CPSC 322: Introduction to Artificial Intelligence

Uncertainty: Inference by Enumeration and Independence

Textbook reference: [8.2]

Instructor: Varada Kolhatkar
University of British Columbia

Credit: These slides are adapted from the slides of the previous offerings of the course. Thanks to all instructors for creating and improving the teaching material and making it available!

Announcements

- Assignment 4 has been released.
  - Due date: **Nov 29th, 11:59 PM**
- Final exam scheduled:
  - **Time:** Dec 9 at 7:00pm
  - **Location:** SRC A

A rough CPSC 322 overview

Representation and reasoning

Environment

	Problem	Deterministic	Stochastic
Static	Constraint satisfaction	Variables + constraints Arc consistency Search	
	Query	Logics Search	Belief networks Variable elimination
Sequential	Planning	STRIPS Search	Decision networks Variable elimination Markov decision processes Value iteration

Lecture outline

- Recap 👈
- Inference with enumeration
- Marginal independence
- Conditional independence
- Bayesian networks (time permitting)

Recap: Random variables

Random variables: Some aspect of the world about which we may have uncertainty. Like CSP variables, random variables have **domains**.

Examples:

S: Is it sunny? $S \in \{yes, no\}$

M: What will be my mood tomorrow?
$M \in \{happy, sad, disgusted\}$

B: How many bikes are parked at the moment on UBC campus? $B \in \{0, 1, 2, \ldots, 100000\}$

Recap: Probability distributions

Unobserved random variables have distributions.

P(M)

Mood (M)	P
happy	0.60
sad	0.20
disgusted	0.20

P(W)

Weather (W)	P
sunny	0.40
cloudy	0.60

A distribution is a table of probabilities of values.

$\forall x P(X = x) \geq 0$ and $\sum_x P(X = x) = 1$

Recap: Marginal distributions

Marginal distributions are sub-tables for a subset of variables.

What are the marginal distributions for P(Activity) and P(Mood)?

Activity (A)	Mood (M)	P
study	happy	0.45
cry	happy	0.05
study	sad	0.15
cry	sad	0.35

P(Activity)

Activity (A)	P

P(Mood)

Mood (M)	P

Recap: Marginal distributions

Marginal distributions are sub-tables for a subset of variables.

What are the marginal distributions for P(Activity) and P(Mood)?

Activity (A)	Mood (M)	P
study	happy	0.45
cry	happy	0.05
study	sad	0.15
cry	sad	0.35

P(Activity)

Activity (A)	P
study	0.60
cry	0.40

P(Mood)

Mood (M)	P
happy	0.50
sad	0.50

Recap: Conditional probability (pair-share)

What is the **conditional probability** P(Activity = study|Mood = happy)?

Whats the **conditional distribution** of P(Activity|Mood = happy)?

Activity (A)	Mood (M)	P
study	happy	0.45
cry	happy	0.05
study	sad	0.15
cry	sad	0.35

P(Activity | Mood = happy)

Activity (A)	P(A | M = happy)
study	
cry	

Recap: Conditional probability (pair-share)

What is the **conditional probability** P(Activity = study|Mood = happy)?

Whats the **conditional distribution** of P(Activity|Mood = happy)?

Activity (A)	Mood (M)	P
study	happy	0.45
cry	happy	0.05
~~study~~	~~sad~~	~~0.15~~
~~cry~~	~~sad~~	~~0.35~~

P(Activity | Mood = happy)

Activity (A)	P(A | M = happy)
study	0.45/(0.45 + 0.05) = 0.90
cry	0.05/(0.45 + 0.05) = 0.10

Today: Learning outcomes

From this lecture, students are expected to be able to:

- Use inference by enumeration
    - to compute joint posterior probability distributions over any subset of variables given evidence
- Define and use marginal independence
- Define and use conditional independence

Lecture outline

- Recap
- Inference with enumeration 👈
- Marginal independence
- Conditional independence
- Bayesian networks (time permitting)

Probabilistic inference

- Compute a desired probability from other known probabilities
- We generally compute conditional probabilities
  Example: P(study | happy) = 0.90
  - These represent the agent's beliefs given the evidence
- Probabilities change with new evidence:
  Example: P(study | happy, exams_finished) = 0.60

Inference by enumeration

With conditional probability, we can compute arbitrary probabilities now!

