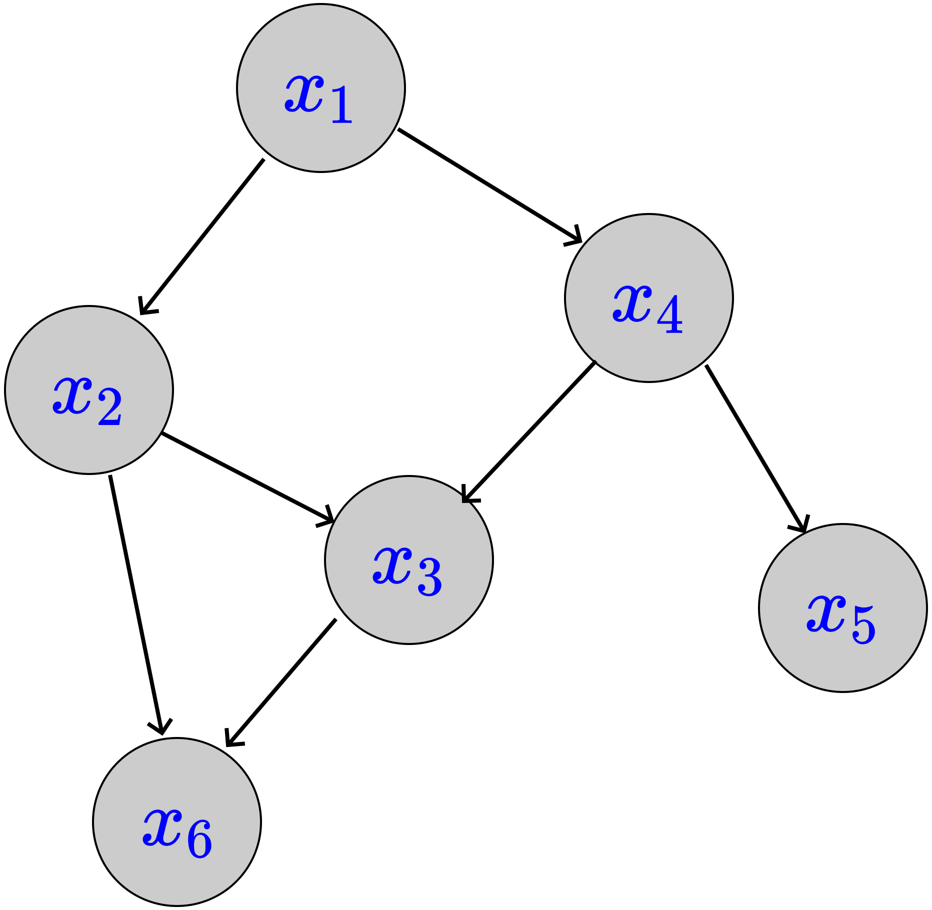
50.007
Machine Learning

Lu, Wei

SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

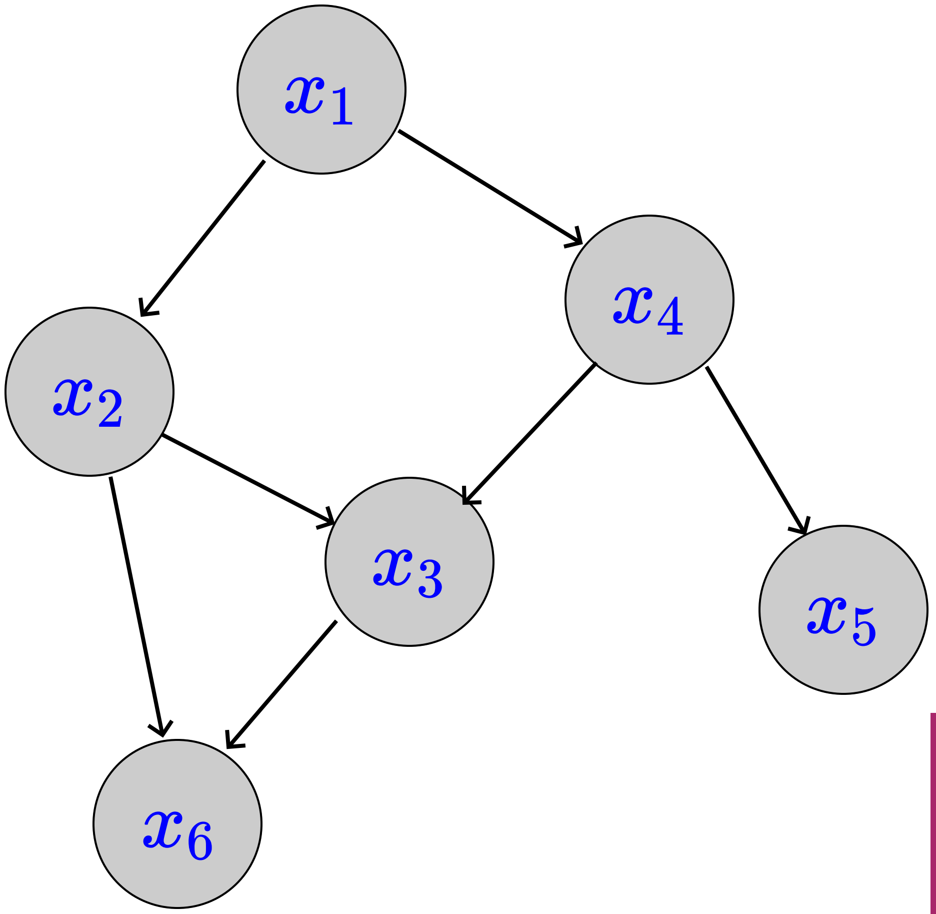
Bayesian Networks (II)

