

# Bayesian Networks and Belief Propagation

Mohammad Emtiyaz Khan  
EPFL

Nov 26, 2015



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

©Mohammad Emtiyaz Khan 2015

# Motivation

When the data is *structured*, e.g. time-series data, prediction can be improved by incorporate the structure into the model. This can be accomplished by using ideas from both probability theory and graph theory. The resulting field is called [probabilistic graphical model](#).

We will lean about Bayesian network where the distribution *respects* a directed graphical structure. Our main goal is to learn *inference* of the latent variables using [belief propagation](#).

## Bayesian network

Given a [directed graph](#)  $G$  and parameters  $\theta$ , a Bayesian network defines the joint distribution as follows,

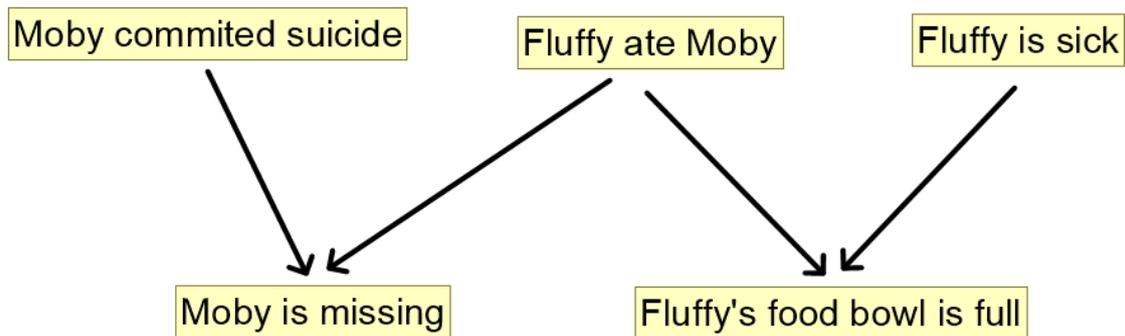
$$p_{\theta}(\mathbf{x}) = \prod_{k=1}^K p_{\theta}(x_k | \text{pa}_k),$$

where  $\text{pa}_k$  are the parents of  $x_k$ .

## A simple example

Moby is the pet fish and Fluffy is the pet cat. One day, you observed that Moby is missing. At the same time, you notice that Fluffy's food bowl is full. You wonder what happened.

There are three possibilities: (1) Moby committed suicide (2) Fluffy ate Moby (3) Fluffy is sick.



Given  $y_a = 1$ ,  $y_b = 0$ , and  $\theta$ , we wish to find:

1. Probability that “Moby committed suicide” i.e.  $p(z_1|y_a, y_b)$ .
2. Probability that “Fluffy ate Moby” i.e.  $p(z_2|y_a, y_b)$ .
3. Probability that “Fluffy is sick” i.e.  $p(z_3|y_a, y_b)$ .

There are the [marginal posterior probability](#).

# Posterior computation

Define  $\mathbf{y} := \{y_a, y_b\}$ . Using Bayes' rule, we can write the following:

$$p(\mathbf{y}, z_1, z_2, z_3) = p(\mathbf{y}|z_1, z_2, z_3)p(z_1, z_2, z_3) = p(z_1, z_2, z_3|\mathbf{y})p(\mathbf{y})$$

