Introduction to Probabilistic Machine Learning

Ralf Herbrich, Rainer Schlosser

Tutorial 4 – Recap Theory Unit 4

Overview

1. **Questions and Updates**
2. Recap: Main Concepts of Unit 4
3. Example: Message Passing in Factor Graphs
4. Hints for Exercise 2 (to be handed in May 19)

Course Overview

Week	Topic Lecture	Tutorial	Exercises
07.04. & 08.04.	1 Probability Theory	Intro Julia	
14.04. & 15.04.	2 Inference Methods and Decision-Making	no tutorial	Exercise 1
21.04. & 22.04.	**no lecture**	Theory Unit 1 & 2	(14.04. – **08.05.**)
28.04. & 29.04.	3 Graphical Models: Independence	Theory Unit 3	
05.05. & 06.05.	**4 Graphical Models: Exact Inference**	**Theory Unit 4**	**Exercise 2**
12.05. & 13.05.	5 Graphical Models: Approximate Inference	Theory Unit 5	(05.05. – 19.05.)
19.05. & 20.05.	6 Bayesian Ranking	Theory Unit 6	Exercise 3
26.05. & 27.05.	7 Linear Basis Function Models	Theory Unit 7	(19.05. – 02.06.)
02.06. & 03.06.	8 Bayesian Regression	Theory Unit 8	Exercise 4
09.06. & 10.06.	**no lecture**	9 Bayesian Classification	(02.06. – 23.06.)
16.06. & 17.06.	10 Non-Bayesian Classification Learning	Theory Unit 9 & 10	
23.06. & 24.06.	11 Gaussian Processes	Theory Unit 11	Exercise 5
30.06. & 01.07.	12 Information Theory	Theory Unit 12	(23.06. – 07.07.)
07.07. & 08.07.	13 Real-World Applications		

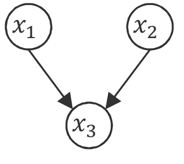
Overview

1. Questions and Updates
2. **Recap: Main Concepts of Unit 4**
3. Example: Message Passing in Factor Graphs
4. Hints for Exercise 2 (to be handed in Monday May 19)

Factor Graphs

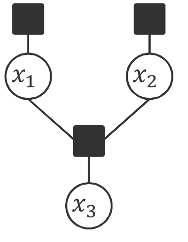
- **Factor Graph (Frey, 1998).** *Given a product of m functions $f_1, f_2, \ldots, f_m$, each over a subset of n variables $x_1, x_2, \ldots, x_n$, a factor graph if a bipartite graphical model with m factor nodes and n variable nodes where an undirected edge connects f_i and x_j if and only if the function f_i depends on x_j.*
- Factor graphs are more expressive than a Bayesian network!

Bayesian network



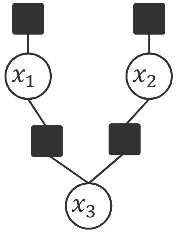
$p(x_1, x_2, x_3) = p(x_1) \cdot p(x_2) \cdot p(x_3|x_1, x_2)$

Corresponding factor graph



$p(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3, x_1, x_2)$

Factor graph with more structure



$p(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_1, x_3) \cdot f_4(x_2, x_3)$

Structure in $p(x_3|x_1, x_2)$

Brendan Frey
(1968 –)

Introduction to Probabilistic Machine Learning

Unit 4 – Graphical Models: Inference

Inference by Importance Sampling

- The key operation is summing-out all but one variable (marginalization).
- **Idea**: If we can get J samples $x_{j,1}, \dots, x_{j,n}$, $j \in \{1, \dots, J\}$ which are drawn according to

$$p(x_1, \dots, x_n) \propto \prod_{i=1}^{m} f_i(x_1, \dots, x_n)$$

then we can approximate the marginals $p(x_k)$ arbitrarily well via

$$p(x_k) \approx \frac{1}{J} \sum_{j=1}^{J} \delta\left(x_k - x_{j,k}\right)$$

- **Challenge**: How do we sample $p(x_1, \dots, x_n)$ if we only have access to (a few) known efficient samplers $q(x_1, \dots, x_n)$ such as (pseudo-random) numbers from the uniform distribution or normal distribution over each X_k?
- **Importance Sampling**: We get J samples $x_{j,1}, \dots, x_{j,n}$, $j \in \{1, \dots, J\}$ from $q(x_1, \dots, x_n)$ and re-weight them with

$$\boxed{\frac{p\left(x_{j,1}, \dots, x_{j,n}\right)}{q\left(x_{j,1}, \dots, x_{j,n}\right)}}$$

Importance weight

Marginalization by Sampling a Factor Graph

Importance Sampling

Given: Proposal distributions $q_1(\cdot), \dots, q_n(\cdot)$ for $X_1, \dots, X_n$

For $j \in \{1, \dots, J\}$

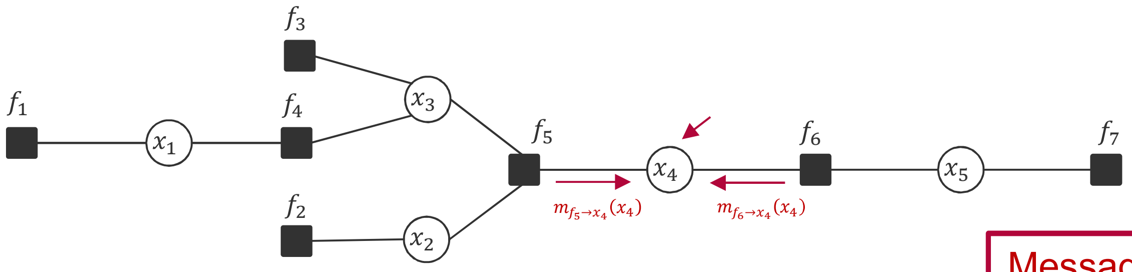
1. Sample $x_{j,1} \sim q_1$, $x_{j,2} \sim q_2$ up to $x_{j,n} \sim q_n$ into $\boldsymbol{x}_j = (x_{j,1}, \dots, x_{j,n})$
2. Compute

$$w_j = \prod_{i=1}^{m} f_i(x_{j,1}, \dots, x_{j,n}) \Big/ \prod_{k=1}^{n} q_k(x_{j,k})$$

Return: $\{\boldsymbol{x}_j\} \in \mathbb{R}^{J \times n}$ and $\boldsymbol{w} \in \mathbb{R}^{+J}$ as weighted empirical distribution

- **Pros**
  1. Sampling of the $\boldsymbol{x}_j$ is parallel rather than sequential (as in a Bayesian network)!
  2. The weights can also be computed in parallel!
- **Cons**
  1. If the proposal $\prod_{k=1}^{n} q_k(\cdot)$ is far from the marginal of $\prod_{i=1}^{m} f_i(\cdot, \dots, \cdot)$ then convergence is slow

Sum-Product Algorithm: Marginals



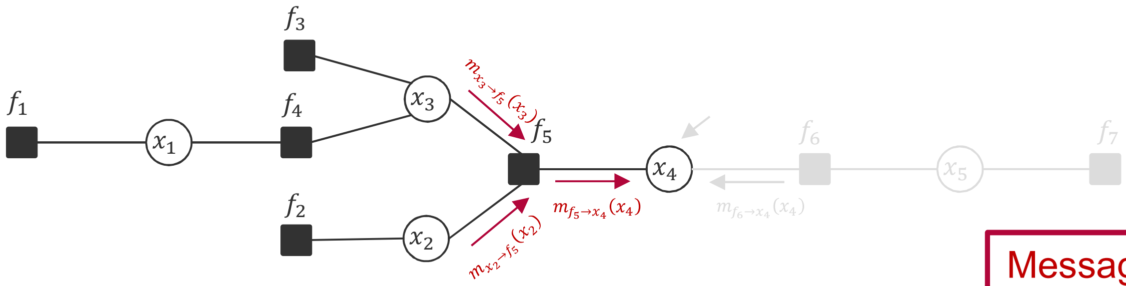
Message $m_{f_j \to x_i}(x_i)$ is the sum over all variables in the subtree rooted at f_j

$$p(x_4) = \sum_{\{x_1\}} \sum_{\{x_2\}} \sum_{\{x_3\}} \sum_{\{x_5\}} f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_2) \cdot f_4(x_1, x_3) \cdot f_5(x_2, x_3, x_4) \cdot f_6(x_4, x_5) \cdot f_7(x_5)$$

$$= \underbrace{\left[\sum_{\{x_1\}} \sum_{\{x_2\}} \sum_{\{x_3\}} f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_1, x_3) \cdot f_5(x_2, x_3, x_4) \right]}_{m_{f_5 \to x_4}(x_4)} \cdot \underbrace{\left[\sum_{\{x_5\}} f_6(x_4, x_5) \cdot f_7(x_5) \right]}_{m_{f_6 \to x_4}(x_4)}$$

Marginals are the product of all incoming messages from neighbouring factors!

