

# 50.007

# Machine Learning

Lu, Wei



# Bayesian Networks (I)

# Soft EM for HMM

## E-Step

Run forward-backward algorithm  
to collect fractional counts  
from each instance

## M-Step

$$a_{u,v} = \frac{\text{count}(u,v)}{\text{count}(u)} \qquad b_u(o) = \frac{\text{count}(u \rightarrow o)}{\text{count}(u)}$$

# Soft EM for HMM

Finding the fractional count

$$\begin{aligned}\text{count}(u, v) &= \sum_{i=1}^m \text{count}^{(i)}(u, v) \\ &= \sum_{i=1}^m \sum_{\mathbf{y}} p(\mathbf{y} | \mathbf{x}^{(i)}) \text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \rightarrow v) \\ &= \sum_{i=1}^m \sum_{j=0}^n p(y_j = u, y_{j+1} = v | \mathbf{x})\end{aligned}$$

$$\begin{aligned}\text{count}(u) &= \sum_{i=1}^m \text{count}^{(i)}(u) \\ &= \sum_{i=1}^m \sum_{j=0}^n p(y_j = u | \mathbf{x})\end{aligned}$$

# Soft EM for HMM

Finding the fractional count

$$\begin{aligned}
 \text{count}(u, v) &= \sum_{i=1}^m \text{count}^{(i)}(u, v) \\
 &= \sum_{i=1}^m \sum_{\mathbf{y}} p(\mathbf{y} | \mathbf{x}^{(i)}) \text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \rightarrow v) \\
 &= \sum_{i=1}^m \sum_{j=0}^n \frac{\alpha_u(j) \cdot b_u(x_j) \cdot a_{u,v} \cdot \beta_v(j+1)}{\sum_v \alpha_v(k) \beta_v(k)} p(y_j = u, y_{j+1} = v | \mathbf{x})
 \end{aligned}$$

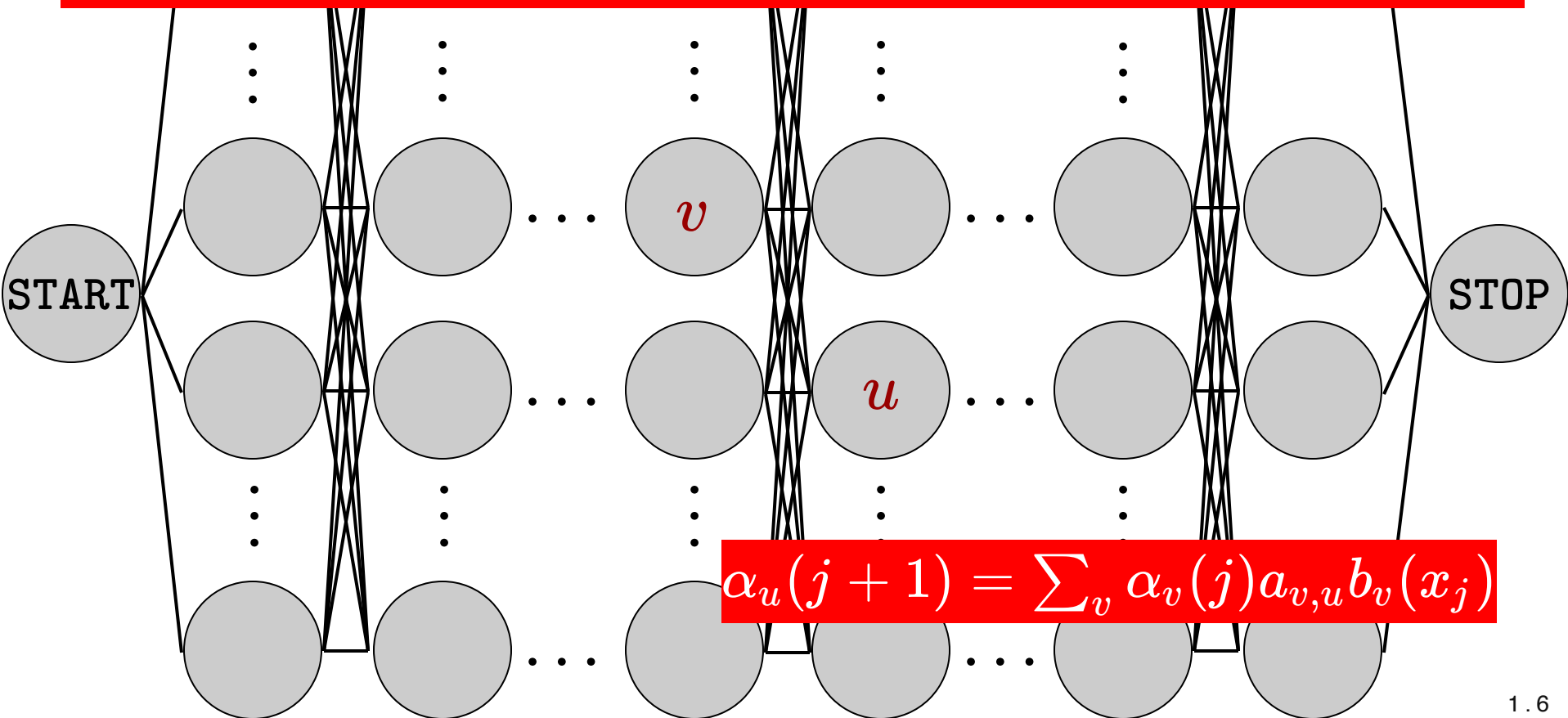
$$\begin{aligned}
 \text{count}(u) &= \sum_{i=1}^m \text{count}^{(i)}(u) \\
 &= \sum_{i=1}^m \sum_{j=0}^n \frac{\alpha_u(j) \beta_u(j)}{\sum_v \alpha_v(k) \beta_v(k)} p(y_j = u | \mathbf{x})
 \end{aligned}$$

# Forward-Backward Algorithm

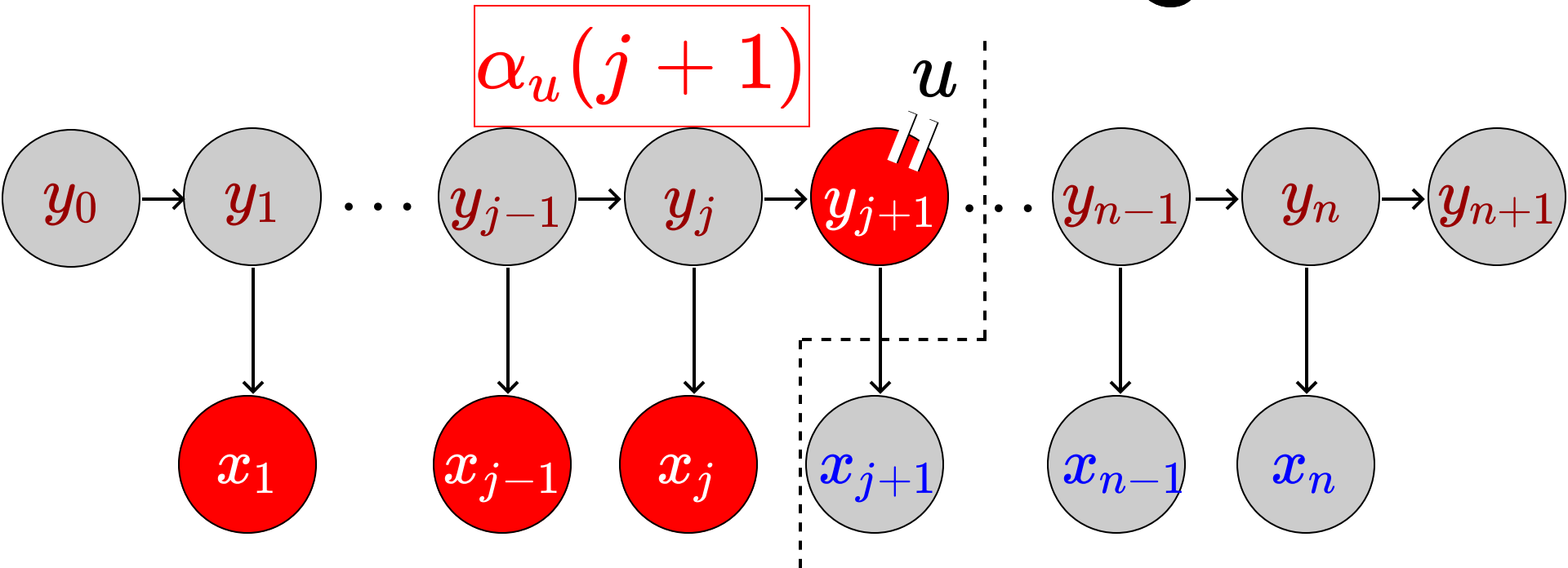
0      1      2       $j$        $j+1$        $n-1$        $n$        $n+1$

$$\alpha_u(j) = p(x_1, \dots, x_{j-1}, y_j = u)$$

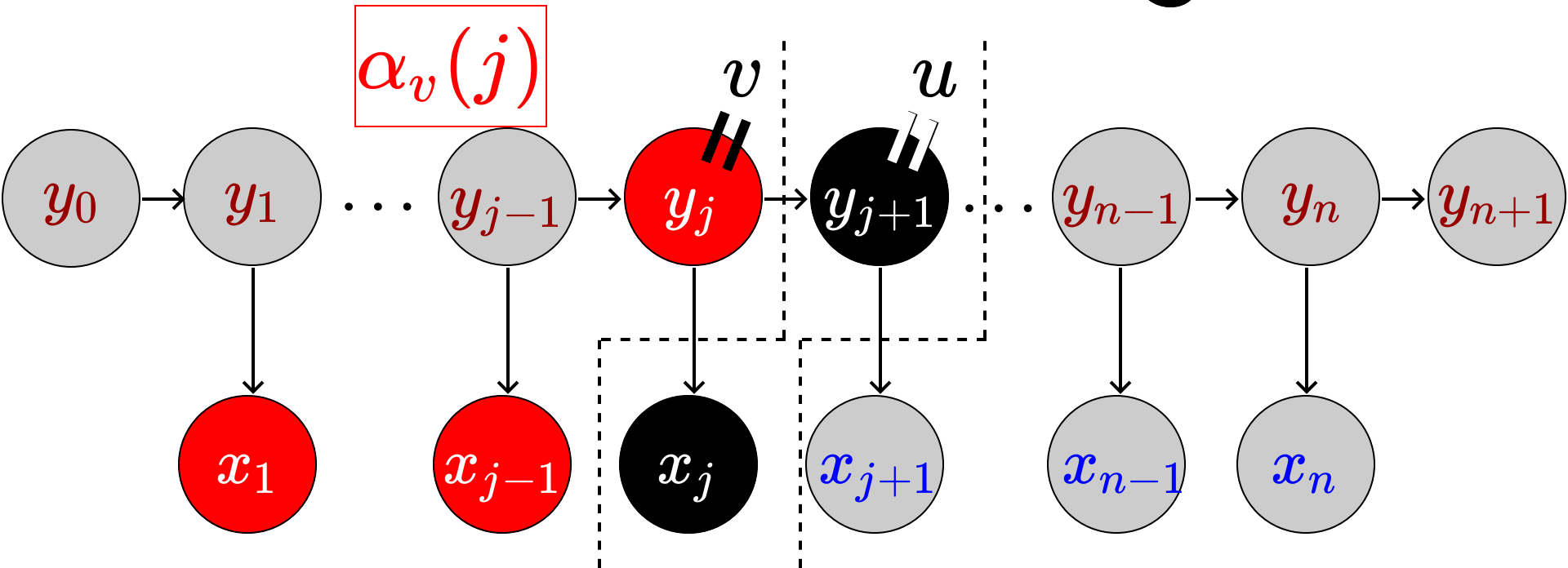
The sum of the scores of all paths from START to node  $u$  at  $j$