Bayesian Networks



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

Bayesian Networks

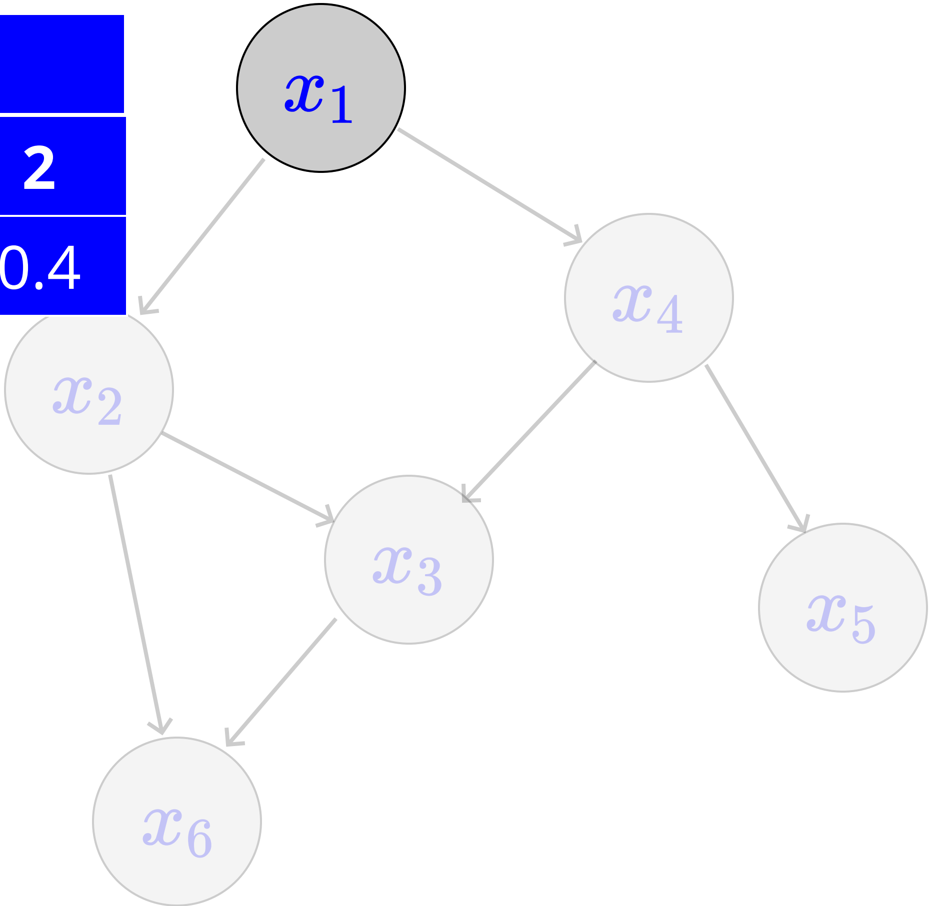


The set of parent nodes for x_i.

$$p(x_1, \dots, x_d) = \prod_{i=1}^{d} p(x_i | \mathbf{x}_{pa_i})$$

Bayesian Networks

x_1	
1	**2**
0.6	0.4



$$x_1 \in \{1,2\}$$
$$x_2 \in \{1,2\}$$
$$x_3 \in \{1,2,3\}$$
$$x_4 \in \{1,2,3\}$$
$$x_5 \in \{1,2,3,4,5\}$$
$$x_6 \in \{1,2,3,4\}$$