Given: **Prior** joint probability distribution (JPD) on set of variables X and specific values e for the **evidence** variables E (subset of X).

We want to compute: **Posterior** joint distribution of query variables Y (a subset of X) given evidence e.

Step 1: Condition to get distribution $P(X \mid e)$

Step 2: Marginalize to get distribution $P(Y \mid e)$

Inference by enumeration

Given $P(X)$ as JPD below, and evidence e = "Wind=yes"
What is the probability it is **hot**?
Step 1: condition to get distribution $P(X|e)$

P(Temperature=hot | Wind=yes)?

Wind (N)	Weather (W)	Temperature (T)	P(N,W,T)
yes	sunny	hot	0.04
yes	sunny	mild	0.09
yes	sunny	cold	0.07
yes	cloudy	hot	0.01
yes	cloudy	mild	0.10
yes	cloudy	cold	0.12
no	sunny	hot	0.06
no	sunny	mild	0.11
no	sunny	cold	0.03
no	cloudy	hot	0.04
no	cloudy	mild	0.25
no	cloudy	cold	0.08

Inference by enumeration

Given $P(X)$ as JPD below, and evidence e = "Wind=yes"
What is the probability it is **hot**?
Step 1: condition to get distribution $P(X|e)$

P(Temperature=hot | Wind=yes)?

Wind (N)	Weather (W)	Temperature (T)	P(N,W,T)
yes	sunny	hot	0.04
yes	sunny	mild	0.09
yes	sunny	cold	0.07
yes	cloudy	hot	0.01
yes	cloudy	mild	0.10
yes	cloudy	cold	0.12
~~no~~	~~sunny~~	~~hot~~	~~0.06~~
~~no~~	~~sunny~~	~~mild~~	~~0.11~~
~~no~~	~~sunny~~	~~cold~~	~~0.03~~
~~no~~	~~cloudy~~	~~hot~~	~~0.04~~
~~no~~	~~cloudy~~	~~mild~~	~~0.25~~
~~no~~	~~cloudy~~	~~cold~~	~~0.08~~

Weather (W)	Temperature (T)	P(W,T | N = yes)
sunny	hot	
sunny	mild	
sunny	cold	
cloudy	hot	
cloudy	mild	
cloudy	cold	

$$P(W = w \wedge T = t \,|\, N = yes)$$

$$= \frac{P(W = w \wedge T = t \wedge N = yes)}{P(N = yes)}$$

Inference by enumeration

Given $P(X)$ as JPD below, and evidence e = "Wind=yes"
What is the probability it is **hot**?
Step 1: condition to get distribution $P(X \mid e)$

P(Temperature=hot | Wind=yes)?

Wind (N)	Weather (W)	Temperature (T)	P(N,W,T)
yes	sunny	hot	0.04
yes	sunny	mild	0.09
yes	sunny	cold	0.07
yes	cloudy	hot	0.01
yes	cloudy	mild	0.10
yes	cloudy	cold	0.12
~~no~~	~~sunny~~	~~hot~~	~~0.06~~
~~no~~	~~sunny~~	~~mild~~	~~0.11~~
~~no~~	~~sunny~~	~~cold~~	~~0.03~~
~~no~~	~~cloudy~~	~~hot~~	~~0.04~~
~~no~~	~~cloudy~~	~~mild~~	~~0.25~~
~~no~~	~~cloudy~~	~~cold~~	~~0.08~~

Weather (W)	Temperature (T)	P(W,T | N = yes)
sunny	hot	0.04/0.43 ≅ 0.10
sunny	mild	0.09/0.43 ≅ 0.21
sunny	cold	0.07/0.43 ≅ 0.16
cloudy	hot	0.01/0.43 ≅ 0.02
cloudy	mild	0.10/0.43 ≅ 0.23
cloudy	cold	0.12/0.43 ≅ 0.28

$$P(W = w \wedge T = t \mid N = yes)$$

$$= \frac{P(W = w \wedge T = t \wedge N = yes)}{P(N = yes)}$$

Inference by enumeration

Given $P(X)$ as JPD below, and evidence e = "Wind=yes"
What is the probability it is **hot**?
Step 2: marginalize to get distribution $P(Y \mid e)$

P(Temperature=hot | Wind=yes)?