# Forward-Backward Algorithm



# Forward-Backward Algorithm



$$\alpha_u(j+1) = \sum_v \alpha_v(j) a_{v,u} b_v(x_j)$$

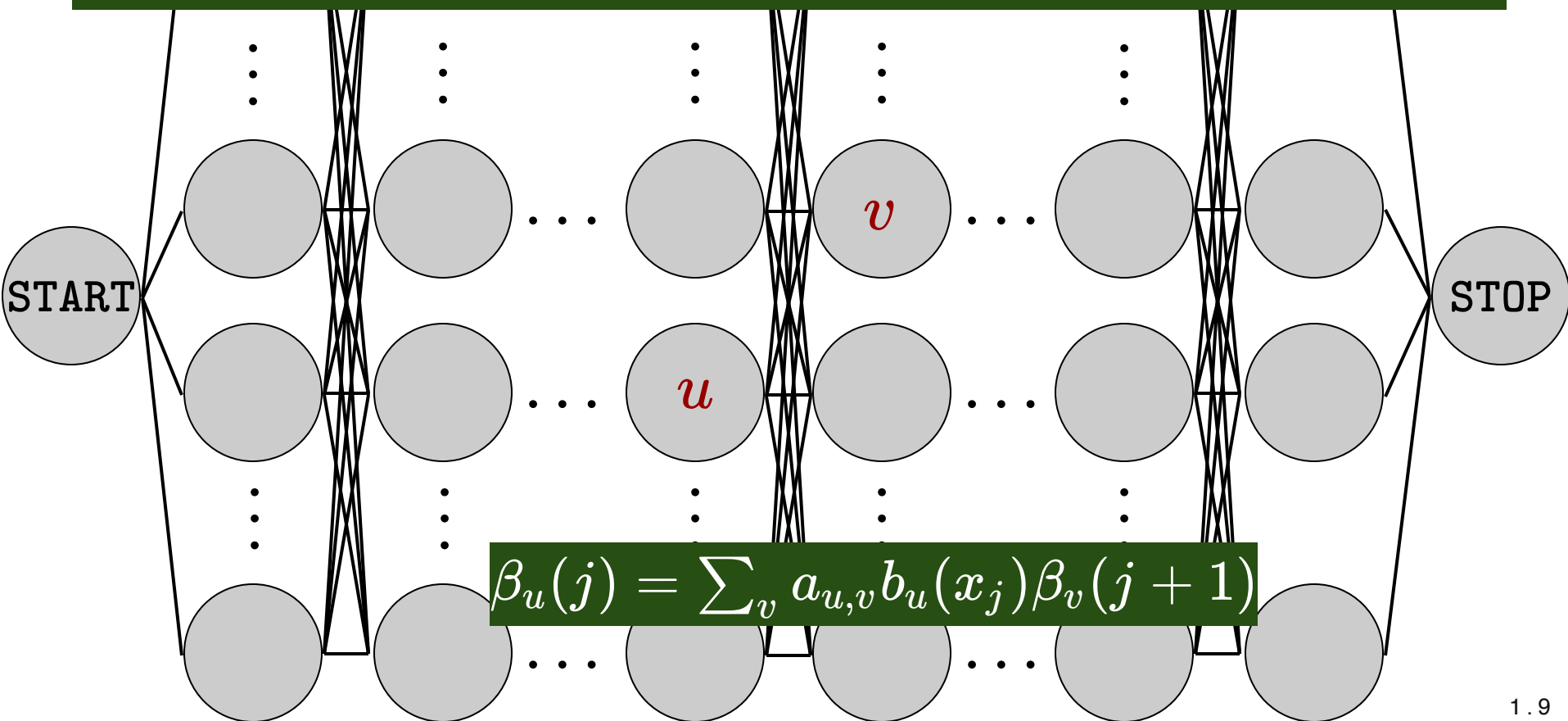


# Forward-Backward Algorithm

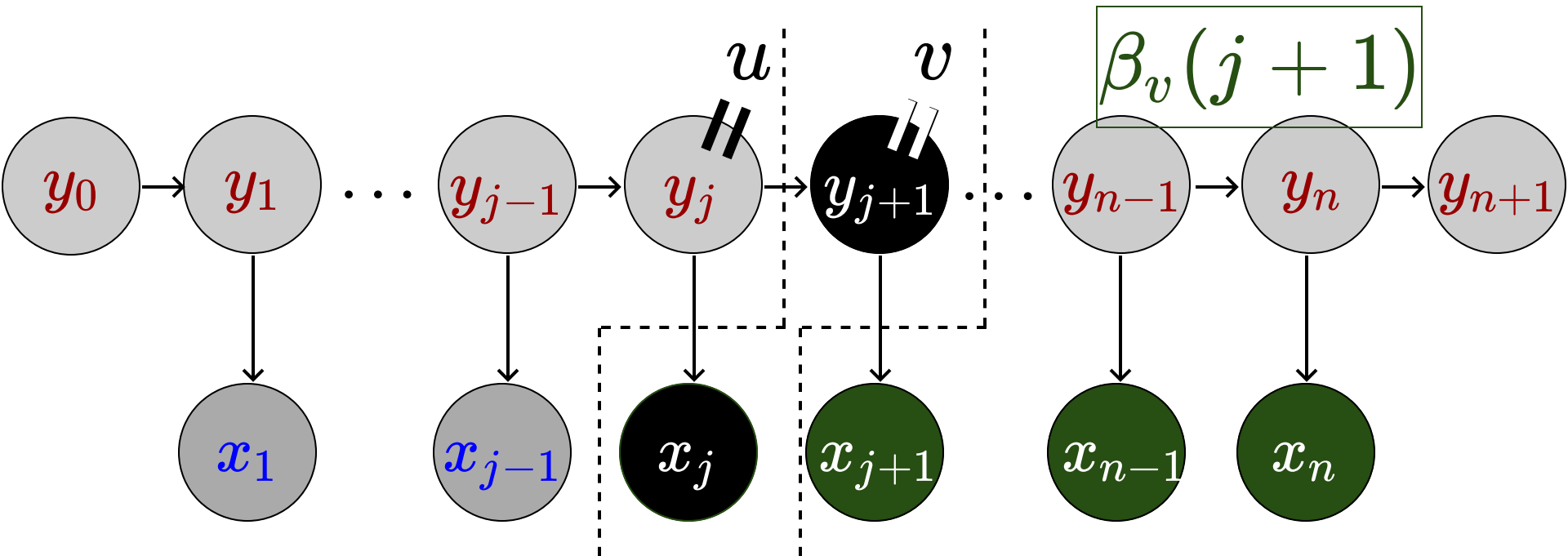
0      1      2       $j$        $j+1$        $n-1$        $n$        $n+1$

$$\beta_u(j) = p(x_j, \dots, x_n | y_j = u)$$

The sum of the scores of all paths from node  $u$  at  $j$  to STOP



# Forward-Backward Algorithm



$$\beta_u(j) = \sum_v a_{u,v} b_u(x_j) \beta_v(j+1)$$

# Soft EM for HMM

Finding the fractional count

$$\begin{aligned}\text{count}(u, v) &= \sum_{i=1}^m \text{count}^{(i)}(u, v) \\ &= \sum_{i=1}^m \sum_{\mathbf{y}} p(\mathbf{y} | \mathbf{x}^{(i)}) \text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \rightarrow v) \\ &= \sum_{i=1}^m \sum_{j=0}^n \frac{\alpha_u(j) \cdot b_u(x_j) \cdot a_{u,v} \cdot \beta_v(j+1)}{\sum_v \alpha_v(k) \beta_v(k)}\end{aligned}$$

$$\begin{aligned}\text{count}(u) &= \sum_{i=1}^m \text{count}^{(i)}(u) \\ &= \sum_{i=1}^m \sum_{j=0}^n \frac{\alpha_u(j) \beta_u(j)}{\sum_v \alpha_v(k) \beta_v(k)}\end{aligned}$$

In the M-Step:

$$a_{u,v} = \frac{\text{count}(u, v)}{\text{count}(u)}$$

# Soft EM for HMM

Finding the fractional count

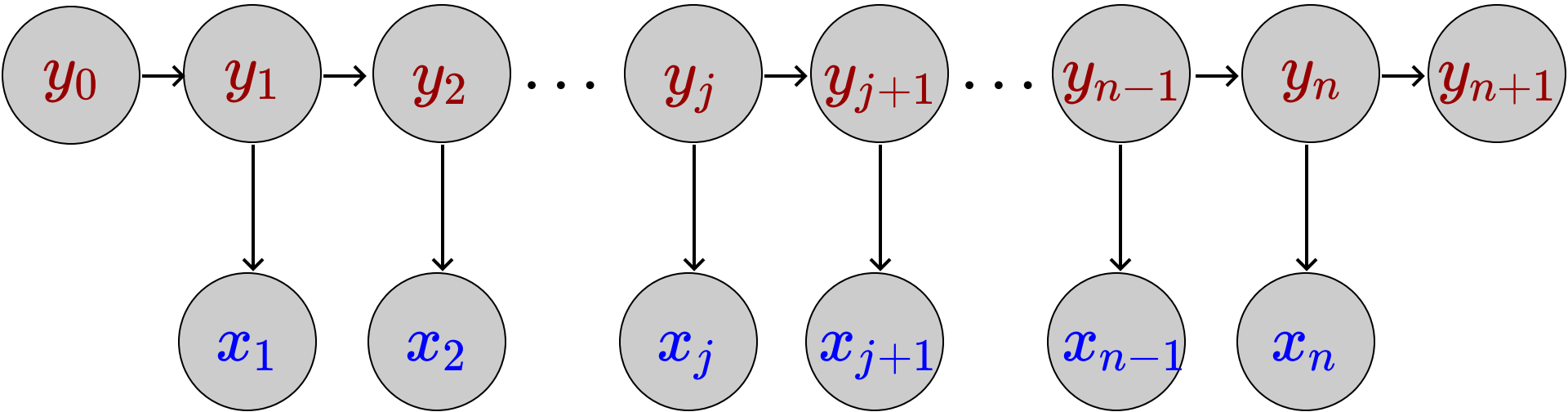
$$\begin{aligned}\text{count}(u \rightarrow o) &= \sum_{i=1}^m \text{count}^{(i)}(u \rightarrow o) \\ &= \sum_{i=1}^m \sum_{\mathbf{y}} p(\mathbf{y} | \mathbf{x}^{(i)}) \text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \rightarrow o) \\ &= \sum_{i=1}^m \sum_{j \text{ s.t. } x_j=o} \frac{\alpha_u(j) \beta_u(j)}{\sum_v \alpha_v(k) \beta_v(k)}\end{aligned}$$

$$\begin{aligned}\text{count}(u) &= \sum_{i=1}^m \text{count}^{(i)}(u) \\ &= \sum_{i=1}^m \sum_{j=0}^n \frac{\alpha_u(j) \beta_u(j)}{\sum_v \alpha_v(k) \beta_v(k)}\end{aligned}$$

In the M-Step:

$$b_u(o) = \frac{\text{count}(u \rightarrow o)}{\text{count}(u)}$$

# Hidden Markov Model



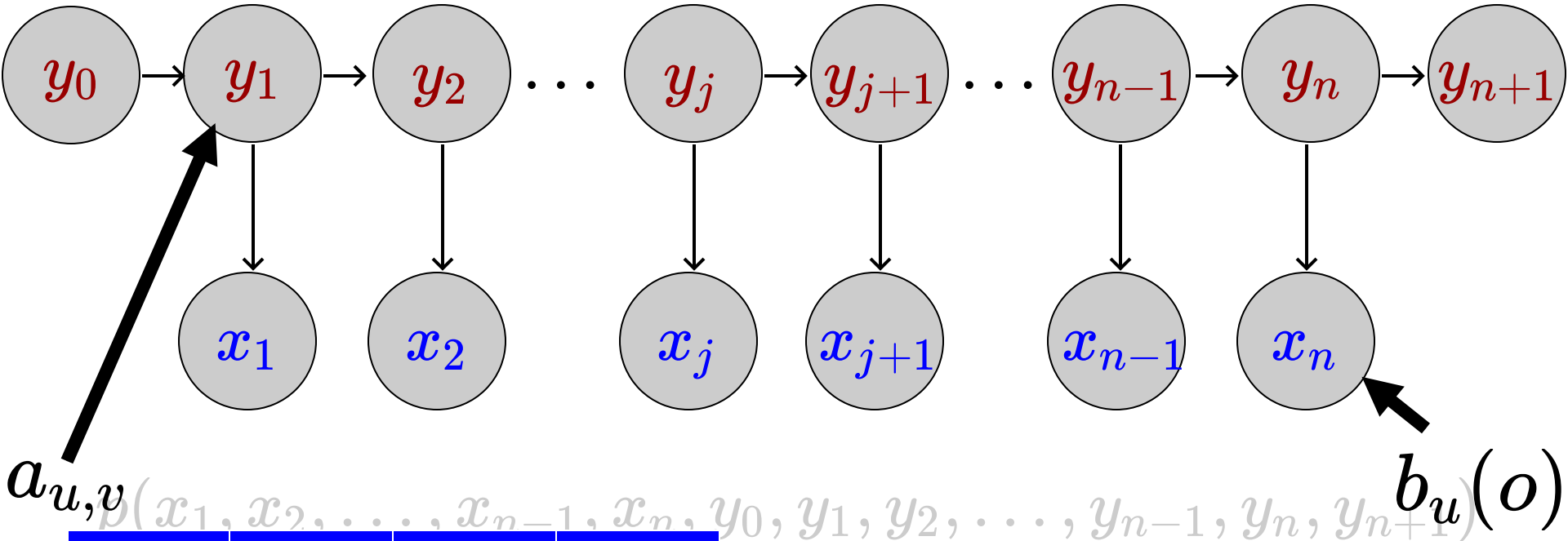
$$p(x_1, x_2, \dots, x_{n-1}, x_n, y_0, y_1, y_2, \dots, y_{n-1}, y_n, y_{n+1})$$

$$\underbrace{\prod_{j=0}^n a_{y_j, y_{j+1}}}_{\text{Transition probabilities}} \times \underbrace{\prod_{j=1}^n b_{y_j}(x_j)}_{\text{Emission probabilities}}$$

Transition probabilities

Emission probabilities

# Hidden Markov Model



| $u \backslash v$ | A   | B   | STOP |
|------------------|-----|-----|------|
| START            | 1.0 | 0.0 | 0.0  |
| A                | 0.5 | 0.5 | 0.0  |
| B                | 0.0 | 0.8 | 0.2  |

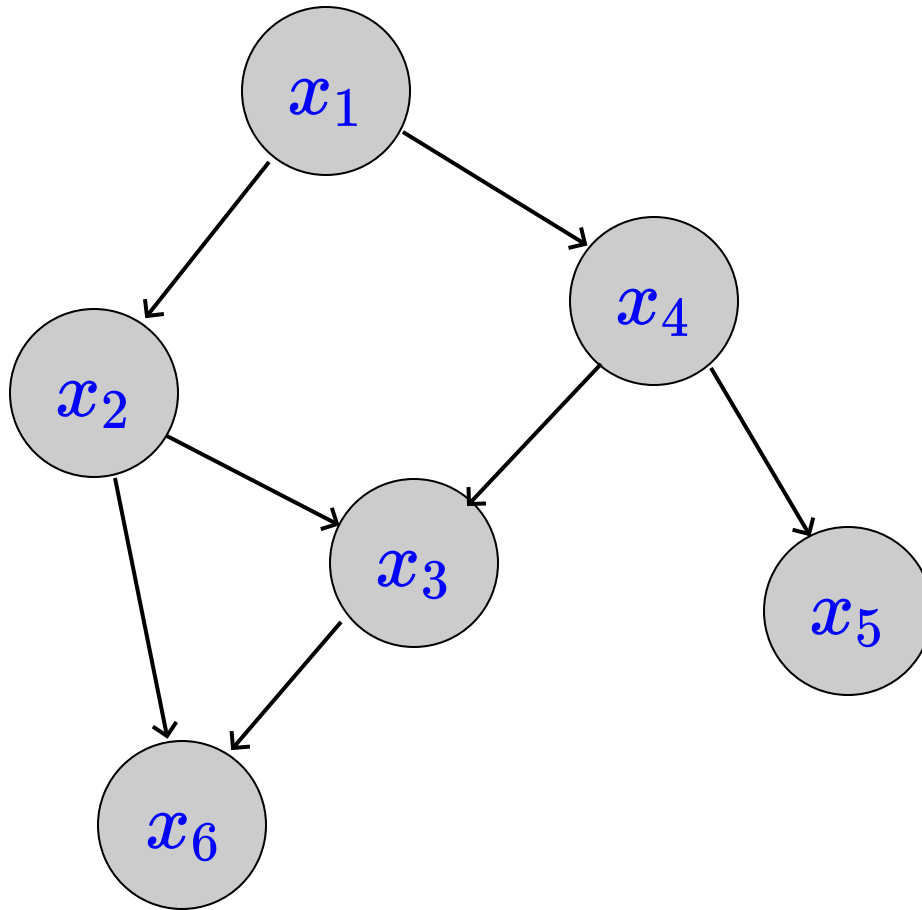
Transition probabilities

×

| $u \backslash o$ | “the” | “dog” |
|------------------|-------|-------|
| A                | 0.9   | 0.1   |
| B                | 0.1   | 0.9   |