Sum-Product Algorithm: Message from Factor to Variable



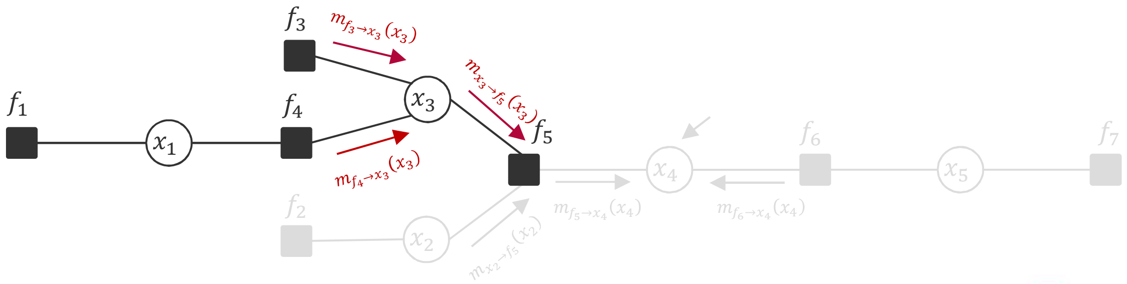
Message $m_{x_i \to f_j}(x_i)$ is the sum over all variables in the subtree rooted at x_i

$$m_{f_5 \to x_4}(x_4) = \sum_{\{x_1\}} \sum_{\{x_2\}} \sum_{\{x_3\}} f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_1, x_3) \cdot f_5(x_2, x_3, x_4)$$

$$= \sum_{\{x_2\}} \sum_{\{x_3\}} f_5(x_2, x_3, x_4) \cdot \underbrace{[f_2(x_2)]}_{m_{x_2 \to f_5}(x_2)} \cdot \underbrace{\left[\sum_{\{x_1\}} f_1(x_1) \cdot f_3(x_3) \cdot f_4(x_1, x_3)\right]}_{m_{x_3 \to f_5}(x_3)}$$

Messages from a factor to a variable sum out all neighboring variables weighted by their incoming message

Sum-Product Algorithm: Message from Variable to Factor



- $$m_{x \to f}(x) = \frac{p(x)}{m_{f \to x}(x)}$$

$$m_{x_3 \to f_5}(x_3) = \sum_{\{x_1\}} f_1(x_1) \cdot f_3(x_3) \cdot f_4(x_1, x_3)$$

$$= \underbrace{[f_3(x_3)]}_{m_{f_3 \to x_3}(x_3)} \cdot \underbrace{\left[\sum_{\{x_1\}} f_1(x_1) \cdot f_4(x_1, x_3)\right]}_{m_{f_4 \to x_3}(x_3)}$$

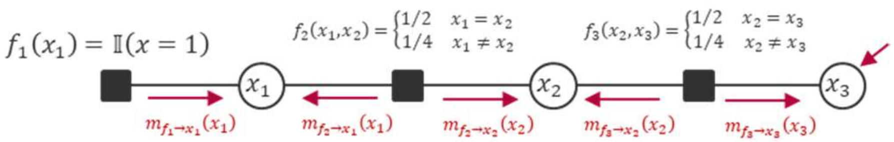
Messages from a variable to a factor multiply incoming message from neighboring factors

Overview

1. Questions and Updates
2. Recap: Main Concepts of Unit 4
3. **Example: Message Passing in Factor Graphs**
4. Hints for Exercise 2 (to be handed in Monday May 19)

Example Chain (Slide 13): Model & Dynamics

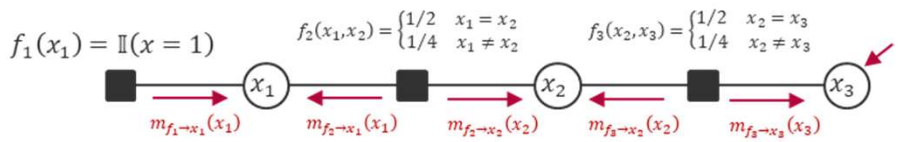
Chain Graph:



$$x_1 \in \{1,2,3\},\ \ x_2 \in \{1,2,3\},\ \ x_3 \in \{1,2,3\}$$

Example Chain (Slide 13): Model & Dynamics

Chain Graph:

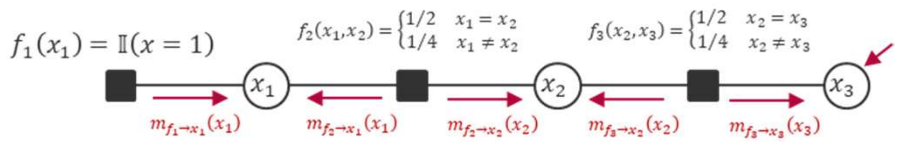


Initial Probability: $X_1 = 1 \quad \Rightarrow \quad P(X_1 = x_1) = 1_{\{x_1 = 1\}} \xrightarrow{x_1 = 1,2,3} (1,0,0)$ $\qquad x_1 \in \{1,2,3\},\ x_2 \in \{1,2,3\},\ x_3 \in \{1,2,3\}$

Transition Probability: $P(X_{i+1} = x_{i+1} \mid X_i = x_i) = f(x_i, x_{i+1}) = \begin{cases} 1/2 & x_{i+1} = x_i \\ 1/4 & x_{i+1} \neq x_i \end{cases} \qquad i = 1,2$

Example Chain (Slide 13): Computing Marginals as Usual

Chain Graph:



$$p(x_1, x_2, x_3) = p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2)$$

Initial Probability: $X_1 = 1 \quad \Rightarrow \quad P(X_1 = x_1) = 1_{\{x_1 = 1\}} \xrightarrow{x_1 = 1,2,3} (1,0,0)$ $\qquad x_1 \in \{1,2,3\},\ x_2 \in \{1,2,3\},\ x_3 \in \{1,2,3\}$

Transition Probability: $P(X_{i+1} = x_{i+1} \mid X_i = x_i) = f(x_i, x_{i+1}) = \begin{cases} 1/2 & x_{i+1} = x_i \\ 1/4 & x_{i+1} \neq x_i \end{cases} \qquad i = 1,2$

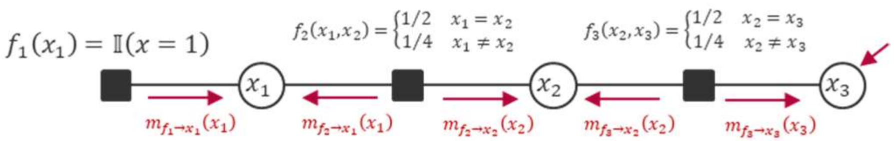
Marginal x1:

Marginal x2:

Marginal x3:

Example Chain (Slide 13): Computing Marginals as Usual

Chain Graph:



$$p(x_1, x_2, x_3) = p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2)$$

Initial Probability: $X_1 = 1 \quad \Rightarrow \quad P(X_1 = x_1) = 1_{\{x_1 = 1\}} \xrightarrow{x_1 = 1,2,3} (1,0,0)$ $\qquad x_1 \in \{1,2,3\},\ x_2 \in \{1,2,3\},\ x_3 \in \{1,2,3\}$

Transition Probability: $P(X_{i+1} = x_{i+1} \mid X_i = x_i) = f(x_i, x_{i+1}) = \begin{cases} 1/2 & x_{i+1} = x_i \\ 1/4 & x_{i+1} \neq x_i \end{cases} \qquad i = 1,2$

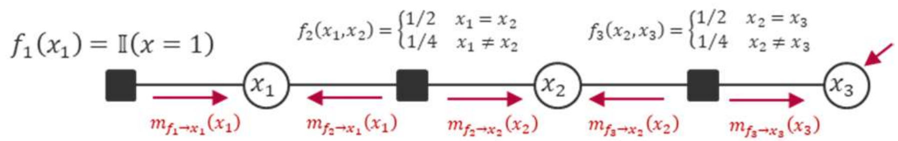
Marginal x1: $p(x_1) = \sum_{x_2} \sum_{x_3} p(x_1, x_2, x_3) = ??$

Marginal x2: $p(x_2) = \sum_{x_1} \sum_{x_3} p(x_1, x_2, x_3) = ??$

Marginal x3: $p(x_3) = \sum_{x_1} \sum_{x_2} p(x_1, x_2, x_3) = ??$

Example Chain (Slide 13): Computing Marginals as Usual

Chain Graph:



$$p(x_1,x_2,x_3) = p(x_1)\cdot p(x_2 \mid x_1)\cdot p(x_3 \mid x_2)$$

Initial Probability: $X_1 = 1 \quad \Rightarrow \quad P(X_1 = x_1) = 1_{\{x_1=1\}} \xrightarrow{x_1=1,2,3} (1,0,0)$ $\qquad x_1 \in \{1,2,3\},\ x_2 \in \{1,2,3\},\ x_3 \in \{1,2,3\}$

Transition Probability: $P(X_{i+1} = x_{i+1} \mid X_i = x_i) = f(x_i, x_{i+1}) = \begin{cases} 1/2 & x_{i+1} = x_i \\ 1/4 & x_{i+1} \neq x_i \end{cases} \qquad i = 1,2$

Marginal x1:

$$p(x_1) = \sum_{x_2}\sum_{x_3} p(x_1,x_2,x_3) = p(x_1)\cdot \underbrace{\sum_{x_2} p(x_2 \mid x_1)}_{1}\cdot \underbrace{\sum_{x_3} p(x_3 \mid x_2)}_{1} = p(x_1) = \boxed{f_1(x_1)}$$

Marginal x2:

$$p(x_2) = \sum_{x_1}\sum_{x_3} p(x_1,x_2,x_3) = \sum_{x_1}\sum_{x_3} p(x_1)\cdot p(x_2 \mid x_1)\cdot p(x_3 \mid x_2)$$

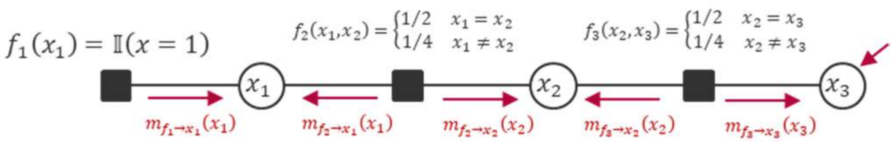
$$= \sum_{x_1} p(x_1)\cdot p(x_2 \mid x_1)\cdot \underbrace{\sum_{x_3} p(x_3 \mid x_2)}_{=1} = \boxed{\sum_{x_1} p(x_1)\cdot p(x_2 \mid x_1)}$$

Marginal x3:

$$p(x_3) = \sum_{x_1}\sum_{x_2} p(x_1,x_2,x_3) = \boxed{\sum_{x_1} p(x_1)\cdot \sum_{x_2} p(x_2 \mid x_1)\cdot p(x_3 \mid x_2)}$$