$= 0.12$

Weather (W)	Temperature (T)	P(W,T | N = yes)
sunny	hot	0.10
sunny	mild	0.21
sunny	cold	0.16
cloudy	hot	0.02
cloudy	mild	0.23
cloudy	cold	0.28

Temperature (T)	P(T | N = yes)
hot	0.10 + 0.02 = 0.12
mild	0.21 + 0.23 = 0.44
cold	0.16 + 0.28 = 0.44

Space complexity for inference by enumeration

If all entries are Boolean, how many entries does the joint probability distribution (JPD) $P(X_1, X_2, \ldots, X_n)$ have?

A. 2^n ✅

B. $2 + n$

C. $2n$

D. n^2

Problems of inference by enumeration

- If we have n variables and d is the size of the the largest domain, what's the space complexity to store the joint distribution?
  - We need to store the probability for each possible world
  - There are $O(d^n)$ possible worlds, so the space complexity is $O(d^n)$
- How do we find the numbers for $O(d^n)$ entries?
- Time complexity $O(d^n)$
- We have some of our basic tools, but to gain computational efficiency we need to do more
- We will exploit (conditional) independence between variables

Lecture outline

- Recap
- Inference with enumeration
- Marginal independence 👈
- Conditional independence
- Bayesian networks (time permitting)

Marginal independence: example

Some variables are independent: Knowing the value of one does not tell you anything about the other.

<u>Example</u>: variables W(weather) and R (result of a die throw)

Let's compare
P(W) vs. P(W | R =6).

Weather (W)	Result (R)	P(W,R)
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Marginal independence: example

Some variables are independent: Knowing the value of one does not tell you anything about the other.

<u>Example</u>: variables W (weather) and R (result of a die throw)

Let's compare
P(W) vs. P(W | R = 6).

What's P(W = cloudy)?

Weather (W)	Result (R)	P(W,R)
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Marginal independence: example

Some variables are independent: Knowing the value of one does not tell you anything about the other.

Example: variables W(weather) and R (result of a die throw)

Let's compare
P(W) vs. P(W | R =6).

P(W = cloudy) = $\sum_{r \in domR} P(W = cloudy, R = r)$
$= 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1$
$= 0.6$

Weather (W)	Result (R)	P(W,R)
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Marginal independence: example

Some variables are independent: Knowing the value of one does not tell you anything about the other.

<u>Example</u>: variables W(weather) and R (result of a die throw)

Let's compare
P(W) vs. P(W | R = 6).

P(W = cloudy | R = 6)

$$= \frac{P(W = cloudy \wedge R = 6)}{P(R = 6)} = \frac{0.1}{0.1 + 0.066}$$

$= 0.6$

Weather (W)	Result (R)	P(W,R)
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Marginal independence: example

Some variables are independent:
Knowing the value of one does not tell you anything about the other.

Example: variables W(weather) and R (result of a die throw)

Let's compare
P(W) vs. P(W | R =6).

P(W = cloudy) = P(W = cloudy | R =6) = **0.6**

Weather (W)	Result (R)	P(*W*,*R*)
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Marginal independence: example

Some variables are independent: Knowing the value of one does not tell you anything about the other.

Example: variables W (weather) and R (result of a die throw)

Let's compare P(W) vs. P(W | R = 6).