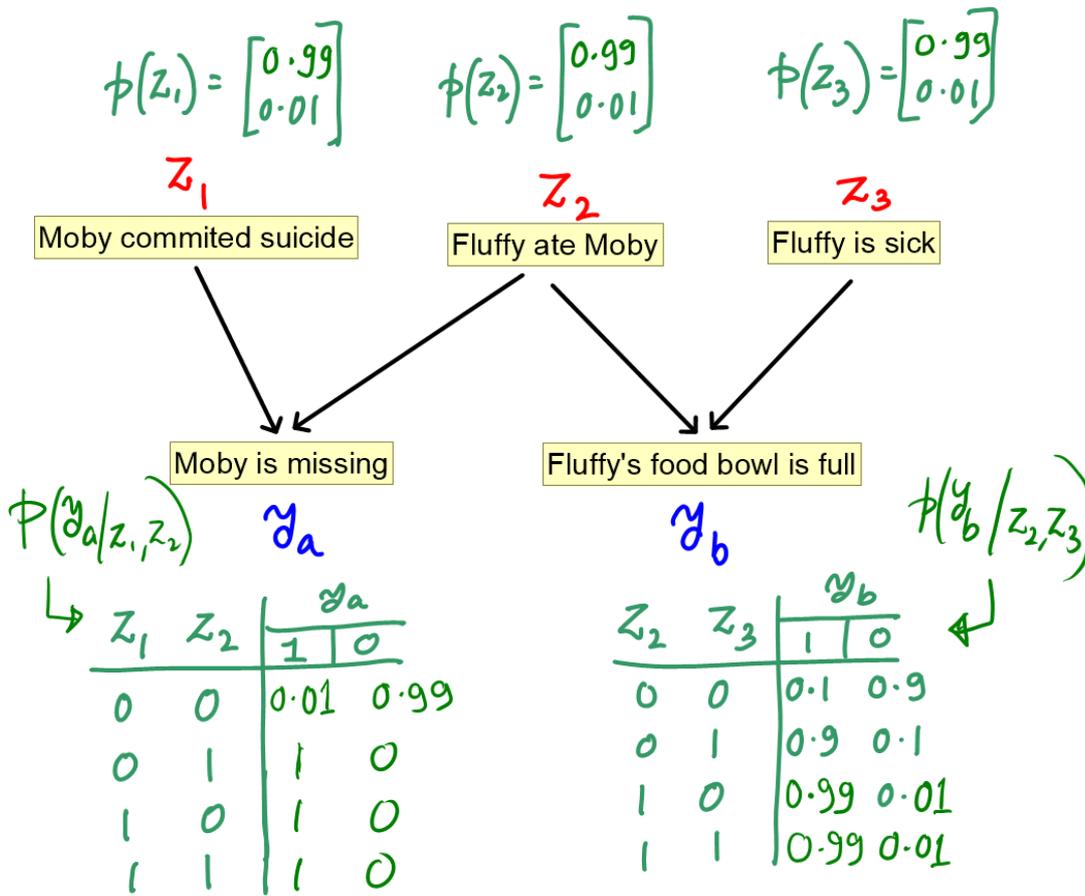
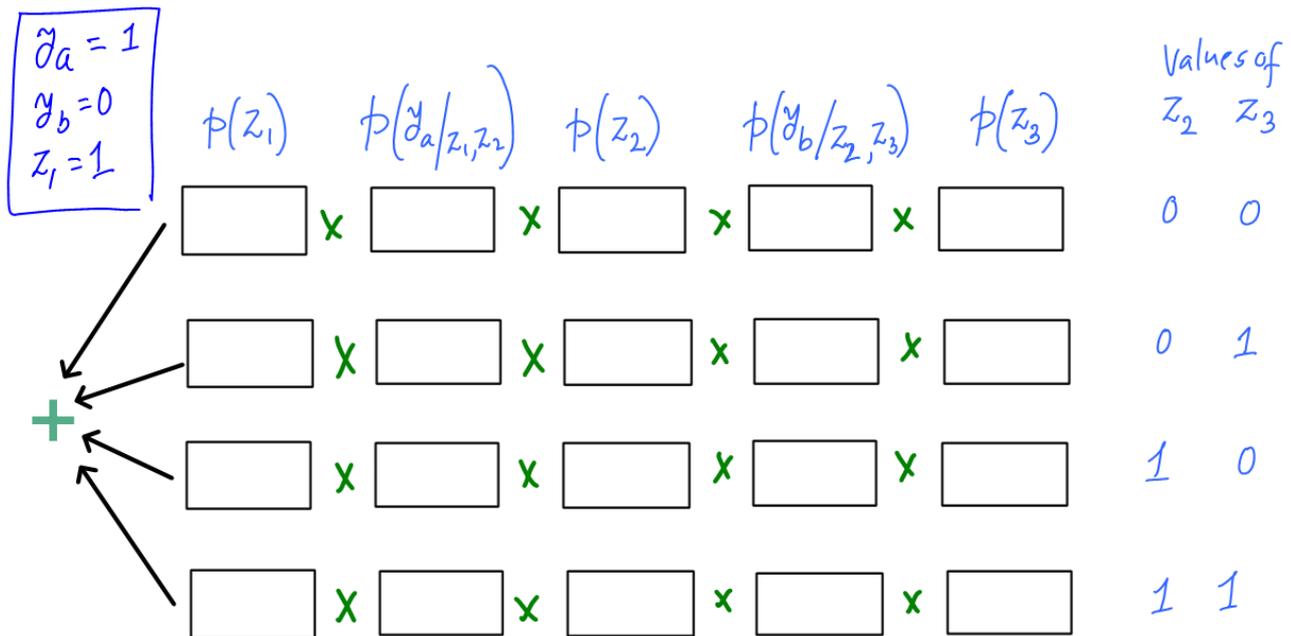
To compute  $p(z_1|y_a, y_b, y_c)$ , we can marginalize  $z_2$  and  $z_3$ .

