# Exercise Sheet 10

Robin Schmidt
Probabilistic Inference & Learning

January 11, 2019

## Graphical Models

### 1. The Sum-Product Algorithm

We get the following six messages to the root:

$$
\begin{aligned}
\mu_{x_1 \to f_a}(x_1) &= 1 \\
\mu_{f_a \to x_2}(x_2) &= \sum_{x_1} f_a(x_1, x_2) \\
\mu_{x_2 \to f_b}(x_2) &= \mu_{f_a \to x_2}(x_2)\, \mu_{f_c \to x_2}(x_2) \\
\mu_{f_b \to x_3}(x_3) &= \sum_{x_2} f_b(x_2, x_3)\, \mu_{x_2 \to f_b}(x_2) \\
\mu_{x_4 \to f_c}(x_4) &= 1 \\
\mu_{f_c \to x_2}(x_2) &= \sum_{x_4} f_c(x_2, x_4)
\end{aligned}
$$

Aswell as the following six messages to the leaves:

$$
\begin{aligned}
\mu_{x_3 \to f_b}(x_3) &= 1 \\
\mu_{f_b \to x_2}(x_2) &= \sum_{x_3} f_b(x_2, x_3) \\
\mu_{x_2 \to f_a}(x_2) &= \mu_{f_b \to x_2}(x_2)\, \mu_{f_c \to x_2}(x_2) \\
\mu_{f_a \to x_1}(x_1) &= \sum_{x_2} f_a(x_1, x_2)\, \mu_{x_2 \to f_a}(x_2) \\
\mu_{x_2 \to f_c}(x_2) &= \mu_{f_a \to x_2}(x_2)\, \mu_{f_b \to x_2}(x_2) \\
\mu_{f_c \to x_4}(x_4) &= \sum_{x_2} f_c(x_2, x_4)\, \mu_{x_2 \to f_c}(x_2)
\end{aligned}
$$

### (a)

*Proof.* Using the messages stated above, we can confirm that the product on incoming messages gives the correct marginal for $x_2$. We do this by combining all incoming messages to $x_2$ and keeping $p\left(x_1, \ldots, x_4\right) = f_a\left(x_1, x_2\right) \cdot f_b\left(x_2, x_3\right) \cdot f_c\left(x_2, x_4\right)$ in mind like:

$$\begin{aligned}
p(x_2) &= \mu_{f_a \to x_2}\left(x_2\right) \cdot \mu_{f_b \to x_2}\left(x_2\right) \cdot \mu_{f_c \to x_2}\left(x_2\right) \\
&= \sum_{x_1} f_a(x_1, x_2) \cdot \sum_{x_3} f_b\left(x_2, x_3\right) \cdot \sum_{x_4} f_c\left(x_2, x_4\right) \\
&= \sum_{x_1} \sum_{x_3} \sum_{x_4} f_a(x_1, x_2) \cdot f_b\left(x_2, x_3\right) \cdot f_c\left(x_2, x_4\right) \\
&= \sum_{x_1, x_3, x_4} p(x_1, x_2, x_3, x_4)
\end{aligned}$$

□

### (b)

*Proof.* Using an analogous method and combining all incoming messages to $x_3$ and $p\left(x_1, \ldots, x_4\right) = f_a\left(x_1, x_2\right) \cdot f_b\left(x_2, x_3\right) \cdot f_c\left(x_2, x_4\right)$ we get:

$$\begin{aligned}
p(x_3) &= \mu_{f_b \to x_3}\left(x_3\right) \\
&= \sum_{x_2} f_b\left(x_2, x_3\right) \mu_{x_2 \to f_b}\left(x_2\right) \\
&= \sum_{x_2} f_b\left(x_2, x_3\right) \mu_{f_a \to x_2}\left(x_2\right) \mu_{f_c \to x_2}\left(x_2\right) \\
&= \sum_{x_2} f_b\left(x_2, x_3\right) \sum_{x_1} f_a\left(x_1, x_2\right) \sum_{x_4} f_c\left(x_2, x_4\right) \\
&= \sum_{x_1} \sum_{x_2} \sum_{x_4} f_a\left(x_1, x_2\right) f_b\left(x_2, x_3\right) f_c\left(x_2, x_4\right) \\
&= \sum_{x_1, x_2, x_4} p(x_1, x_2, x_3, x_4)
\end{aligned}$$