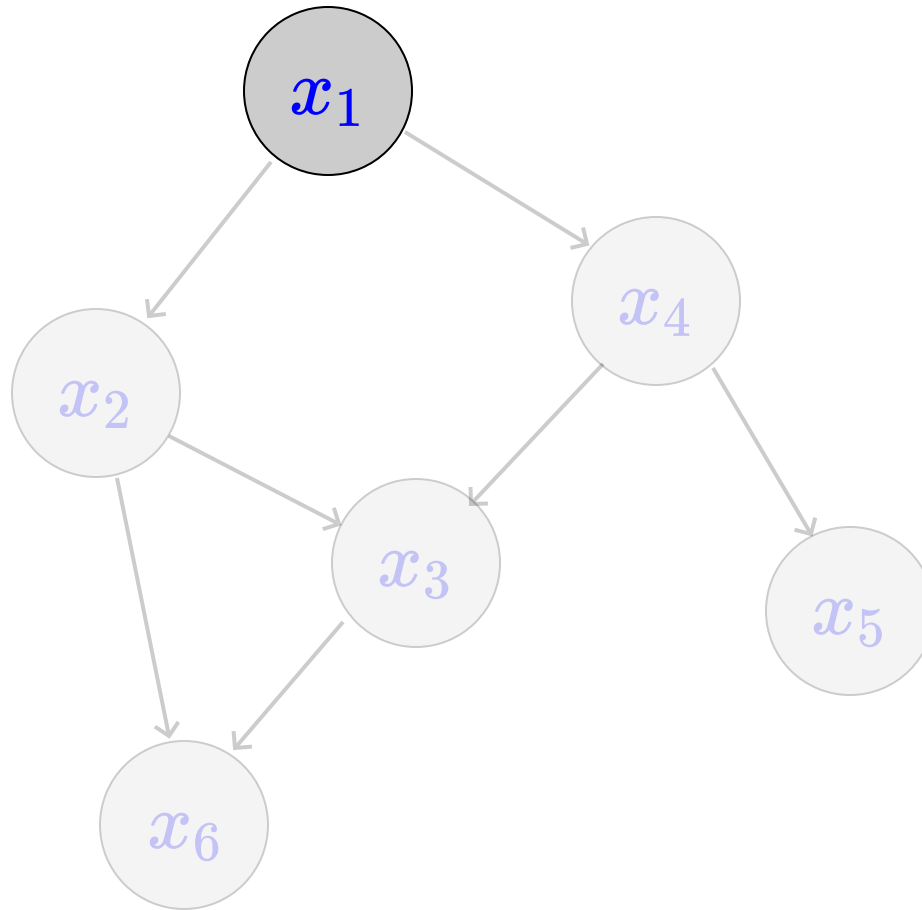
Emission probabilities

# Generative Model



$$p(x_1, x_2, x_3, x_4, x_5, x_6) = ?$$

# Generative Model

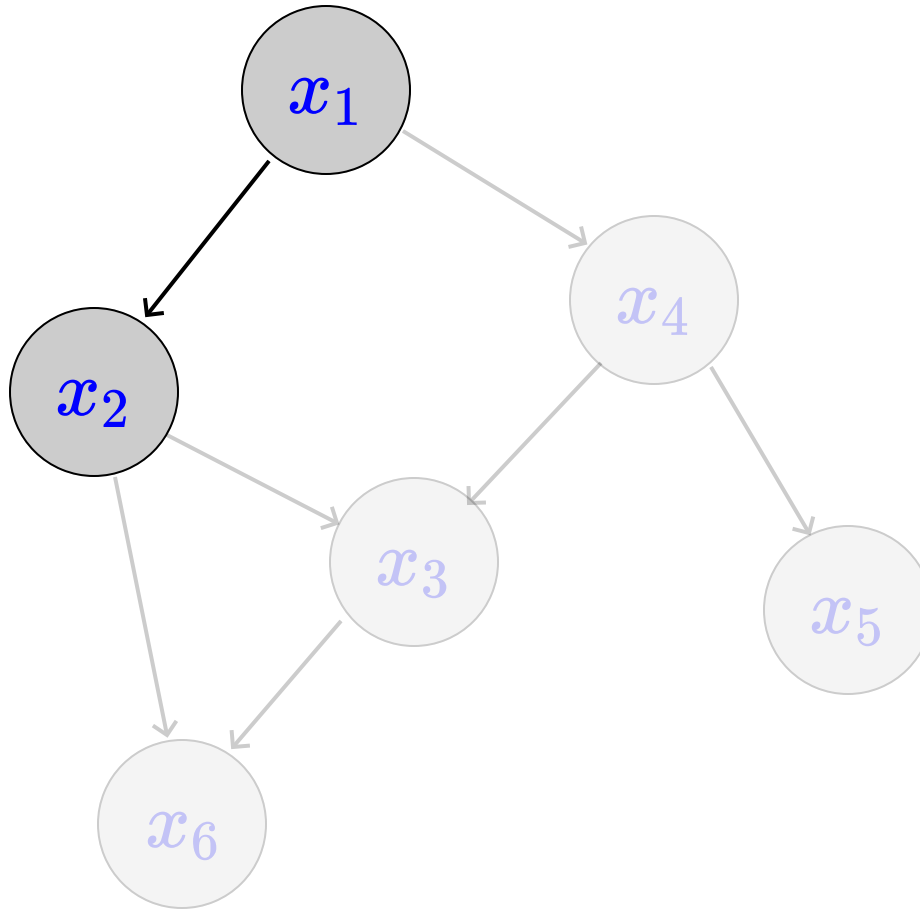


$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)$$



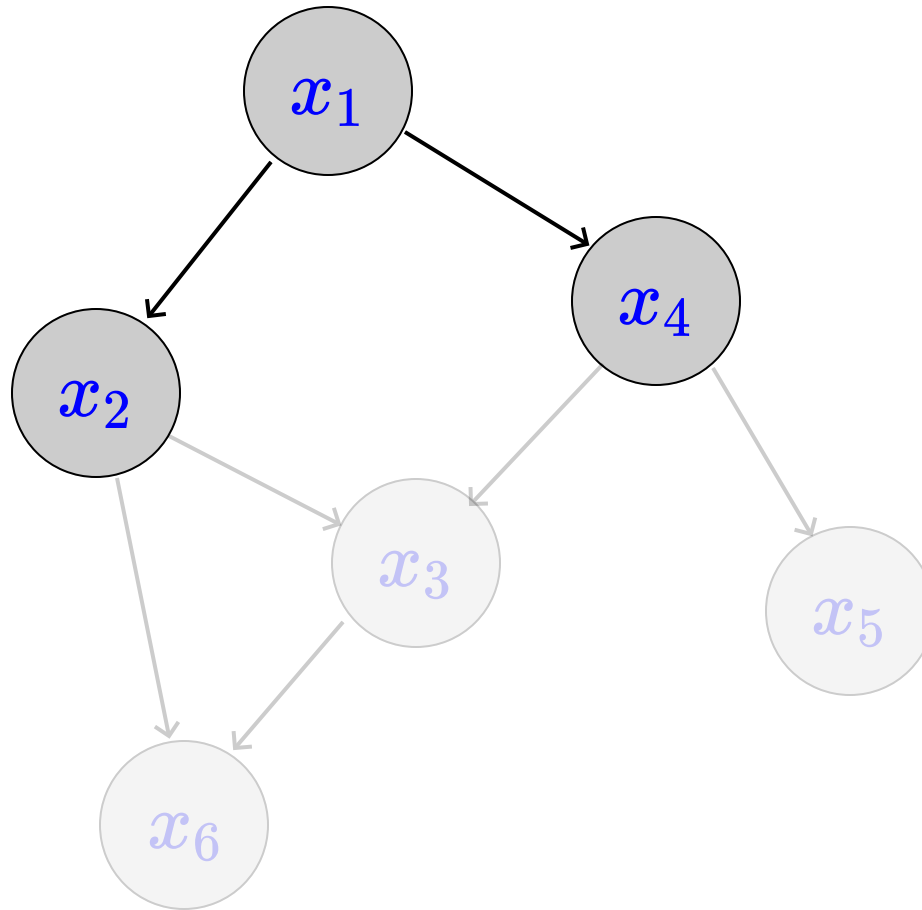
# Generative Model



$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)$$

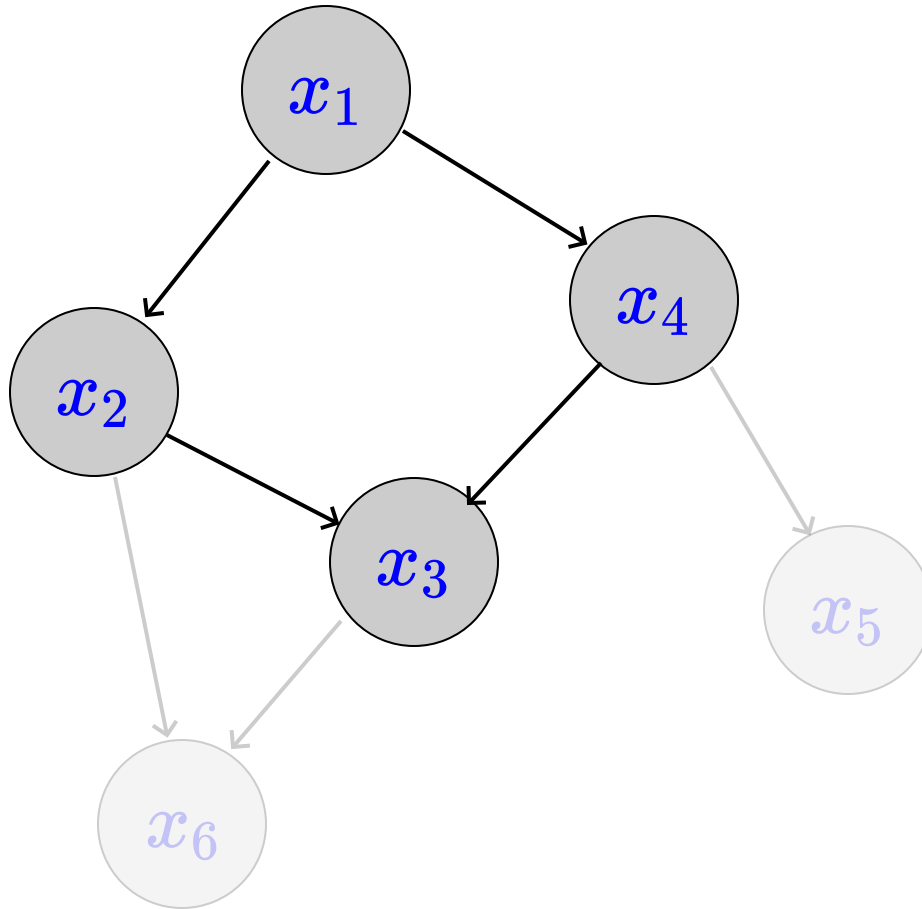
# Generative Model



$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)p(x_4|x_1)$$

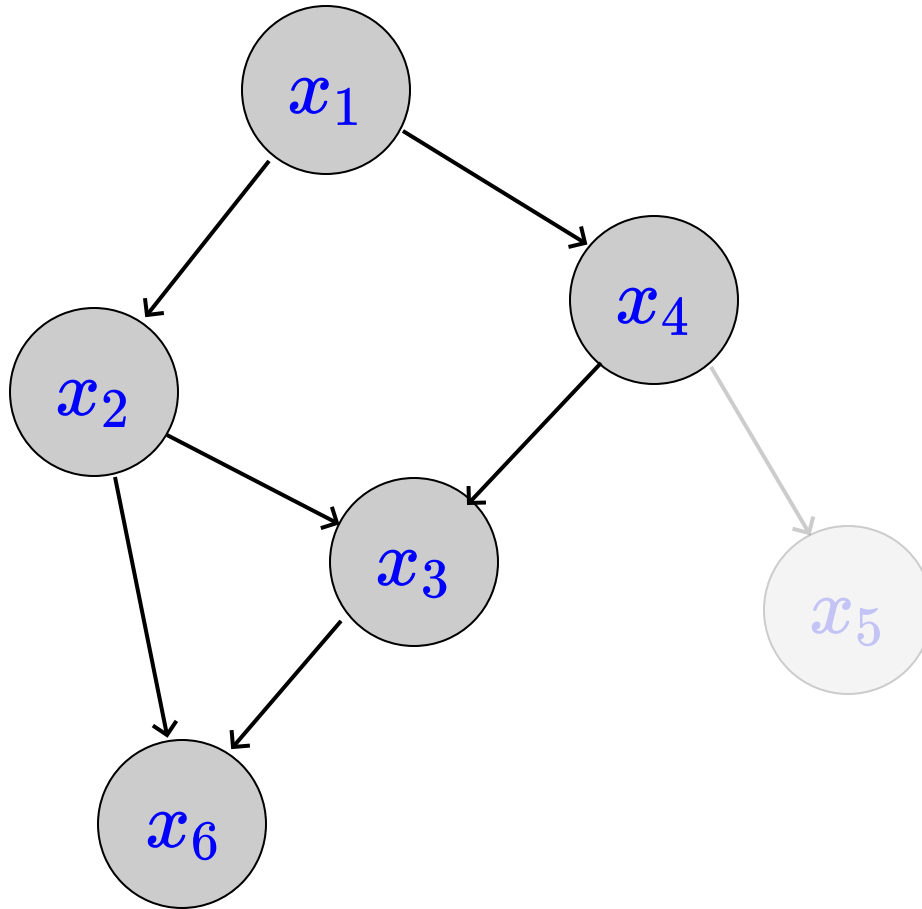
# Generative Model



$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_3|x_2, x_4)$$

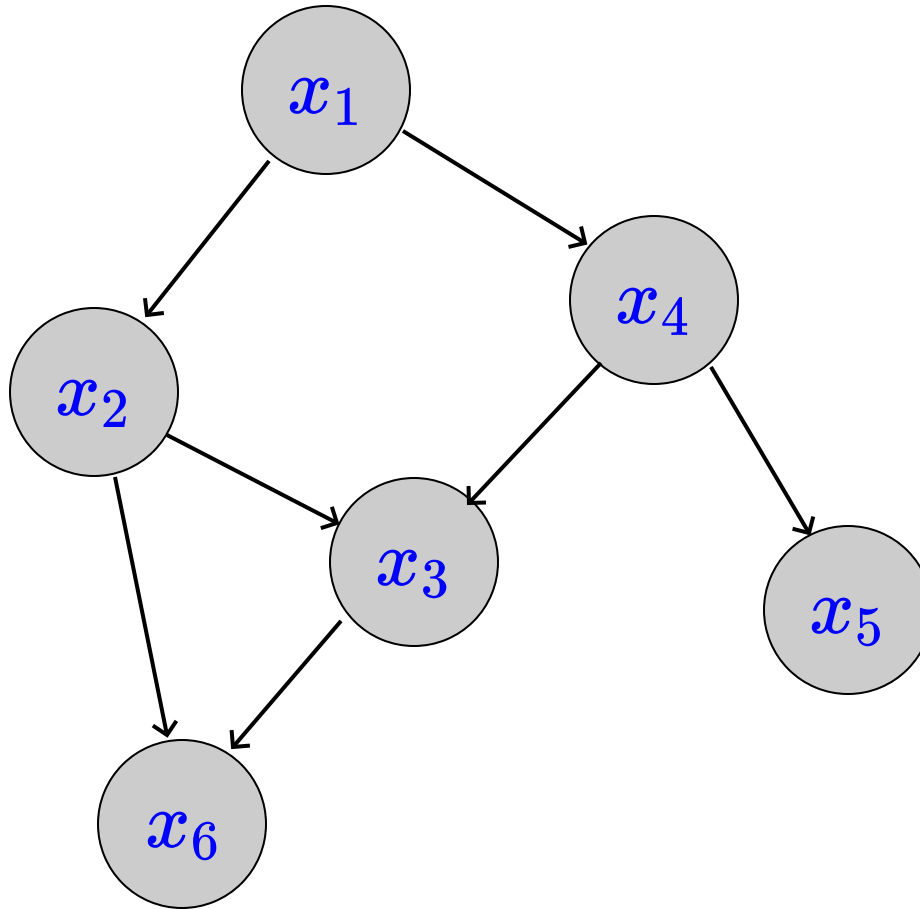
# Generative Model



$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_3|x_2, x_4)p(x_6|x_2, x_3)$$

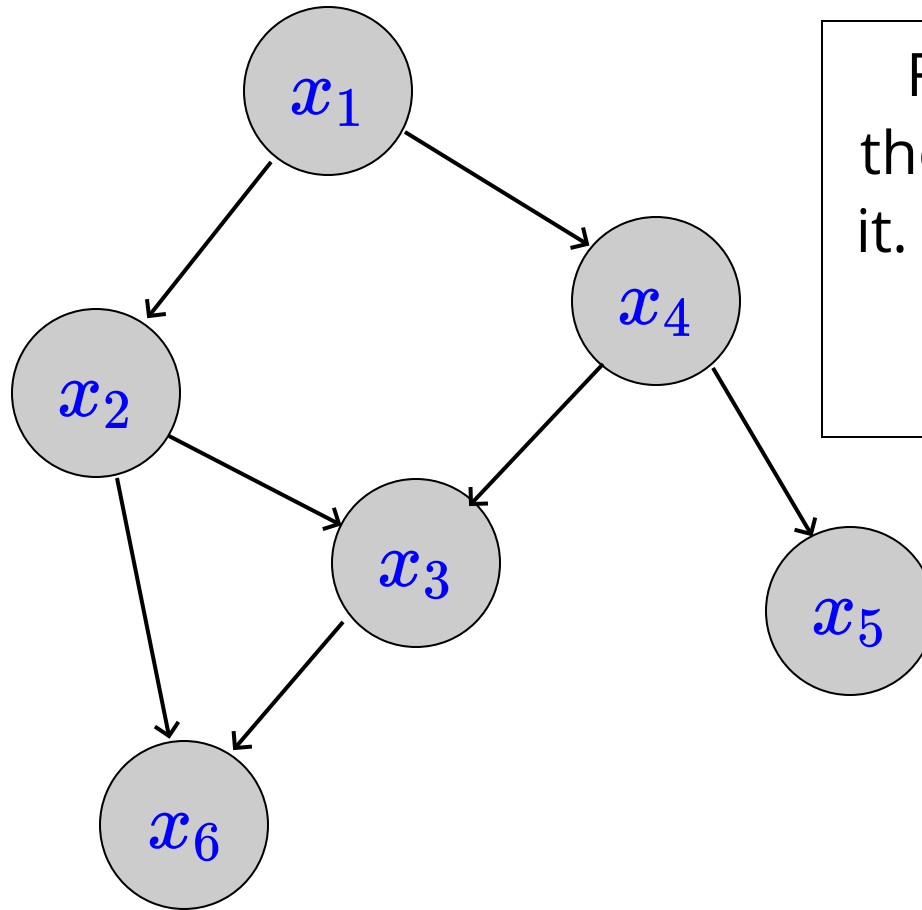
# Generative Model



$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_3|x_2, x_4)p(x_6|x_2, x_3)p(x_5|x_4)$$