Example Chain (Slide 13): Results for Marginals

Chain Graph:



$$p(x_1, x_2, x_3) = p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2)$$

Initial Probability: $X_1 = 1 \quad \Rightarrow \quad P(X_1 = x_1) = 1_{\{x_1 = 1\}} \xrightarrow{x_1 = 1,2,3} (1,0,0)$

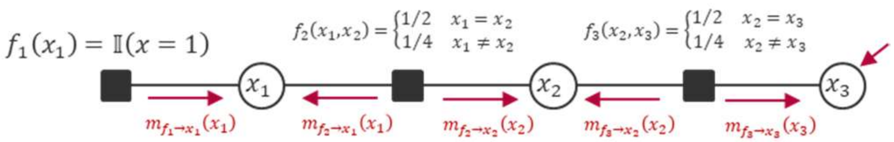
Transition Probability: $P(X_{i+1} = x_{i+1} \mid X_i = x_i) = \begin{cases} 1/2 & x_{i+1} = x_i \\ 1/4 & x_{i+1} \neq x_i \end{cases}$

Marginal x1: $p(x_1) = P(X_1 = x_1) = f_1(x_1) = 1_{\{x_1 = 1\}} \boxed{\xrightarrow{x_1 = 1,2,3} (1,0,0)}$

Marginal x2:
$$p(x_2) = \sum_{x_1} \sum_{x_3} p(x_1, x_2, x_3) = \sum_{x_1} \sum_{x_3} p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2)$$
$$= \sum_{x_1 = 1,2,3} p(x_1) \cdot p(x_2 \mid x_1)$$

Example Chain (Slide 13): Results for Marginals

Chain Graph:



$$p(x_1, x_2, x_3) = p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2)$$

Initial Probability: $X_1 = 1 \quad \Rightarrow \quad P(X_1 = x_1) = 1_{\{x_1=1\}} \xrightarrow{x_1=1,2,3} (1,0,0)$

Transition Probability: $P(X_{i+1} = x_{i+1} \mid X_i = x_i) = \begin{cases} 1/2 & x_{i+1} = x_i \\ 1/4 & x_{i+1} \neq x_i \end{cases}$

Marginal x1: $p(x_1) = P(X_1 = x_1) = f_1(x_1) = 1_{\{x_1=1\}} \xrightarrow{x_1=1,2,3} (1,0,0)$

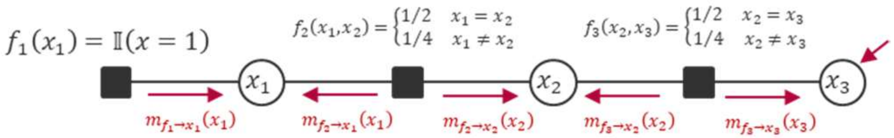
Marginal x2:

$$p(x_2) = \sum_{x_1} \sum_{x_3} p(x_1, x_2, x_3) = \sum_{x_1} \sum_{x_3} p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2)$$

$$= \sum_{x_1=1,2,3} p(x_1) \cdot p(x_2 \mid x_1) \overset{X_1=1}{=} p(x_2 \mid X_1 = 1) \boxed{\xrightarrow{x_2=1,2,3} \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4} \right)}$$

Example Chain (Slide 13): Computing Probabilities as Usual

Chain Graph:



$$p(x_1, x_2, x_3) = p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2)$$

Initial Probability: $X_1 = 1 \quad \Rightarrow \quad P(X_1 = x_1) = 1_{\{x_1 = 1\}} \xrightarrow{x_1 = 1,2,3} (1,0,0)$

Transition Probability: $P(X_{i+1} = x_{i+1} \mid X_i = x_i) = \begin{cases} 1/2 & x_{i+1} = x_i \\ 1/4 & x_{i+1} \neq x_i \end{cases}$

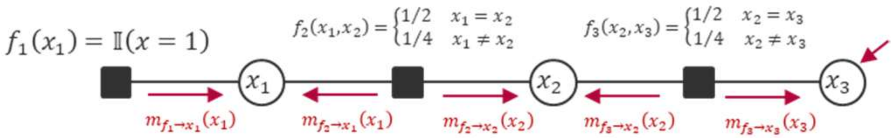
Marginal x3:

$$p(x_3) = \sum_{x_1 = 1,2,3} \sum_{x_2 = 1,2,3} p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2)$$

$$\overset{X_1 = 1}{=} \sum_{x_2 = 1,2,3} P(X_2 = x_2 \mid X_1 = 1) \cdot P(X_3 = x_3 \mid X_2 = x_2)$$

Example Chain (Slide 13): Computing Probabilities as Usual

Chain Graph:



$$p(x_1, x_2, x_3) = p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2)$$

Initial Probability: $X_1 = 1 \quad \Rightarrow \quad P(X_1 = x_1) = 1_{\{x_1 = 1\}} \xrightarrow{x_1 = 1,2,3} (1,0,0)$

Transition Probability: $P(X_{i+1} = x_{i+1} \mid X_i = x_i) = \begin{cases} 1/2 & x_{i+1} = x_i \\ 1/4 & x_{i+1} \neq x_i \end{cases}$

Marginal x3:

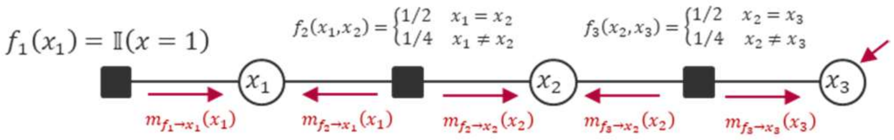
$$p(x_3) = \sum_{x_1 = 1,2,3} \sum_{x_2 = 1,2,3} p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2)$$

$$\overset{X_1 = 1}{=} \sum_{x_2 = 1,2,3} P(X_2 = x_2 \mid X_1 = 1) \cdot P(X_3 = x_3 \mid X_2 = x_2)$$

$$= \left(\underbrace{\overbrace{\frac{1}{2} \cdot \frac{1}{2}}^{x_2=1} + \overbrace{\frac{1}{4} \cdot \frac{1}{4}}^{x_2=2} + \overbrace{\frac{1}{4} \cdot \frac{1}{4}}^{x_2=3}}_{x_3=1}, \underbrace{\overbrace{\frac{1}{2} \cdot \frac{1}{4}}^{x_2=1} + \overbrace{\frac{1}{4} \cdot \frac{1}{2}}^{x_2=2} + \overbrace{\frac{1}{4} \cdot \frac{1}{4}}^{x_2=3}}_{x_3=2}, \underbrace{\overbrace{\frac{1}{2} \cdot \frac{1}{4}}^{x_2=1} + \overbrace{\frac{1}{4} \cdot \frac{1}{4}}^{x_2=2} + \overbrace{\frac{1}{4} \cdot \frac{1}{2}}^{x_2=3}}_{x_3=3} \right) \boxed{\xrightarrow{x_3 = 1,2,3} \left(\frac{6}{16}, \frac{5}{16}, \frac{5}{16} \right)}$$

Example Chain (Slide 13): Impact of Observations (x3=2)

Chain Graph:

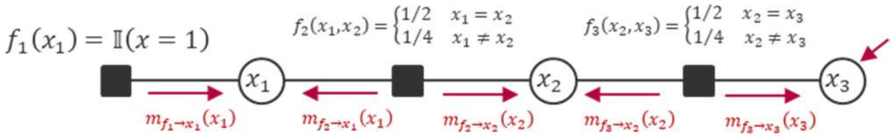


$Observation: \; X_3 = 2$

Marginals (**conditioned on x3=2**): ??

Example Chain (Slide 13): Impact of Observations (x3=2)

Chain Graph:



$Observation: \; X_3 = 2$

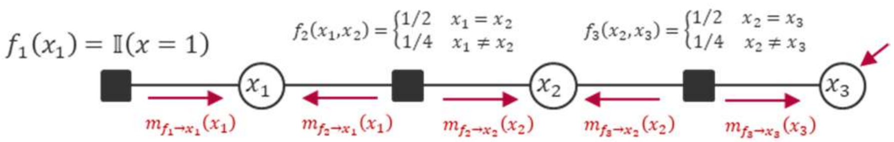
Marginals (conditioned):

$$P(X_1 = x_1 \mid X_3 = 2) = 1_{\{x_1 = 1\}} \xrightarrow{x_1 = 1,2,3} (1,0,0)$$

$$P(X_3 = x_3 \mid X_3 = 2) = 1_{\{x_3 = 2\}} \xrightarrow{x_3 = 1,2,3} (0,1,0)$$

Example Chain (Slide 13): Impact of Observations (x3=2)

Chain Graph:



$Observation:\ X_3 = 2$

Marginals (conditioned):

$$P(X_1 = x_1 \mid X_3 = 2) = 1_{\{x_1=1\}} \xrightarrow{x_1=1,2,3} (1,0,0)$$

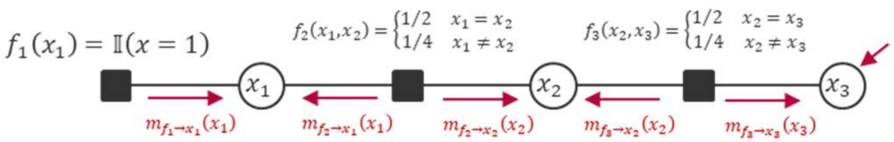
$$P(X_3 = x_3 \mid X_3 = 2) = 1_{\{x_3=2\}} \xrightarrow{x_3=1,2,3} (0,1,0)$$

$$P(X_2 = x_2 \mid X_3 = 2) = \frac{P(X_2 = x_2, X_3 = 2)}{P(X_3 = 2)} = \frac{\sum_{x_1} \sum_{x_3 : x_3 = 2} p(x_1, x_2, x_3)}{\sum_{x_1} \sum_{x_2} \sum_{x_3 : x_3 = 2} p(x_1, x_2, x_3)} = \frac{\sum_{x_1} p(x_1) \cdot p(x_2 \mid x_1) \cdot \sum_{x_3 : x_3 = 2} p(x_3 \mid x_2)}{\sum_{x_1} p(x_1) \cdot \sum_{x_2} p(x_2 \mid x_1) \cdot \sum_{x_3 : x_3 = 2} p(x_3 \mid x_2)}$$

$$\overset{\substack{X_1=1 \\ X_3=2}}{=} \quad \frac{??}{5/16}$$

Example Chain (Slide 13): Impact of Observations (x3=2)