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

Bayesian Networks

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

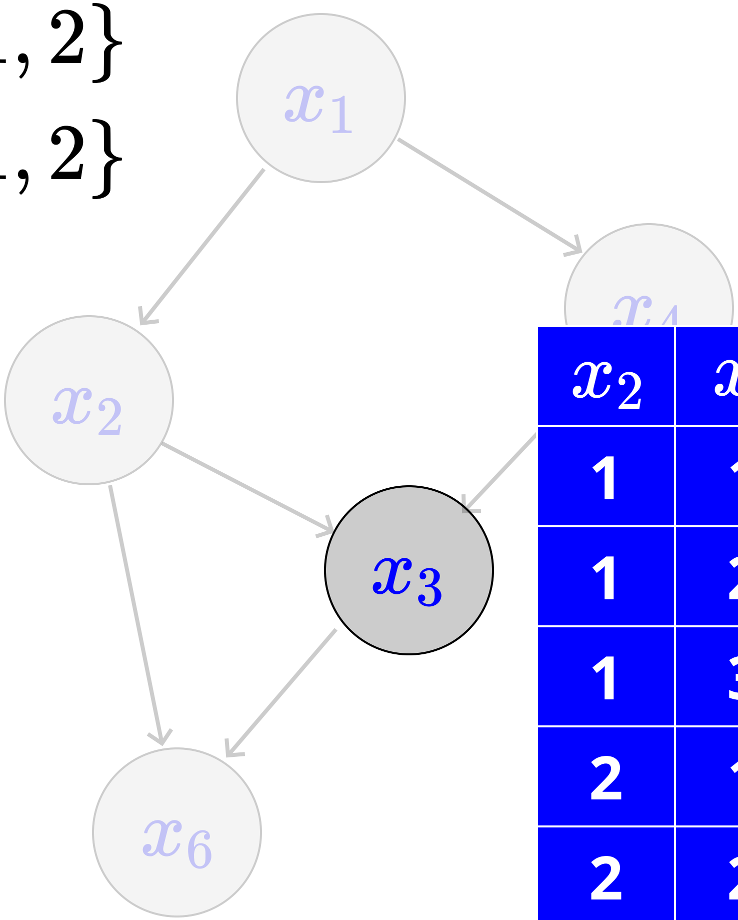
$x_2 \in \{1,2\}$

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

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

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

$1,2,3,4\}$

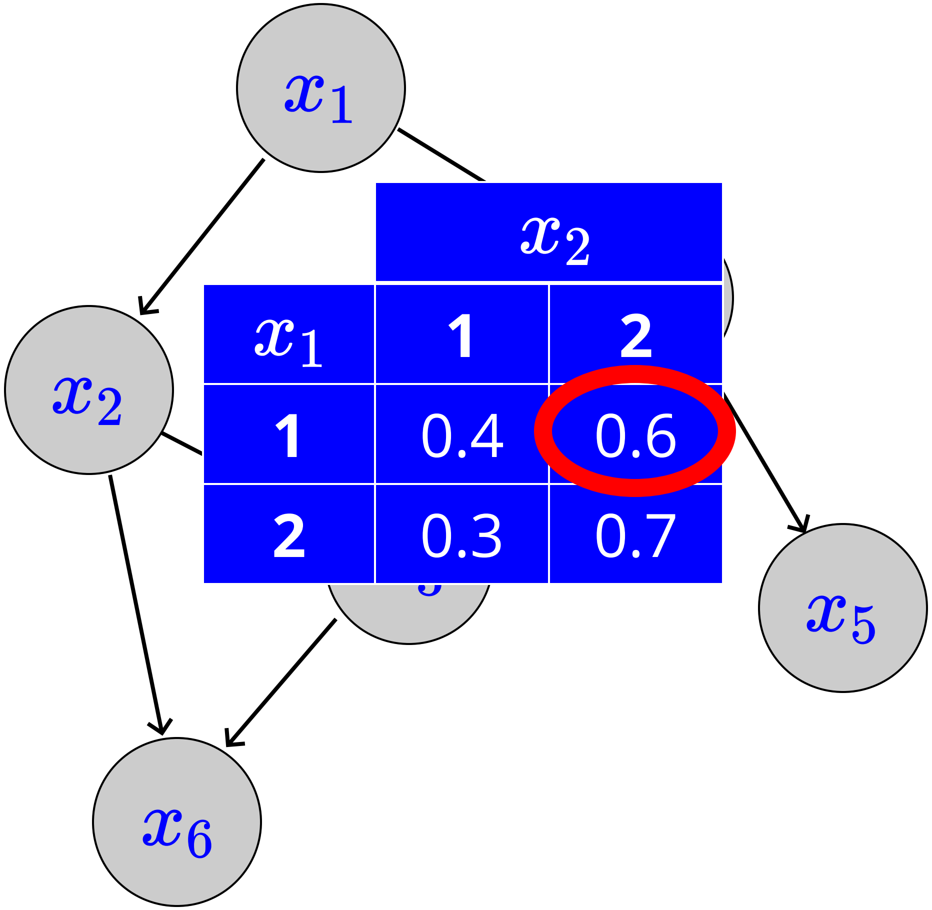


		x_3		
x_2	x_4	1	2	3
1	1	0.4	0.5	0.1
1	2	0.3	0.4	0.3
1	3	0.1	0.8	0.1
2	1	0.7	0.3	0.0
2	2	0.2	0.6	0.2
2	3	0.5	0.3	0.2

$p(x_1, x$

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

Bayesian Networks



	x_2	
x_1	1	2
1	0.4	0.6
2	0.3	0.7

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

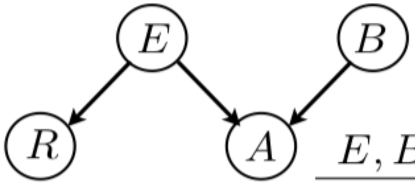
Bayesian Networks

E:

T	F
0.01	0.99

B:

T	F
0.01	0.99



R:

E	T	F
T	1	0
F	0	1

$A|E,B$:

E,B	T	F
T,T	1	0
T,F	1	0
F,T	1	0
F,F	0	1

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

$$P(B=T|A=T) = \frac{P(B=T,A=T)}{\sum_{b\in\{T,F\}} P(B=b,A=T)} = 0.5025\ldots$$

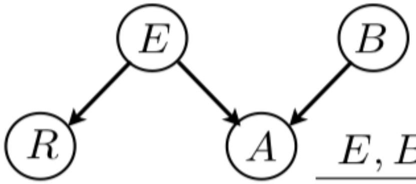
Bayesian Networks

E:

T	F
0.01	0.99

B:

T	F
0.01	0.99



R:

E	T	F
T	1	0
F	0	1

$A|E,B$:

E, B	T	F
T, T	1	0
T, F	1	0
F, T	1	0
F, F	0	1

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

$$P(B=T|A=T,E=T) = \frac{P(B{=}T,A{=}T,E{=}T)}{\sum_{b\in\{T,F\}} P(B{=}b,A{=}T,E{=}T)} = 0.01$$

