Reasoning under Uncertainty Part I

Artificial Intelligence, 2015 TIN172/DIT410

Olof Mogren
based on slides by

Poole, Mackworth

Chalmers University of Technology

April 29, 2015

Learning Objectives

At the end of the class you should be able to:

- justify the use and **semantics** of probability
- know how to compute **marginals** and apply **Bayes' theorem**
- build a **belief network** for a domain
- perform **inference** in a **belief network**
- explain the predictions of a **causal model**

Using Uncertain Knowledge

- Complete knowledge about the world not possible.
- Decisions are still needed!
- **Example:** wearing a seat belt.

Why Probability?

- Prediction approaches:
  - definitive: you will be run over tomorrow
  - point probabilities:
    $P(\text{you will be run over tomorrow}) = 0.002$
  - probability ranges:
    $P(\text{you will be run over tomorrow}) \in [0.001, 0.34]$
- Acting is gambling! Dutch books.

Byesian Probability

- Probabilities can be learned from data.
- Bayes' rule specifies how to combine data and prior knowledge.

Probability

- Probability can model one's belief in some proposition — **subjective probability.**
- An agent's belief depends on its prior assumptions and on observations.

Numerical Measures of Belief

- a - a proposition
- $P(a)$ - **probability of** a, or the belief in a, is a number between 0 and 1
  - $P(a) = 0$ - a is believed to be definitely false.
  - $P(a) = 1$ - a is believed to be definitely true.
- Using 0 and 1 is purely a convention.

Random Variables

- A **random variable** is a variable that can take one of a number of different values.
- The **range** of a variable X, written $range(X)$, is the set of values X can take.
- Assignment $X = x$ means variable X has value x.
- Each assignment to a random variable is associated to a probability, $P(X = x)$.
- A **proposition** is a Boolean formula made from assignments of values to variables.

Axioms of Probability

Three axioms define what follows from a set of probabilities:

Axiom 1 $0 \leq P(a)$ for any proposition a.

Axiom 2 $P(true) = 1$

Axiom 3 $P(a \vee b) = P(a) + P(b)$ if a and b cannot both be true.

Probability Distributions

- A probability distribution $P(X)$ on a random variable X is a function $range(X) \rightarrow [0, 1]$.
- Joint distribution of several variables: $P(X, Y, Z)$.
- A (discrete) distribution always has to sum to one:

$$\sum_{x \in range(X)} P(X = x) = 1$$

.

- For continuous random variables, the distribution has a probability density function (PDF).

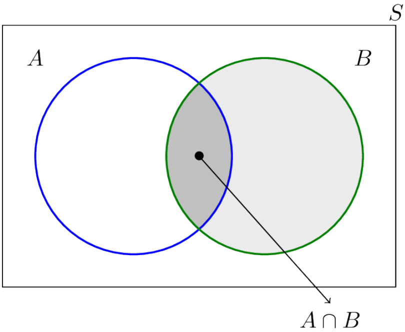
Probabilistic Conditioning

- How to revise beliefs based on new information.
- **Prior probability**: the belief before observing evidence
- Let e be the observed **evidence**, the **conditional probability** $P(h|e)$ of h given e is the **posterior probability** of h.

Conditional Probability

- The conditional probability of h given evidence e is

$$P(h|e) = \frac{P(h \wedge e)}{P(e)}$$



$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Conditional Probability: Example

We toss a die.
Someone tells you that the outcome is an even number.
What is the probability that the outcome is 6?

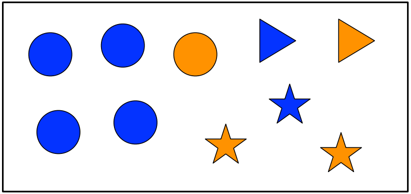
Conditional Probability: Example

We toss a die.
Someone tells you that the outcome is an even number.
What is the probability that the outcome is 6?