Chain Graph:



$$\mathit{Observation}: \; X_3 = 2$$

Marginals (conditioned):

$$P(X_1 = x_1 \mid X_3 = 2) = 1_{\{x_1 = 1\}} \overset{x_1 = 1,2,3}{\rightarrow} (1,0,0)$$

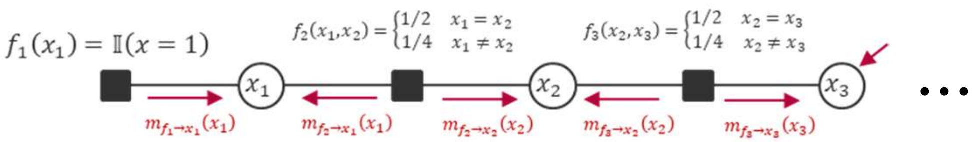
$$P(X_3 = x_3 \mid X_3 = 2) = 1_{\{x_3 = 2\}} \overset{x_1 = 1,2,3}{\rightarrow} (0,1,0)$$

$$P(X_2 = x_2 \mid X_3 = 2) = \frac{P(X_2 = x_2, X_3 = 2)}{P(X_3 = 2)} = \frac{\sum_{x_1} \sum_{x_3 : x_3 = 2} p(x_1, x_2, x_3)}{\sum_{x_1} \sum_{x_2} \sum_{x_3 : x_3 = 2} p(x_1, x_2, x_3)} = \frac{\sum_{x_1} p(x_1) \cdot p(x_2 \mid x_1) \cdot \sum_{x_3 : x_3 = 2} p(x_3 \mid x_2)}{\sum_{x_1} p(x_1) \cdot \sum_{x_2} p(x_2 \mid x_1) \cdot \sum_{x_3 : x_3 = 2} p(x_3 \mid x_2)}$$

$$\overset{\substack{X_1 = 1 \\ X_3 = 2}}{=} \frac{P(X_2 = x_2 \mid X_1 = 1) \cdot P(X_3 = 2 \mid X_2 = x_2)}{5/16} \overset{x_2 = 1,2,3}{\rightarrow} \frac{16}{5} \cdot \left(\frac{1}{2} \cdot \frac{1}{4}, \frac{1}{4} \cdot \frac{1}{2}, \frac{1}{4} \cdot \frac{1}{4} \right) = \left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5} \right)$$

Example Chain (Slide 13): Are Longer Chains Tractable?

Chain with K nodes:



Initial Probability: $X_1 = 1 \quad \Rightarrow \quad P(X_1 = x_1) = f_1(x_1) = 1_{\{x_1 = 1\}} \xrightarrow{x_1 = 1,2,3} (1,0,0)$

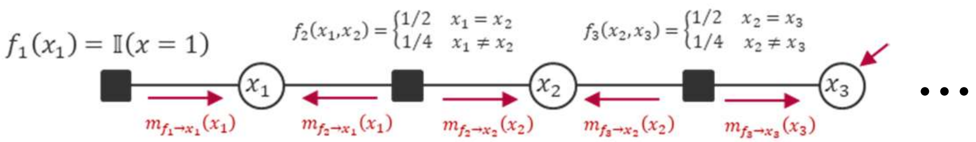
Transition Probability: $P(X_{i+1} = x_{i+1} \mid X_i = x_i) = \begin{cases} 1/2 & x_{i+1} = x_i \\ 1/4 & x_{i+1} \neq x_i \end{cases} \qquad i = 1, \ldots, K-1$

Marginals: ??

Marginals (conditioned): ??

Example Chain (Slide 13): Are Longer Chains Tractable?

Chain with K nodes:



Initial Probability: $X_1 = 1 \quad \Rightarrow \quad P(X_1 = x_1) = f_1(x_1) = 1_{\{x_1=1\}} \overset{x_1=1,2,3}{\longrightarrow} (1,0,0)$

Transition Probability: $P(X_{i+1} = x_{i+1} \mid X_i = x_i) = \begin{cases} 1/2 & x_{i+1} = x_i \\ 1/4 & x_{i+1} \neq x_i \end{cases} \qquad i = 1,\ldots,K-1$

Marginals: $P(X_k = x_k) = p(x_k) = \sum_{x_1} \cdots \sum_{x_{k-1}} \sum_{x_{k+1}} \cdots \sum_{x_K} p(x_1,\ldots,x_K) \overset{x_k=1,2,3}{\longrightarrow} (?,?,?)$

Marginals (conditioned):

$$P(X_k = x_k \mid X_K = 2) = \frac{P(X_k = x_k, X_K = 2)}{P(X_K = 2)}$$

$$= \frac{\sum_{x_1} \cdots \sum_{x_{k-1}} \sum_{x_{k+1}} \cdots \sum_{x_K : x_K = 2} p(x_1,\ldots,x_K)}{\sum_{x_1} \cdots \sum_{x_K : x_K = 2} p(x_1,\ldots,x_K)} \overset{x_k=1,2,3}{\longrightarrow} (?,?,?)$$

Towards Alternatives: Sums & Products of Probabilities

Discrete events: $P(X \leq 2) = P(X = 1) + P(X = 2)$

Expected Value: $\sum_k k \cdot P(X = k)$

Towards Alternatives: Sums & Products of Probabilities

Discrete events: $P(X \leq 2) = P(X = 1) + P(X = 2)$

Expected Value: $\sum_k k \cdot P(X = k)$

Joint Probability: $P(X = 2, Y = 30) \overset{indep.}{=} P(X = 2) \cdot P(Y = 30)$

Cond. Probability: $P(X = 2 \mid Y = 30) = \dfrac{P(X = 2, Y = 30)}{P(Y = 30)}$

Towards Alternatives: Sums & Products of Probabilities

Discrete events: $P(X \le 2) = P(X=1) + P(X=2)$

Expected Value: $\sum_k k \cdot P(X=k)$

Joint Probability: $P(X=2, Y=30) \overset{indep.}{=} P(X=2) \cdot P(Y=30)$

Cond. Probability: $P(X=2 \mid Y=30) = \dfrac{P(X=2, Y=30)}{P(Y=30)}$

Likelihood x Prior: $P(\theta \mid D) = \dfrac{P(D \mid \theta) \cdot P(\theta)}{P(D)}$ $\quad P(D) := \int_{-\infty}^{+\infty} P(D \mid \theta) \cdot P(\theta)\ d\theta$

Towards Alternatives: Sums & Products of Probabilities

Discrete events: $P(X \le 2) = P(X = 1) + P(X = 2)$

Expected Value: $\sum_k k \cdot P(X = k)$

Joint Probability: $P(X = 2, Y = 30) \overset{indep.}{=} P(X = 2) \cdot P(Y = 30)$

Cond. Probability: $P(X = 2 \mid Y = 30) = \dfrac{P(X = 2, Y = 30)}{P(Y = 30)}$

Likelihood x Prior: $P(\theta \mid D) = \dfrac{P(D \mid \theta) \cdot P(\theta)}{P(D)}$ $\qquad P(D) := \int_{-\infty}^{+\infty} P(D \mid \theta) \cdot P(\theta)\ d\theta$

$$\Rightarrow \int_{-\infty}^{+\infty} P(\theta \mid D)\ d\theta = \frac{1}{P(D)} \cdot \int_{-\infty}^{+\infty} P(D \mid \theta) \cdot P(\theta)\ d\theta = \frac{P(D)}{P(D)} = 1$$

Towards Alternatives: Sums & Products of Probabilities

Discrete events: $P(X \leq 2) = P(X=1) + P(X=2)$

Expected Value: $\sum_k k \cdot P(X=k)$

Joint Probability: $P(X=2, Y=30) \overset{indep.}{=} P(X=2) \cdot P(Y=30)$

Cond. Probability: $P(X=2 \mid Y=30) = \dfrac{P(X=2, Y=30)}{P(Y=30)}$

Likelihood x Prior: $P(\theta \mid D) = \dfrac{P(D \mid \theta) \cdot P(\theta)}{P(D)}$ $\qquad P(D) := \int_{-\infty}^{+\infty} P(D \mid \theta) \cdot P(\theta)\ d\theta$

$$\Rightarrow \int_{-\infty}^{+\infty} P(\theta \mid D)\ d\theta = \frac{1}{P(D)} \cdot \int_{-\infty}^{+\infty} P(D \mid \theta) \cdot P(\theta)\ d\theta = \frac{P(D)}{P(D)} = 1$$

Product of Probabilities: $h(x,y) = f(x) \cdot g(y)?$ $\qquad h(y) = f(g(y))?$ $\qquad h(x) = f(x) \cdot g(x)?$

Towards Alternatives: Sums & Products of Probabilities

Discrete events: $P(X \le 2) = P(X=1) + P(X=2)$

Expected Value: $\sum_k k \cdot P(X=k)$

Joint Probability: $P(X=2, Y=30) \overset{indep.}{=} P(X=2) \cdot P(Y=30)$

Cond. Probability: $P(X=2 \mid Y=30) = \dfrac{P(X=2, Y=30)}{P(Y=30)}$

Likelihood x Prior: $P(\theta \mid D) = \dfrac{P(D \mid \theta) \cdot P(\theta)}{P(D)}$ $\quad P(D) := \int_{-\infty}^{+\infty} P(D \mid \theta) \cdot P(\theta)\ d\theta$

