequential	Planning	*STRIPS*: Search; As CSP (using arc consistency)	*Decision Networks*: Variable Elimination; *Markov Processes*: Value Iteration	Decision Theory

This concludes the uncertainty module

Course Overview

Course Module: *Representation* / Reasoning Technique

Environment

Problem Type		Deterministic	Stochastic	
Static	Constraint Satisfaction	*Variables + Constraints* — Arc Consistency, Search		
Static	Logic	*Logics* — Search	*Bayesian Networks* — Variable Elimination	Uncertainty
Sequential	Planning	*STRIPS* — Search; As CSP (using arc consistency)	*Decision Networks* — Variable Elimination; *Markov Processes* — Value Iteration	Decision Theory

But uncertainty is also at the core of decision theory: now we're acting under uncertainty

Lecture Overview

- Variable elimination: recap and some more details
- Summary of Reasoning under Uncertainty
- Decision Theory
  - Intro
  - Time-permitting: Single-Stage Decision Problems

Decisions Under Uncertainty: Intro

- Earlier in the course, we focused on decision making in deterministic domains
  - Search/CSPs: single-stage decisions
  - Planning: sequential decisions

- Now we face stochastic domains
  - so far we've considered how to represent and update beliefs
  - What if an agent has to make decisions under uncertainty?

- Making decisions under uncertainty is important
  - We mainly represent the world probabilistically so we can use our beliefs as the basis for making decisions

Decisions Under Uncertainty: Intro

- An agent's decision will depend on
  - What actions are available
  - What beliefs the agent has
  - Which goals the agent has

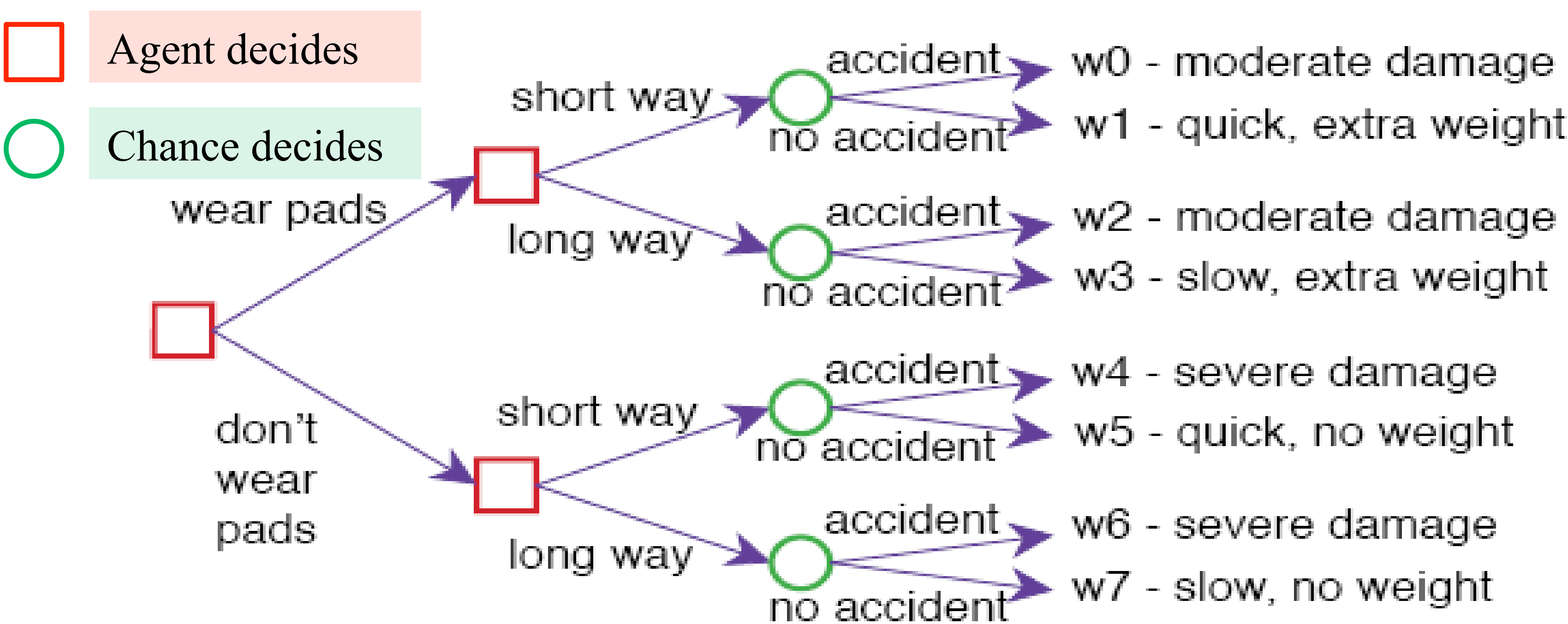
- Differences between deterministic and stochastic setting
  - Obvious difference in representation: need to represent our uncertain beliefs
  - Now we'll speak about representing actions and goals
    - Actions will be pretty straightforward: decision variables
    - Goals will be interesting: we'll move from all-or-nothing goals to a richer notion: rating how happy the agent is in different situations.
    - Putting these together, we'll extend Bayesian Networks to make a new representation called Decision Networks