Weather (W)	P(W)
sunny	0.4
cloudy	0.6

Weather (W)	P(W|R=6)
sunny	0.066/0.166 = 0.4
cloudy	0.1/0.166 = 0.6

The two distributions are identical!! Knowing the result of the die does not change our belief in the weather

Weather (W)	Result (R)	P(W,R)
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Marginal independence

Definition: Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds.
$P(X = x_i \mid Y = y_j) = P(X = x_i \mid Y = y_k) = P(X = x_i)$

Intuitively, if X and Y are **marginally independent**, then

- Learning that $Y = y$ does not change your belief in X
- And this is true for **all** values y that Y could take.

In the example we saw before, weather is marginally independent of the results of a dice throw.

Marginal independence (pair-share)

Definition: Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds.
$P(X = x_i \,|\, Y = y_j) = P(X = x_i \,|\, Y = y_k) = P(X = x_i)$

Are weather and temperature marginally independent? Why or why not?

Weather (W)	Temperature (T)	P(W,T)
sunny	hot	0.10
sunny	mild	0.20
sunny	cold	0.10
cloudy	hot	0.05
cloudy	mild	0.35
cloudy	cold	0.20

Marginal independence (pair-share)

Definition: Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds.

$$P(X = x_i \,|\, Y = y_j) = P(X = x_i \,|\, Y = y_k) = P(X = x_i)$$

Are weather and temperature marginally independent? Why or why not?

No we saw before that knowing weather changes our belief about the temperature. For example:

$$P(cold \,|\, sunny) \neq p(cold)$$

Weather (W)	Temperature (T)	P(W,T)
sunny	hot	0.10
sunny	mild	0.20
sunny	cold	0.10
cloudy	hot	0.05
cloudy	mild	0.35
cloudy	cold	0.20

Marginal independence (pair-share)

Definition: Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds.
$P(X = x_i \mid Y = y_j) = P(X = x_i \mid Y = y_k) = P(X = x_i)$

A and B are results of two tosses of a fair coin.

A	B	P(A, B)
heads	heads	0.25
heads	tails	0.25
tails	heads	0.25
tails	tails	0.25

Are A and B marginally independent?

Yes.

Marginal independence (pair-share)

Definition: Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds.
$P(X = x_i \mid Y = y_j) = P(X = x_i \mid Y = y_k) = P(X = x_i)$

Intuitively (without numbers)
Boolean random variable W: "Canucks win the Stanley Cup this season"
Numerical random variable R: "Canucks' revenue last season"

Are W and R marginally independent?

Marginal independence (pair-share)

Definition: Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds.

$$P(X = x_i \mid Y = y_j) = P(X = x_i \mid Y = y_k) = P(X = x_i)$$

Intuitively (without numbers)
Boolean random variable W: "Canucks win the Stanley Cup this season"
Numerical random variable R: "Canucks' revenue last season"

Are W and R marginally independent?
No because without the revenue they cannot afford to keep the best players. (Assuming that the revenue does influence how much money you can pay to the players.)

Exploiting marginal independence

- Recall the product rule: $P(f_2 \wedge f_1) = P(f_2 \mid f_1) \times P(f_1)$
- Thus $P(X = x \wedge Y = y) = P(X = x \mid Y = y) \times P(Y = y)$
- If X and Y are marginally independent,
- $P(X = x) = P(X = x \mid Y = y)$
- So $P(X = x \wedge Y = y) = P(X = x) \times P(Y = y)$
- In distribution form: $P(X, Y) = P(X) \times P(Y)$

Exploiting marginal independence

In general, if $X_1, X_2, \ldots, X_n$ are marginally independent, then we can represent their JPD as a **product of marginal distributions**

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i)$$

Exploiting marginal independence

iclicker.

In general, if $X_1, X_2, \ldots, X_n$ are marginally independent, then we can represent their JPD as a **product of marginal distributions**: $P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i)$

If all entries are Boolean, how many entries would all **marginal distributions** have combined?