1/3

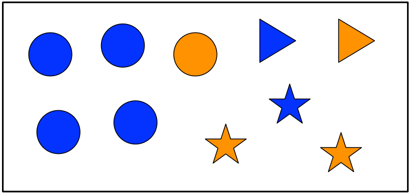
Conditioning

Possible values before evidence:

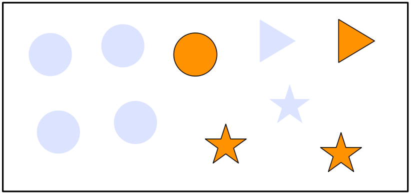


Conditioning

Possible values before evidence:



Observe $Color = orange$:



Marginals

If you have a joint probability distribution $P(X, Y)$ over some random variables X, Y, the marginal distribution $P(X)$ can be computed by summing over all values of Y:

- $P(X) = \sum_Y P(X, Y)$

Exercise

Flu	$Sneeze$	$Snore$	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$

(b) $P(flu \wedge \neg sneeze)$

(c) $P(flu)$

(d) $P(sneeze \mid flu)$

(e) $P(\neg flu \wedge sneeze)$

(f) $P(flu \mid sneeze)$

(g) $P(sneeze \mid flu \wedge snore)$

(h) $P(flu \mid sneeze \wedge snore)$

Chain Rule

$$P(f_1 \land f_2 \land \ldots \land f_n)$$
$$=$$

Chain Rule

$$
\begin{aligned}
&P(f_1 \wedge f_2 \wedge \ldots \wedge f_n) \\
&\quad = \quad P(f_n | f_1 \wedge \cdots \wedge f_{n-1}) \times \\
&\qquad\quad\; P(f_1 \wedge \cdots \wedge f_{n-1}) \\
&\quad =
\end{aligned}
$$

Chain Rule

$$
\begin{aligned}
& P(f_1 \wedge f_2 \wedge \ldots \wedge f_n) \\
& \quad = \quad P(f_n | f_1 \wedge \cdots \wedge f_{n-1}) \times \\
& \qquad\quad\; P(f_1 \wedge \cdots \wedge f_{n-1}) \\
& \quad = \quad P(f_n | f_1 \wedge \cdots \wedge f_{n-1}) \times \\
& \qquad\quad\; P(f_{n-1} | f_1 \wedge \cdots \wedge f_{n-2}) \times \\
& \qquad\quad\; P(f_1 \wedge \cdots \wedge f_{n-2}) \\
& \quad = \quad P(f_n | f_1 \wedge \cdots \wedge f_{n-1}) \times \\
& \qquad\quad\; P(f_{n-1} | f_1 \wedge \cdots \wedge f_{n-2}) \\
& \qquad\quad\; \times \cdots \times P(f_3 | f_1 \wedge f_2) \times P(f_2 | f_1) \times P(f_1) \\
& \quad = \quad \prod_{i=1}^{n} P(f_i | f_1 \wedge \cdots \wedge f_{i-1})
\end{aligned}
$$

Bayes' theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$P(h \wedge e) \quad =$$

Bayes' theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$P(h \wedge e) = P(h|e) \times P(e)$$

Bayes' theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$\begin{aligned} P(h \wedge e) &= P(h|e) \times P(e) \\ &= P(e|h) \times P(h). \end{aligned}$$

Bayes' theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$\begin{aligned} P(h \wedge e) &= P(h|e) \times P(e) \\ &= P(e|h) \times P(h). \end{aligned}$$

If $P(e) \neq 0$, divide the right hand sides by $P(e)$:

$$P(h|e) =$$

Bayes' theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$\begin{aligned} P(h \wedge e) &= P(h|e) \times P(e) \\ &= P(e|h) \times P(h). \end{aligned}$$

If $P(e) \neq 0$, divide the right hand sides by $P(e)$:

$$P(h|e) = \frac{P(e|h) \times P(h)}{P(e)}.$$

This is **Bayes' theorem.**

Why is Bayes' theorem interesting?

