CS 3630!

Lecture 5:
Bayes Nets

Topics

1. **A Grid-world Agent**
2. **Modeling Sensors**
3. **Joint Distribution**
4. **Bayes Nets**
5. **Ancestral Sampling**
6. **Dynamic Bayes Nets**
7. **Bayes' rule**
8. **Inference in Bayes nets**
9. **MPE**
10. **MAP**

Motivation

- Probability -> simulate robots !
- Our example: grid world
- Probabilistic statements about state: Bayesian inference

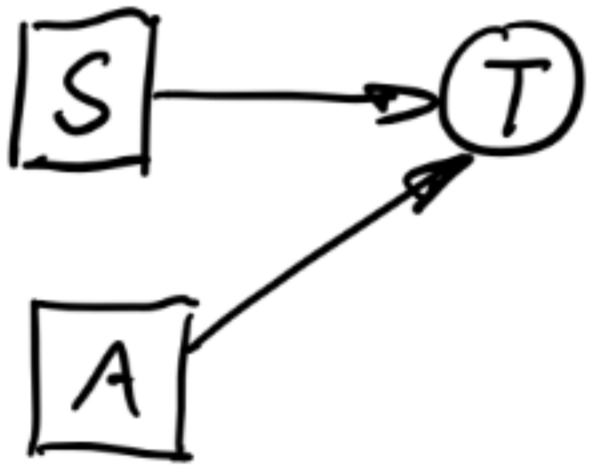
1. Grid world example

				.1					
	.1								
			.1						
.1					.1			.1	
		.1				.1			
							.1		
									.1

Table 2.2: A PMF describing where a robot might be in a grid.

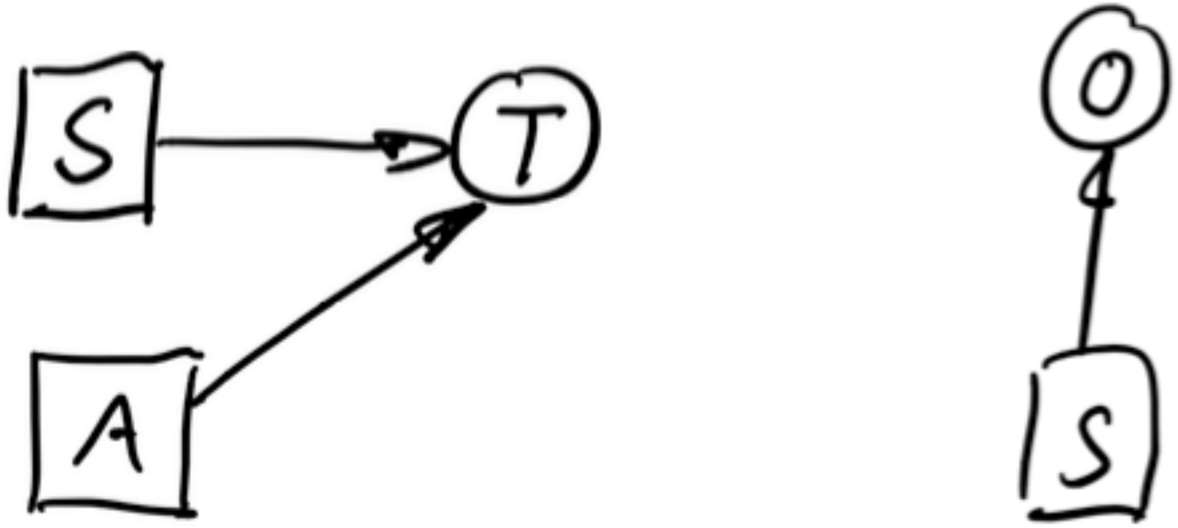
- 100 outcomes!

Recap: Modeling action



- We can do the same for modeling action:
- State transition model $P(T|S,A)$
- Exercise: come up with a fairly realistic model for grid world.

2. Modeling Sensors



(a) State transition model $P(T|S, A)$

(b) Sensor model $P(O|S)$

Figure 1: Conditional distributions to model acting and sensing.

- Conditionals are cool: we can use them to model *and* sensing.

Parametric descriptions

$$\begin{cases} P(O=k|S=i,j)=1 & \text{iff } k=j \\ P(O=k|S=i,j)=0 & \text{otherwise} \end{cases}$$

$$\begin{cases} P(O=k|S=i,j)=0.91 & \text{iff } k=i \\ P(O=k|S=i,j)=0.01 & \text{otherwise} \end{cases}$$

- Sensor model P(O|S)
- CPTs can become very big
- Uniformity: implement as a function!
- Exercise: what are the above models? *S = (i,j)* is robot location in grid.