Lecture Overview

- Variable elimination: recap and some more details
- Variable elimination: pruning irrelevant variables
- Summary of Reasoning under Uncertainty
- Decision Theory
  - Intro
  - Time-permitting: Single-Stage Decision Problems

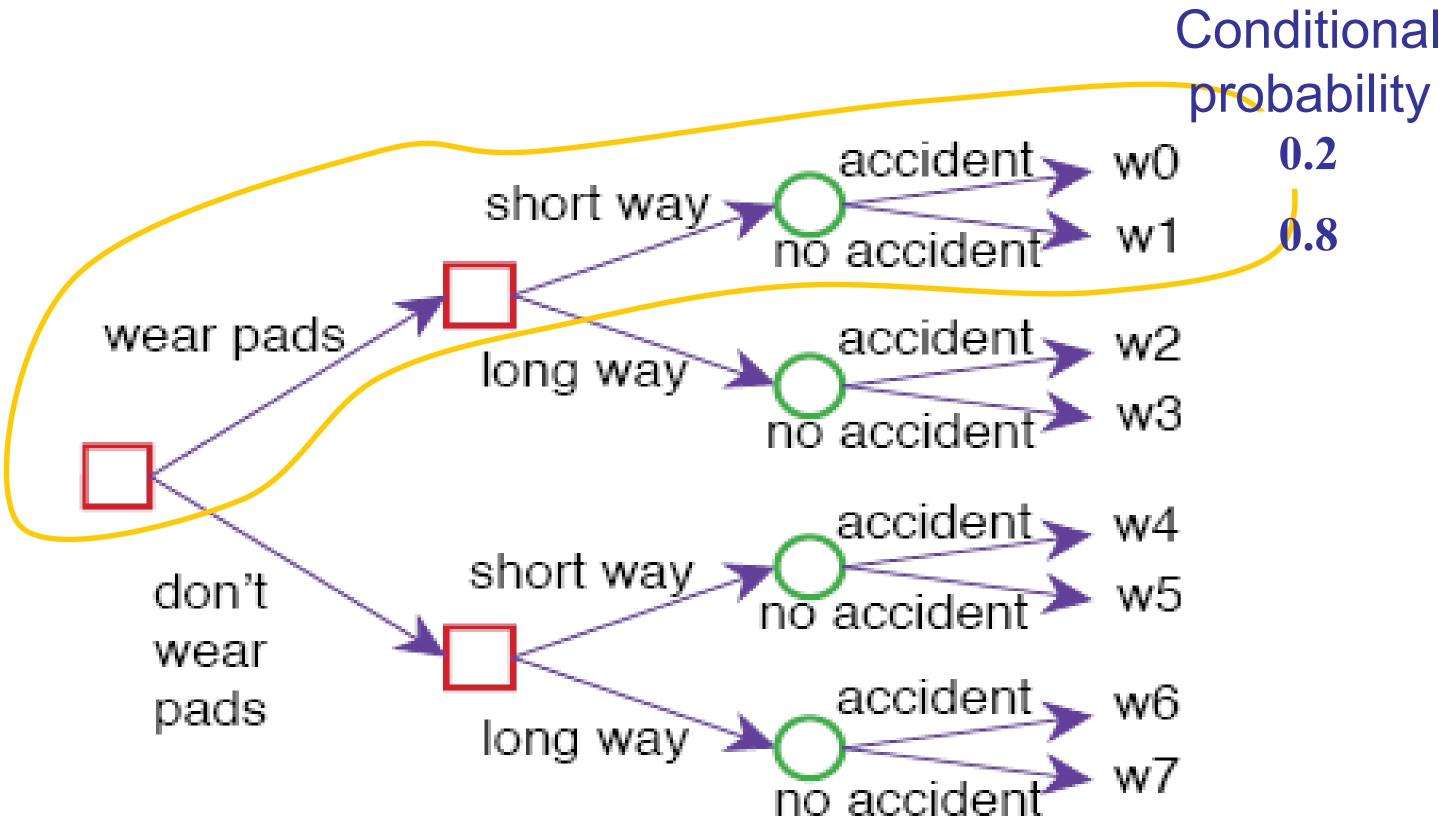
Delivery Robot Example

- Decision variable 1: the robot can choose to wear pads
  - Yes: protection against accidents, but extra weight
  - No: fast, but no protection