# Generative Model



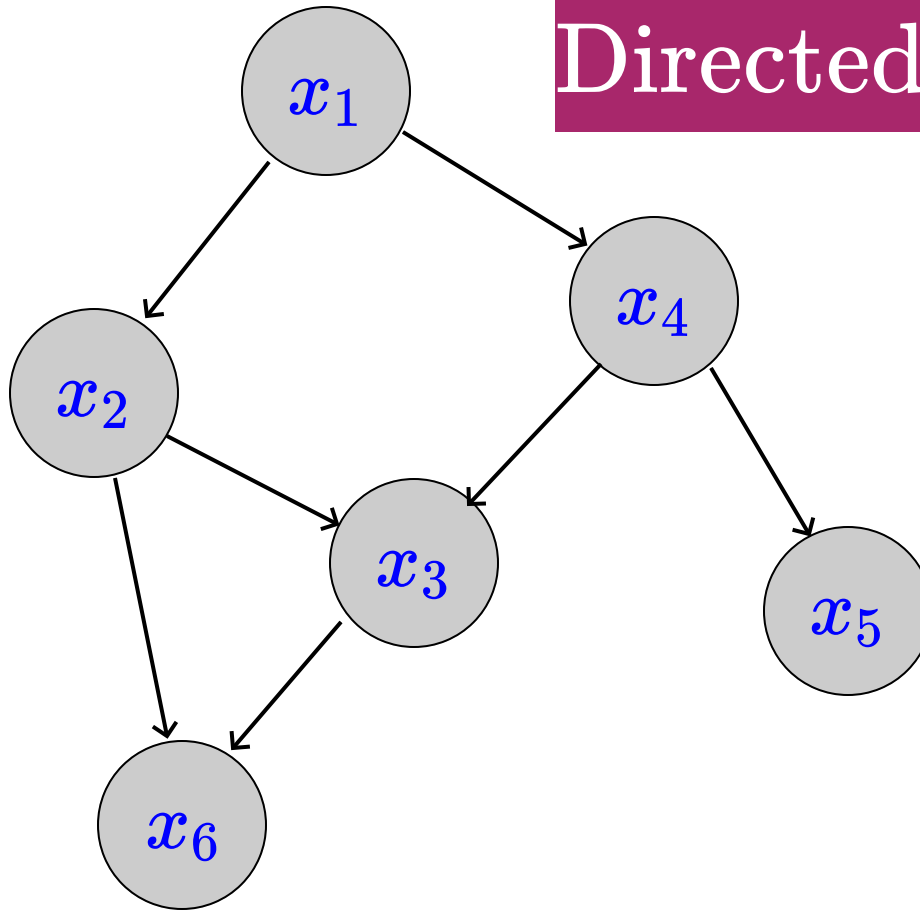
For a generative model, there is a graph underlying it. What types of properties do we have for such a graph?

$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_3|x_2, x_4)p(x_6|x_2, x_3)p(x_5|x_4)$$

# Generative Model

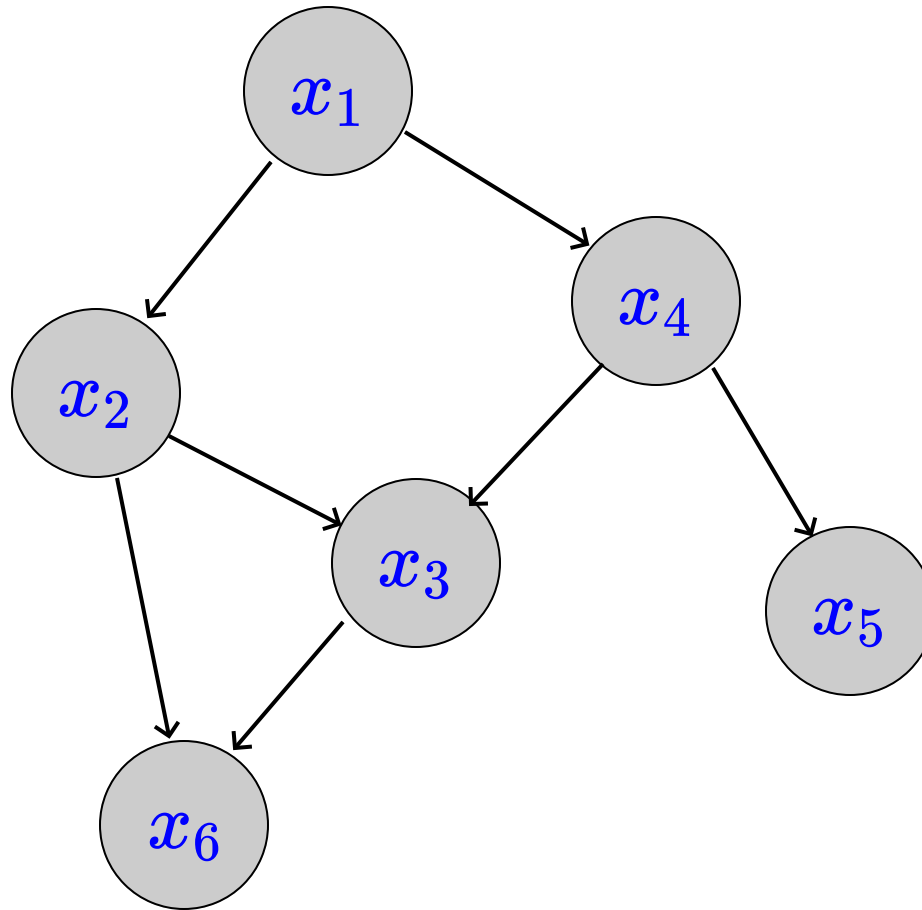
Directed Acyclic Graphs



$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_3|x_2, x_4)p(x_6|x_2, x_3)p(x_5|x_4)$$

# Generative Model



$$x_1 \in \{1, 2\}$$

$$x_2 \in \{1, 2\}$$

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

$$x_4 \in \{1, 2, 3\}$$

$$x_5 \in \{1, 2, 3, 4, 5\}$$

$$x_6 \in \{1, 2, 3, 4\}$$

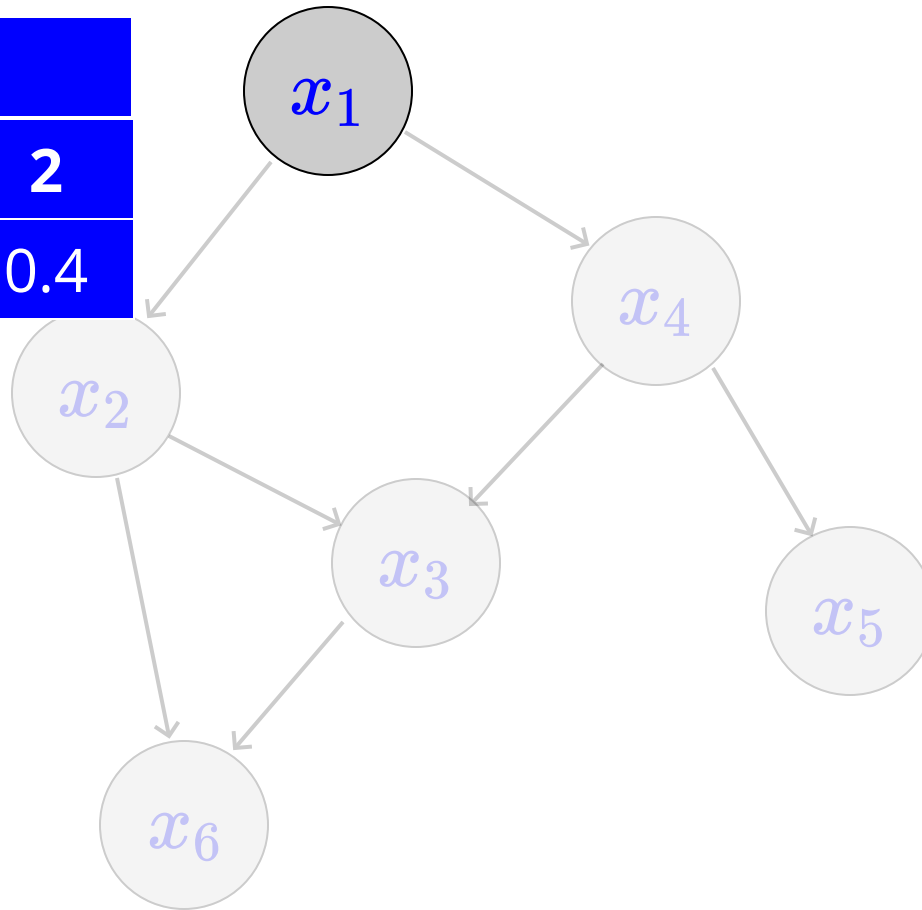
$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_3|x_2, x_4)p(x_6|x_2, x_3)p(x_5|x_4)$$



# Generative Model

| $x_1$ |     |
|-------|-----|
| 1     | 2   |
| 0.6   | 0.4 |



$$x_1 \in \{1, 2\}$$

$$x_2 \in \{1, 2\}$$

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

$$x_4 \in \{1, 2, 3\}$$

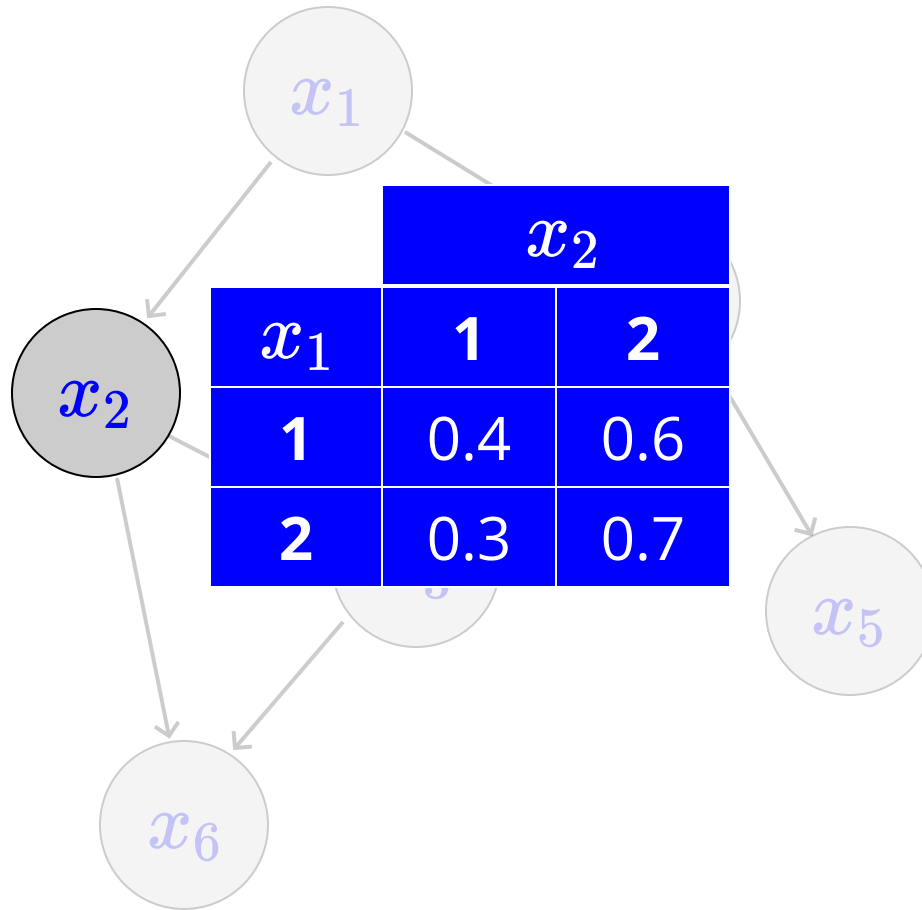
$$x_5 \in \{1, 2, 3, 4, 5\}$$

$$x_6 \in \{1, 2, 3, 4\}$$

$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_3|x_2, x_4)p(x_6|x_2, x_3)p(x_5|x_4)$$

# Generative Model



$$x_1 \in \{1, 2\}$$

$$x_2 \in \{1, 2\}$$

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

$$x_4 \in \{1, 2, 3\}$$

$$x_5 \in \{1, 2, 3, 4, 5\}$$

$$x_6 \in \{1, 2, 3, 4\}$$

$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_3|x_2, x_4)p(x_6|x_2, x_3)p(x_5|x_4)$$

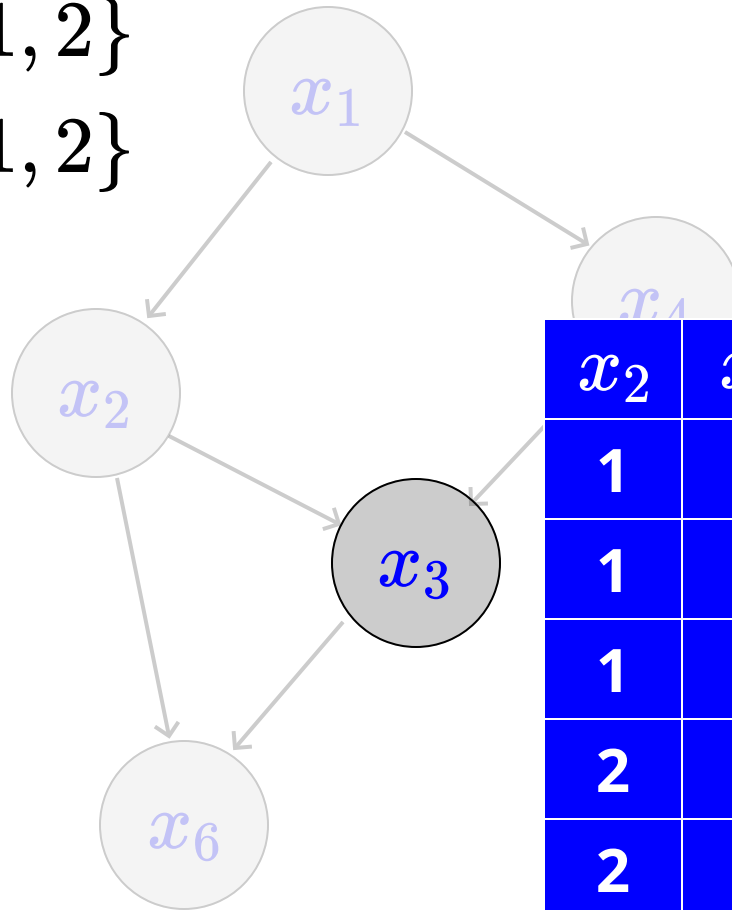
# Generative Model

$$x_1 \in \{1, 2\}$$

$$x_2 \in \{1, 2\}$$

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

$$x_4 \in \{1, 2, 3\}$$



|       |       | $x_3$ |     |     |
|-------|-------|-------|-----|-----|
| $x_2$ | $x_4$ | 1     | 2   | 3   |
| 1     | 1     | 0.4   | 0.5 | 0.1 |
| 1     | 2     | 0.3   | 0.4 | 0.3 |
| 1     | 3     | 0.1   | 0.8 | 0.1 |
| 2     | 1     | 0.7   | 0.3 | 0.0 |
| 2     | 2     | 0.2   | 0.6 | 0.2 |
| 2     | 3     | 0.5   | 0.3 | 0.2 |