Parametric descriptions (cont'd)

$$\begin{cases} P(O=k|S=i,j)=1 & \text{iff } k=j \\ P(O=k|S=i,j)=0 & \text{otherwise} \end{cases}$$

$$\begin{cases} P(O=k|S=i,j)=0.91 & \text{iff } k=i \\ P(O=k|S=i,j)=0.01 & \text{otherwise} \end{cases}$$

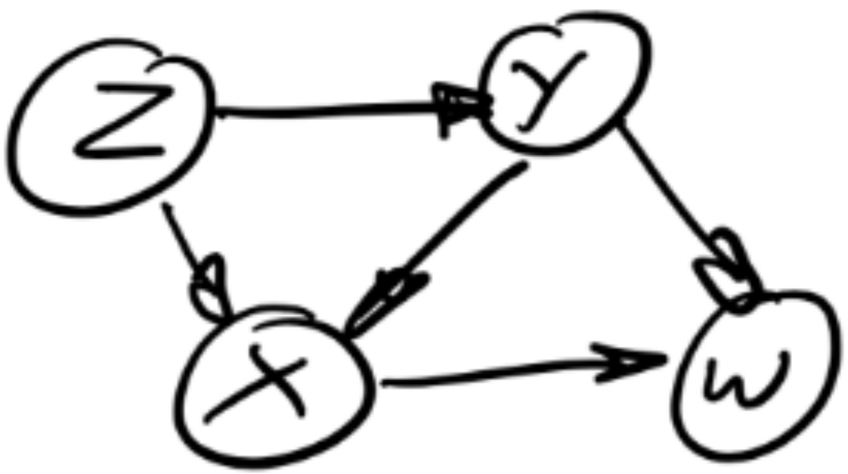
- Uniformity: implement as a function!
- Left: report the horizontal coordinate *j* of the robot faithfully
- Right: reports the vertical coordinate *i* of the robot, but with 9% probability gives a random faulty reading

3. Joint Distribution

Figure 2.4: Joint probability distribution on two variables X and Y.

- What if parameter in conditional is itself a random variable?
- Chain rule: $P(X,Y) = P(X|Y)\,P(Y)$
- Riddle: How do we sample?

4. Bayes Nets

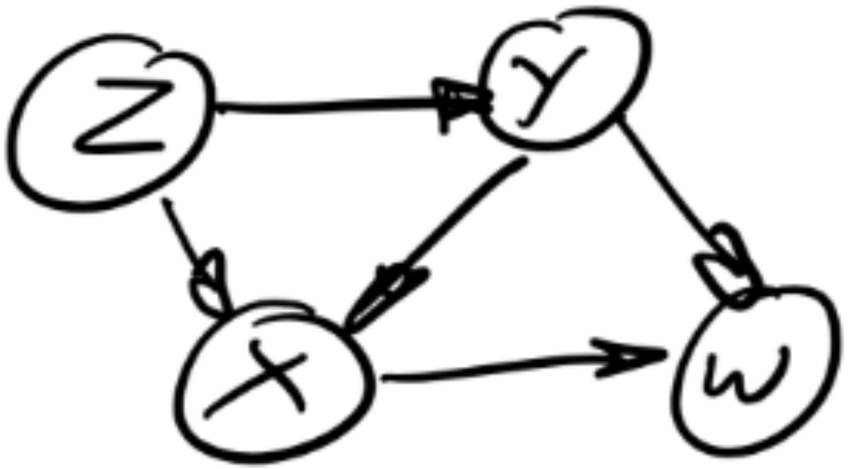


- A Bayes net is a directed acyclic graph (DAG) of conditionals.
- The joint is given by multiplying all conditionals:

$$P(\{X_i\}) = \prod_{i=1}^{n} P(X_i|\Pi_i)$$

- Exercise: *P(W,X,Y,Z,* .
- Note chain rule *P(X,Y) = (X|Y)P(Y)* is a special case.