- Decision variable 2: the robot can choose the way
  - Short way: quick, but higher chance of accident
  - Long way: safe, but slow
- Random variable: is there an accident?



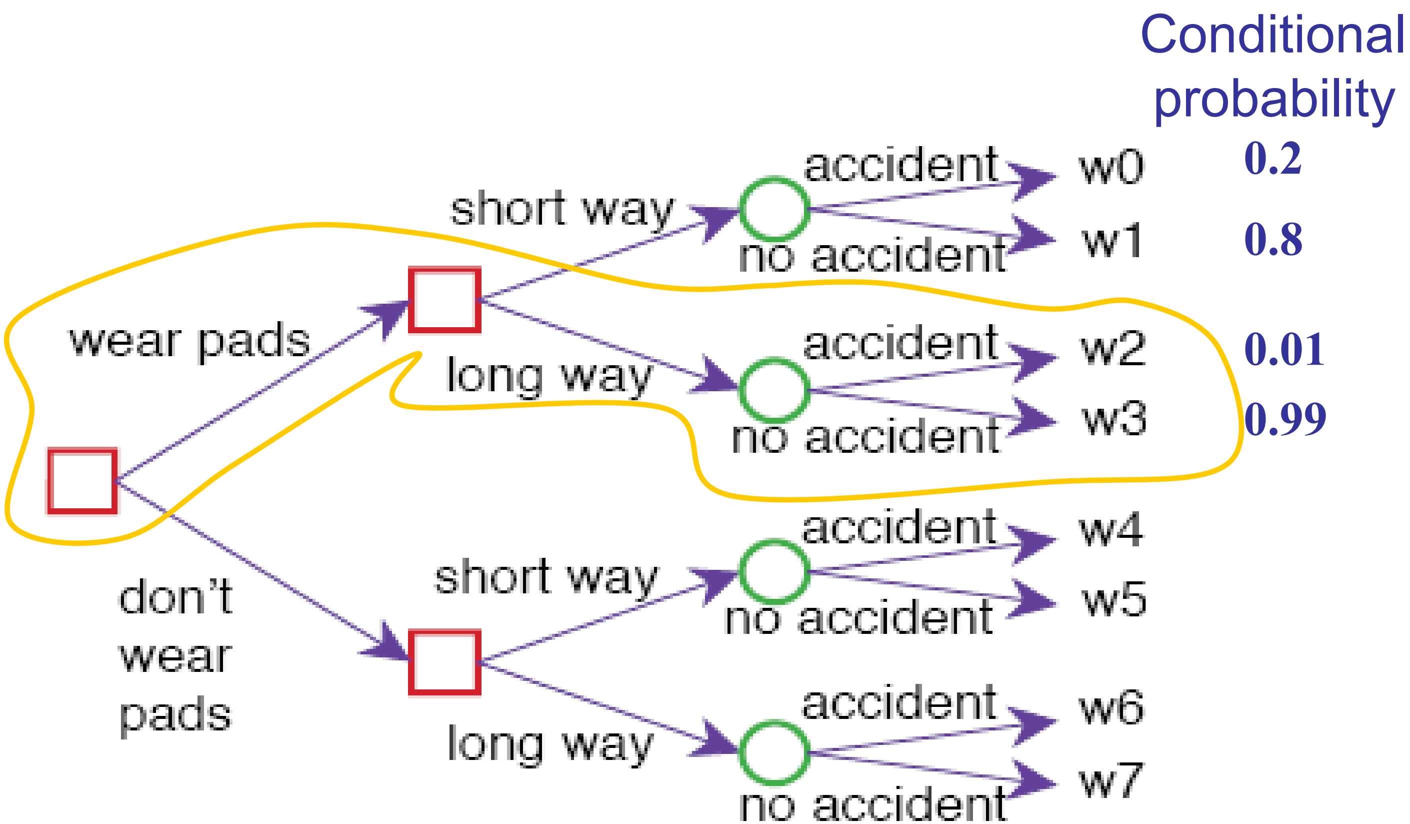
Possible worlds and decision variables

- A possible world specifies a value for each random variable and each decision variable
- For each assignment of values to all decision variables
  - the probabilities of the worlds satisfying that assignment sum to 1.