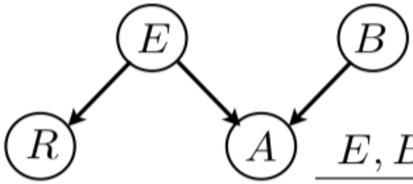
Bayesian Networks

$E:$

T	F
0.01	0.99

$B:$

T	F
0.01	0.99



$R:$

E	T	F
T	1	0
F	0	1

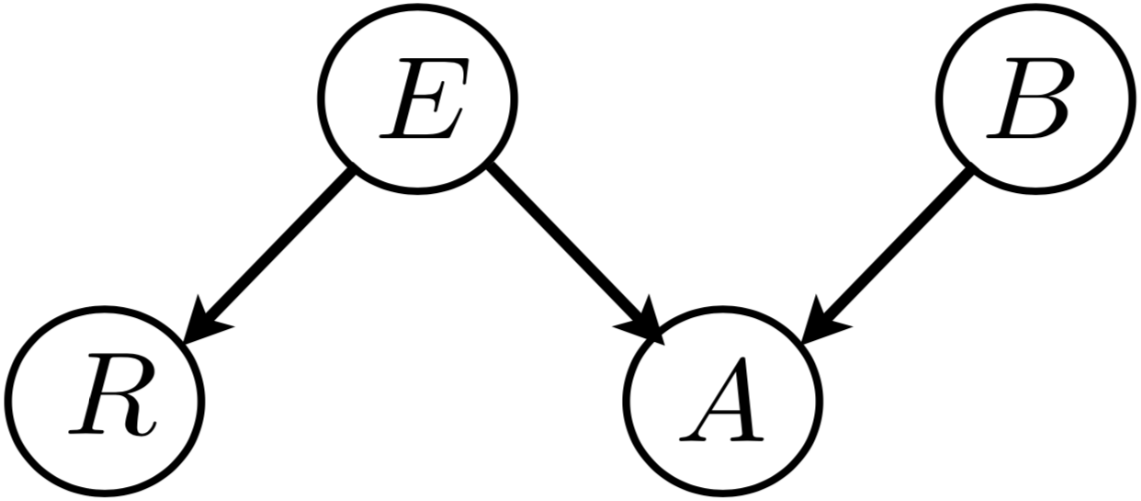
$A|E,B:$

E,B	T	F
T,T	1	0
T,F	1	0
F,T	1	0
F,F	0	1

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

E and B are not conditionally independent given A

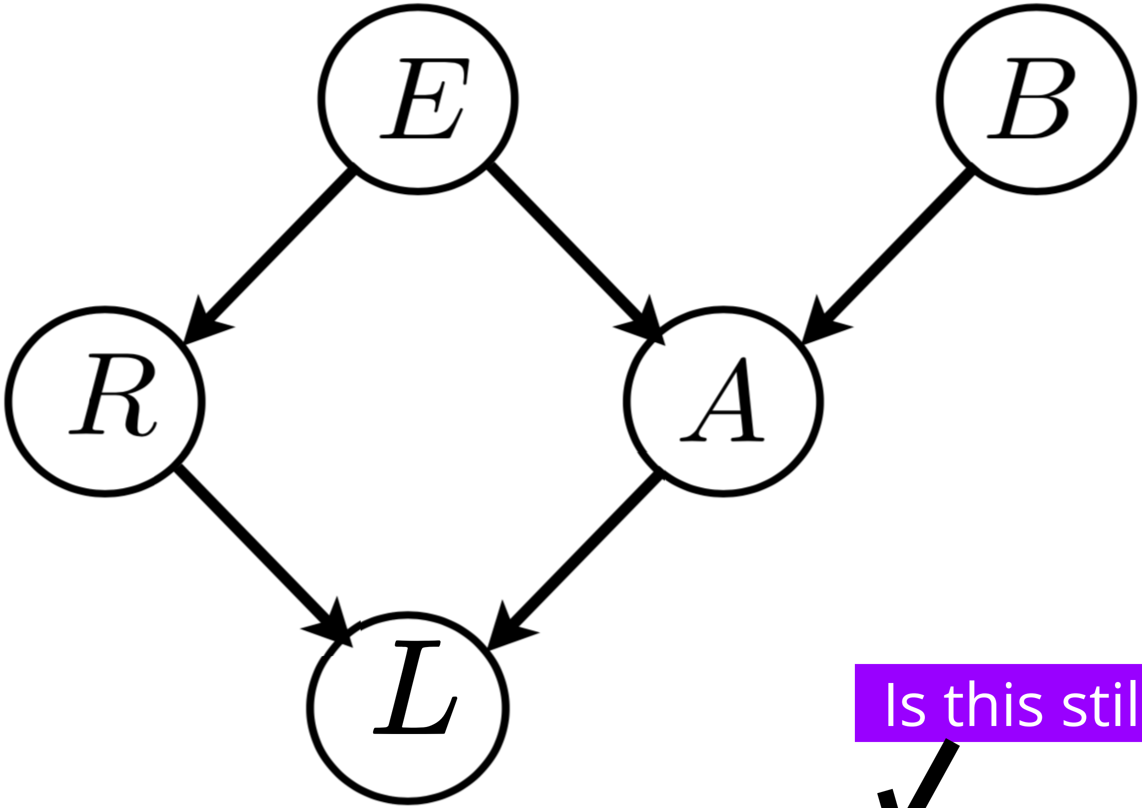
Bayesian Networks



Is this still true?