A. 2^n

B. $2 + n$

C. $2n$ ✅

D. n^2

Exploiting marginal independence

In general, if $X_1, X_2, \ldots, X_n$ are marginally independent, then we can represent their JPD as a **product of marginal distributions**: $P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i)$

If all entries are Boolean, how many entries would all marginal distributions have combined? $2n$
Each of the n tables only has two entries:
$P(X_i) = True, P(X_i) = False$

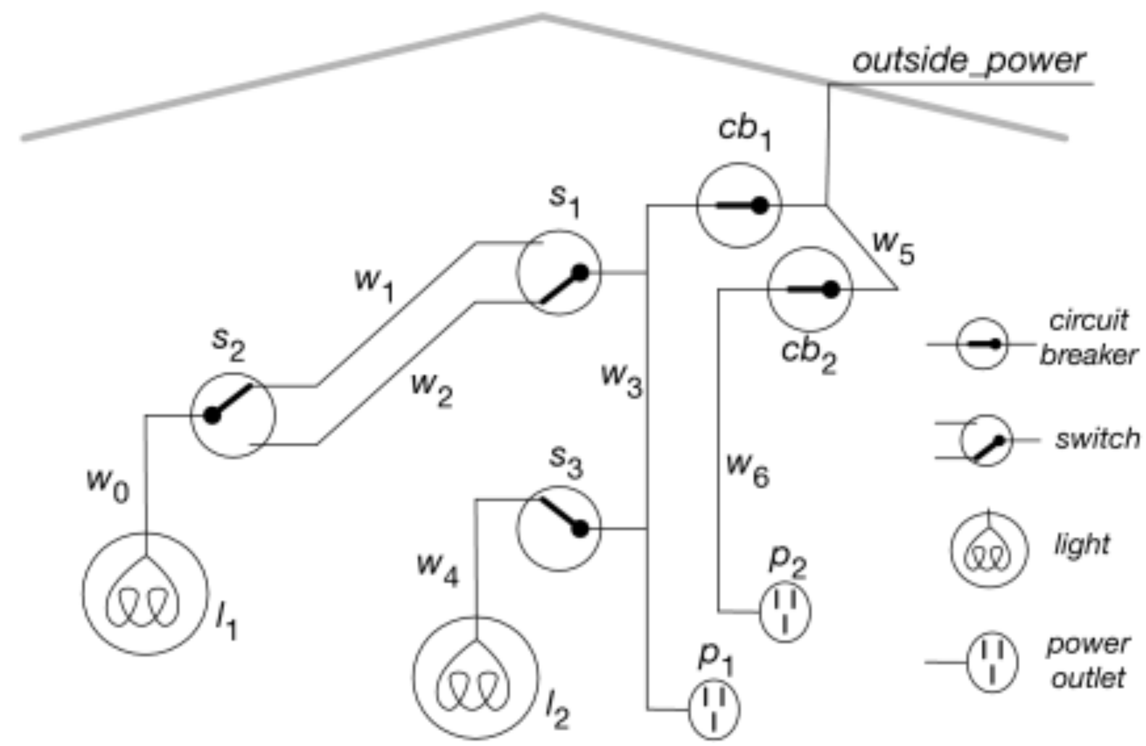
Lecture outline

- Recap
- Inference with enumeration
- Marginal independence
- Conditional independence 👈
- Bayesian networks (time permitting)

Conditional independence: example

Whether light l_1 is lit or not is conditionally independent from the position of the switch s_2 given whether there is power in w_0.

Once we know $Power(w_0)$, learning values from any other variable will not change our beliefs about $Lit(l_1)$.



$Lit(l_1)$ is independent of any other variable given $Power(w_0)$

Conditional independence

Definition: Random variable X is **conditionally independent** of random variable Y given random variable Z if, for all $x_i \in dom(X), y_j \in dom(Y), y_k \in dom(Y)$ and $z_n \in dom(Z)$, the following equation holds.