- Often you have causal knowledge:
  $P(symptom \mid disease)$
- and want to do evidential reasoning:
  $P(disease \mid symptom)$

Exercise

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given:

- 85% of the cabs in the city are Green and 15% are Blue.
- A witness identified the cab as Blue.
- The witness reliability is 80%.

What is the probability that the cab involved in the accident was Blue?

D. Kahneman, Thinking Fast and Slow, 2011, p. 166.

Exercise: Solution

$P(\text{cab is blue}|\text{witness says cab is blue}) =$

$\frac{P(\text{witness says blue}|\text{cab is blue}) \times P(\text{cab is blue})}{P(\text{witness says blue})} =$

$\frac{0.8 \times 0.15}{0.29} \approx$

0.41

(The normalizing constant ($P(\text{witness says blue})$) can be computed by marginalizing (summing over hypotheses): $0.8 * 0.15 + 0.2 * 0.85 = 0.29$.)

Conditional independence

Random variable X is **independent** of random variable Y **given** random variable Z if,

$$P(X|Y,Z) = P(X|Z)$$

Conditional independence

Random variable X is **independent** of random variable Y **given** random variable Z if,

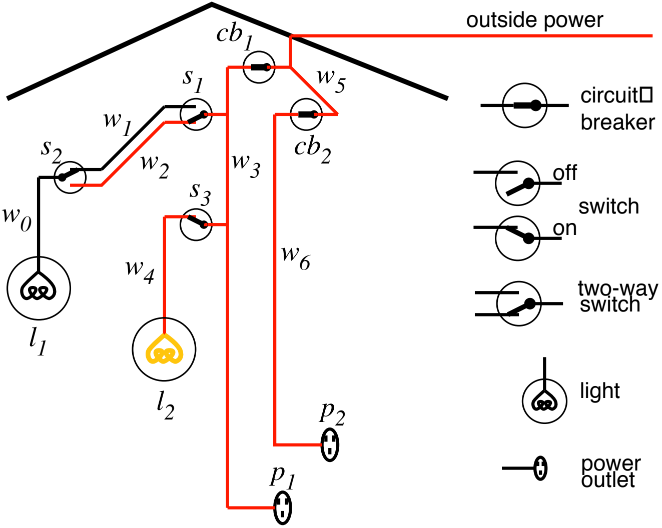
$$P(X|Y,Z) = P(X|Z)$$

i.e. for all $x_i \in dom(X)$, $y_j \in dom(Y)$, $y_k \in dom(Y)$ and $z_m \in dom(Z)$,

$$\begin{aligned} & P(X = x_i|Y = y_j, Z = z_m) \\ = \quad & P(X = x_i|Y = y_k, Z = z_m) \\ = \quad & P(X = x_i|Z = z_m). \end{aligned}$$

That is, knowledge of Y's value doesn't affect the belief in the value of X, given a value of Z.

Example domain (diagnostic assistant)



Examples of conditional independence?

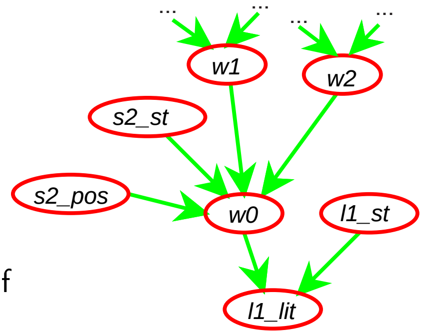
- The identity of the queen of Canada is dependent or independent of whether light $l1$ is lit given whether there is outside power?
- Whether there is someone in a room is independent of whether a light $l2$ is lit given what?
- Whether light $l1$ is lit is independent of the position of light switch $s2$ given what?
- Every other variable may be independent of whether light $l1$ is lit given

Examples of conditional independence?

- The identity of the queen of Canada is dependent or independent of whether light $l1$ is lit given whether there is outside power?
- Whether there is someone in a room is independent of whether a light $l2$ is lit given what?
- Whether light $l1$ is lit is independent of the position of light switch $s2$ given what?
- Every other variable may be independent of whether light $l1$ is lit given whether there is power in wire w_0 and the status of light $l1$ (if it's ok, or if not, how it's broken).