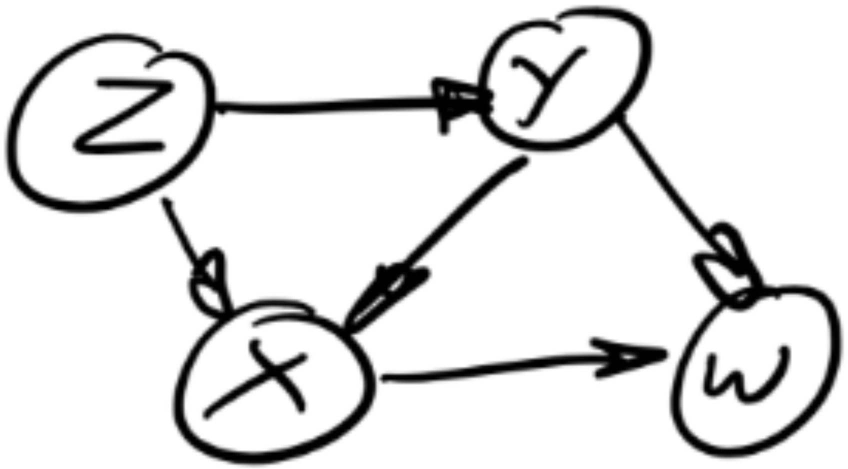
Bayes Nets (cont'd)



- P(W|X,Y)P(X|Y,Z)P(Y|Z)P(Z)
- A Bayes net is very efficient!
- Assume *W,X,Y,Z* are all 10-valued. How many entries in the joint PMF?
- How many entries in CPTs ?

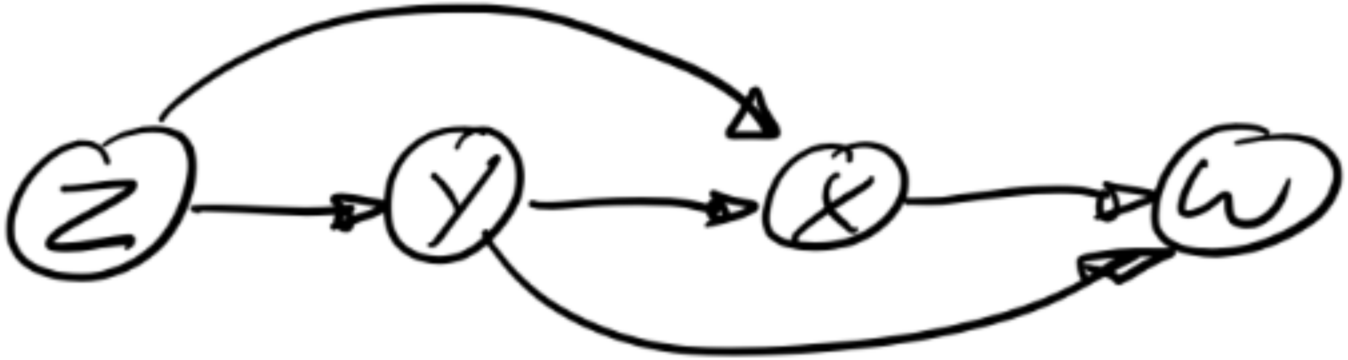
CPT	# entries
$P(Z)$	9
$P(Y\|Z)$	90
$P(X\|Y,Z)$	900
$P(W\|X,Y)$	900