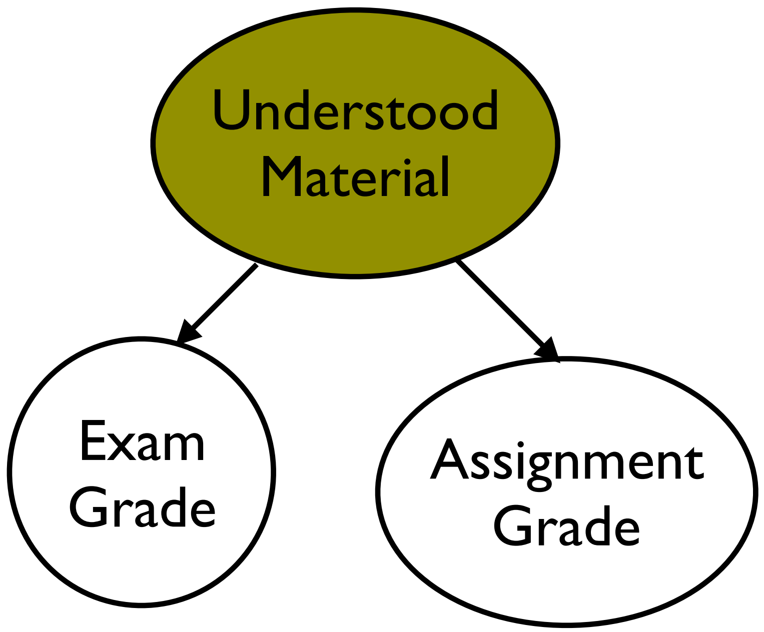
$P(X = x_i | Y = y_j, Z = z_m) = P(X = x_i | Y = y_k, Z = z_m) = P(X = x_i | Z = z_m)$

Intuitively: If X and Y are conditionally independent given Z,

- then learning that $Y = y$ does not change your belief in X when we already know $Z = z$
- and this is true for all values y that Y could take and all values z that Z could take

Conditionally but not marginally independent: example

- ExamGrade and AssignmentGrade are not marginally independent
  - Students who do well on one typically do well on the other
- But conditional on UnderstoodMaterial, ExamGrade and AssignmentGrade are independent
  - Variable UnderstoodMaterial is a common cause of variables ExamGrade and AssignmentGrade
  - UnderstoodMaterial shields any information we could get from AssignmentGrade



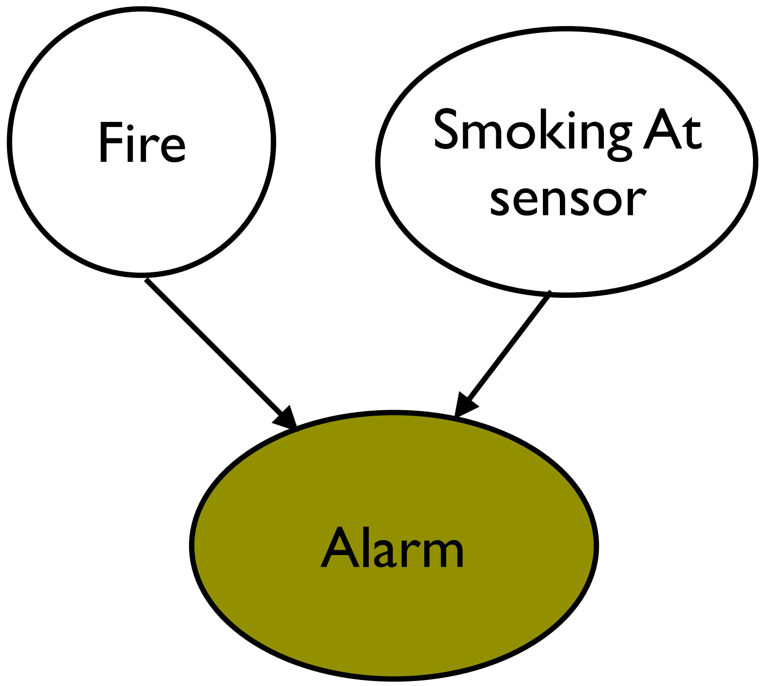
Marginally but not conditionally independent: example

Two variables can be marginally but not conditionally independent

“Smoking At Sensor” (S): resident smokes cigarette next to fire sensor
“Fire” (F): there is a fire somewhere in the building
“Alarm” (A): the fire alarm rings