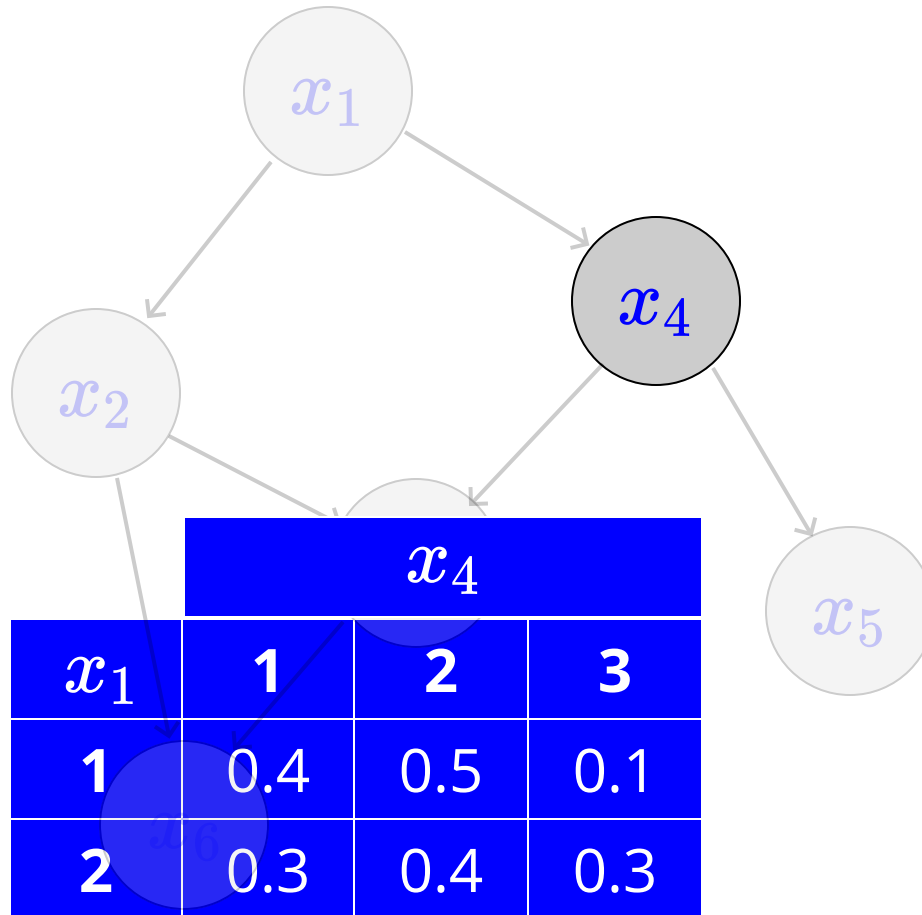
$\{1, 2, 3, 4, 5\}$

$\{1, 2, 3, 4\}$

$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_3|x_2, x_4)p(x_6|x_2, x_3)p(x_5|x_4)$$

# Generative Model



$$x_1 \in \{1, 2\}$$

$$x_2 \in \{1, 2\}$$

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

$$x_4 \in \{1, 2, 3\}$$

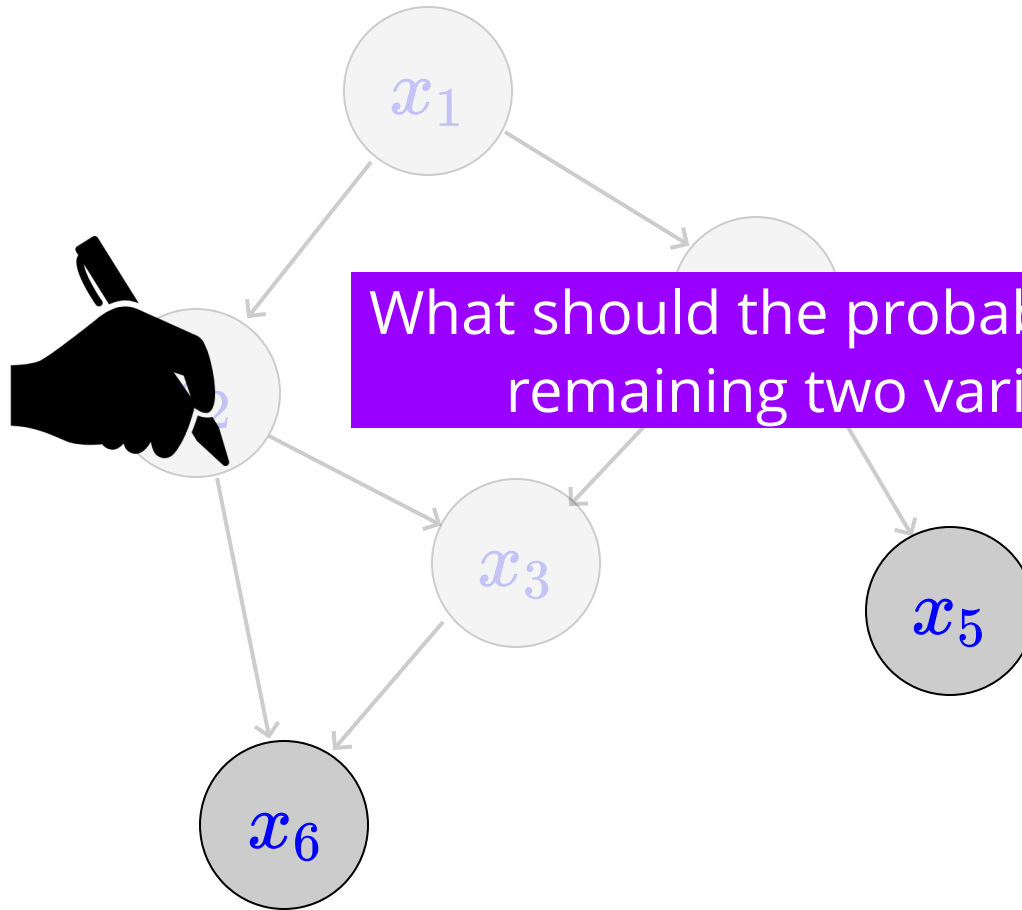
$$x_5 \in \{1, 2, 3, 4, 5\}$$

$$x_6 \in \{1, 2, 3, 4\}$$

$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_3|x_2, x_4)p(x_6|x_2, x_3)p(x_5|x_4)$$

# Generative Model



What should the probability tables for the remaining two variables be like?

$$x_1 \in \{1, 2\}$$

$$x_2 \in \{1, 2\}$$

$$x_4 \in \{1, 2, 3\}$$

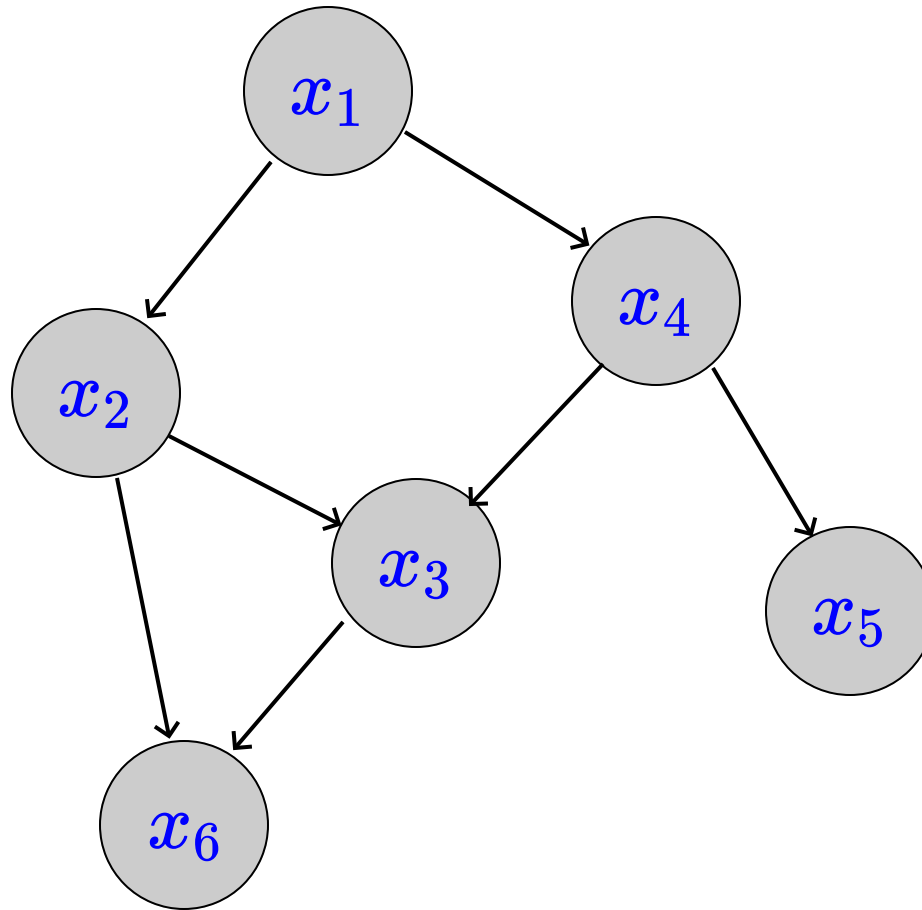
$$x_5 \in \{1, 2, 3, 4, 5\}$$

$$x_6 \in \{1, 2, 3, 4\}$$

$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

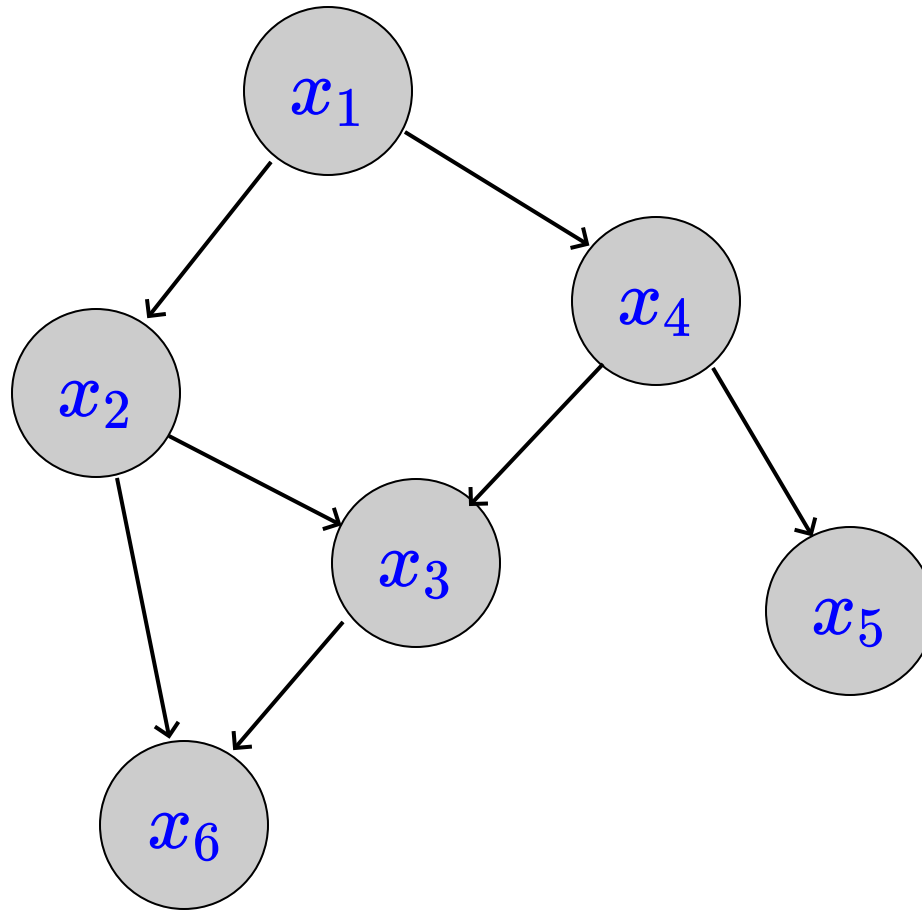
$$= p(x_1)p(x_2|x_1)p(x_4|x_1)p(x_3|x_2, x_4)p(x_6|x_2, x_3)p(x_5|x_4)$$

# Bayesian Networks



$$p(x_1 = 1, x_2 = 2, x_3 = 1, x_4 = 3, x_5 = 5, x_6 = 2)$$

# Bayesian Networks

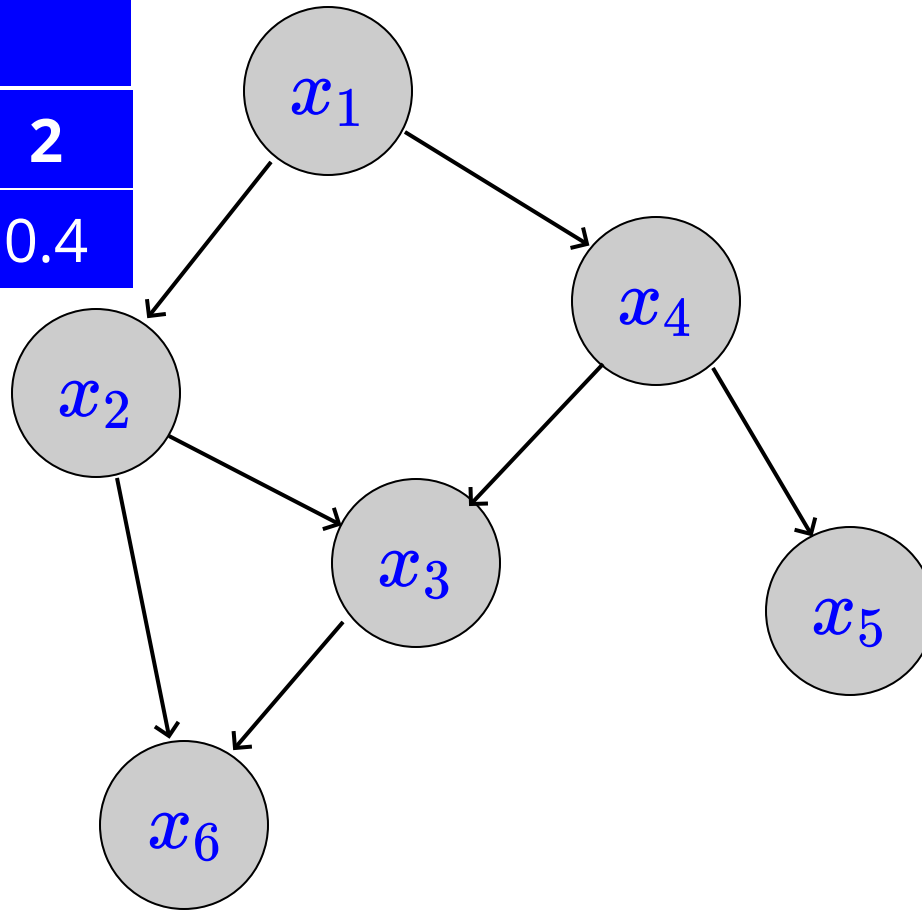


$$\begin{aligned} & p(x_1 = 1, x_2 = 2, x_3 = 1, x_4 = 3, x_5 = 5, x_6 = 2) \\ &= p(x_1 = 1) \times p(x_2 = 2 | x_1 = 1) \times p(x_4 = 3 | x_1 = 1) \end{aligned}$$

$$\times p(x_3 = 1 | x_2 = 2, x_4 = 3) \times p(x_6 = 2 | x_2 = 2, x_3 = 1) \times p(x_5 = 5 | x_4 = 3)$$