$$\begin{aligned} p(z_1|\mathbf{y}) &= \sum_{z_3=\{0,1\}} \sum_{z_2=\{0,1\}} p(z_1, z_2, z_3|\mathbf{y}) \\ &= \sum_{z_3=\{0,1\}} \sum_{z_2=\{0,1\}} \frac{p(\mathbf{y}|z_1, z_2, z_3)p(z_1, z_2, z_3)}{p(\mathbf{y})} \\ &\propto \sum_{z_3=\{0,1\}} \sum_{z_2=\{0,1\}} p(\mathbf{y}|z_1, z_2, z_3)p(z_1, z_2, z_3) \end{aligned}$$

Expanding we get the following:

$$p(z_1|\mathbf{y}) \propto \sum_{z_3} \sum_{z_2} p(y_a|z_1, z_2)p(y_b|z_2, z_3)p(z_1)p(z_2)p(z_3)$$

Flow chart of computation  $p(z_1 = 1 | y_a = 1, y_b = 0)$



What is the computational complexity for  $D$  unknowns?

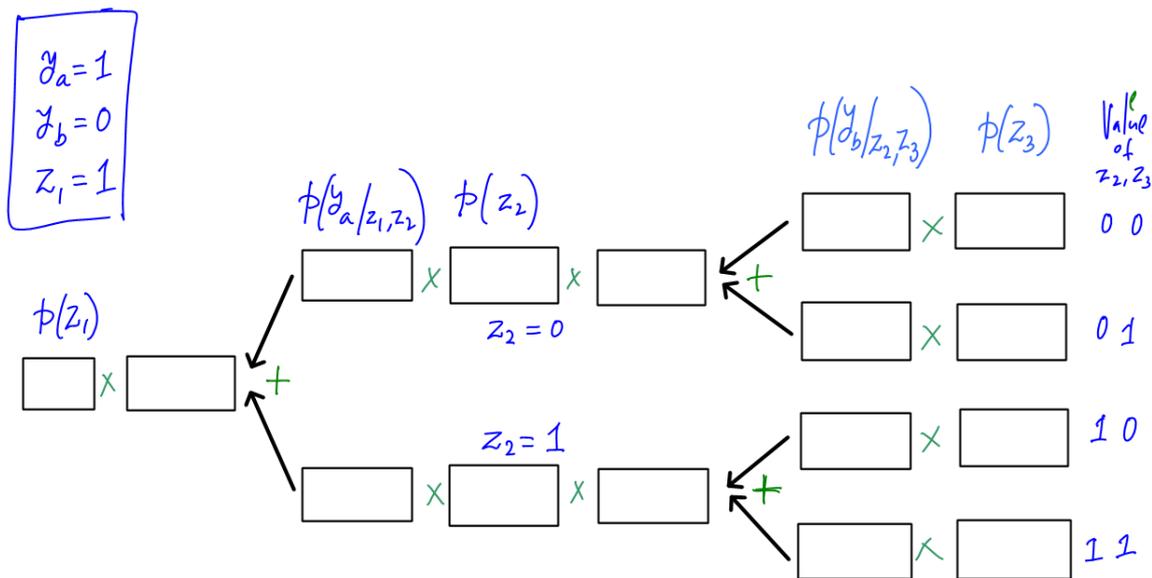
# Sum-product algorithm

We can reduce the complexity drastically by using the structure of the problem. We will use the [sum-product algorithm](#).