S and F are marginally independent: Learning S=true or S=false does not change your belief in F

But they are not conditionally independent given alarm: If the alarm rings and you learn S=true your belief in F decreases



Conditional vs. marginal independence

Two variables can be

- Both marginally and conditionally independent
  CanucksWinStanleyCup and $Lit(l_1)$
  CanucksWinStanleyCup and $Lit(l_1)$ given $Power(w_0)$

- Neither marginally nor conditionally independent
  Temperature and Cloudiness
  Temperature and Cloudiness given Wind

- Conditionally but not marginally independent
  ExamGrade and AssignmentGrade
  ExamGrade and AssignmentGrade given UnderstoodMaterial

- Marginally but not conditionally independent
  SmokingAtSensor and Fire
  SmokingAtSensor and Fire given Alarm

Exploiting conditional independence

Definition: Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X), y_j \in dom(Y), y_k \in dom(Y)$ and $z_n \in dom(Z)$, the following equation holds.

$$P(X = x_i \,|\, Y = y_j, Z = z_m) = P(X = x_i \,|\, Y = y_k, Z = z_m) = P(X = x_i \,|\, Z = z_m)$$

<u>Example 1</u>: Boolean variables A,B,C

C is conditionally independent of A given B

We can then rewrite P(C | A,B) as P(C|B)

Exploiting conditional independence

Definition: Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X), y_j \in dom(Y), y_k \in dom(Y)$ and $z_n \in dom(Z)$, the following equation holds.

$$P(X = x_i \mid Y = y_j, Z = z_m) = P(X = x_i \mid Y = y_k, Z = z_m) = P(X = x_i \mid Z = z_m)$$

<u>Example 2</u>: Consider Boolean variables A,B,C,D

D is conditionally independent of A given C and D is conditionally independent of B given C.

We can then rewrite P(D | A,B,C) as P(D|B,C)

And can further rewrite P(D|B,C) as P(D|C)

Exploiting conditional independence

Recall the chain rule:

$$P(fn \wedge \ldots \wedge f_1) = \prod_{i=1}^{n} P(f_i \,|\, f_{i-1} \wedge \ldots \wedge f_1)$$

Examples:

$$P(A, B, C, D) = P(A) \times P(B \,|\, A) \times P(C \,|\, A, B) \times P(D \,|\, A, B, C)$$

If D, for example, is conditionally independent of A and B given C, we can rewrite this as

$$P(A, B, C, D) = P(A) \times P(B \,|\, A) \times P(C \,|\, A, B) \times P(D \,|\, C)$$

Conditional independence

- Under independence we gain **compactness**
- The chain rule lets us represent the JPD as a product of conditional distributions
- Conditional independence allows us to write them compactly

Lecture outline

- Recap
- Inference with enumeration
- Marginal independence
- Conditional independence
- Bayesian networks (time permitting) 👈🏻

Bayesian networks: Motivation

We want a representation and reasoning system that is based on conditional (and marginal) independence

Compact yet expressive representation

Efficient reasoning procedures

Bayesian networks: Motivation

- Bayes[ian] (Belief) Net[work]s are such a representation.
- Named after Thomas Bayes (1702 –1761)
- Term coined in 1985 by Judea Pearl
- Their invention changed the primary focus of AI from logic to probability!

