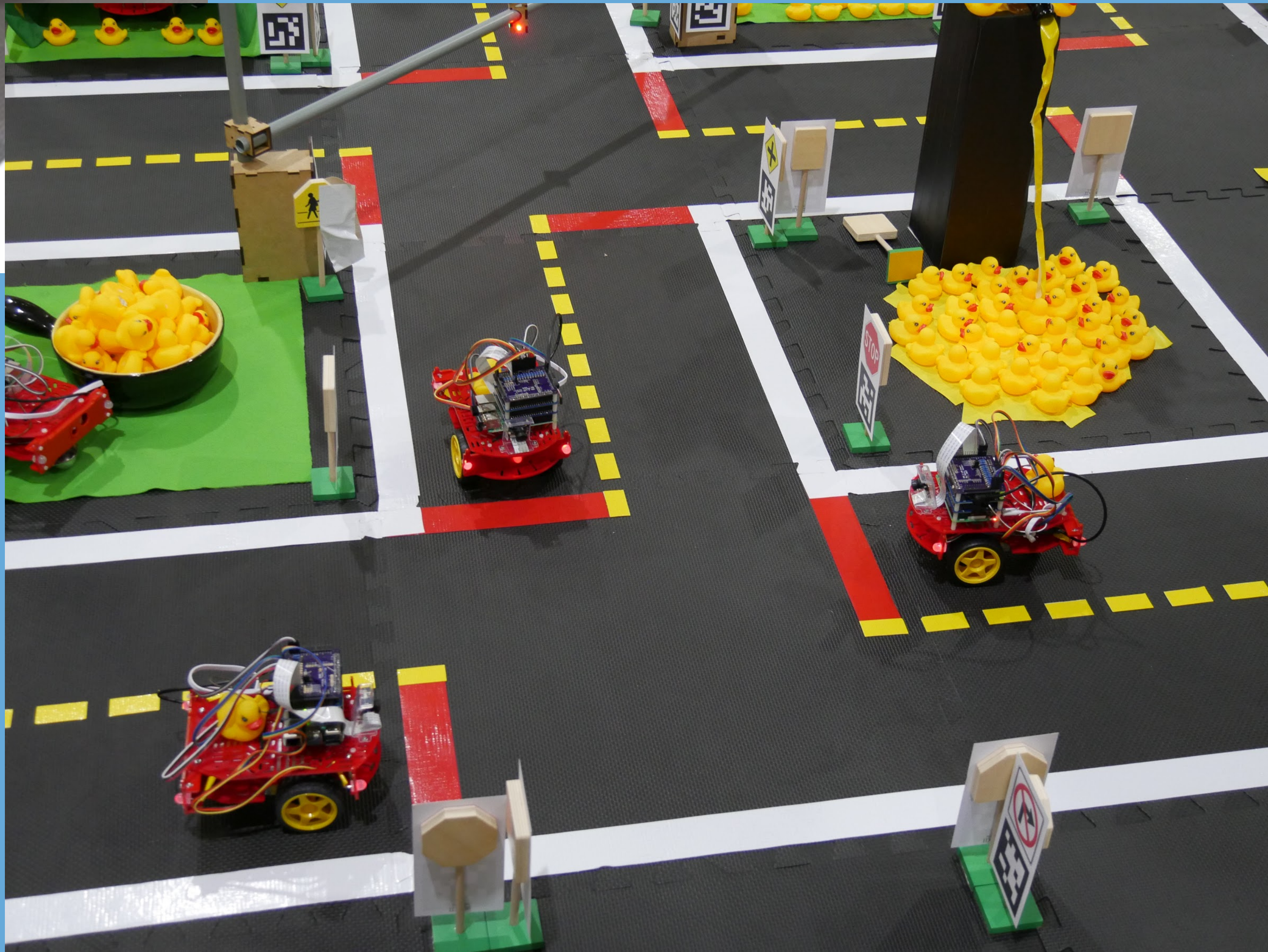
CS 3630!