E and B are not conditionally independent given A

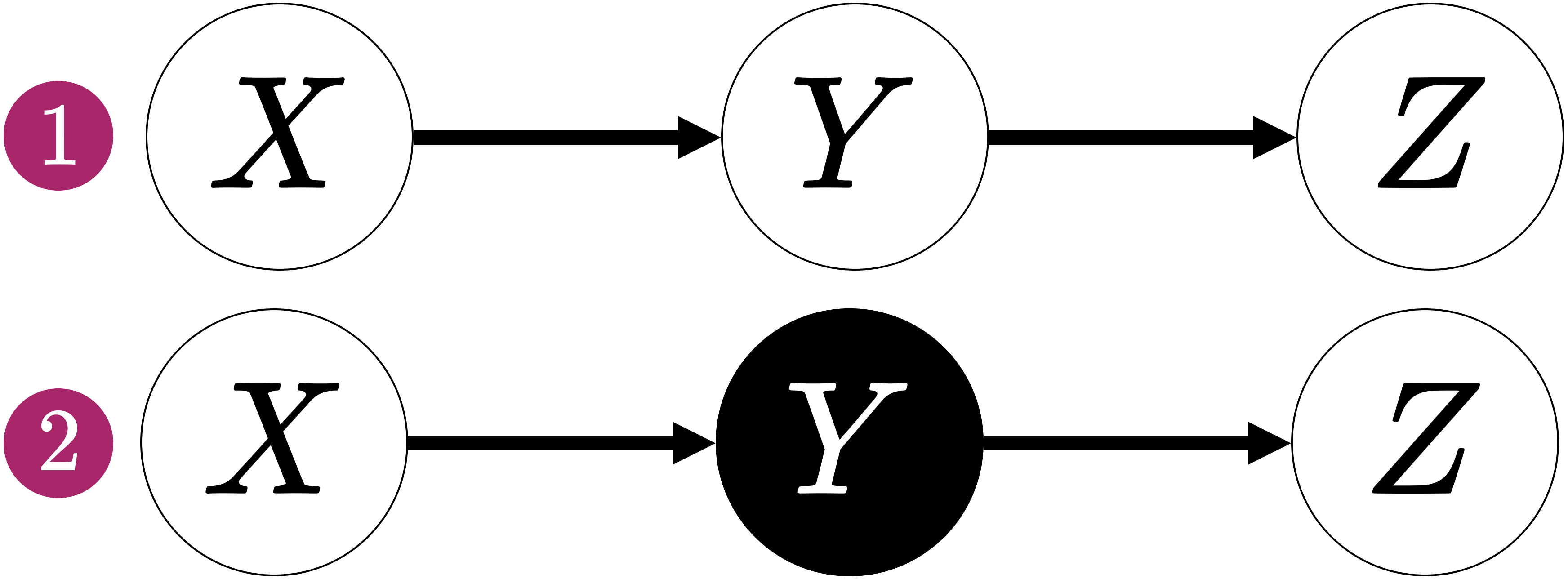
Bayesian Networks



Is this still true?

E and B are not conditionally independent given A

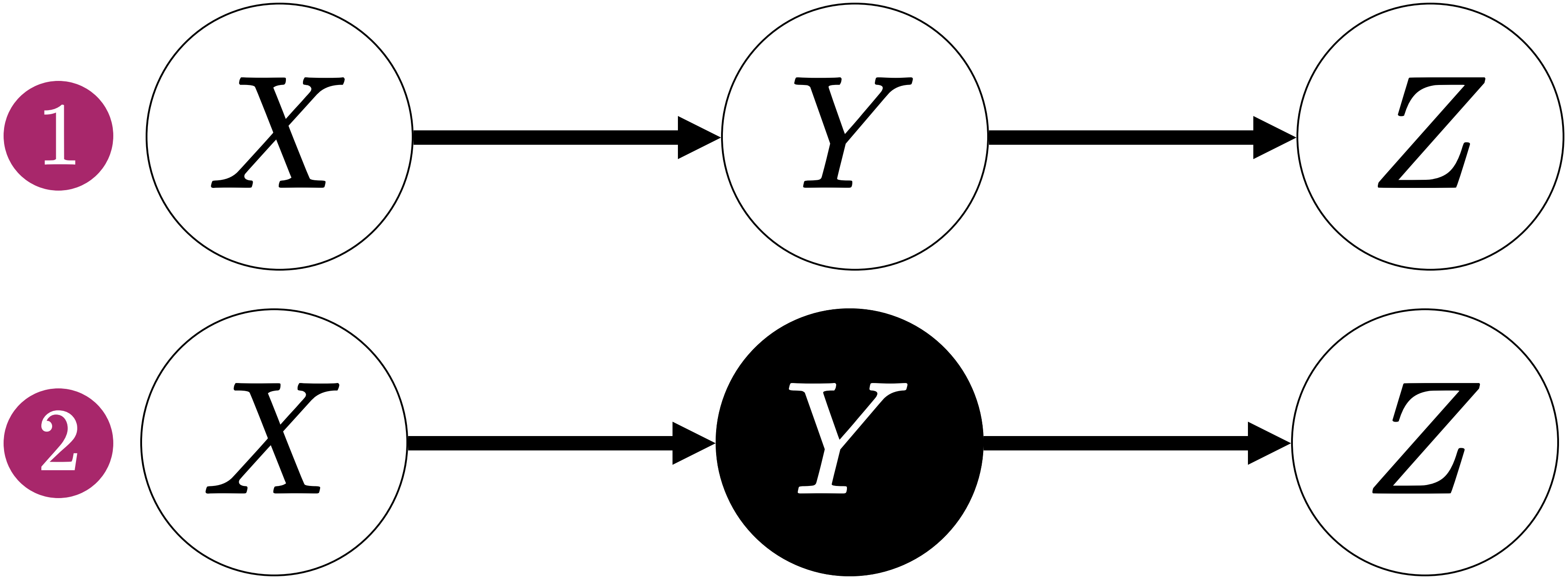
Independence in BN



Chain

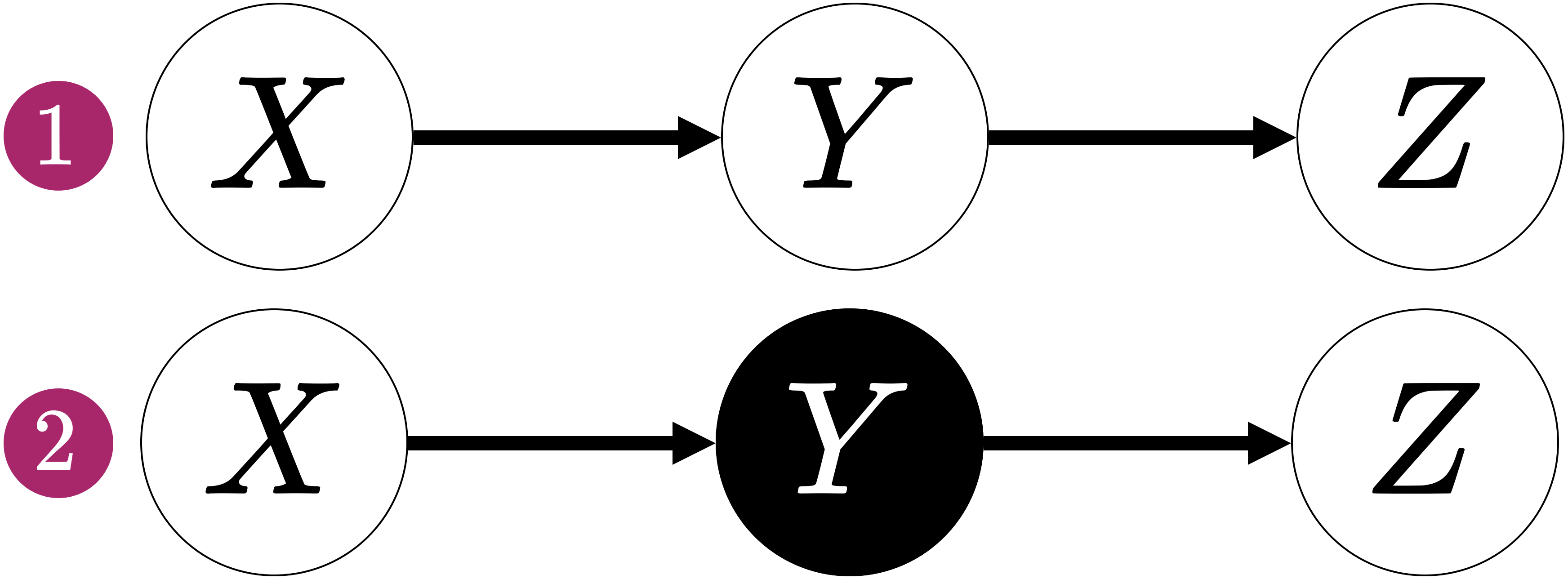
In which case X and Z are independent?