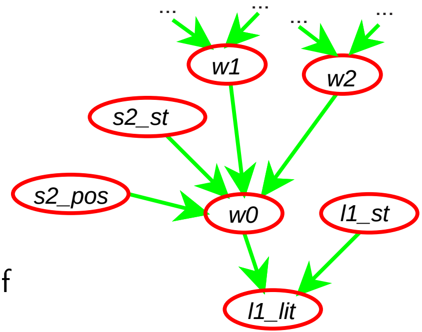
Idea of belief networks

- $l1$ is lit ($L1\_lit$) depends only on the status of the light ($L1\_st$) and whether there is power in wire $w0$.
- In a belief network, $W0$ and $L1\_st$ are **parents** of $L1\_lit$.
- $W0$ depends only on



Idea of belief networks

- $l1$ is lit ($L1\_lit$) depends only on the status of the light ($L1\_st$) and whether there is power in wire $w0$.
- In a belief network, $W0$ and $L1\_st$ are **parents** of $L1\_lit$.



- $W0$ depends only on whether there is power in $w1$, whether there is power in $w2$, the position of switch $s2$ ($S2\_pos$), and the status of switch $s2$ ($S2\_st$).

Belief networks

- Totally order the variables of interest: $X_1, \ldots, X_n$
- Theorem of probability theory (chain rule): $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1})$
- The **parents** $parents(X_i)$ of X_i are those predecessors of X_i that render X_i independent of the other predecessors. That is,

Belief networks

- Totally order the variables of interest: $X_1, \ldots, X_n$
- Theorem of probability theory (chain rule): $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1})$
- The **parents** $parents(X_i)$ of X_i are those predecessors of X_i that render X_i independent of the other predecessors. That is, $parents(X_i) \subseteq X_1, \ldots, X_{i-1}$ and $P(X_i | parents(X_i)) = P(X_i | X_1, \ldots, X_{i-1})$
- So $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | parents(X_i))$
- A **belief network** is a graph: the nodes are random variables; there is an arc from the parents of each node into that node.

Example: fire alarm belief network

Variables:

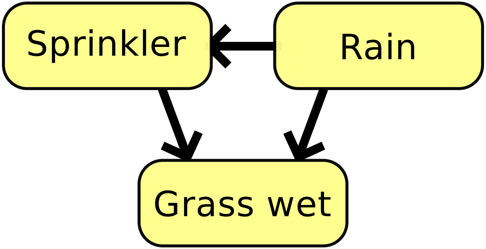
- Fire: there is a fire in the building
- Tampering: someone has been tampering with the fire alarm
- Smoke: what appears to be smoke is coming from an upstairs window
- Alarm: the fire alarm goes off
- Leaving: people are leaving the building *en masse*.
- Report: a colleague says that people are leaving the building *en masse*. (A noisy sensor for leaving.)

Components of a belief network

A belief network consists of:

- a directed acyclic graph with nodes labeled with random variables
- a domain for each random variable
- a set of conditional probability tables for each variable given its parents (including prior probabilities for nodes with no parents).