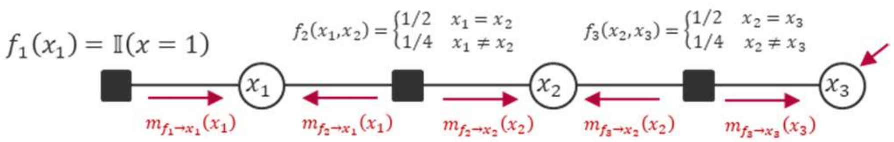
$$\Rightarrow \int_{-\infty}^{+\infty} P(\theta \mid D)\ d\theta = \frac{1}{P(D)} \cdot \int_{-\infty}^{+\infty} P(D \mid \theta) \cdot P(\theta)\ d\theta = \frac{P(D)}{P(D)} = 1$$

Product of Probabilities: $h(x,y) = f(x) \cdot g(y)?$ $\quad h(y) = f(g(y))?$ $\quad h(x) = f(x) \cdot g(x)?$

$\int f(x) \cdot g(x)\ dx = 1?$ $\quad \tilde{h}(x) = f(x) \cdot g(x) / \int f(x) \cdot g(x)\ dx\ ?$

Example Chain (Slide 13): Marginal & Messages (Plan)

Chain Graph:



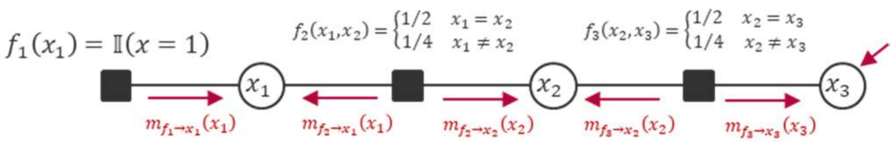
1. Initialize
2. Start at leaf x1

Initial message f1 to x1: $m_{f_1 \to x_1}(x_1) = f_1(x_1)$

then marginal x1: $p(x_1)$

Example Chain (Slide 13): Marginal & Messages (Plan)

Chain Graph:



1. Initialize
2. Start at leaf x1

Initial message f1 to x1: $m_{f_1 \to x_1}(x_1) = f_1(x_1)$

then marginal x1: $p(x_1)$

message from f2 to x2: $m_{f_2 \to x_2}(x_2)$

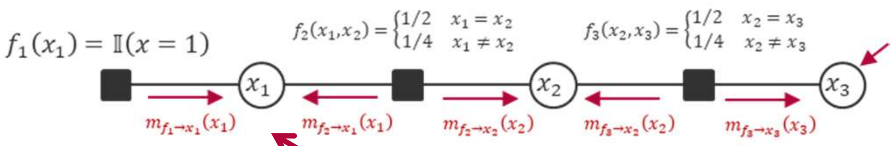
then marginal x2: $p(x_2)$

message from f3 to x3: $m_{f_3 \to x_3}(x_3)$

then marginal x3: $p(x_3)$

Example Chain (Slide 13): Marginal & Messages (Plan)

Chain Graph:



1. Initialize
2. Start at leaf x1

Initial message f1 to x1: $m_{f_1 \to x_1}(x_1) = f_1(x_1)$

$use: \; p(x_1) \triangleq m_{f_1 \to x_1}(x_1) \cdot m_{f_2 \to x_1}(x_1)$

then **marginal x1**: $p(x_1)$

message from f2 to x2: $m_{f_2 \to x_2}(x_2)$

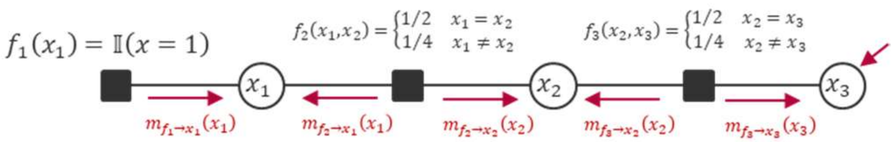
then marginal x2: $p(x_2)$

message from f3 to x3: $m_{f_3 \to x_3}(x_3)$

then marginal x3: $p(x_3)$

Example Chain (Slide 13): Marginal & Messages (Plan)

Chain Graph:



$$\underbrace{m_{x_1 \to f_2}(x_1)}_{p(x_1)/m_{f_2 \to x_1}(x_1)}$$

1. Initialize

2. Start at leaf x1

Initial message f1 to x1: $m_{f_1 \to x_1}(x_1) = f_1(x_1)$

then marginal x1: $p(x_1)$

use: $p(x_1) \triangleq m_{f_1 \to x_1}(x_1) \cdot m_{f_2 \to x_1}(x_1)$

message from f2 to x2: $m_{f_2 \to x_2}(x_2)$

use: $m_{f_2 \to x_2}(x_2) \triangleq \sum_{x_1 = 1,2,3} f_2(x_1, x_2) \cdot m_{x_1 \to f_2}(x_1)$

then marginal x2: $p(x_2)$

and: $p(x_1) = m_{f_2 \to x_1}(x_1) \cdot m_{x_1 \to f_2}(x_1)$

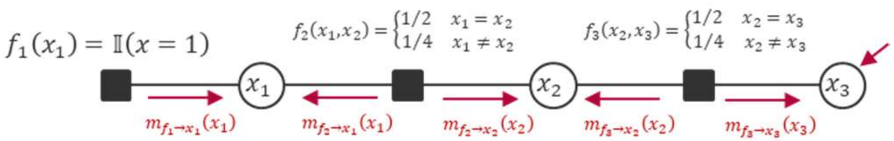
$\Rightarrow m_{x_1 \to f_2}(x_1) = p(x_1) / m_{f_2 \to x_1}(x_1)$

message from f3 to x3: $m_{f_3 \to x_3}(x_3)$

then marginal x3: $p(x_3)$

Example Chain (Slide 13): Marginal & Messages (Plan)

Chain Graph:



1. Initialize
2. Start at leaf x1

Initial message f1 to x1: $m_{f_1 \to x_1}(x_1) = f_1(x_1)$

then marginal x1: $p(x_1)$

message from f2 to x2: $m_{f_2 \to x_2}(x_2)$

then marginal x2: $p(x_2)$

message from f3 to x3: $m_{f_3 \to x_3}(x_3)$

then marginal x3: $p(x_3)$

$use: \; p(x_1) \triangleq m_{f_1 \to x_1}(x_1) \cdot m_{f_2 \to x_1}(x_1)$

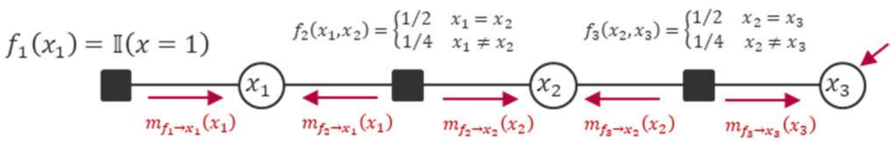
$use: \; m_{f_2 \to x_2}(x_2) \triangleq \sum_{x_1 = 1,2,3} f_2(x_1, x_2) \cdot m_{x_1 \to f_2}(x_1)$

$and: \; p(x_1) = m_{f_2 \to x_1}(x_1) \cdot m_{x_1 \to f_2}(x_1)$

$\Rightarrow m_{x_1 \to f_2}(x_1) = p(x_1) / m_{f_2 \to x_1}(x_1)$

Example Chain (Slide 13): Computing Marginal & Messages

Chain Graph:



1. Initialize (Uniform!): $P(X_k = x_k) = 1/3 \quad m_{f_k \to x_k}(x_k) = 1/3 \quad m_{f_{k+1} \to x_k}(x_k) = 1/3 \quad x_k = 1,2,3 \quad k = 1,\dots,3$

2. Start at leaf x1

Initial message f1 to x1: $m_{f_1 \to x_1}(x_1) = f_1(x_1) = 1_{\{x_1 = 1\}} \xrightarrow{x_1 = 1,2,3} (1,0,0)$

then marginal x1:

message from f2 to x2:

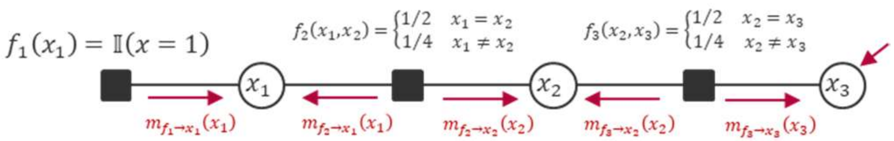
then marginal x2:

message from f3 to x3:

then marginal x3:

Example Chain (Slide 13): Computing Marginal & Messages

Chain Graph:



1. Initialize (Uniform!): $P(X_k = x_k) = 1/3 \quad m_{f_k \to x_k}(x_k) = 1/3 \quad m_{f_{k+1} \to x_k}(x_k) = 1/3 \quad x_k = 1,2,3 \quad k = 1,\ldots,3$

2. Start at leaf x1

Initial message f1 to x1: $m_{f_1 \to x_1}(x_1) = f_1(x_1) = 1_{\{x_1 = 1\}} \xrightarrow{x_1 = 1,2,3} (1,0,0)$

then **marginal x1**:

$$p(x_1) = P(X_1 = x_1) = \frac{m_{f_1 \to x_1}(x_1) \cdot m_{f_2 \to x_1}(x_1)}{\sum_{\tilde{x}_1 = 1,2,3} m_{f_1 \to x_1}(\tilde{x}_1) \cdot m_{f_2 \to x_1}(\tilde{x}_1)} \xrightarrow{x_1 = 1,2,3} (1,0,0)$$

message from f2 to x2:

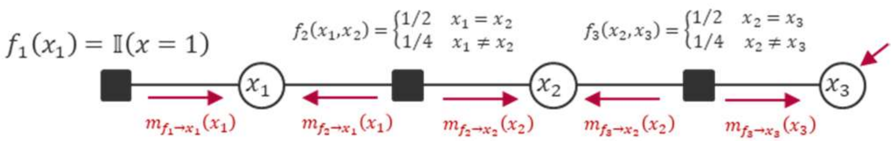
then marginal x2:

message from f3 to x3:

then marginal x3:

Example Chain (Slide 13): Computing Marginal & Messages

Chain Graph:



1. Initialize (Uniform!): $P(X_k = x_k) = 1/3 \quad m_{f_k \to x_k}(x_k) = 1/3 \quad m_{f_{k+1} \to x_k}(x_k) = 1/3 \quad x_k = 1,2,3 \quad k = 1,\ldots,3$

2. Start at leaf x1

Initial message f1 to x1: $m_{f_1 \to x_1}(x_1) = f_1(x_1) = 1_{\{x_1 = 1\}} \xrightarrow{x_1 = 1,2,3} (1,0,0)$

then marginal x1:

$$p(x_1) = P(X_1 = x_1) = \frac{m_{f_1 \to x_1}(x_1) \cdot m_{f_2 \to x_1}(x_1)}{\sum_{\tilde{x}_1 = 1,2,3} m_{f_1 \to x_1}(\tilde{x}_1) \cdot m_{f_2 \to x_1}(\tilde{x}_1)} \xrightarrow{x_1 = 1,2,3} (1,0,0)$$

message from f2 to x2:

then marginal x2:

message from f3 to x3:

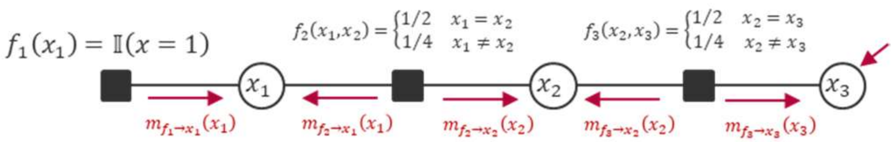
$$m_{f_2 \to x_2}(x_2) = \frac{\sum_{x_1 = 1,2,3} f_2(x_1, x_2) \cdot p(x_1) / m_{f_2 \to x_1}(x_1)}{\sum_{\tilde{x}_2 = 1,2,3} \sum_{\tilde{x}_1 = 1,2,3} f_2(\tilde{x}_1, \tilde{x}_2) \cdot p(\tilde{x}_1) / m_{f_2 \to x_1}(\tilde{x}_1)} \xrightarrow{x_2 = 1,2,3} \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$$

then marginal x3:

…

Example Chain (Slide 13): Forward Results

Chain Graph:



initial message from f1 to x1: $m_{f_1 \to x_1}(x_1) = f_1(x_1) \xrightarrow{x_1=1,2,3} (1,0,0)$

then marginal x1: $p(x_1) \xrightarrow{x_1=1,2,3} (1,0,0)$

then message from f2 to x2: $m_{f_2 \to x_2}(x_2) \xrightarrow{x_2=1,2,3} \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$

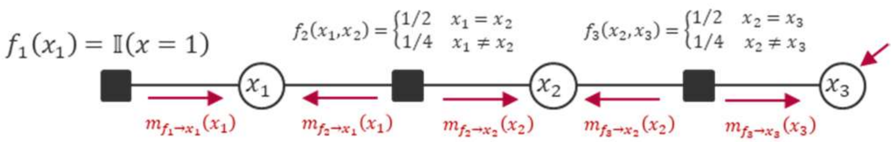
then **marginal x2**: should be ??

then message from f3 to x3:

then **marginal x3**: should be ??

Example Chain (Slide 13): Forward Results (Confirmed)

Chain Graph:



initial message from f1 to x1: $m_{f_1 \to x_1}(x_1) = f_1(x_1) \overset{x_1=1,2,3}{\to} (1,0,0)$

then marginal x1: $p(x_1) \overset{x_1=1,2,3}{\to} (1,0,0)$

then message from f2 to x2: $m_{f_2 \to x_2}(x_2) \overset{x_2=1,2,3}{\to} \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$

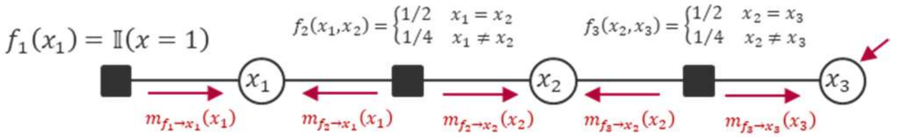
then **marginal x2**: $p(x_2) \overset{x_2=1,2,3}{\to} \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$

then message from f3 to x3: $m_{f_3 \to x_3}(x_3) \overset{x_3=1,2,3}{\to} \left(\frac{6}{16}, \frac{5}{16}, \frac{5}{16}\right)$

then **marginal x3**: $p(x_3) \overset{x_3=1,2,3}{\to} \left(\frac{6}{16}, \frac{5}{16}, \frac{5}{16}\right)$

Example Chain (Slide 13): Impact of Observations (x3=2)

Chain Graph:



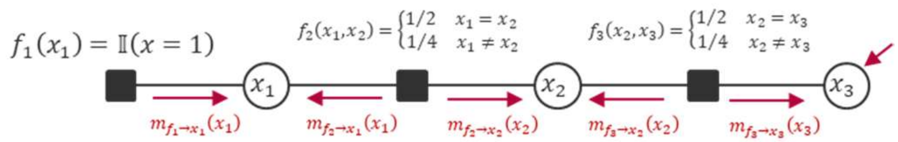
$Observation: \; X_3 = 2$

3. Observation: $X_3 = 2$

4. What now?

Example Chain (Slide 13): Impact of Observations (x3=2)

Chain Graph:



$Observation: \; X_3 = 2$

3. Observation: $X_3 = 2$

4. New factor f4 to x3: $f_4(x_3) = 1_{\{x_3=2\}}$

Now messages **backwards**: $m_{f_4 \to x_3}(x_3) = f_4(x_3)$

then marginal x3: $p(x_3)$

message from f3 to x2: $m_{f_3 \to x_2}(x_2)$

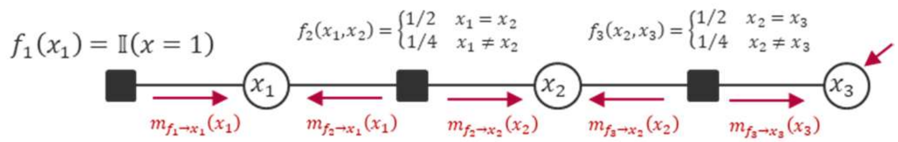
then marginal x2: $p(x_2)$

message from f2 to x1: $m_{f_2 \to x_1}(x_1)$

then marginal x1: $p(x_1)$

Example Chain (Slide 13): Impact of Observations (x3=2)

Chain Graph:



$Observation: \ X_3 = 2$

3. Observation: $X_3 = 2$

4. New factor f4 to x3: $f_4(x_3) = 1_{\{x_3=2\}}$

Now messages **backwards**: $m_{f_4\to x_3}(x_3) = f_4(x_3) = 1_{\{x_3=2\}} \xrightarrow{x_3=1,2,3} (0,1,0)$

then **marginal x3**: $p(x_3) \longleftarrow$ $use: \ p(x_3) \triangleq m_{f_3\to x_3}(x_3)\cdot m_{f_4\to x_3}(x_3)$

message from f3 to x2: $m_{f_3\to x_2}(x_2)$

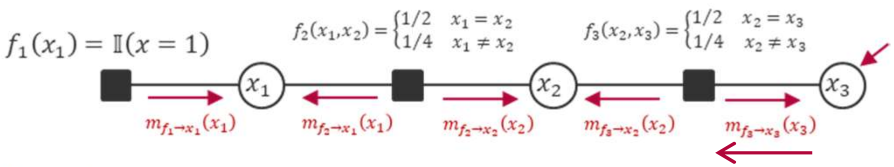
then marginal x2: $p(x_2)$

message from f2 to x1: $m_{f_2\to x_1}(x_1)$

then marginal x1: $p(x_1)$

Example Chain (Slide 13): Impact of Observations (x3=2)

Chain Graph:



$Observation:\ X_3 = 2$

$\underbrace{m_{x_3\to f_3}(x_3)}_{p(x_3)/m_{f_3\to x_3}(x_3)}$

3. Observation: $X_3 = 2$

4. New factor f4 to x3: $f_4(x_3) = 1_{\{x_3=2\}}$

Now messages **backwards**: $m_{f_4\to x_3}(x_3) = f_4(x_3) = 1_{\{x_3=2\}} \overset{x_3=1,2,3}{\rightarrow} (0,1,0)$

then marginal x3: $p(x_3)$ $\quad$ $use:\ p(x_3) \triangleq m_{f_3\to x_3}(x_3)\cdot m_{f_4\to x_3}(x_3)$

message from f3 to x2: $m_{f_3\to x_2}(x_2)$

$use:\ m_{f_3\to x_2}(x_2) \triangleq \sum_{x_3=1,2,3} f_3(x_2,x_3)\cdot m_{x_3\to f_3}(x_3)$

then marginal x2: $p(x_2)$

$and:\ p(x_3) = m_{f_3\to x_3}(x_3)\cdot m_{x_3\to f_3}(x_3)$

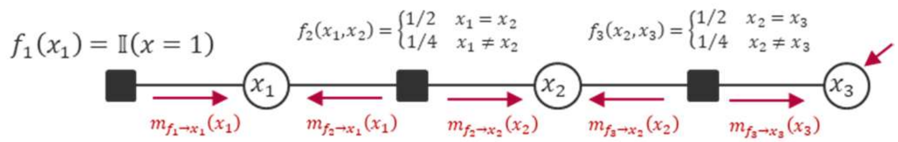
message from f2 to x1: $m_{f_2\to x_1}(x_1)$