$$\sum_{z_3} \sum_{z_2} p(y_a|z_1, z_2)p(y_b|z_2, z_3)p(z_1)p(z_2)p(z_3)$$

$$= p(z_1) \left\{ \sum_{z_2} p(y_a|z_1, z_2)p(z_2) \left[ \sum_{z_3} p(y_b|z_2, z_3)p(z_3) \right] \right\}$$

Flow chart of computation for  $p(z_1 = 1|y_a = 1, y_b = 0)$



We can repeat the same procedure for  $z_2$  and  $z_3$ .

$$p(z_1|\mathbf{y}) \propto p(z_1) \left\{ \sum_{z_2} p(y_a|z_1, z_2)p(z_2) \left[ \sum_{z_3} p(y_b|z_2, z_3)p(z_3) \right] \right\}$$

$$p(z_2|\mathbf{y}) \propto p(z_2) \left\{ \sum_{z_1} p(y_a|z_1, z_2)p(z_1) \left[ \sum_{z_3} p(y_b|z_2, z_3)p(z_3) \right] \right\}$$

$$p(z_3|\mathbf{y}) \propto p(z_3) \left\{ \sum_{z_2} p(y_b|z_2, z_3)p(z_2) \left[ \sum_{z_1} p(y_a|z_1, z_2)p(z_1) \right] \right\}$$

Notice that there are many redundant computations. We can reduce computation by reusing the redundant ones.

$$p(z_1|\mathbf{y}) \propto p(z_1) \underbrace{\sum_{z_2} p(y_a|z_1, z_2) p(z_2)}_{m_{a \rightarrow 1}(z_1)} \underbrace{\sum_{z_3} p(y_b|z_2, z_3) \underbrace{p(z_3)}_{m_{3 \rightarrow b}(z_3)}}_{m_{b \rightarrow 2}(z_2)} \underbrace{\quad}_{m_{2 \rightarrow a}(z_2)}$$

## Messages

Suppose, we have many latent variables  $z_i$  and many observations  $y_a$ . Also, assume that the graph between the observations and the variables is [bi-partite](#). In this graph, neighbours of observations are variables, and vice-versa. Define  $\mathbb{N}(i)$  to be the set of all neighbourhood of  $z_i$  and  $\mathbb{N}(a)$  to be the set of all neighbours of observation  $y_a$ . Also, define the set  $\mathbb{N}(i) \setminus a$  to be the set of “all neighbours of  $i$  except  $a$ ”.

Define the messages from variables to observations as shown below:

$$m_{i \rightarrow a}(z_i) = p(z_i) \prod_{b \in \mathbb{N}(i) \setminus a} m_{b \rightarrow i}(z_i).$$

Therefore, in our example, messages from variables to observations are,

$$\begin{aligned} m_{1 \rightarrow a}(z_1) &= p(z_1), \\ m_{2 \rightarrow a}(z_2) &= p(z_2)m_{b \rightarrow 2}(z_2), \\ m_{2 \rightarrow b}(z_2) &= p(z_2)m_{a \rightarrow 2}(z_2), \\ m_{3 \rightarrow b}(z_3) &= p(z_3). \end{aligned}$$

Similarly, the messages from observations to variables are shown below:

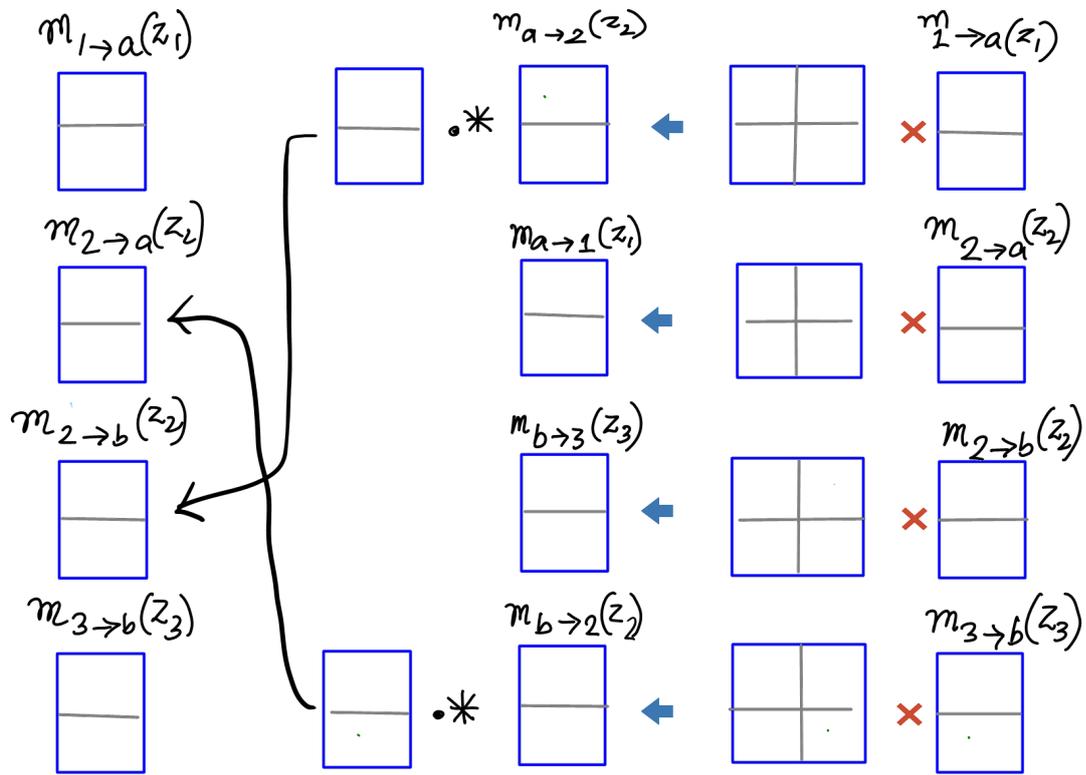
$$m_{a \rightarrow i}(z_i) = \sum_j p(y_a | \text{Pa}_a) \prod_{j \in \mathbb{N}(a) \setminus i} m_{j \rightarrow a}(z_j).$$

In our example, these can be written as shown below:

$$\begin{aligned} m_{a \rightarrow 1}(z_1) &= \sum_{z_2} p(y_a | z_1, z_2) m_{2 \rightarrow a}(z_2), \\ m_{a \rightarrow 2}(z_2) &= \sum_{z_1} p(y_a | z_1, z_2) m_{1 \rightarrow a}(z_1), \\ m_{b \rightarrow 2}(z_2) &= \sum_{z_3} p(y_b | z_2, z_3) m_{3 \rightarrow b}(z_3), \\ m_{b \rightarrow 3}(z_3) &= \sum_{z_2} p(y_b | z_2, z_3) m_{2 \rightarrow b}(z_2). \end{aligned}$$

The first set of messages perform a product operation, while the second set performs the sum.

Calculation of these messages for our example are visualized in this figure.