□

*Proof.* Using an analogous method and combining all incoming messages to $x_1$ and $p(x_1, \ldots, x_4) = f_a(x_1, x_2) \cdot f_b(x_2, x_3) \cdot f_c(x_2, x_4)$ we get:

$$
\begin{aligned}
p(x_1) &= \mu_{f_a \to x_1}(x_1) \\
&= \sum_{x_2} f_a(x_1, x_2)\, \mu_{x_2 \to f_a}(x_2) \\
&= \sum_{x_2} f_a(x_1, x_2)\, \mu_{f_b \to x_2}(x_2)\, \mu_{f_c \to x_2}(x_2) \\
&= \sum_{x_2} f_a(x_1, x_2) \sum_{x_3} f_b(x_2, x_3) \sum_{x_4} f_c(x_2, x_4) \\
&= \sum_{x_2} \sum_{x_3} \sum_{x_4} f_a(x_1, x_2)\, f_b(x_2, x_3)\, f_c(x_2, x_4) \\
&= \sum_{x_2, x_3, x_4} p(x_1, x_2, x_3, x_4)
\end{aligned}
$$

□

### (c)

Following the logic from (b) we can construct the joint as follows:

*Proof.*

$$
\begin{aligned}
p(x_1, x_2) &= \sum_{x_3, x_4} p(x_1, x_2, x_3, x_4) \\
&= \sum_{x_3, x_4} f_a(x_1, x_2) \cdot f_b(x_2, x_3) \cdot f_c(x_2, x_4) \\
&= f_a(x_1, x_2) \sum_{x_3} f_b(x_2, x_3) \sum_{x_4} f_c(x_2, x_4) \\
&= f_a(x_1, x_2)\, \mu_{f_b \to x_2}(x_2)\, \mu_{f_c \to x_2}(x_2) \\
&= f_a(x_1, x_2)\, \mu_{x_2 \to f_a}(x_2)\, \underbrace{\mu_{x_1 \to f_a}(x_1)}_{=1} \\
&= f_a(x_1, x_2) \cdot \prod_{i=1,2} \mu_{x_i \to f_a}(x_i)
\end{aligned}
$$

□

## 2.

If the variables $x_a$ and $x_b$ do not have a common factor, we can evaluate the joint distribution $p(x_a, x_b)$ by introducing a common factor $\delta(x_a - x_b)$ into the graph. It holds that $p(x_a, x_b) \propto p(x_a | x_b)$.

### 3.

For the discrete variables $x$ and $y$ with $x, y \in \{0; 1; 2\}$ we can construct a joint distribution $p(x, y)$ having the property that the value $\hat{x}$ that maximizes the marginal $p(x)$ and $\hat{y}$ that maximizes the marginal $p(y)$ together have probability zero under the joint distribution $p(\hat{x}, \hat{y}) = 0$. In the following example this is true for $\hat{x} = 1$ and $\hat{y} = 1$:

| | | x | | | |
|---|---|---|---|---|---|
| | | **0** | **1** | **2** | **marginal** |
| y | **0** | $\frac{1}{36}$ | $\frac{2}{9}$ | $\frac{1}{36}$ | $\frac{10}{36}$ |
| | **1** | $\frac{2}{9}$ | 0 | $\frac{2}{9}$ | $\frac{4}{9}$ |
| | **2** | $\frac{1}{36}$ | $\frac{2}{9}$ | $\frac{1}{36}$ | $\frac{10}{36}$ |
| | **marginal** | $\frac{10}{36}$ | $\frac{4}{9}$ | $\frac{10}{36}$ | |

If $x, y$ are binary values such a situation is not possible, due to the following problem. We try to construct an example, where $\hat{x} = 1$ and $\hat{y} = 1$ maximize the marginal and are zero in joint. In this example we can see, that $z$ can't be equal to 0 because then the marginal of $\hat{x}$

| | | x | | |
|---|---|---|---|---|
| | | **0** | **1** | **marginal** |
| y | **0** | z | $\frac{1}{2} - \frac{z}{2}$ | $z + \frac{1}{2} - \frac{z}{2}$ |
| | **1** | $\frac{1}{2} - \frac{z}{2}$ | 0 | $\frac{1}{2} - \frac{z}{2}$ |
| | **marginal** | $z + \frac{1}{2} - \frac{z}{2}$ | $\frac{1}{2} - \frac{z}{2}$ | |

and $\hat{y}$ wouldn't maximize and would just be of equal value. Therefore we set $z$ to a really small value greater than 0. This causes the other marginals to get a higher value than the marginal of $\hat{x}$ or $\hat{y}$. This can't be changed, due to the zero at $\hat{x}, \hat{y}$ and therefore causes these marginals to always be smaller by the amount of $z$.