$\Rightarrow m_{x_3\to f_3}(x_3) = p(x_3)/m_{f_3\to x_3}(x_3)$

then marginal x1: $p(x_1)$

Example Chain (Slide 13): Impact of Observations (x3=2)

Chain Graph:



Observation: $X_3 = 2$

3. Observation: $X_3 = 2$

4. New factor f4 to x3: $f_4(x_3) = 1_{\{x_3=2\}}$

Now messages **backwards**: $m_{f_4 \to x_3}(x_3) = f_4(x_3) = 1_{\{x_3=2\}} \overset{x_3=1,2,3}{\to} (0,1,0)$

then marginal x3: $p(x_3)$ — *use*: $p(x_3) \triangleq m_{f_3 \to x_3}(x_3) \cdot m_{f_4 \to x_3}(x_3)$

message from f3 to x2: $m_{f_3 \to x_2}(x_2)$

then **marginal x2**: $p(x_2)$

use: $m_{f_3 \to x_2}(x_2) \triangleq \sum_{x_3=1,2,3} f_3(x_2, x_3) \cdot m_{x_3 \to f_3}(x_3)$

and: $p(x_3) = m_{f_3 \to x_3}(x_3) \cdot m_{x_3 \to f_3}(x_3)$

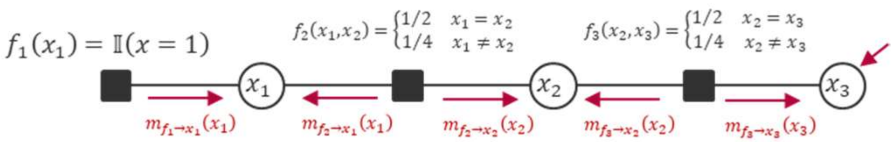
$\Rightarrow m_{x_3 \to f_3}(x_3) = p(x_3) / m_{f_3 \to x_3}(x_3)$

message from f2 to x1: $m_{f_2 \to x_1}(x_1)$

then **marginal x1**: $p(x_1)$

Example Chain (Slide 13): Impact of Observations (x3=2)

Chain Graph:



$Observation: \; X_3 = 2$

3. Observation: $X_3 = 2$

4. New factor f4 to x3: $f_4(x_3) = 1_{\{x_3=2\}}$

Now messages **backwards**: $m_{f_4 \to x_3}(x_3) = f_4(x_3) = 1_{\{x_3=2\}} \overset{x_3=1,2,3}{\rightarrow} (0,1,0)$

then **marginal x3**: $$p(x_3) = P(X_3 = x_3) = \frac{m_{f_3 \to x_3}(x_3) \cdot m_{f_4 \to x_3}(x_3)}{\sum_{\tilde{x}_3=1,2,3} m_{f_3 \to x_3}(\tilde{x}_3) \cdot m_{f_4 \to x_3}(\tilde{x}_3)} \overset{x_3=1,2,3}{\rightarrow} (0,1,0)$$

message from f3 to x2: $m_{f_3 \to x_2}(x_2)$

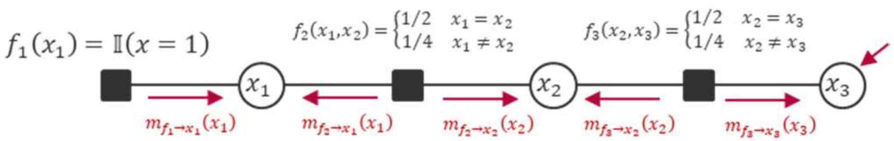
then marginal x2: $p(x_2)$

message from f2 to x1: $m_{f_2 \to x_1}(x_1)$

then marginal x1: $p(x_1)$

Example Chain (Slide 13): Impact of Observations (x3=2)

Chain Graph:



$Observation: \; X_3 = 2$

3. Observation: $X_3 = 2$

4. New factor f4 to x3: $f_4(x_3) = 1_{\{x_3=2\}}$

Now messages **backwards**: $m_{f_4\to x_3}(x_3) = f_4(x_3) = 1_{\{x_3=2\}} \overset{x_3=1,2,3}{\longrightarrow} (0,1,0)$

then marginal x3: $p(x_3) = P(X_3 = x_3) = \dfrac{m_{f_3\to x_3}(x_3)\cdot m_{f_4\to x_3}(x_3)}{\sum_{\tilde{x}_3=1,2,3} m_{f_3\to x_3}(\tilde{x}_3)\cdot m_{f_4\to x_3}(\tilde{x}_3)} \overset{x_3=1,2,3}{\longrightarrow} (0,1,0)$

message from f3 to x2:

$$m_{f_3\to x_2}(x_2) = \frac{\sum_{x_3=1,2,3} f_3(x_2,x_3)\cdot p(x_3)/m_{f_3\to x_3}(x_3)}{\sum_{\tilde{x}_2=1,2,3}\sum_{\tilde{x}_3=1,2,3} f_3(\tilde{x}_2,\tilde{x}_3)\cdot p(\tilde{x}_3)/m_{f_3\to x_3}(\tilde{x}_3)} \overset{x_2=1,2,3}{\longrightarrow} \left(\frac{1}{4},\frac{1}{2},\frac{1}{4}\right)$$

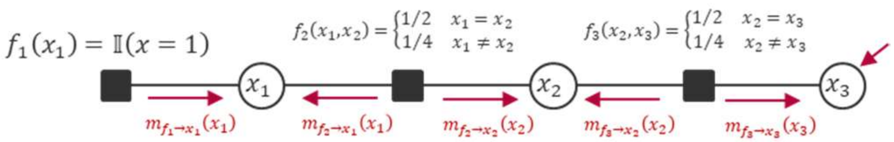
then marginal x2:

message from f2 to x1:

then marginal x1: $\cdots$

Example Chain (Slide 13): Backward Results (x3=2)

Chain Graph:



$Observation: \; X_3 = 2$

initial message: $m_{f_4 \to x_3}(x_3) := 1_{\{x_3=2\}} \overset{x_3=1,2,3}{\to} (0,1,0)$

then marginal x3: $p(x_3) \overset{x_3=1,2,3}{\to} (0,1,0)$

then message from x3 to x2: $m_{f_3 \to x_2}(x_2) \overset{x_2=1,2,3}{\to} \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$

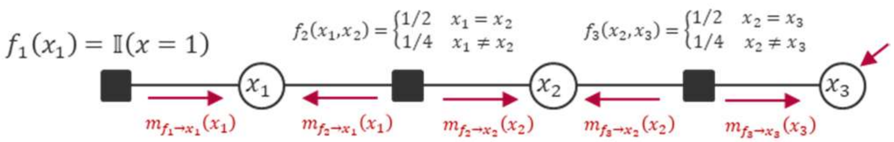
then **marginal x2**: should be ??

then message from x2 to x1:

then **marginal x1**:

Example Chain (Slide 13): Backward Results (x3=2)

Chain Graph:



$Observation: \; X_3 = 2$

initial message: $m_{f_4 \to x_3}(x_3) := 1_{\{x_3=2\}} \overset{x_3=1,2,3}{\rightarrow} (0,1,0)$

then marginal x3: $p(x_3) \overset{x_3=1,2,3}{\rightarrow} (0,1,0)$

then message from x3 to x2: $m_{f_3 \to x_2}(x_2) \overset{x_2=1,2,3}{\rightarrow} \left(\frac{1}{4},\frac{1}{2},\frac{1}{4}\right)$

then **marginal x2**: $p(x_2) \overset{x_2=1,2,3}{\rightarrow} \left(\frac{2}{5},\frac{2}{5},\frac{1}{5}\right)$

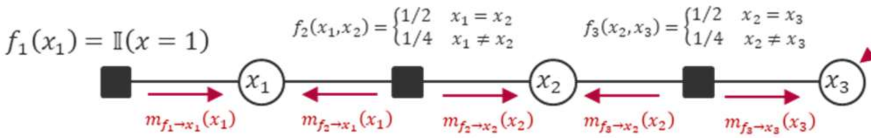
then message from x2 to x1: $m_{f_2 \to x_1}(x_1) \overset{x_1=1,2,3}{\rightarrow} \left(\frac{5}{16},\frac{6}{16},\frac{5}{16}\right)$

then **marginal x1**: $p(x_1) \overset{x_1=1,2,3}{\rightarrow} (1,0,0)$

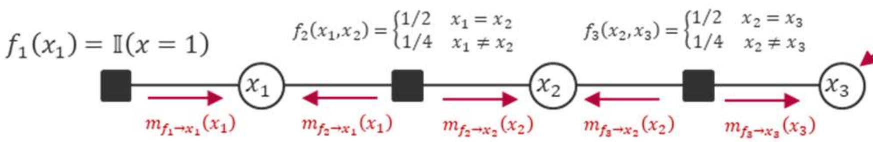
Chains: Generalizations (More Nodes & Any Dynamics)

Chain (K nodes):

$$f_{K+1}(x_K) = (*,*,*)$$



Chain (K nodes):

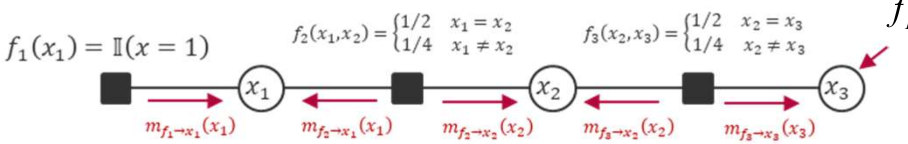


$$f_{K+1}(x_K) = (*,*,*)$$

$$P(X_k = x_k \mid X_{k-1} = x_{k-1}) = f_k(x_{k-1}, x_k) = \begin{array}{c|ccc} x_{k-1} \setminus x_k & 1 & 2 & 3 \\ \hline 1 & * & * & * \\ 2 & * & * & * \\ 3 & * & * & * \end{array}$$