# Belief Propagation

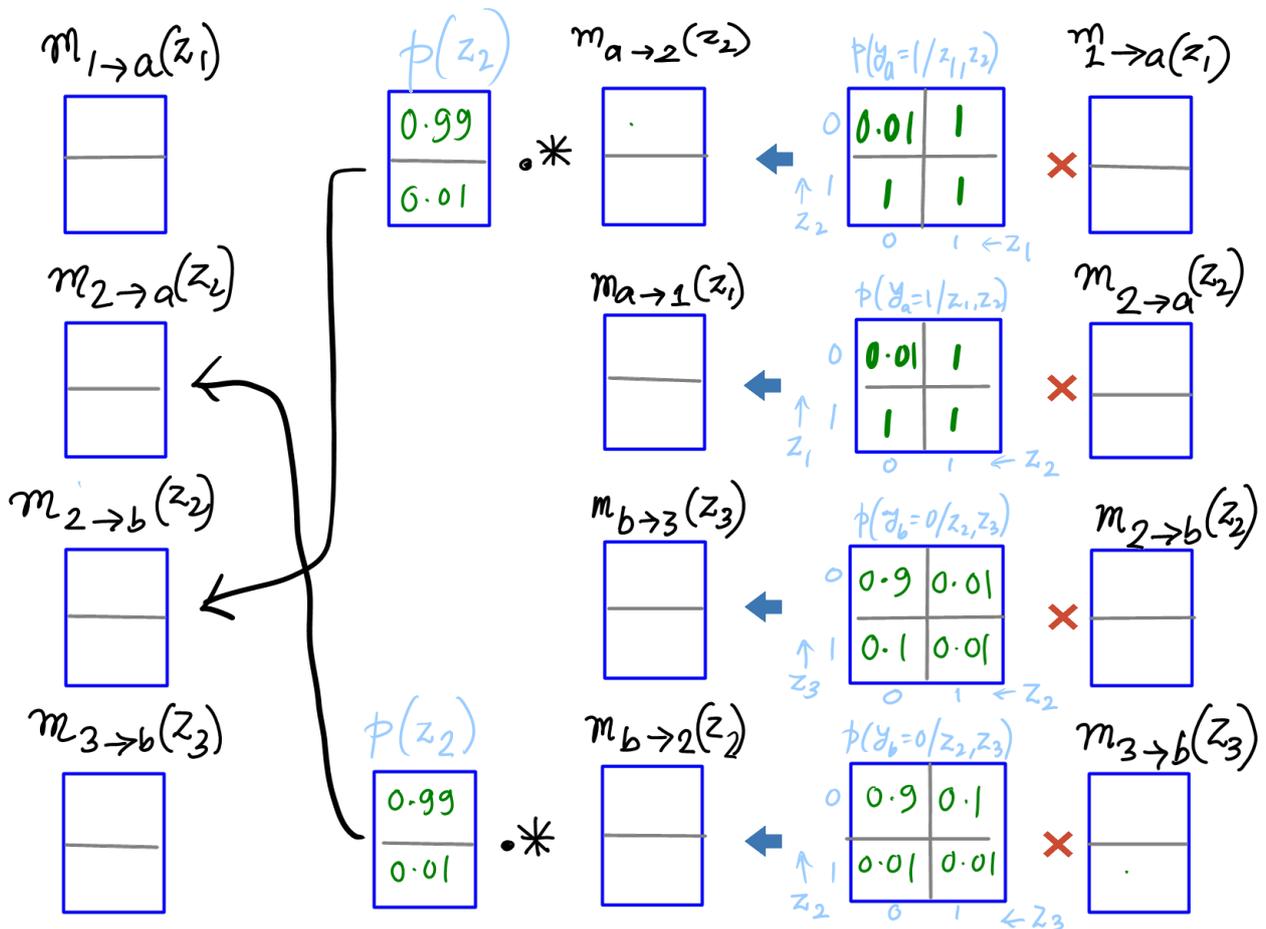
Here is a Belief propagation algorithm for our example.

1. Initialize messages of variables  $i$  to  $p(z_i)$  and iterate until messages do not change:
  - (a) Send messages from variables to observations.
  - (b) Send messages from observations to variables.
2. Compute marginals by multiplying all the message received at node  $i$ .

$$p(z_i|\mathbf{y}) = p(z_j) \prod_{j \in \mathbb{N}(a)} m_{j \rightarrow a}(z_j)$$