# Bayesian Networks

| $x_1$ |     |
|-------|-----|
| 1     | 2   |
| 0.6   | 0.4 |



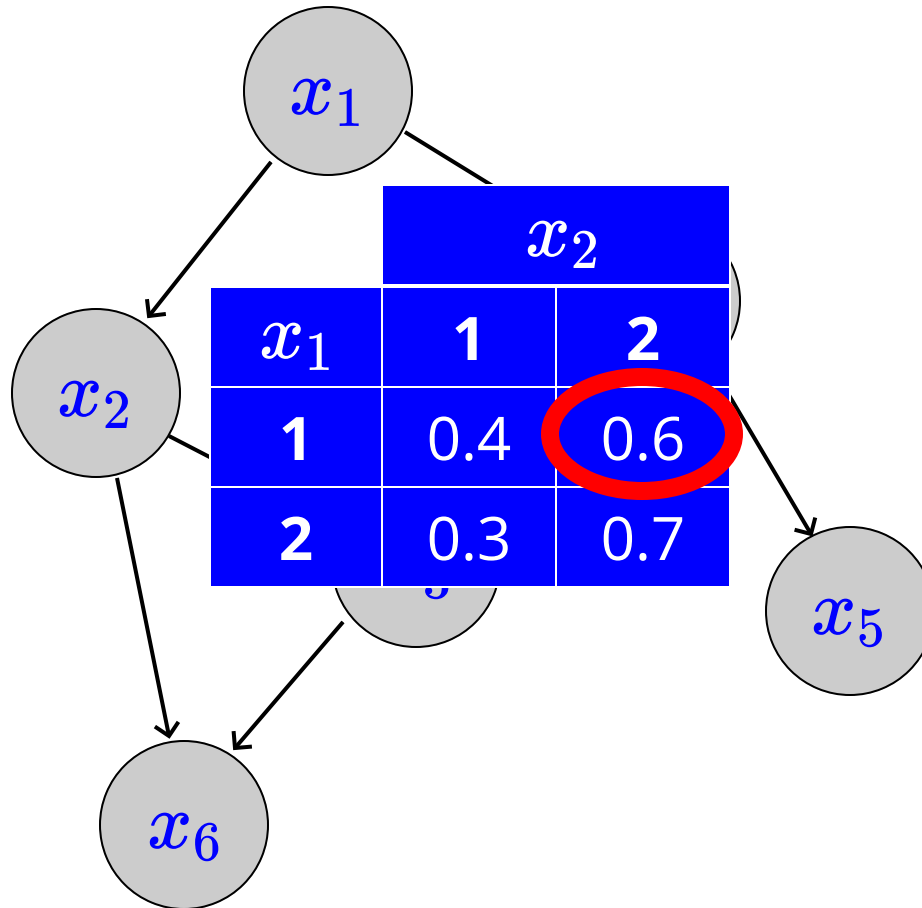
$$p(x_1 = 1, x_2 = 2, x_3 = 1, x_4 = 3, x_5 = 5, x_6 = 2)$$

$$= p(x_1 = 1) \times p(x_2 = 2 | x_1 = 1) \times p(x_4 = 3 | x_1 = 1)$$

$$\times p(x_3 = 1 | x_2 = 2, x_4 = 3) \times p(x_6 = 2 | x_2 = 2, x_3 = 1) \times p(x_5 = 5 | x_4 = 3)$$



# Bayesian Networks



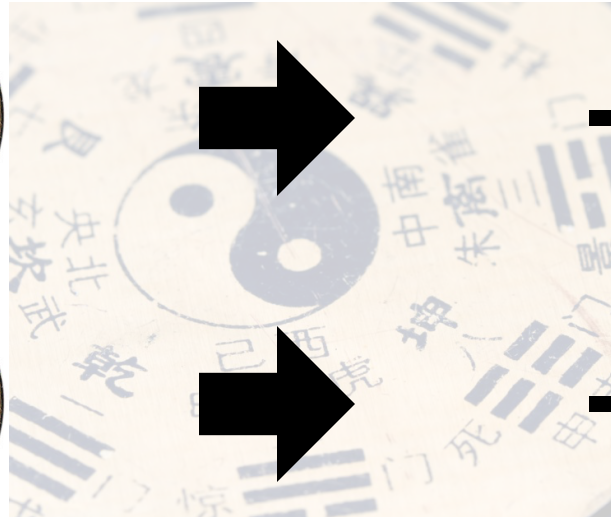
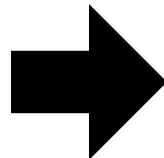
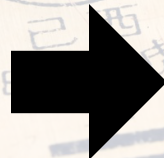
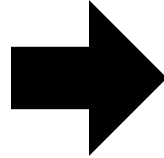
$$\begin{aligned}
 & p(x_1 = 1, x_2 = 2, x_3 = 1, x_4 = 3, x_5 = 5, x_6 = 2) \\
 &= p(x_1 = 1) \times p(x_2 = 2 | x_1 = 1) \times p(x_4 = 3 | x_1 = 1)
 \end{aligned}$$

$$\times p(x_3 = 1 | x_2 = 2, x_4 = 3) \times p(x_6 = 2 | x_2 = 2, x_3 = 1) \times p(x_5 = 5 | x_4 = 3)$$

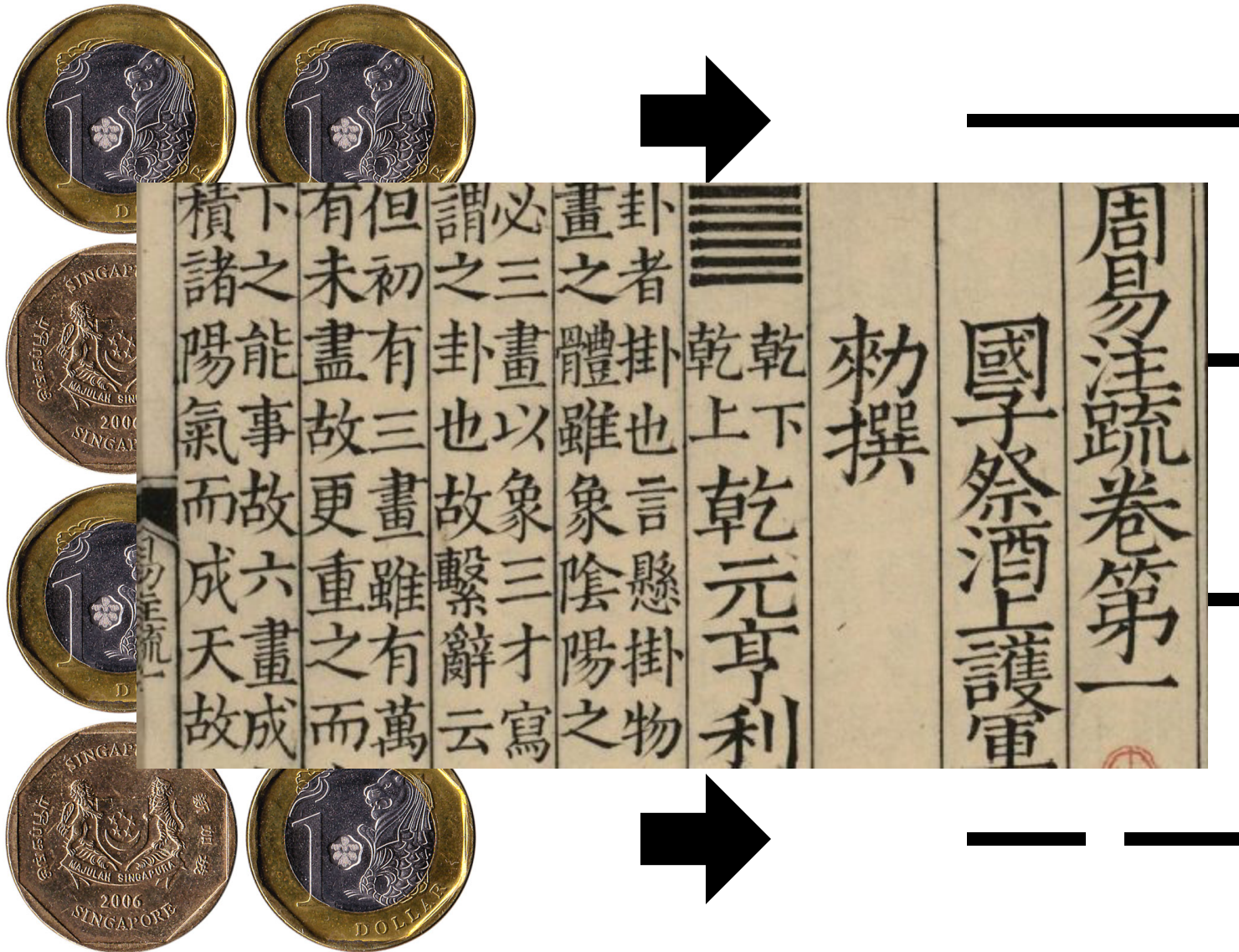
# Bayesian Networks



# Bayesian Networks



# Bayesian Networks





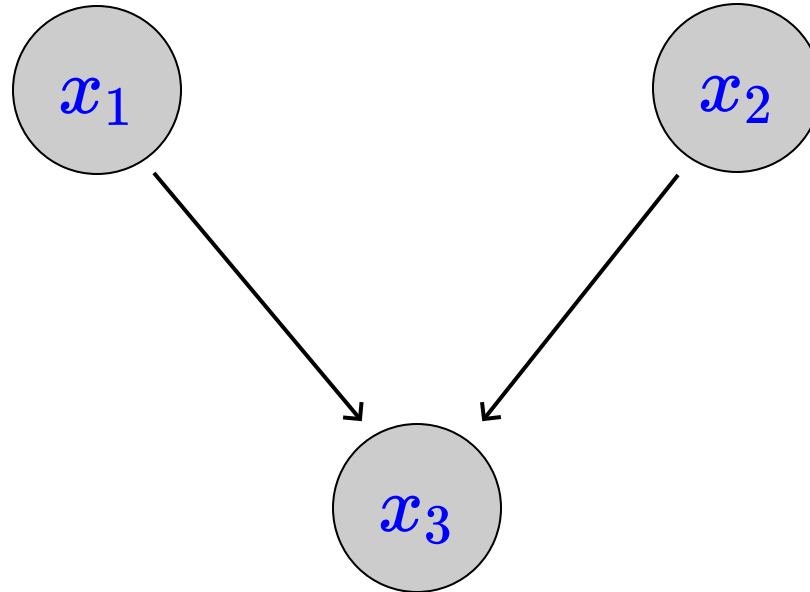
# Bayesian Networks



Can we represent this generative process  
with a Bayesian Network?

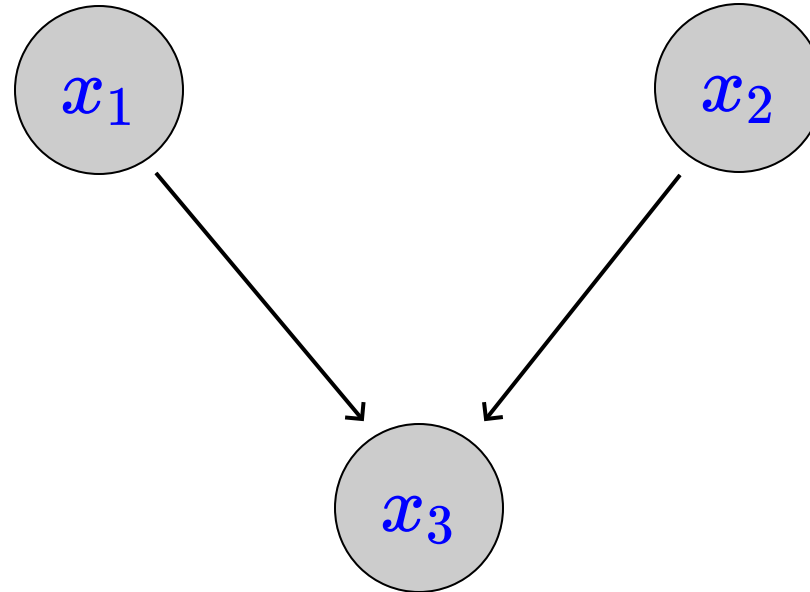


# Bayesian Networks



# Bayesian Networks

| $x_1$ |     |
|-------|-----|
| 1     | 2   |
| 0.5   | 0.5 |

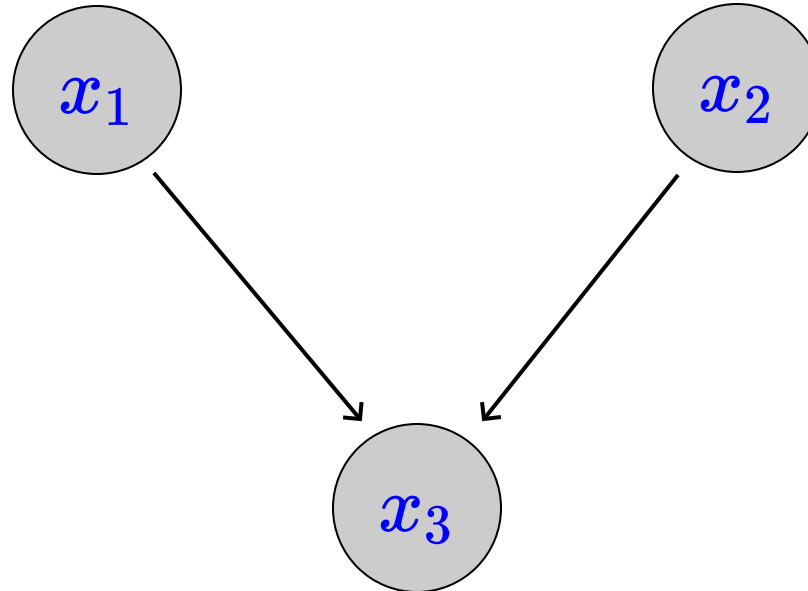


| $x_2$ |     |
|-------|-----|
| 1     | 2   |
| 0.5   | 0.5 |

|       |       | $x_3$ |     |
|-------|-------|-------|-----|
| $x_1$ | $x_2$ | —     | - - |
| H     | H     | 1     | 0   |
| H     | T     | 0     | 1   |
| T     | H     | 0     | 1   |
| T     | T     | 1     | 0   |