Example belief network

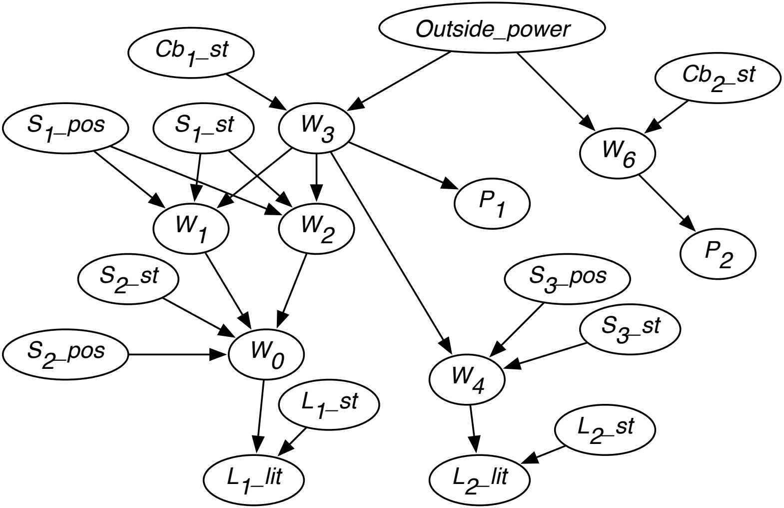


Components of a belief network

A belief network consists of:

- a directed acyclic graph with nodes labeled with random variables
- a domain for each random variable
- a set of conditional probability tables for each variable given its parents (including prior probabilities for nodes with no parents).

Example belief network



Example belief network (continued)

The belief network also specifies:

- The domain of the variables:
  $W_0, \ldots, W_6$ have domain $\{live, dead\}$
  S_1_pos, S_2_pos, and S_3_pos have domain $\{up, down\}$
  S_1_st has
  $\{ok, upside\_down, short, intermittent, broken\}$.
- Conditional probabilities, such as:
  $P(W_1 = live | s_1\_pos = up \wedge S_1\_st = ok \wedge W_3 = live)$

Belief network summary

- A belief network is a directed acyclic graph (DAG).
- Its nodes are random variables.
- The **parents** of X are those that X directly depends on.
- Acyclic by construction.
- A representation of **dependence** and **independence**:
  - X is independent of its non-descendants given its parents.

Constructing belief networks

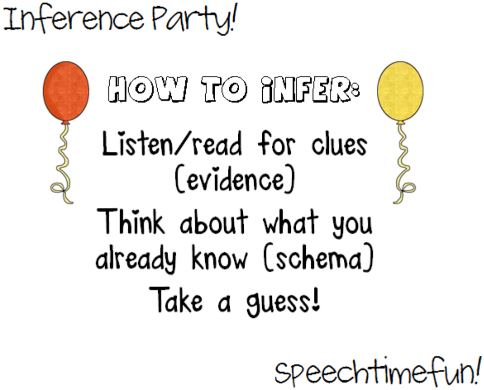
- What are the relevant variables?
  - Observed?
  - Query variables?
  - Variables that make the model simpler?
- What values should these variables take?
- What is the relationship between them?
- How does the value of each variable depend on its parents?

Using belief networks

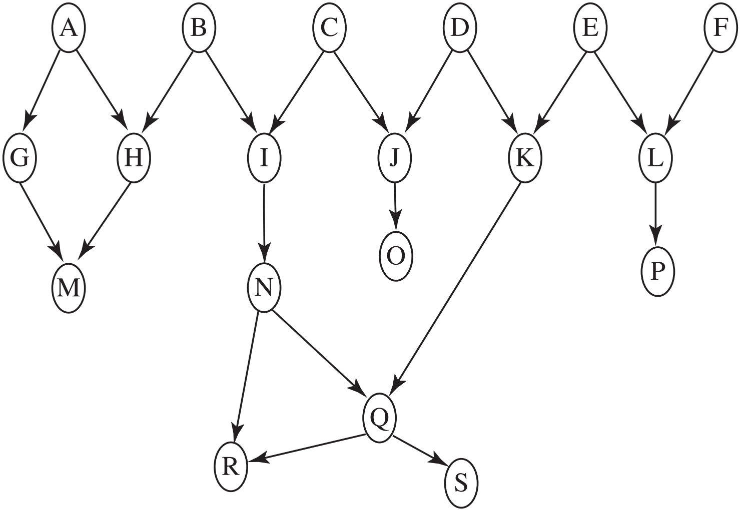
An example of how the power network can be used:

- Given values for:
  - switches,
  - outside power,
  - whether the lights are lit,
- you can determine the posterior probability that each switch or circuit breaker is ok or not.

This is called **inference**.