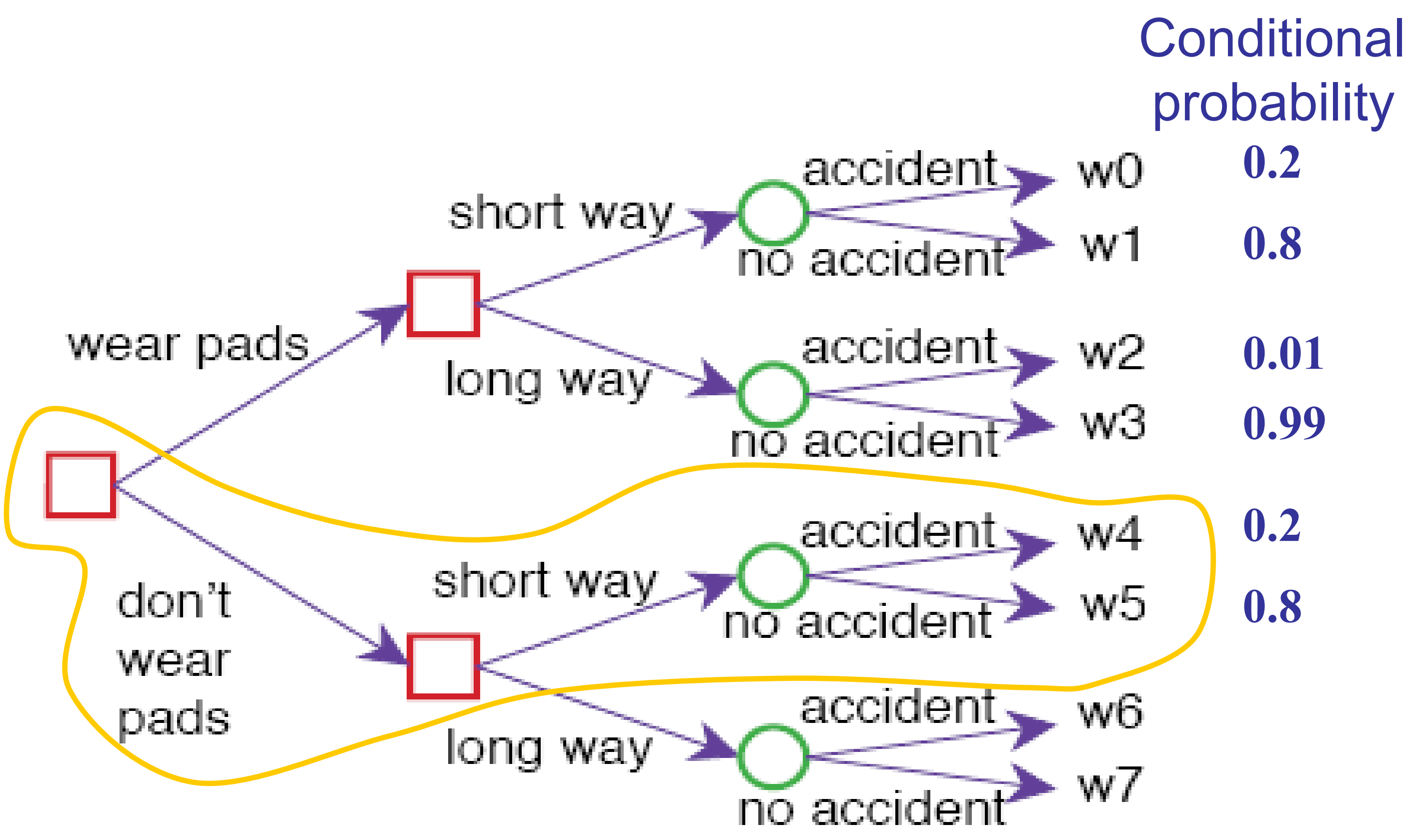
Possible worlds and decision variables

- A possible world specifies a value for each random variable and each decision variable
- For each assignment of values to all decision variables
  - the probabilities of the worlds satisfying that assignment sum to 1.