# Bayesian Networks

| $x_1$ |     |
|-------|-----|
| 1     | 2   |
| 0.5   | 0.5 |



| $x_2$ |     |
|-------|-----|
| 1     | 2   |
| 0.5   | 0.5 |

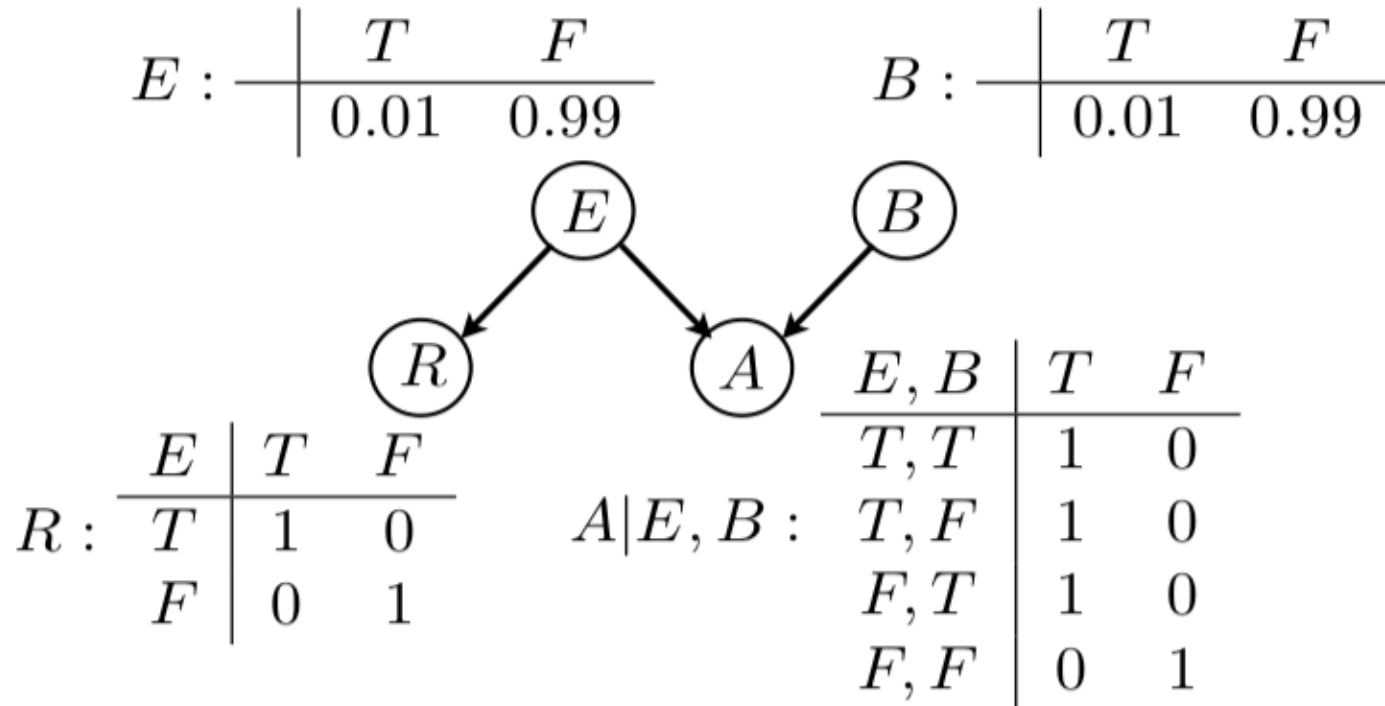
|       |       | $x_3$ |    |
|-------|-------|-------|----|
| $x_1$ | $x_2$ | —     | -- |
| H     | H     | 1     | 0  |
| H     | T     | 0     | 1  |
| T     | H     | 0     | 1  |
| T     | T     | 1     | 0  |

Are  $x_1$  and  $x_2$  independent?

What if  $x_3$  is given?

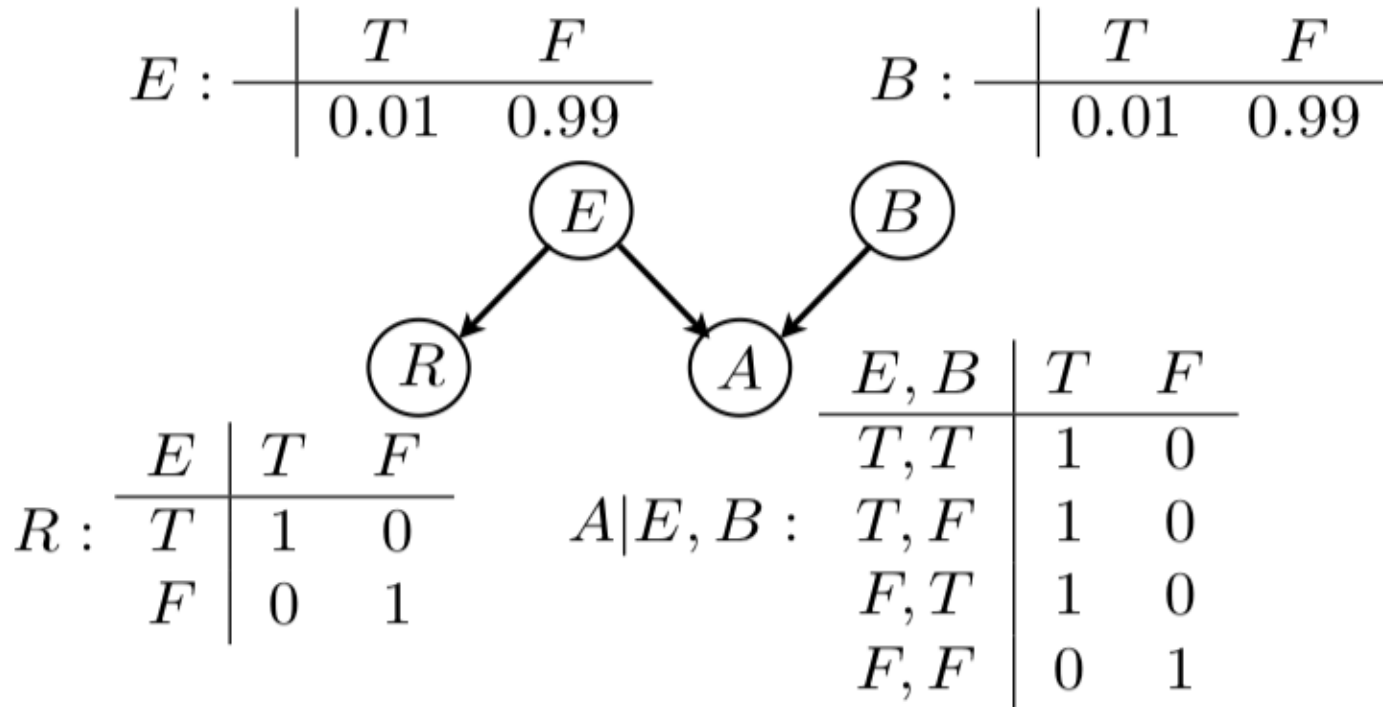


# Bayesian Networks



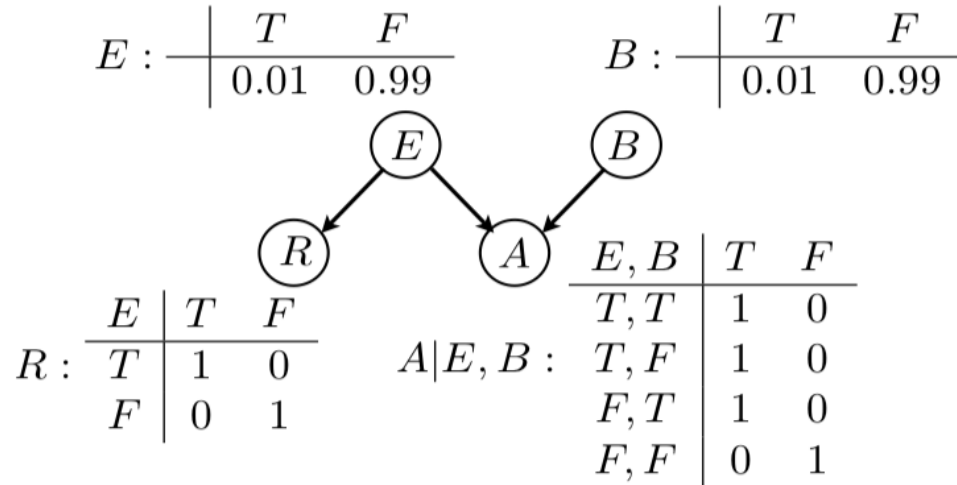
$$P(E = e, B = b, A = a, R = r)$$

# Bayesian Networks



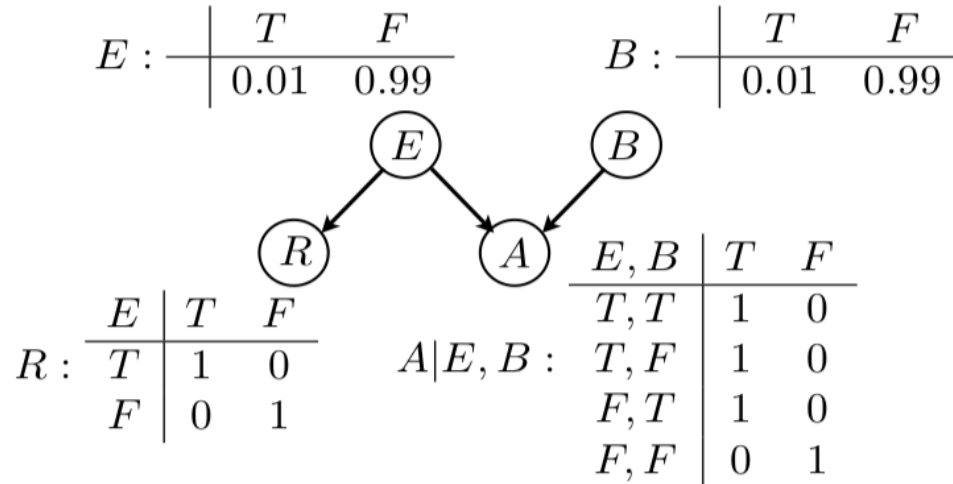
$$\begin{aligned}
 &P(E = e, B = b, A = a, R = r) \\
 &= P(E = e)P(B = b)P(A = a|E = e, B = b)P(R = r|E = e)
 \end{aligned}$$

# Bayesian Networks



$$P(B = T | A = T) = ?$$

# Bayesian Networks

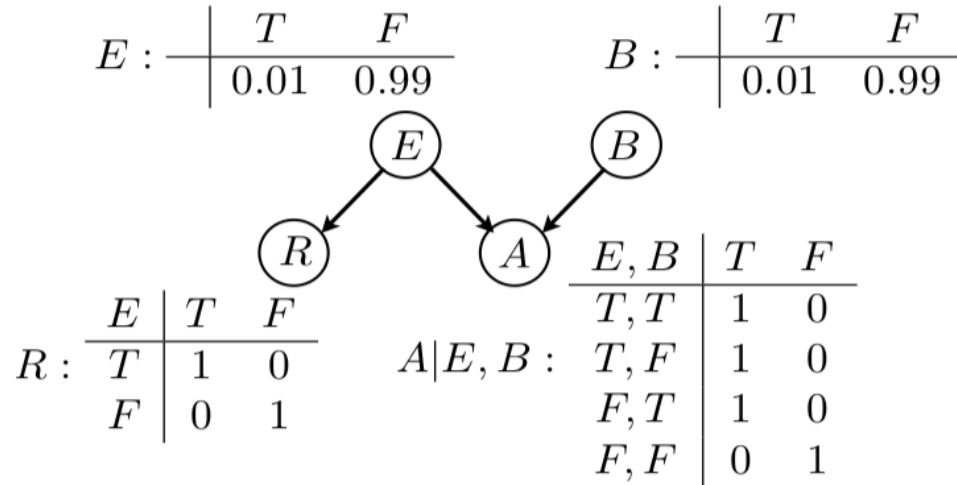


$$P(B = T | A = T) = \frac{P(B=T, A=T)}{\sum_{b \in \{T, F\}} P(B=b, A=T)}$$



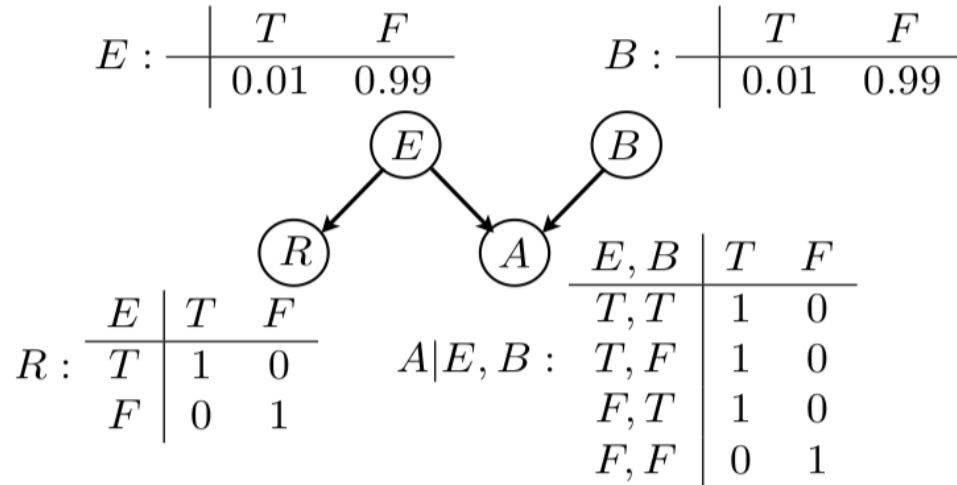
Let us look at  
this problem

# Bayesian Networks



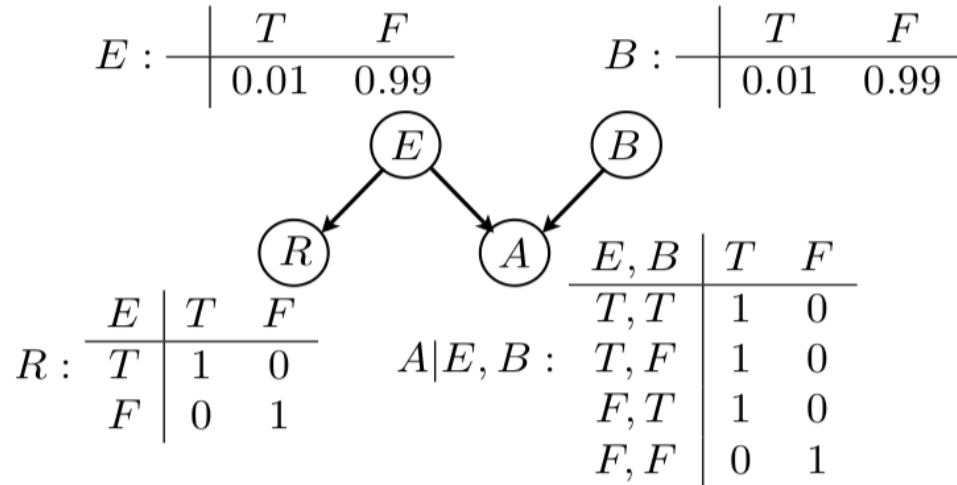
$$\begin{aligned}
 & P(B = b, A = T) \\
 = & \sum_{e \in \{T, F\}} \sum_{r \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b) P(R = r | E = e)
 \end{aligned}$$