Independence in BN



$$\boxed{P(X,Y,Z) = P(X)P(Y|X)P(Z|Y)}$$

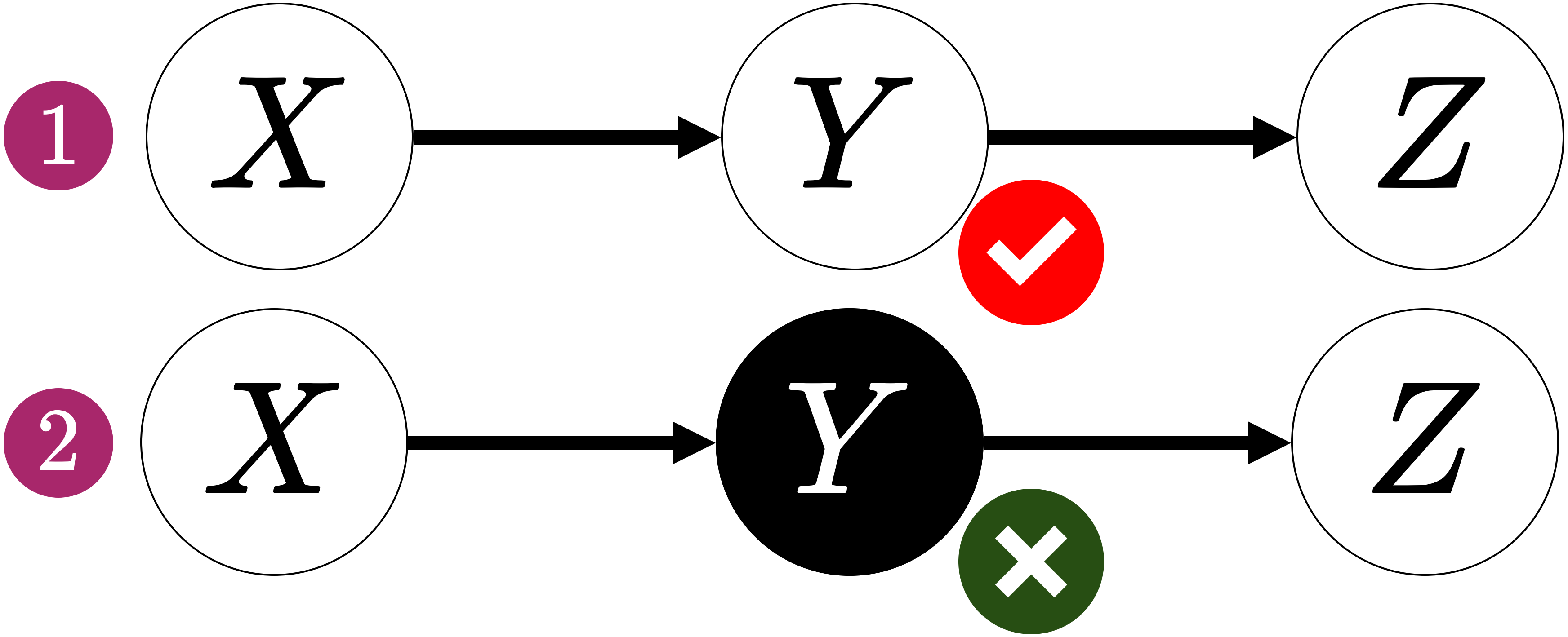
Independence in BN



$$\boxed{P(X,Y,Z) = P(X)P(Y|X)P(Z|Y)}$$

$$P(Z|X,Y) = \frac{P(X,Y,Z)}{P(X,Y)} = \frac{P(X)P(Y|X)P(Z|Y)}{P(X)P(Y|X)} = P(Z|Y)$$