Lecture 5: The Variable Elimination Algorithm

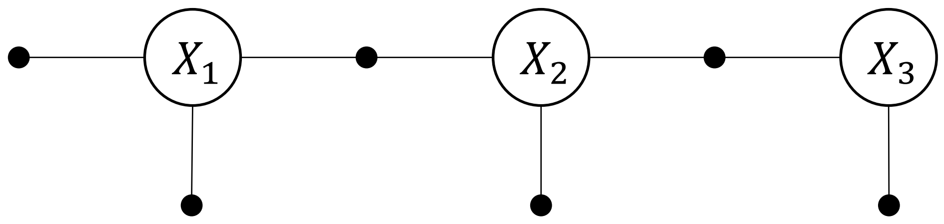
Topics

- **1. The Sum-Product Algorithm for HMMs**
- **2. Variable Elimination Algorithm**
- **3. Complexity of Sparse Inference**
- **4. MAP vs MPE**

Motivation

- We saw MPE in HMMs: Max-Product
- Now: get full probabilistic picture
- Get an idea about complexity

1. The Sum-product Algorithm for HMMs



Replace elimination with the chain rule.

Done in class:

- Eliminate X_1
- Eliminate X_2
- Eliminate X_3

Eliminate X_1

- Form product factor

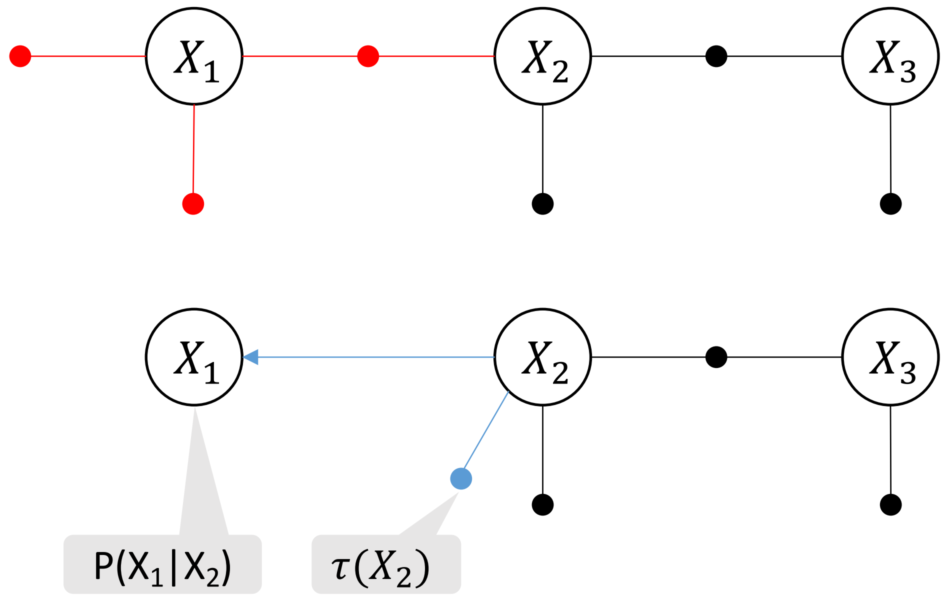
$$\psi(X_1, X_2) = \phi_1(X_1)\phi_2(X_1)\phi_3(X_1, X_2).$$

- Calculate marginal:

$$\tau(X_{j+1}) = \sum_{x_j} \psi(x_j, X_{j+1})$$

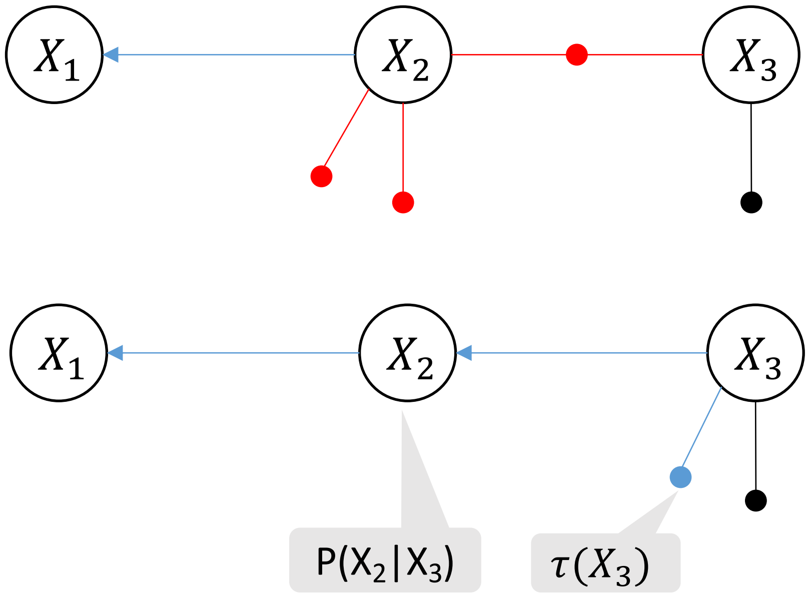
- Calculate conditional:

$$P(X_j|X_{j+1}) = \frac{\psi(X_j, X_{j+1})}{\tau(X_{j+1})}$$



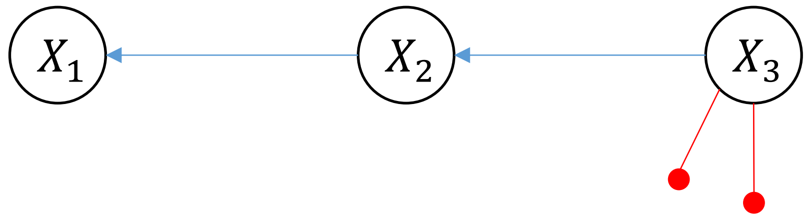
Eliminate X_2

- Form product factor
- Calculate marginal:
- Calculate conditional:

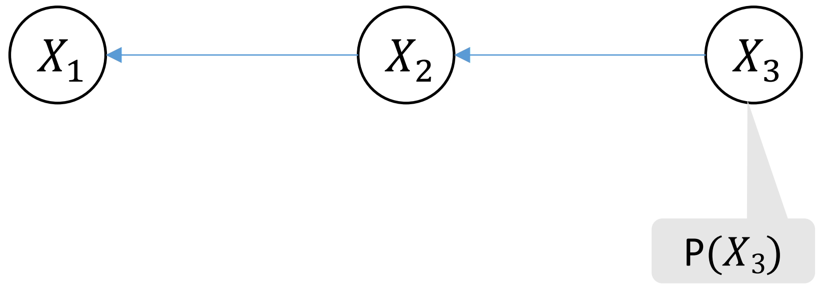


Eliminate X_3

- Form product factor

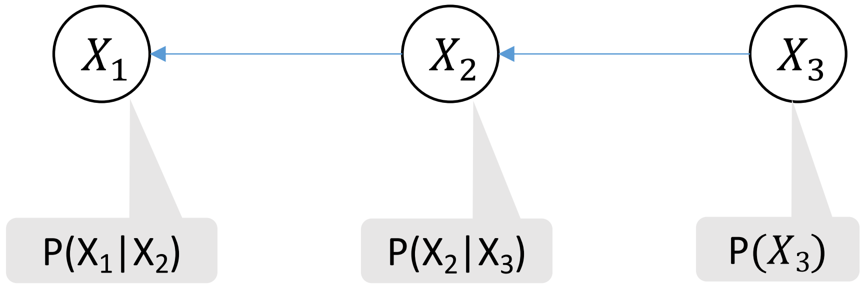


- Calculate $P(X_3)$



Back-substitute? Sample!

- The reverse elimination order is *always* a topological sort.
- Ancestral sampling yields samples from posterior.



- Sidebar: Can estimate posterior means for any real-valued function.

What, another Bayes net?

- Bayes net -> factor graph -> Bayes net again???
- Bayes net 1, all variables:

$$\text{joint distribution} \;\; P(\mathcal{X}, \mathcal{Z})$$

- Bayes net 2, only unknowns:

$$\text{posterior} \; P(\mathcal{X}|\mathcal{Z})$$

- In practice: build factor graph directly from the measurements.

2. The Variable Elimination Algorithm

Algorithm 5 The Variable Elimination Algorithm

function ELIMINATE($\Phi_{1:n}$) ▷ given a factor graph on n variables
  for $j = 1...n$ **do** ▷ for all variables
    $p(X_j|\mathcal{S}_j), \Phi_{j+1:n} \leftarrow \text{EliminateOne}(\Phi_{j:n}, X_j)$ ▷ eliminate X_j
  return $p(X_1|\mathcal{S}_1)p(X_2|\mathcal{S}_2)\dots p(X_n)$ ▷ return Bayes net

Algorithm 6 Eliminate variable X_j from a factor graph $\Phi_{j:n}$.