5. Ancestral Sampling



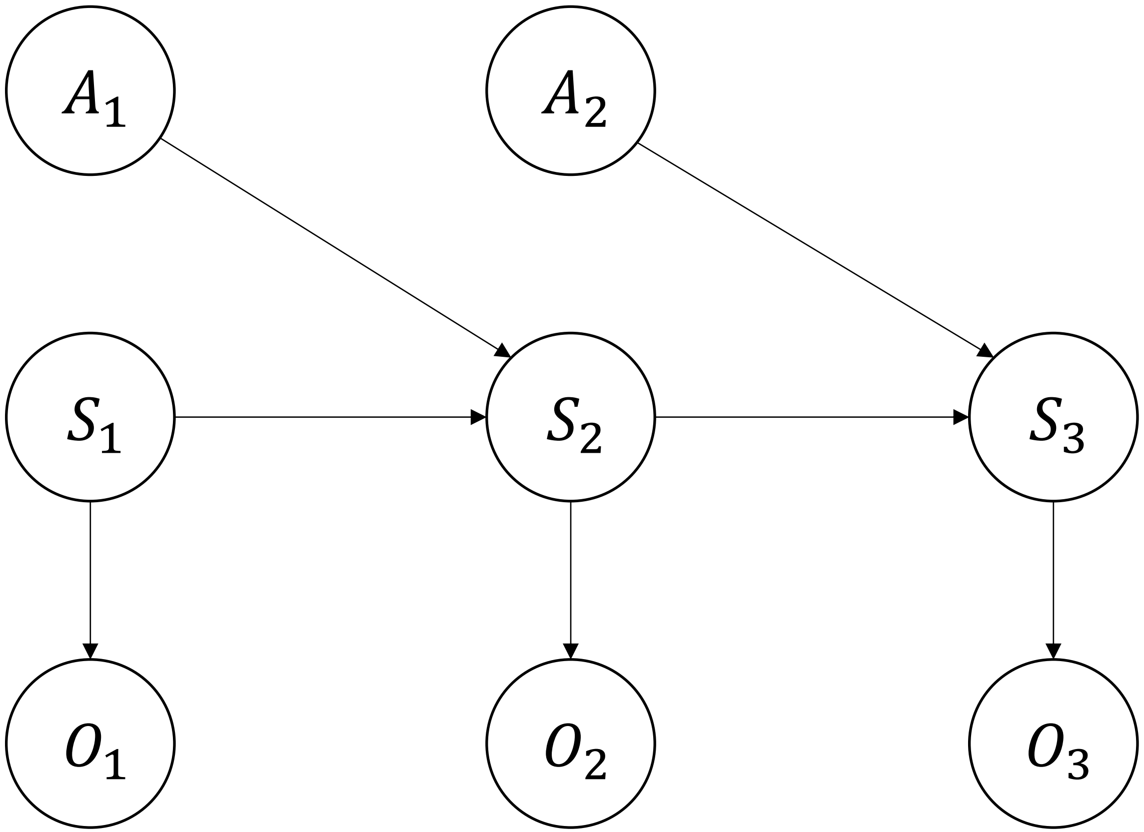
- How do we sample from a Bayes net?
- Recall: sampling from *P(X|Y) P(Y)* ?

- Generalize:
  1. topological sort (Kahn's algorithm)
  2. Sample in topological sort order