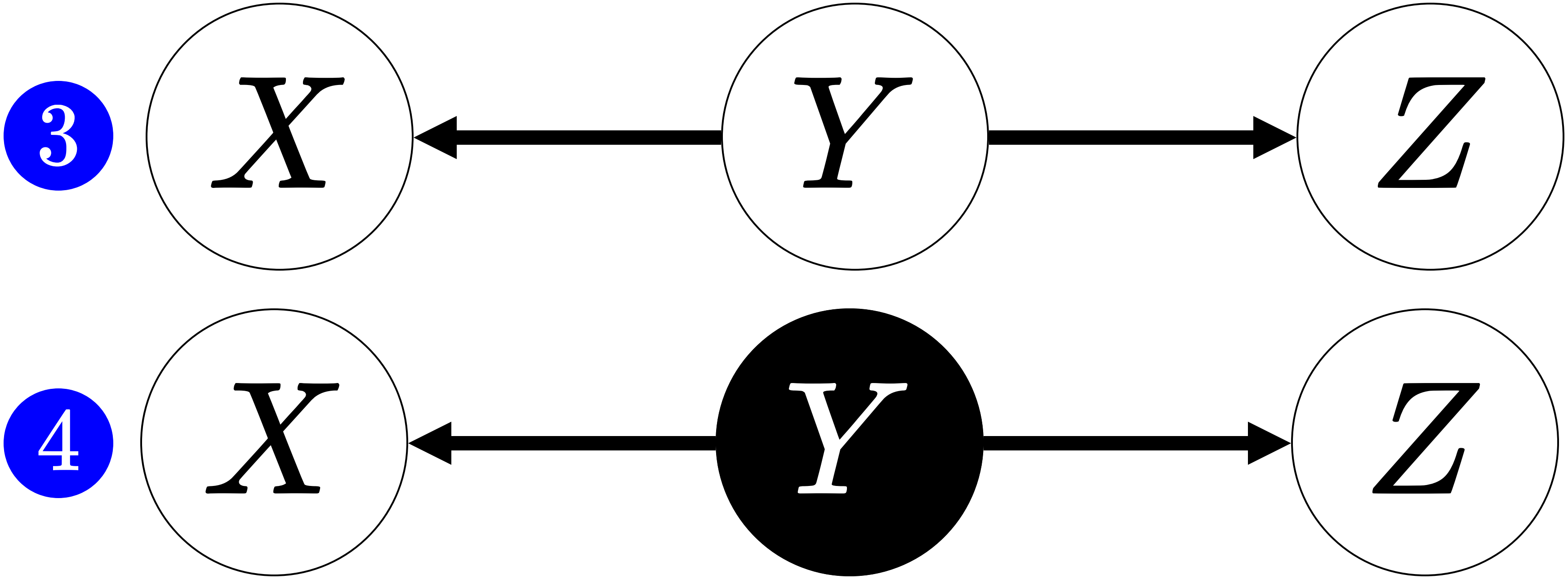
Independence in BN



$$\boxed{P(X,Y,Z) = P(X)P(Y|X)P(Z|Y)}$$

$$P(Z|X,Y) = \frac{P(X,Y,Z)}{P(X,Y)} = \frac{P(X)P(Y|X)P(Z|Y)}{P(X)P(Y|X)} = P(Z|Y)$$