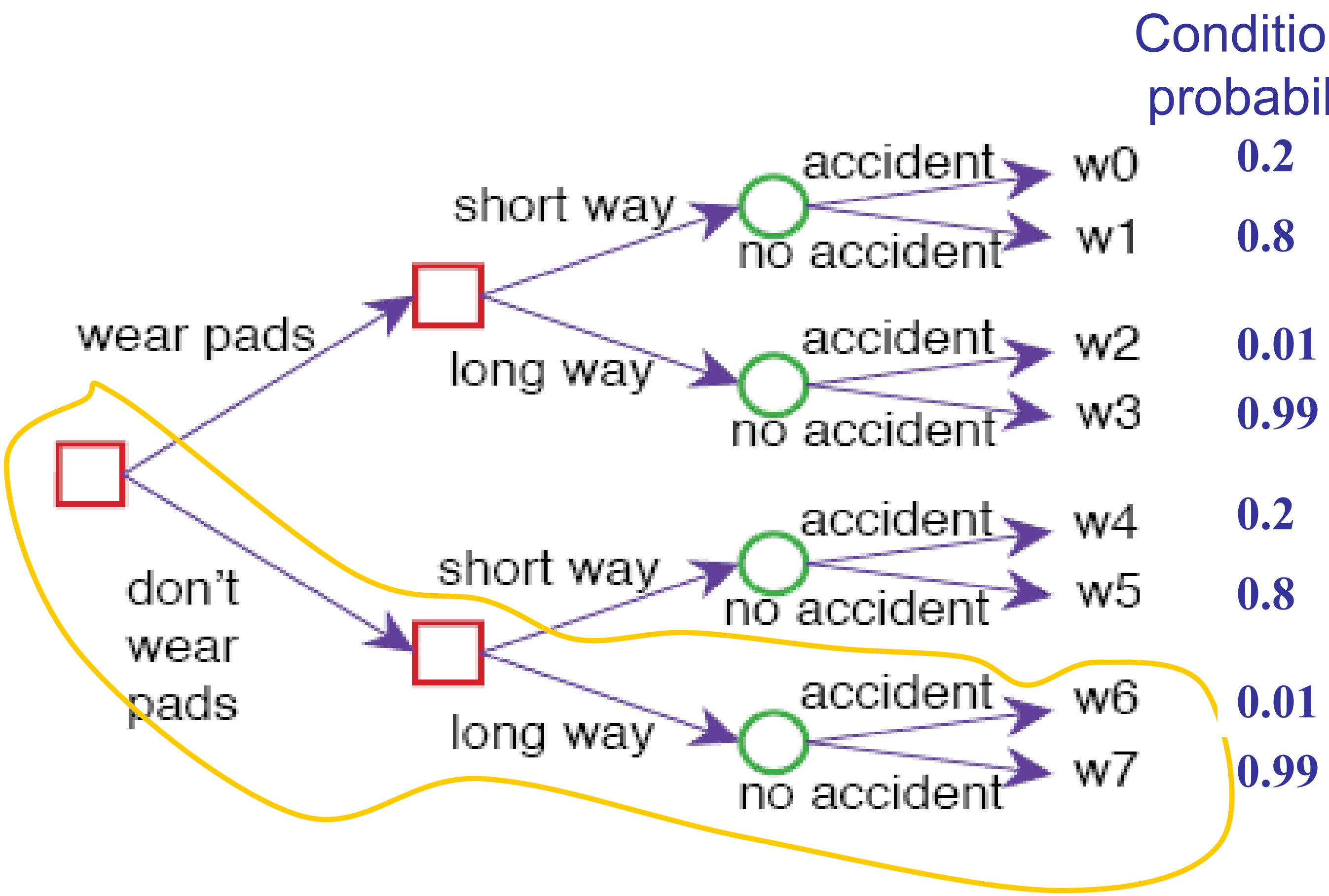
Possible worlds and decision variables

- A possible world specifies a value for each random variable and each decision variable
- For each assignment of values to all decision variables
  - the probabilities of the worlds satisfying that assignment sum to 1.