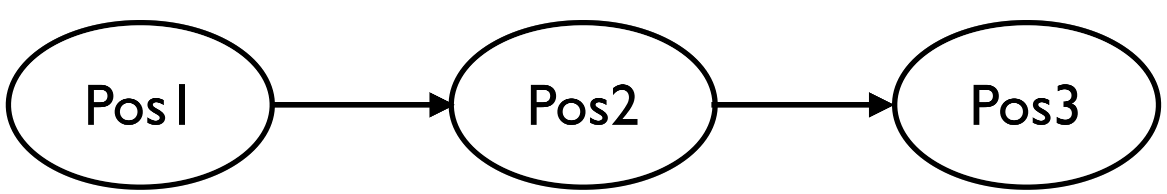
Thomas Bayes

Judea Pearl

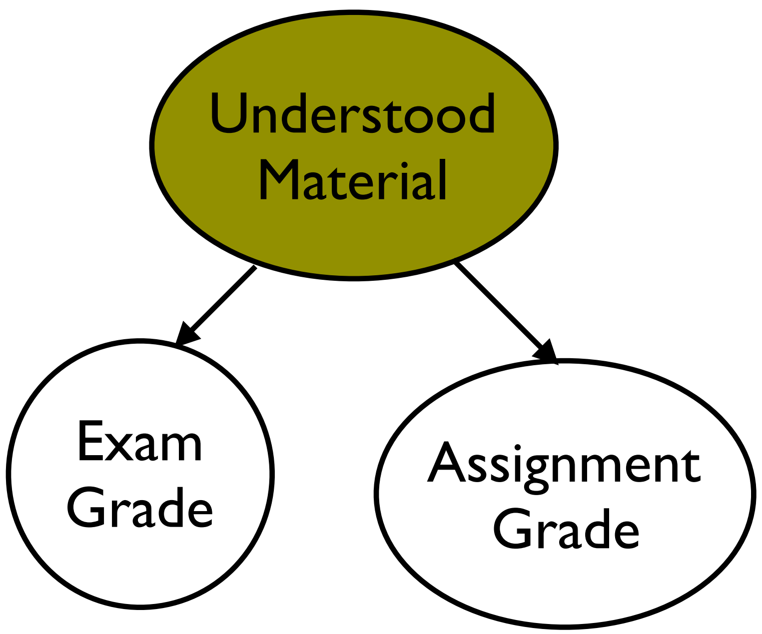
Bayesian networks intuition

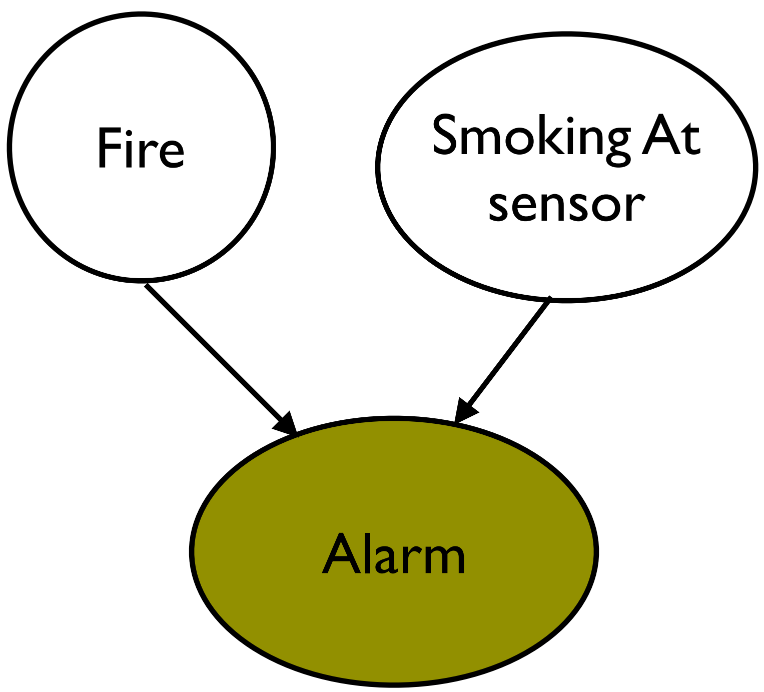
A **graphical representation** for a joint probability distribution

- Nodes are random variables
- Directed edges between nodes reflect dependence



Robot position





Summary

- Probability is a rigorous formalism for uncertain knowledge
- Joint probability distribution specifies probability of every possible world
- Queries can be answered by summing over possible worlds
- For nontrivial domains, we must find a way to reduce the joint distribution size
- Independence *(rare)* and conditional independence *(frequent)* provide the tools

Coming up

8.3 Belief Networks

8.4 Probabilistic Inference