# Bayesian Networks



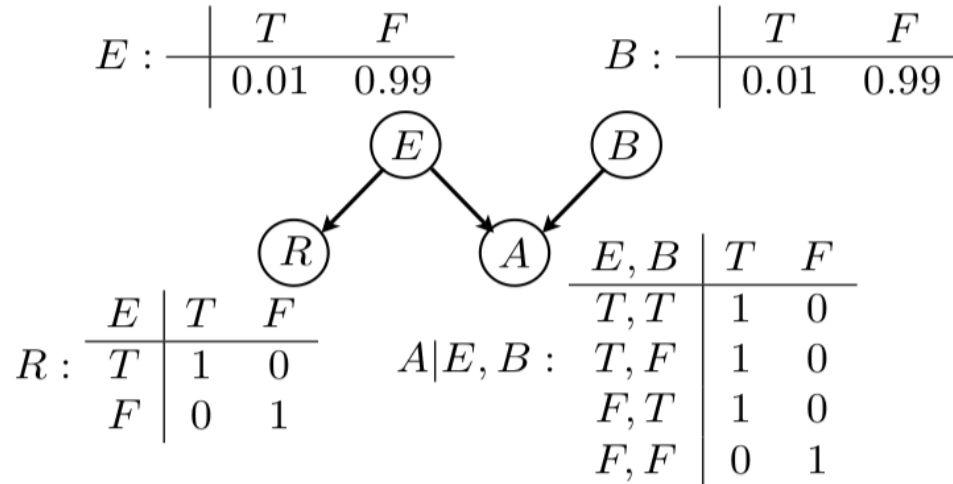
$$\begin{aligned}
 & P(B = b, A = T) \\
 = & \sum_{e \in \{T, F\}} \sum_{r \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b) P(R = r | E = e) \\
 = & \sum_{e \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b) \sum_{r \in \{T, F\}} P(R = r | E = e)
 \end{aligned}$$

# Bayesian Networks



$$\begin{aligned}
 & P(B = b, A = T) \\
 = & \sum_{e \in \{T, F\}} \sum_{r \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b) P(R = r | E = e) \\
 = & \sum_{e \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b) \sum_{r \in \{T, F\}} P(R = r | E = e) \\
 = & \sum_{e \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b)
 \end{aligned}$$

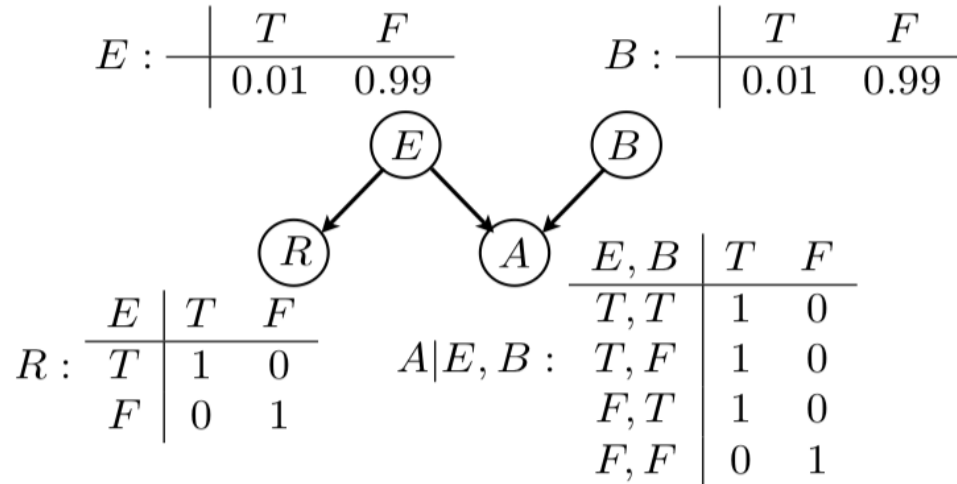
# Bayesian Networks



$$\begin{aligned}
 & P(B = b, A = T) \\
 = & \sum_{e \in \{T, F\}} \sum_{r \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b) P(R = r | E = e) \\
 = & \sum_{e \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b) \sum_{r \in \{T, F\}} P(R = r | E = e) \\
 = & \sum_{e \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b) \\
 = & P(B = b) \sum_{e \in \{T, F\}} P(E = e) P(A = T | E = e, B = b)
 \end{aligned}$$



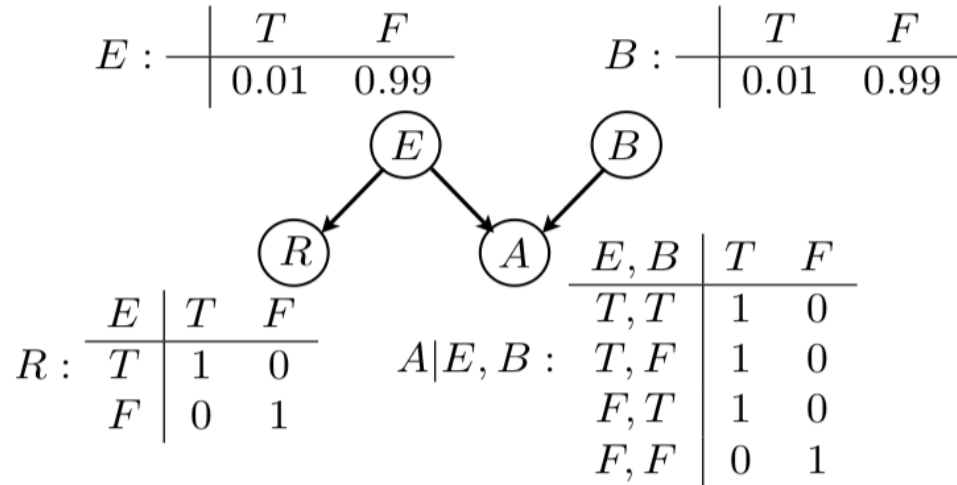
# Bayesian Networks



$$\begin{aligned}
 & P(B = b, A = T) \\
 = & \sum_{e \in \{T, F\}} \sum_{r \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b) P(R = r | E = e) \\
 = & \sum_{e \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b) \sum_{r \in \{T, F\}} P(R = r | E = e) \\
 = & \sum_{e \in \{T, F\}} P(E = e) P(B = b) P(A = T | E = e, B = b) \\
 = & P(B = b) \sum_{e \in \{T, F\}} P(E = e) P(A = T | E = e, B = b)
 \end{aligned}$$

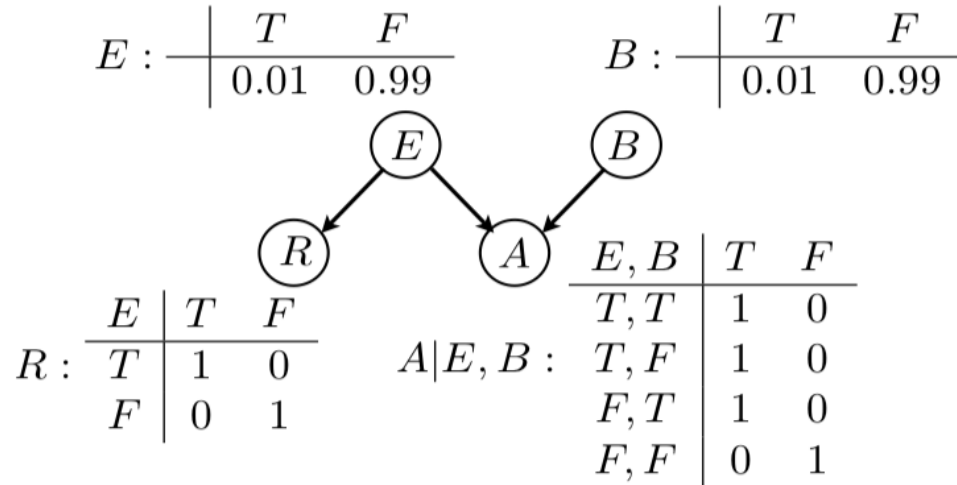
$$P(B = T | A = T) = \frac{P(B=T, A=T)}{\sum_{b \in \{T, F\}} P(B=b, A=T)} = ?$$

# Bayesian Networks



$$P(B = T | A = T, R = T) = ?$$

# Bayesian Networks

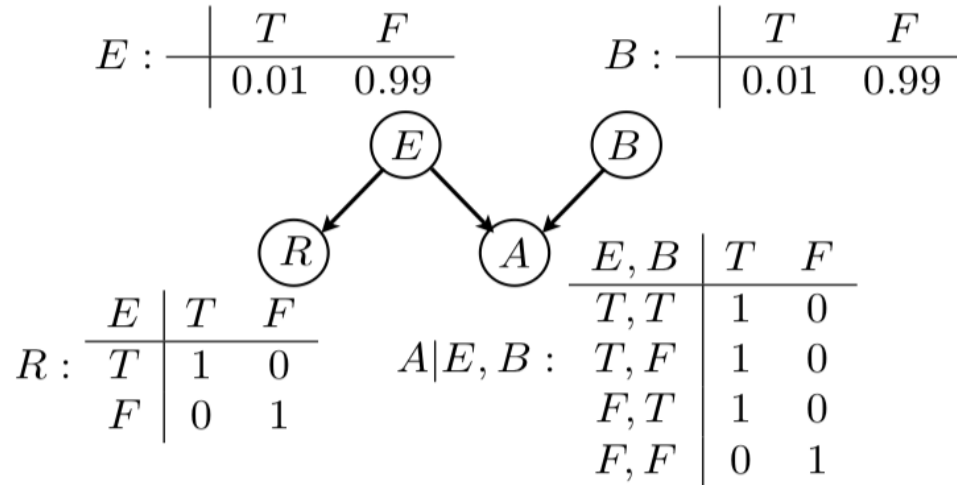


$$P(B = T | A = T, R = T) = ?$$

$$P(B = T | A = T, E = T) = ?$$



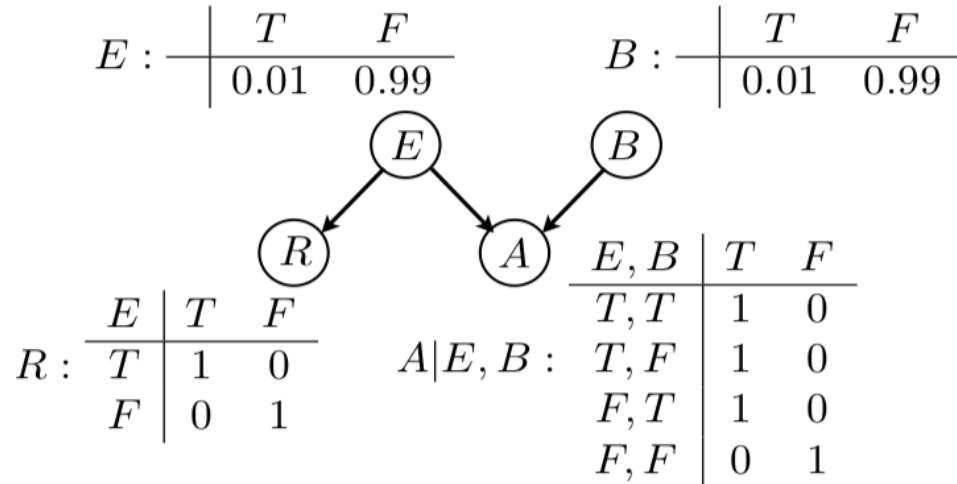
# Bayesian Networks



$$P(B = T | A = T, R = T) = ?$$

$$P(B = T | A = T, E = T) = 0.01$$

# Bayesian Networks



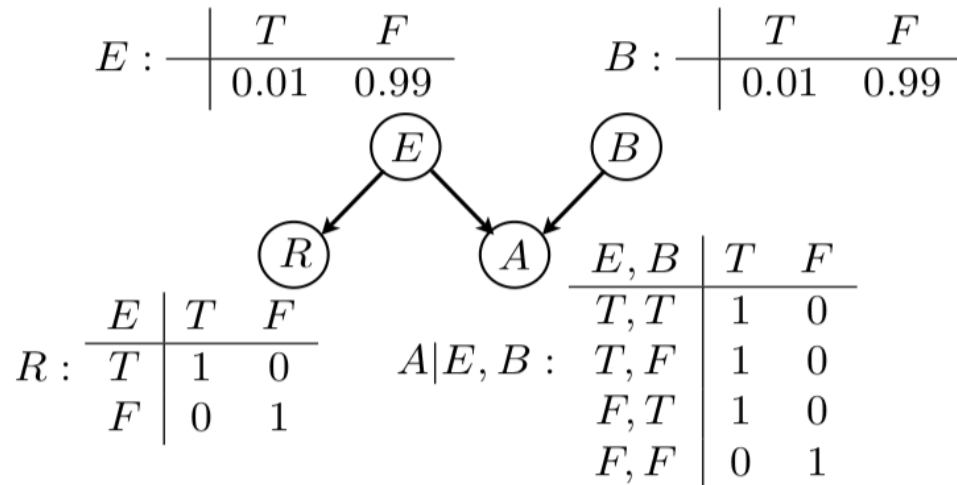
$\approx 0.5$

0.01

$$P(B = T | A = T) \neq P(B = T | A = T, E = T)$$

$B$  and  $E$  are not independent given  $A$

# Explaining Away



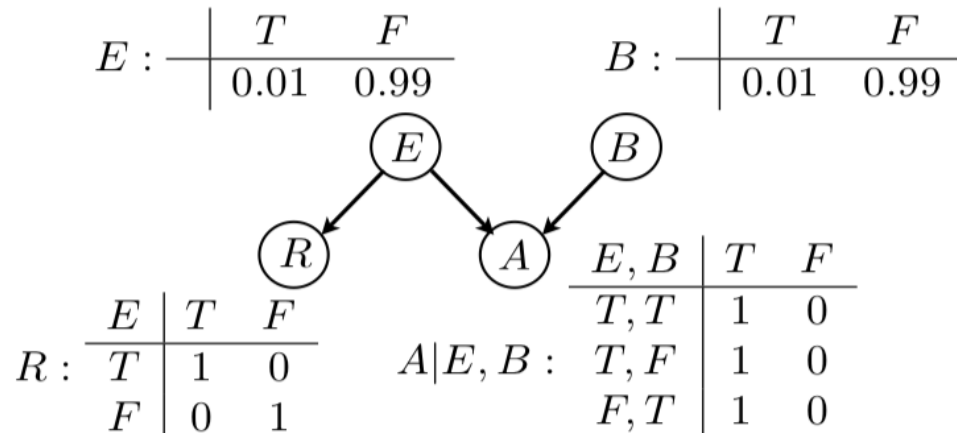
$\approx 0.5$

0.01

$$P(B = T | A = T) \neq P(B = T | A = T, E = T)$$

$B$  and  $E$  are not independent given  $A$

# Bayesian Networks

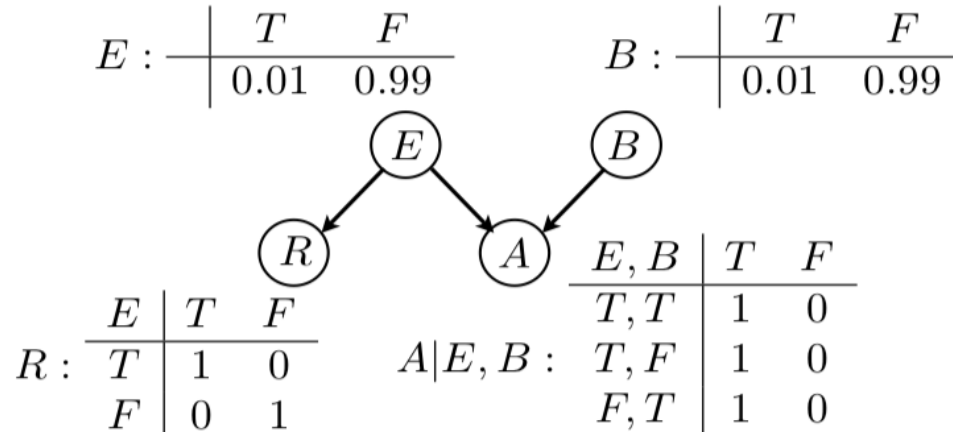


Can we read off such independence information from the network directly without involving calculation?

$$P(B = T | A = T) \neq P(B = T | A = T, E = T)$$

$B$  and  $E$  are not independent given  $A$

# Bayesian Networks



Can we read off such independence information from the network directly without involving probability?

**Next Lecture!**

$$P(B = T | A = T) \neq P(B = T | A = T, E = T)$$

$B$  and  $E$  are not independent given  $A$