Understanding independence: example



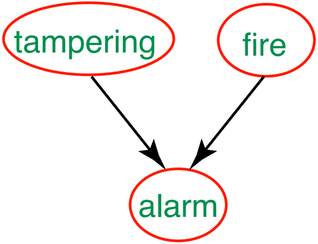
Understanding independence: questions

- On which given probabilities does $P(N)$ depend?
- If you were to observe a value for B, which variables' probabilities will change?
- If you were to observe a value for N, which variables' probabilities will change?
- Suppose you had observed a value for M; if you were to then observe a value for N, which variables' probabilities will change?
- Suppose you had observed B and Q; which variables' probabilities will change when you observe N?

What variables are affected by observing?

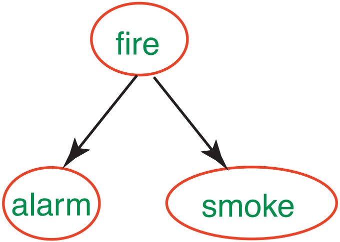
- If you observe variable $\overline{Y}$, the variables whose posterior probability is different from their prior are:
  - The ancestors of $\overline{Y}$ and
  - their descendants.
- Intuitively (if you have a causal belief network):
  - You do **abduction** to possible causes and
  - **prediction** from the causes.

Common descendants



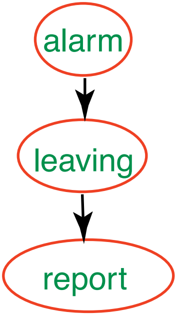
- $tampering$ and $fire$ are independent
- $tampering$ and $fire$ are dependent given $alarm$
- Intuitively, $tampering$ can **explain away** $fire$

Common ancestors



- $alarm$ and $smoke$ are dependent
- $alarm$ and $smoke$ are independent given $fire$
- Intuitively, $fire$ can **explain** $alarm$ and $smoke$; learning one can affect the other by changing your belief in $fire$.

Chain



- $alarm$ and $report$ are dependent
- $alarm$ and $report$ are independent given $leaving$
- Intuitively, the only way that the $alarm$ affects $report$ is by affecting $leaving$.

Belief network inference

- Variable Elimination: exploit the structure of the network to eliminate (sum out) the non-observed, non-query variables one at a time.
- Search-based approaches: enumerate some of the possible assignments, and estimate posterior probabilities.
- Stochastic simulation: generate random assignments according to the probability distributions.
- Variational methods: find the closest tractable distribution to the (posterior) distribution.

Factors

Function from a set of random variables to a number.
$f(X_1, \ldots, X_j)$.
Some or all of the variables of a factor can be assigned:

- $f(X_1 = v_1, X_2, \ldots, X_j)$, is a factor on $X_2, \ldots, X_j$.
- $f(X_1 = v_1, X_2 = v_2, \ldots, X_j = v_j)$ is a number that is the value of f when each X_i has value v_i.

Example factors

$r(X, Y, Z)$:

X	Y	Z	val
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

$r(X{=}t, Y, Z)$:

Y	Z	val
t	t	0.1
t	f	
f	t	
f	f	

Example factors

$r(X, Y, Z)$:

X	Y	Z	val
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

$r(X{=}t, Y, Z)$:

Y	Z	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

$r(X{=}t, Y, Z{=}f)$:

Example factors

$r(X, Y, Z)$:

X	Y	Z	val
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

$r(X{=}t, Y, Z)$:

Y	Z	val
t	t	0.1
t	f	
f	t	
f	f	

$r(X{=}t, Y, Z{=}f)$:

Y	val
t	
f	

$r(X{=}t, Y{=}f, Z{=}f) =$

Example factors

$r(X, Y, Z)$:

X	Y	Z	val
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

$r(X{=}t, Y, Z)$:

Y	Z	val
t	t	0.1
t	f	
f	t	
f	f	

$r(X{=}t, Y, Z{=}f)$:

Y	val
t	0.9
f	0.8

$$r(X{=}t, Y{=}f, Z{=}f) = 0.8$$

Multiplying factors

The **product** of factor $f_1(\overline{X}, \overline{Y})$ and $f_2(\overline{Y}, \overline{Z})$, where $\overline{Y}$ are the variables in common, is the factor $(f_1 \times f_2)(\overline{X}, \overline{Y}, \overline{Z})$ defined by:

$$(f_1 \times f_2)(\overline{X}, \overline{Y}, \overline{Z}) \quad = \quad f_1(\overline{X}, \overline{Y}) f_2(\overline{Y}, \overline{Z}).$$

Multiplying factors example

f_1:

A	B	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

f_2:

B	C	val
t	t	0.3
t	f	0.7
f	t	0.6
f	f	0.4

$f_1 \times f_2$:

A	B	C	val
t	t	t	0.03
t	t	f	
t	f	t	
t	f	f	
f	t	t	
f	t	f	
f	f	t	
f	f	f	

Multiplying factors example

f_1:

A	B	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

f_2:

B	C	val
t	t	0.3
t	f	0.7
f	t	0.6
f	f	0.4

$f_1 \times f_2$:

A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

Summing out variables

We can **sum out** a variable, say X_1 with domain $\{v_1, \ldots, v_k\}$, from factor $f(X_1, \ldots, X_j)$, resulting in a factor on $X_2, \ldots, X_j$ defined by:

$$
\begin{aligned}
& (\sum_{X_1} f)(X_2, \ldots, X_j) \\
& = \quad f(X_1 = v_1, \ldots, X_j) + \cdots + f(X_1 = v_k, \ldots, X_j)
\end{aligned}
$$

Summing out a variable example

f_3:

A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

$\sum_B f_3$:

A	C	val
t	t	0.57
t	f	
f	t	
f	f	

Summing out a variable example

f_3:

A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

$\sum_B f_3$:

A	C	val
t	t	0.57
t	f	0.43
f	t	0.54
f	f	0.46

Exercise

Given factors:

s:

A	val
t	0.75
f	0.25

t:

A	B	val
t	t	0.6
t	f	0.4
f	t	0.2
f	f	0.8

o:

A	val
t	0.3
f	0.1

What is?

(a) $s \times t$

(b) $\sum_A s \times t$

(c) $\sum_B s \times t$

(d) $s \times t \times o$

(e) $\sum_A s \times t \times o$

(f) $\sum_b s \times t \times o$

Evidence

If we want to compute the posterior probability of Z given evidence $Y_1 = v_1 \wedge \ \ldots \wedge Y_j = v_j$:

$$P(Z|Y_1 = v_1, \ldots, Y_j = v_j)$$

$$=$$

Evidence

If we want to compute the posterior probability of Z given evidence $Y_1 = v_1 \wedge \ldots \wedge Y_j = v_j$:

$$
\begin{aligned}
& P(Z | Y_1 = v_1, \ldots, Y_j = v_j) \\
= \quad & \frac{P(Z, Y_1 = v_1, \ldots, Y_j = v_j)}{P(Y_1 = v_1, \ldots, Y_j = v_j)} \\
= \quad &
\end{aligned}
$$

Evidence

If we want to compute the posterior probability of Z given evidence $Y_1 = v_1 \wedge \ldots \wedge Y_j = v_j$:

$$\begin{aligned} & P(Z|Y_1 = v_1, \ldots, Y_j = v_j) \\ = \quad & \frac{P(Z, Y_1 = v_1, \ldots, Y_j = v_j)}{P(Y_1 = v_1, \ldots, Y_j = v_j)} \\ = \quad & \frac{P(Z, Y_1 = v_1, \ldots, Y_j = v_j)}{\sum_Z P(Z, Y_1 = v_1, \ldots, Y_j = v_j).} \end{aligned}$$

So the computation reduces to the probability of $P(Z, Y_1 = v_1, \ldots, Y_j = v_j)$.
We normalize at the end.