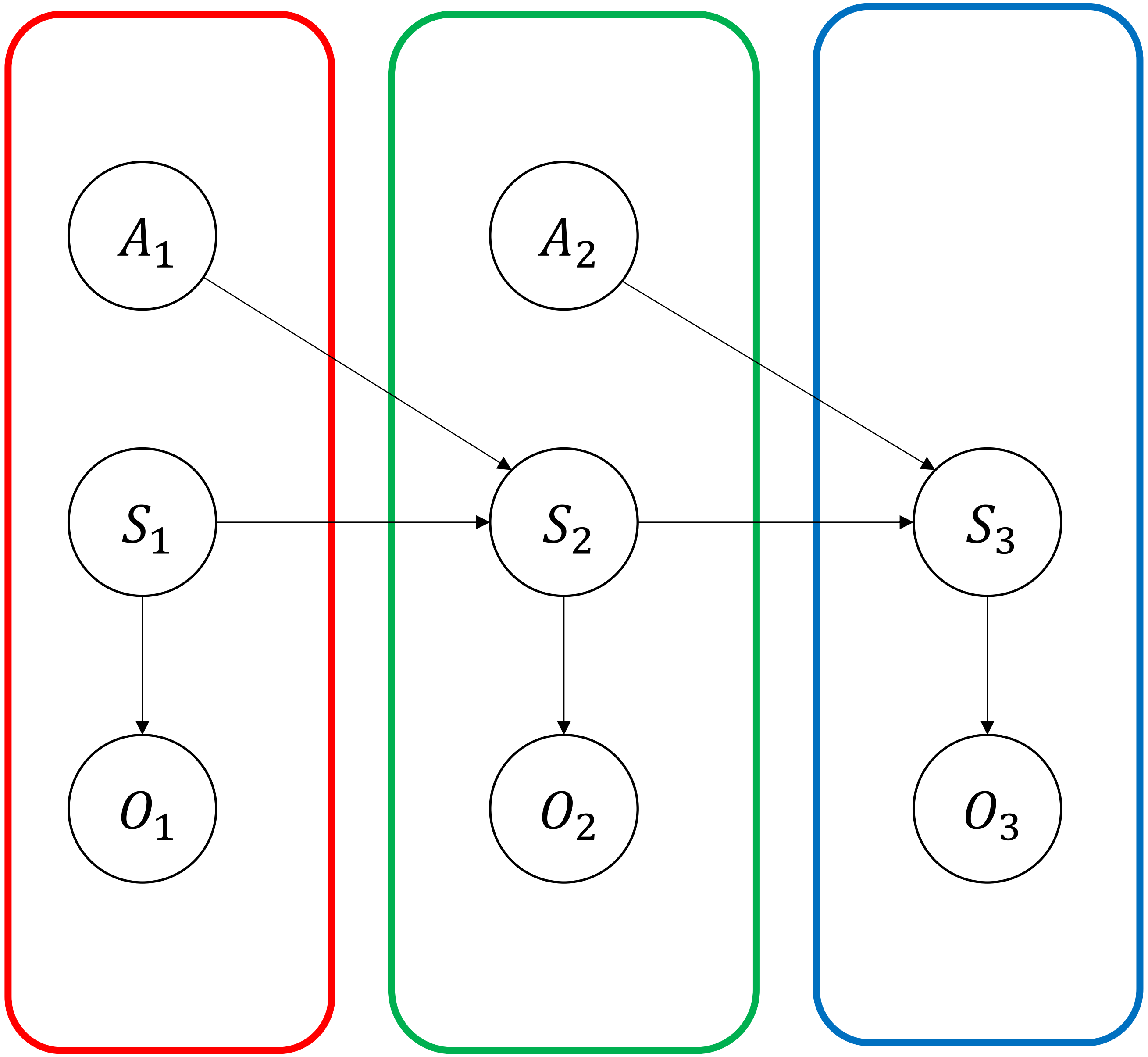
6. Dynamic Bayes Nets

- DBN or dynamic Bayes net: roll out *time*.
- Applied to agents/robots: sequence of sensing and acting!



Simulation of Agents

1. Slice 1:
   a) Sample from $P(S_1)$
   b) Sense $P(O_1|S_1)$
   c) Sample from $P(A_1)$
2. Slice 2:
   a) Act $P(S_2|S_1, A_1)$
   b) Sense $P(O_2|S_2)$
   c) Sample from $P(A_2)$
3. Slice 3:
   a) …



7. Bayes' Rule

- Inference:
  - probabilistic statements about what we *know*
- Given: we observe a sensor measurement $O=o$
- What can we say about the state S ?
- You need:
  - Sensor model $P(O|S)$
  - Prior probability distribution $P(S)$
- What we want:
  - Posterior probability distribution $P(S|O=o)$

Bayes' Rule (cont'd)

- $P(S|O) = P(S,O) / P(O) = P(O|S)\, P(S) / P(O)$
- Hence:

$$P(S|O=o) = \frac{P(O=o|S)P(S)}{P(O=o)}$$

- This is known as Bayes' rule (or: Bayes' law….)

Exercise

				.1					
	.1								
			.1						
.1					.1			.1	
		.1				.1			
							.1		
									.1

$$\begin{cases} P(O = k|S = i, j) = 0.91 & \text{iff } k = i \\ P(O = k|S = i, j) = 0.01 & \text{otherwise} \end{cases}$$

- Apply Bayes' rule to calculate the posterior *P(S|O=5)*
- *First think about the representation of the result: what is it?*

Likelihood functions

- In Bayes' law, given *O=o*, all are functions of S

$$P(S|O=o) = \frac{P(O=o|S)P(S)}{P(O=o)}$$

- Introduce the likelihood function:

$$L(S;o) \triangleq P(O=o|S)$$

- Bayes' law:

$$P(S|O=o) \propto L(S;o)P(S)$$

The many ways of Bayes

- Classic:

$$P(S|O=o) = \frac{P(O=o|S)P(S)}{P(O=o)}$$

- Intuitive:

$$P(S|O=o) \propto L(S;o)P(S)$$

- Bare-bones:

$$P(S|O=o) \propto P(S,O=o)$$

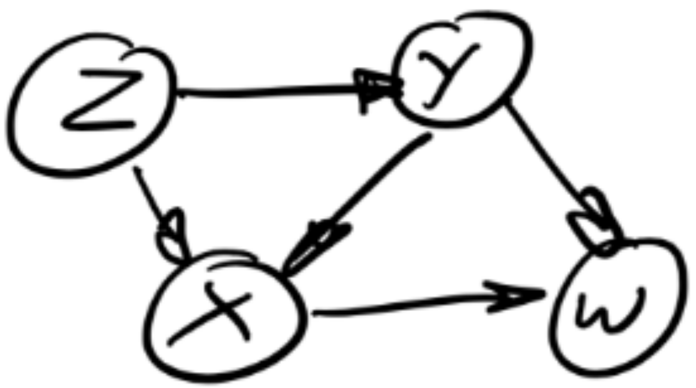
8. Inference in Bayes Nets

- Posterior Probability = Complete knowledge of X given some Z values

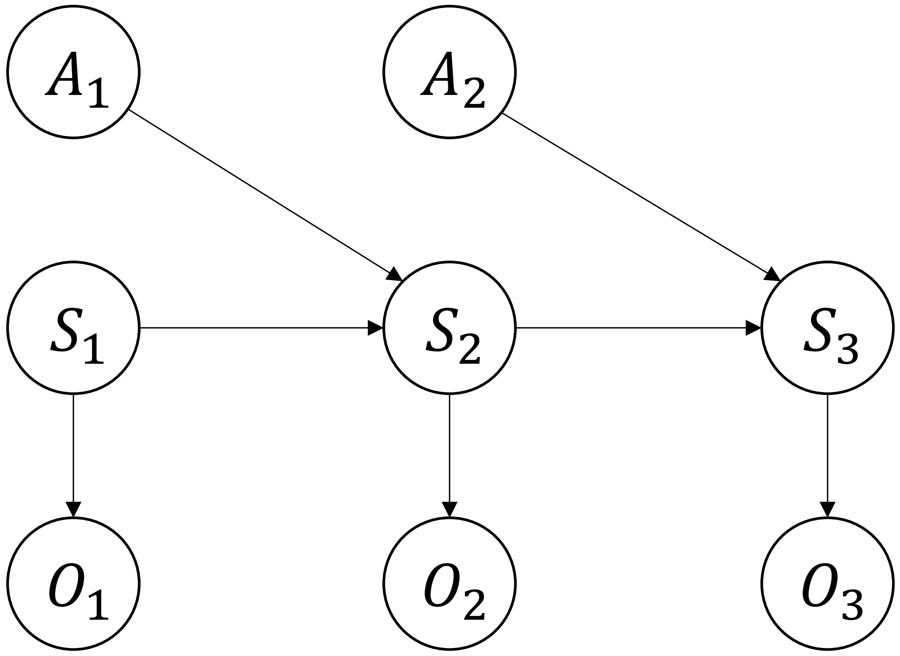
$$P(\mathcal{X}|\mathcal{Z} = \mathfrak{z}) \propto P(\mathcal{X}, \mathcal{Z} = \mathfrak{z}).$$

- Simple algorithm:
  1. Enumerate all combinations of X values
  2. Calculate posterior

Naïve inference, exercises



- Exercise 1:
  - condition on W, Z
  - How big is the table?
- Exercise 2:
  - Condition on Y, X
  - Try branch & bound
- Exercise 3:
  - In DBN, assume states are given
  - What is complexity of inferring the actions?



9. Most Probable Explanation

- MPE or most probable explanation, given some Z values

$$P(\mathcal{X}|\mathcal{Z} = \mathfrak{z}) \propto P(\mathcal{X}, \mathcal{Z} = \mathfrak{z}).$$

- Find assignment to remaining *X* values such that above is maximized!
- Simple algorithm:
  1. Enumerate all combinations of X values
  2. Calculate posterior
  3. Pick maximum

- More sophisticated algorithm: branch & bound. Discuss !

10. MAP Estimate

- MAP or maximum a posteriori estimate, given some Z values

$$P(\mathcal{X}|\mathcal{Z}=\mathfrak{z}) = \sum_{\mathfrak{y}} P(\mathcal{X},\mathcal{Y}=\mathfrak{y}|\mathcal{Z}=\mathfrak{z}) \propto \sum_{\mathfrak{y}} P(\mathcal{X},\mathcal{Y}=\mathfrak{y},\mathcal{Z}=\mathfrak{z}).$$

- We now have nuisance variables Y, which we need to marginalize out.
- At least as expensive as MPE, in many cases much more so.

Summary

- **A Grid world** is a more realistic robotics example
- **Models for sensing and acting** can be built using parametric conditional distributions.
- We can compute a **joint probability** distribution, and marginal and conditionals from it.
- **Bayes nets** allow us to encode more general joint probability distributions over many variables.
- **Ancestral sampling** is a technique to simulate from any Bayes net.
- **Dynamic Bayes nets** unroll time and can be used to simulate robots over time.
- **Bayes' rule** allows us to infer knowledge about a state from a given observation.
- **Inference in Bayes nets** is a simple matter of enumeration, but this can be expensive.
- The **maximum probable explanation** singles out one estimate.
- Marginalizing over some variables leads to **MAP inference**.