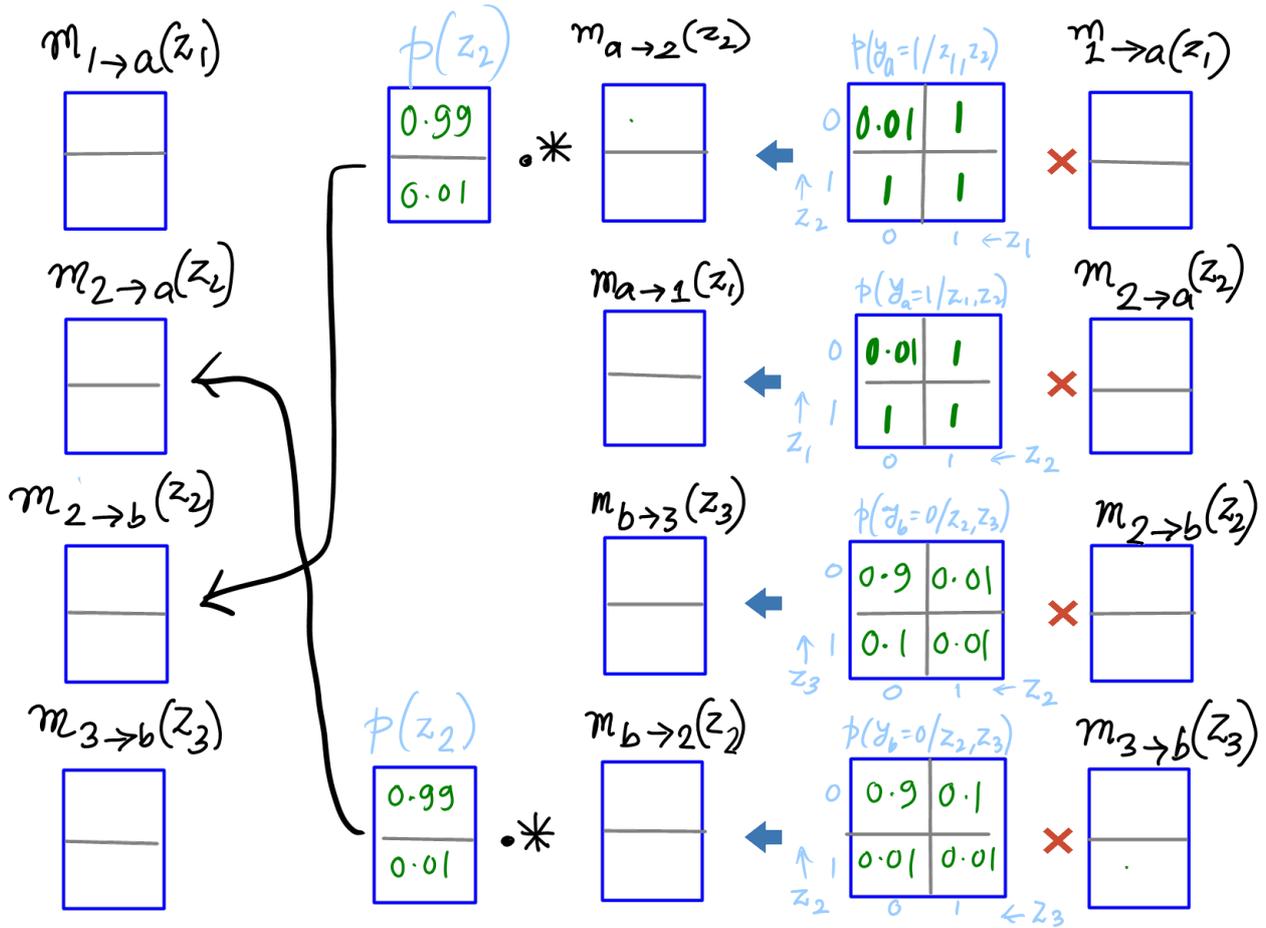
Why will this work? See the path  $1 \leftarrow a \leftarrow 2 \leftarrow b \leftarrow 3$ .

$$p(z_1|\mathbf{y}) \propto p(z_1) \underbrace{\sum_{z_2} p(y_a|z_1, z_2) p(z_2)}_{m_{a \rightarrow 1}(z_1)} \underbrace{\sum_{z_3} p(y_b|z_2, z_3) \underbrace{p(z_3)}_{m_{3 \rightarrow b}(z_3)}}_{m_{b \rightarrow 2}(z_2)} \underbrace{p(z_3)}_{m_{3 \rightarrow b}(z_3)}$$