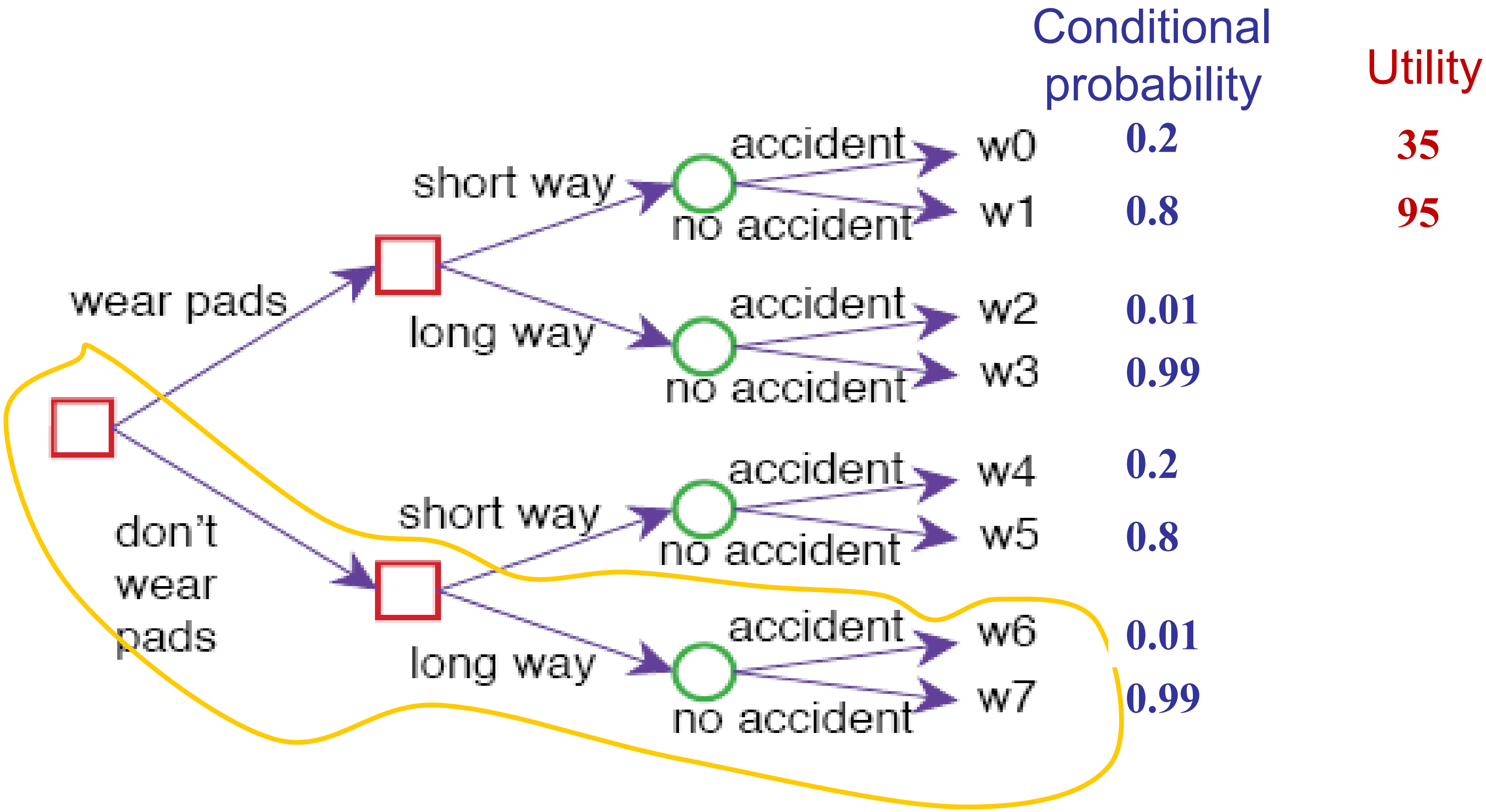
Possible worlds and decision variables

- A possible world specifies a value for each random variable and each decision variable
- For each assignment of values to all decision variables
  - the probabilities of the worlds satisfying that assignment sum to 1.



Possible worlds and decision variables

- A possible world specifies a value for each random variable and each decision variable
- For each assignment of values to all decision variables
  - the probabilities of the worlds satisfying that assignment sum to 1.