Chains: Generalizations (More Nodes & Any Dynamics)

Chain (K nodes):



$f_{K+1}(x_K) = (*,*,*)$

1. Initialize via Uniform

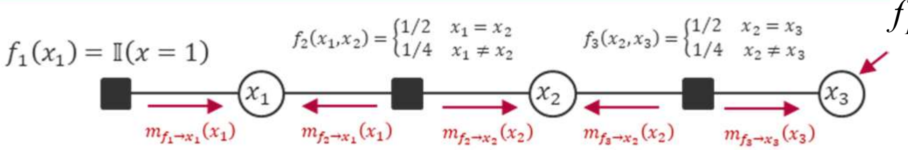
$$P(X_k = x_k \mid X_{k-1} = x_{k-1}) = f_k(x_{k-1}, x_k) =$$

$x_{k-1} \setminus x_k$	1	2	3
1	*	*	*
2	*	*	*
3	*	*	*

2. Update initial message: $f_1(x_1) = (*,*,*)$

Chains: Generalizations (More Nodes & Any Dynamics)

Chain (K nodes):



$f_{K+1}(x_K) = (*,*,*)$

1. Initialize via Uniform

$$P(X_k = x_k \mid X_{k-1} = x_{k-1}) = f_k(x_{k-1}, x_k) =$$

$x_{k-1} \setminus x_k$	1	2	3
1	*	*	*
2	*	*	*
3	*	*	*

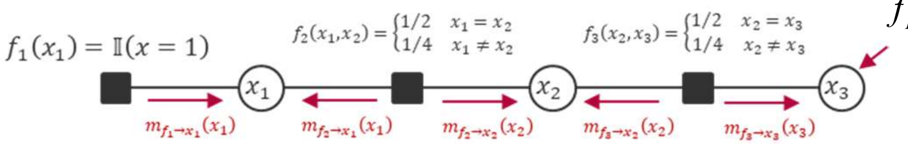
2. Update initial message: $f_1(x_1) = (*,*,*)$

3. Update marginal xk:

$$p(x_k) = P(X_k = x_k) = \frac{m_{f_k \to x_k}(x_k) \cdot m_{f_{k+1} \to x_k}(x_k)}{\sum_{\tilde{x}_k = 1,2,3} m_{f_k \to x_k}(\tilde{x}_k) \cdot m_{f_{k+1} \to x_k}(\tilde{x}_k)} \xrightarrow{x_k = 1,2,3} (*,*,*)$$

Chains: Generalizations (More Nodes & Any Dynamics)

Chain (K nodes):



$f_{K+1}(x_K) = (*,*,*)$

1. Initialize via Uniform

$$P(X_k = x_k \mid X_{k-1} = x_{k-1}) = f_k(x_{k-1}, x_k) =$$

$x_{k-1} \setminus x_k$	1	2	3
1	*	*	*
2	*	*	*
3	*	*	*

2. Update initial message: $f_1(x_1) = (*,*,*)$

3. Update marginal xk: $p(x_k) = P(X_k = x_k) = \dfrac{m_{f_k \to x_k}(x_k) \cdot m_{f_{k+1} \to x_k}(x_k)}{\sum_{\tilde{x}_k = 1,2,3} m_{f_k \to x_k}(\tilde{x}_k) \cdot m_{f_{k+1} \to x_k}(\tilde{x}_k)} \xrightarrow{x_k = 1,2,3} (*,*,*)$

Forward message from fk to xk:

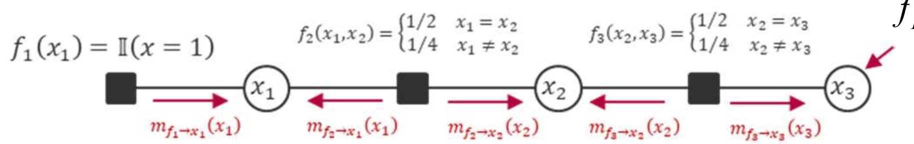
$$m_{f_k \to x_k}(x_k) = \frac{\sum_{x_{k-1}=1,2,3} f_k(x_{k-1}, x_k) \cdot p(x_{k-1}) / m_{f_k \to x_{k-1}}(x_{k-1})}{\sum_{\tilde{x}_k = 1,2,3} \sum_{\tilde{x}_{k-1}=1,2,3} f_k(\tilde{x}_{k-1}, \tilde{x}_k) \cdot p(\tilde{x}_{k-1}) / m_{f_k \to x_{k-1}}(\tilde{x}_{k-1})} \xrightarrow{x_k = 1,2,3} (*,*,*)$$

HPI Hasso Plattner Institut

Chains: Generalizations (More Nodes & Any Dynamics)

$$f_{K+1}(x_K) = (*,*,*)$$

Chain (K nodes):



1. Initialize via Uniform

$$P(X_k = x_k \mid X_{k-1} = x_{k-1}) = f_k(x_{k-1}, x_k) =$$

$x_{k-1} \setminus x_k$	1	2	3
1	*	*	*
2	*	*	*
3	*	*	*

2. Update initial message: $f_1(x_1) = (*,*,*)$

3. Update marginal xk:

$$p(x_k) = P(X_k = x_k) = \frac{m_{f_k \to x_k}(x_k) \cdot m_{f_{k+1} \to x_k}(x_k)}{\sum_{\tilde{x}_k = 1,2,3} m_{f_k \to x_k}(\tilde{x}_k) \cdot m_{f_{k+1} \to x_k}(\tilde{x}_k)} \xrightarrow{x_k = 1,2,3} (*,*,*)$$

Forward message from fk to xk:

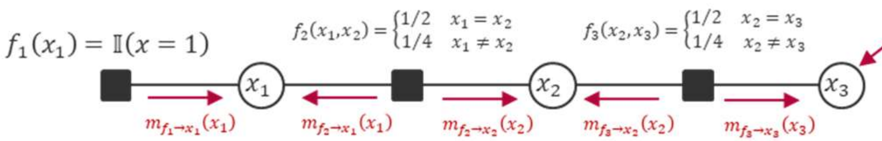
$$m_{f_k \to x_k}(x_k) = \frac{\sum_{x_{k-1}=1,2,3} f_k(x_{k-1}, x_k) \cdot p(x_{k-1}) / m_{f_k \to x_{k-1}}(x_{k-1})}{\sum_{\tilde{x}_k = 1,2,3} \sum_{\tilde{x}_{k-1}=1,2,3} f_k(\tilde{x}_{k-1}, \tilde{x}_k) \cdot p(\tilde{x}_{k-1}) / m_{f_k \to x_{k-1}}(\tilde{x}_{k-1})} \xrightarrow{x_k = 1,2,3} (*,*,*)$$

4. Backward message from fk to xk-1:

$$m_{f_k \to x_{k-1}}(x_{k-1}) = \frac{\sum_{x_k=1,2,3} f_k(x_{k-1}, x_k) \cdot p(x_k) / m_{f_k \to x_k}(x_k)}{\sum_{\tilde{x}_{k-1} = 1,2,3} \sum_{\tilde{x}_k=1,2,3} f_k(\tilde{x}_{k-1}, \tilde{x}_k) \cdot p(\tilde{x}_k) / m_{f_k \to x_k}(\tilde{x}_k)} \xrightarrow{x_{k-1} = 1,2,3} (*,*,*)$$

Chains: Are Longer Chains Tractable? Yes!

Chain (K nodes):



$f_{K+1}(x_K) = (*, *, *)$

$$P(X_k = x_k \mid X_{k-1} = x_{k-1}) = f_k(x_{k-1}, x_k) =$$

$x_{k-1} \setminus x_k$	1	2	3
1	*	*	*
2	*	*	*
3	*	*	*

1. Initialize via Uniform

2. Start again at leaf x1

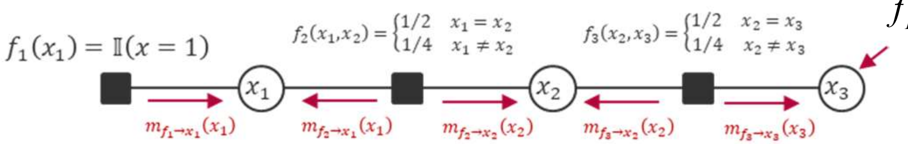
3. Forward Updates: messages & marginals from k=1 to k=K

4. Backward Updates: messages & marginals from k=K to k=1

Do you think you can do that?

Chains: Are Longer Chains Tractable? Yes!

Chain (K nodes):



$f_{K+1}(x_K) = (*,*,*)$

$$P(X_k = x_k \mid X_{k-1} = x_{k-1}) = f_k(x_{k-1}, x_k) = \begin{array}{c|cccc} x_{k-1} \backslash x_k & 1 & 2 & \dots & N \\ 1 & * & * & * & * \\ 2 & * & * & * & * \\ \dots & * & * & * & * \\ N & * & * & * & * \end{array}$$

1. Initialize via Uniform

2. Start again at leaf x1

3. Forward Updates: messages & marginals from k=1 to k=K

4. Backward Updates: messages & marginals from k=K to k=1

Test it: It easily works for K=100.

Note: This has not been possible by using standard summations!

Overview

1. Questions and Updates
2. Recap: Main Concepts of Unit 4
3. Example: Message Passing in Factor Graphs
4. **Hints for Exercise 2 (to be handed in May 19)**

Exercise 2 (until May 19)

- Part I: Sums & Products of Gaussians & Distributions
- Part II: Conditional Independence & D-separation
- Part III: TrueSkill (discrete) via Message Passing

Summary

- Recap I: Conditional Probabilities
- Recap II: Conditional Independence in Bayesian Networks
- Recap III: Messages & Marginals in Factor Graphs (Chain)

See you next Week!