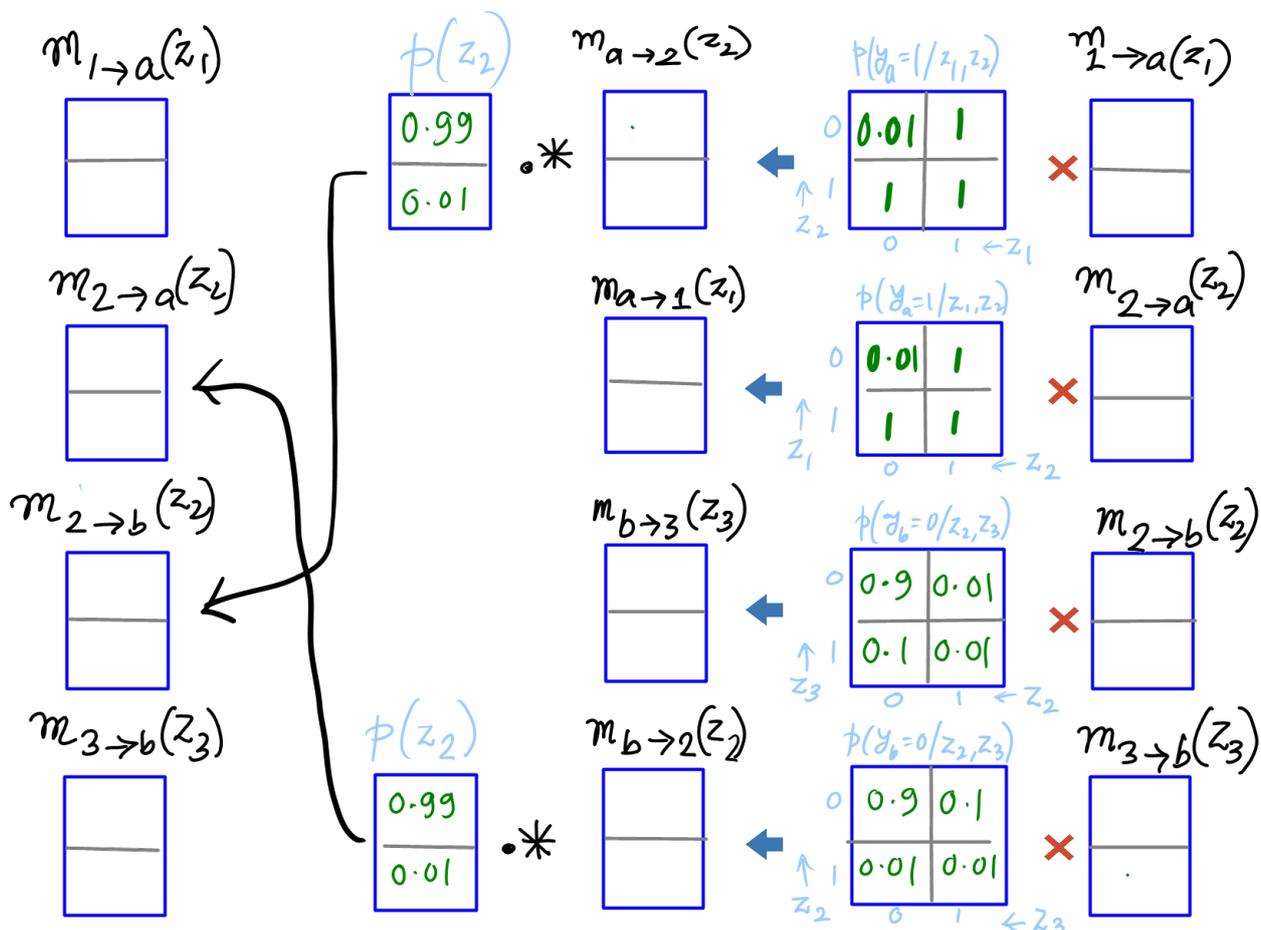


See the path  $1 \rightarrow a \rightarrow 2 \leftarrow b \leftarrow 3$ .

$$p(z_2|\mathbf{y}) \propto p(z_2) \underbrace{\sum_{z_1} p(y_a|z_1, z_2) p(z_1)}_{m_{a \rightarrow 2}(z_2)} \underbrace{\sum_{z_3} p(y_b|z_2, z_3) p(z_3)}_{m_{b \rightarrow 2}(z_2)}$$



Exercise: Work out the example to compute  $p(z_3)$ .



## Discussion

Probabilistic graphical models also contain other models such as undirected graphical models and factor graphs. Belief propagation can be applied to these models as well. The main advantage of this method is that it can be implemented using distributed computation and is very efficient for large scale models.