Utility

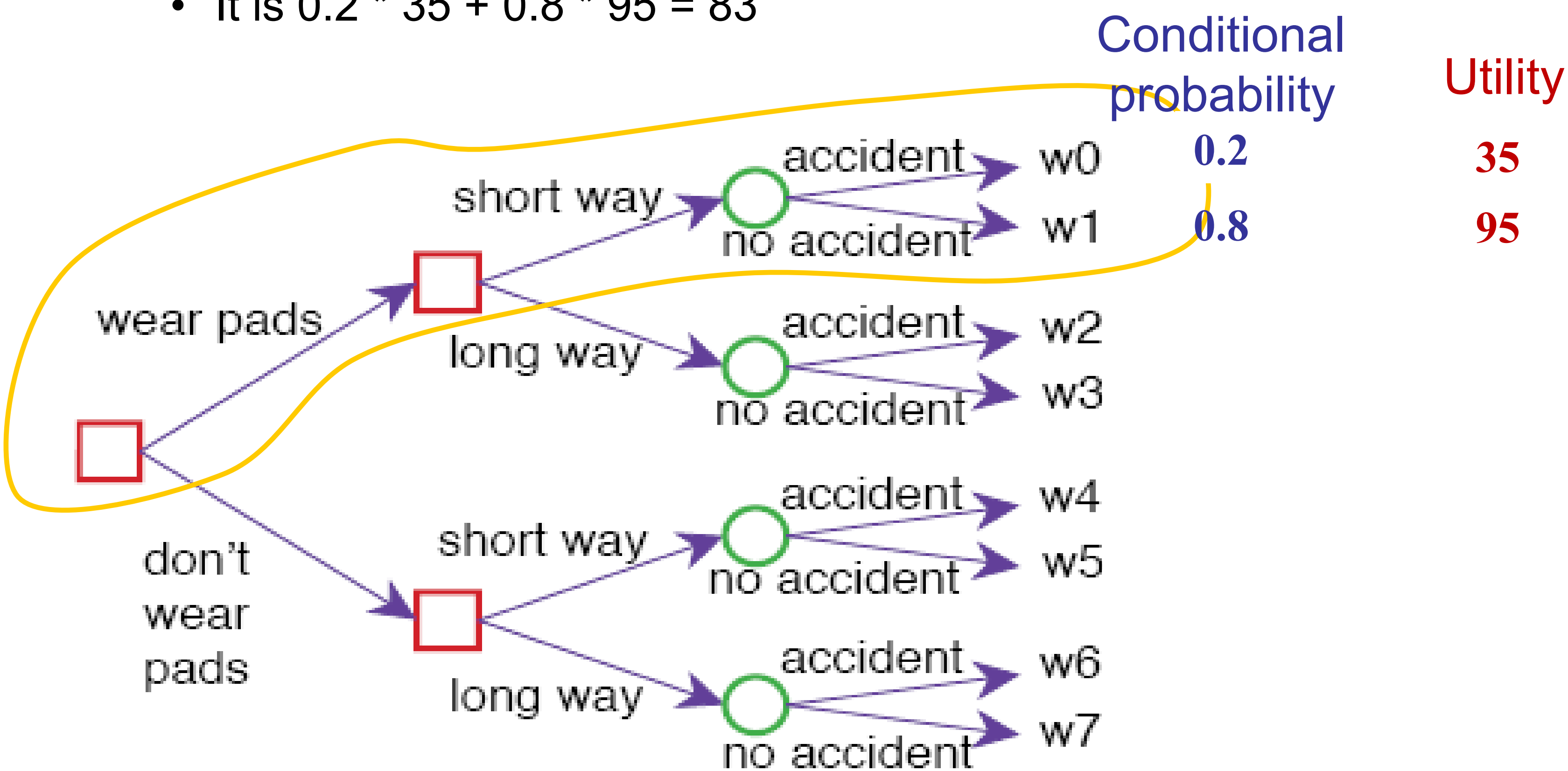
- Utility: a measure of desirability of possible worlds to an agent
  – Let U be a real-valued function such that U(w) represents an agent's degree of preference for world w
  – Expressed by a number in [0,100]

- Simple goals can still be specified
  – Worlds that satisfy the goal have utility 100
  – Other worlds have utility 0

- Utilities can be more complicated
  – For example, in the robot delivery domains, they could involve
    - Amount of damage
    - Reached the target room?
    - Energy left
    - Time taken

Combining probabilities and utilities

- We can combine probability with utility
  - The expected utility of a probability distribution over possible worlds average utility, weighted by probabilities of possible worlds
  - What is the expected utility of Wearpads=yes, Way=short ?
    - It is 0.2 * 35 + 0.8 * 95 = 83



Expected utility

- Suppose U(w) is the utility of possible world w and P(w) is the probability of possible world w