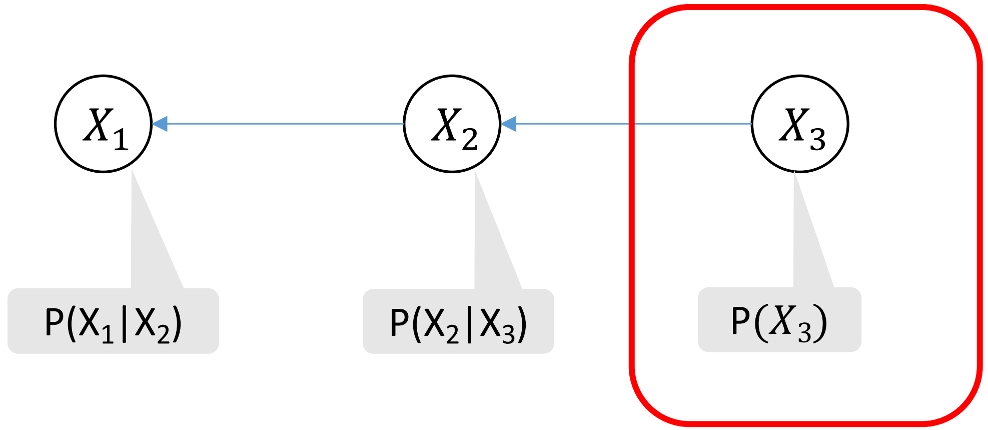
function ELIMINATEONE($\Phi_{j:n}, X_j$) ▷ given reduced graph $\Phi_{j:n}$
  Remove all factors $\phi_i(\mathcal{X}_i)$ that are adjacent to X_j
  $\mathcal{S}(X_j) \leftarrow$ all variables involved excluding X_j ▷ the separator
  $\psi(X_j, \mathcal{S}_j) \leftarrow \prod_i \phi_i(\mathcal{X}_i)$ ▷ create the product factor ψ
  $p(X_j|\mathcal{S}_j)\tau(\mathcal{S}_j) \leftarrow \psi(X_j, \mathcal{S}_j)$ ▷ factorize the product ψ
  Add the new factor $\tau(\mathcal{S}_j)$ back into the graph
  return $p(X_j|\mathcal{S}_j), \Phi_{j+1:n}$ ▷ Conditional and reduced graph

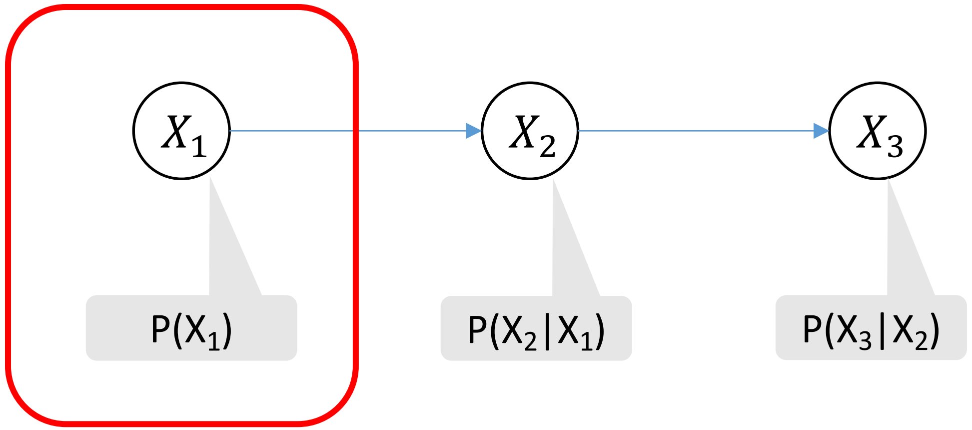
- Works for any factor graph

3. Complexity of Sparse Inference

- Depends on the tree-width.
- Tree-width = size of the largest separator that occurs.
- Examples in class.

4. MAP vs MPE





Map estimate for X_3

Map estimate for X_1

- Eliminate the variables of interest last!
- What does that mean for complexity in this HMM?

Summary

1. The **sum-product** algorithm returns the full Bayesian posterior as a Bayes net.
2. The **variable elimination algorithm** is a generalization that works for *any* factor graph.
3. The **complexity** of variable elimination depends on the elimination order.
4. **MAP estimation** is always as least as expensive, as it constrains the elimination order.