Definition (expected utility)
The expected utility is

$$E[U] = \sum_{w} P(w)U(w)$$

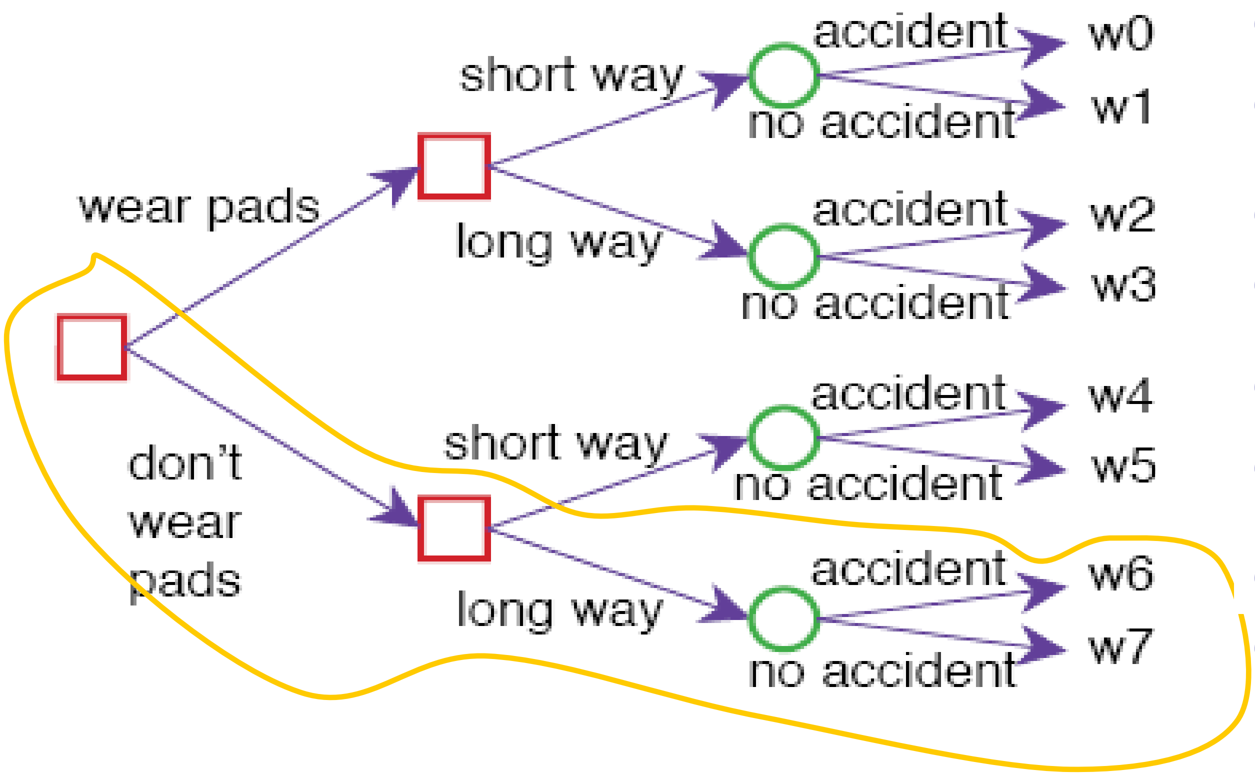
Definition (expected utility)
The conditional expected utility given e is

$$E[U|e] = \sum_{w} P(w|e)U(w)$$

Expected utility of a decision

- We write the expected utility of a decision as:

$$E[U|D = d] = \sum_{w} P(w|D = d)U(w)$$



	Conditional probability	Utility	E[U|D]
w0	0.2	35	83
w1	0.8	95	
w2	0.01	30	74.55
w3	0.99	75	
w4	0.2	3	80.6
w5	0.8	100	
w6	0.01	0	79.2
w7	0.99	80	

Optimal single-stage decision

- Given a single decision variable D
  - the agent can choose $D=d_i$ for any value $d_i \in dom(D)$

Definition (optimal single-stage decision)
An optimal single-stage decision is the decision $D=d_{max}$ whose expected value is maximal:

$$d_{max} \in \underset{d_i \in dom(D)}{\operatorname{argmax}} \; E[U|D=d_i]$$

Learning Goals For Today's Class

- Variable elimination
  - Carry out variable elimination by using factor representation and using the factor operations
- Define a Utility Function on possible worlds
- Define and compute optimal one-off decisions

- Assignment 4 is due next Wednesday
- Please complete the Teaching Evaluation