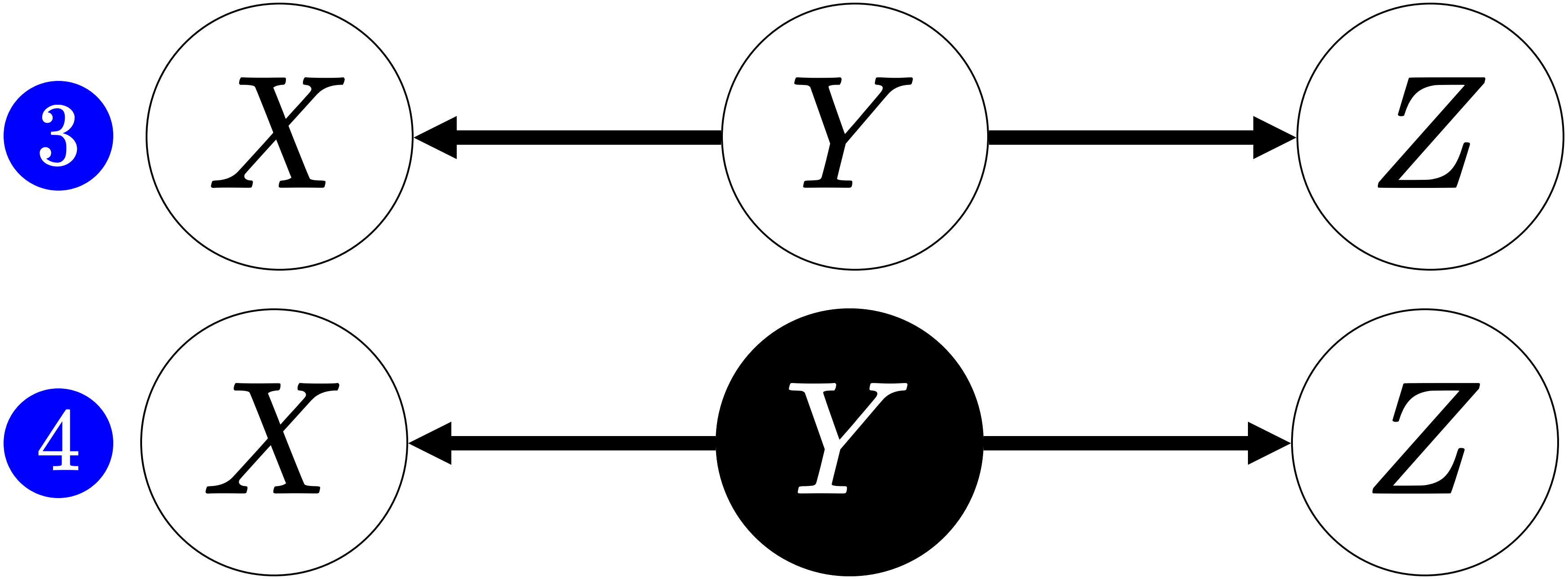
Independence in BN



Common Cause

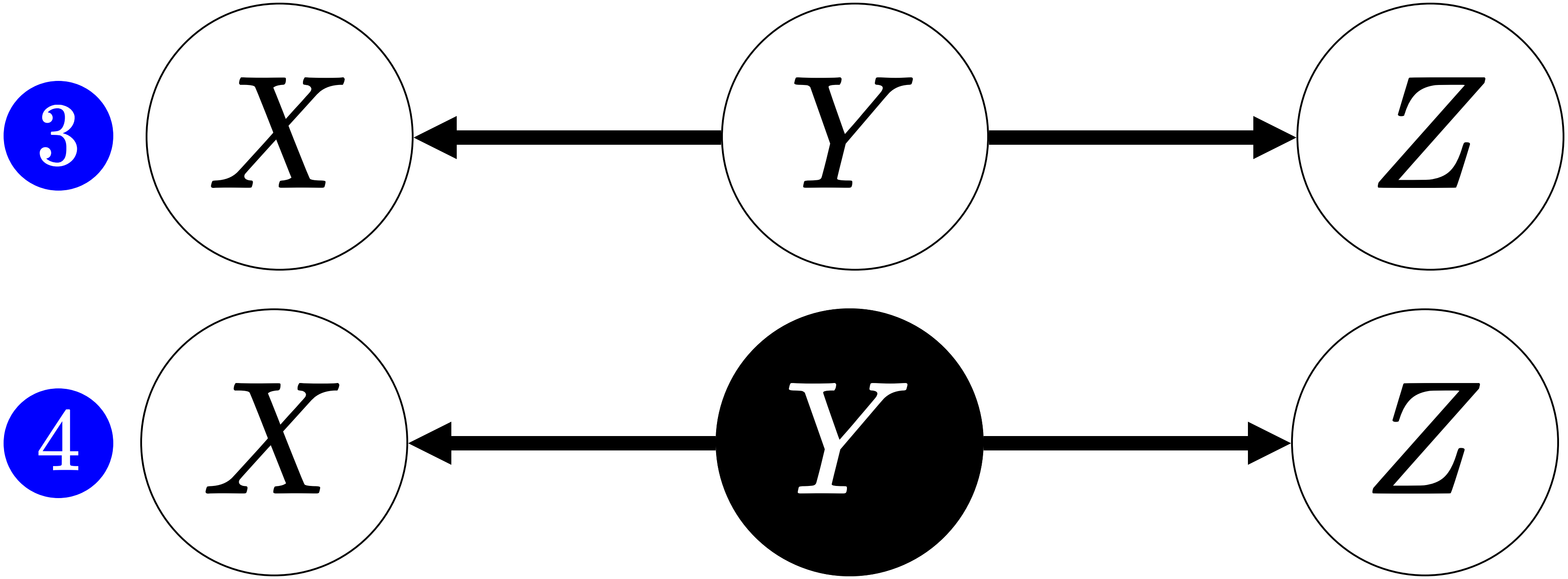
In which case X and Z are independent?

Independence in BN



$$\boxed{P(X,Y,Z) = P(Y)P(X|Y)P(Z|Y)}$$

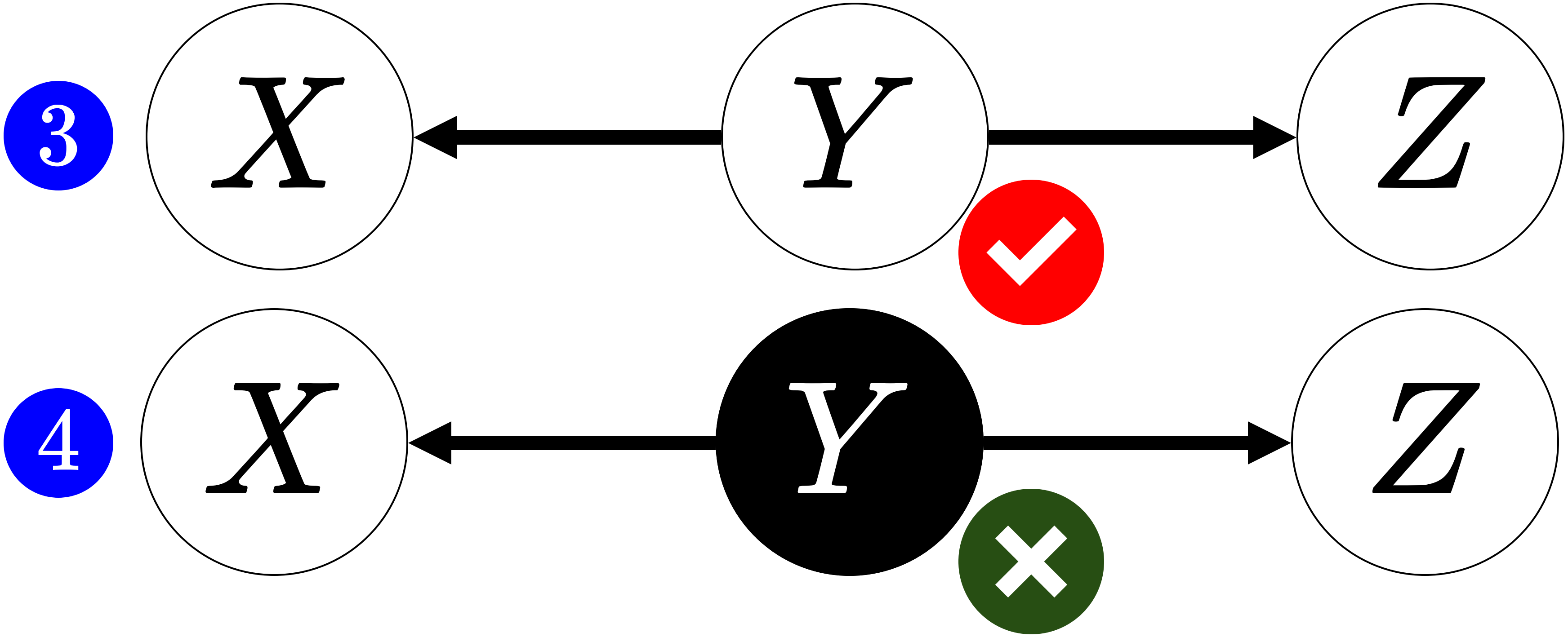
Independence in BN



$$\boxed{P(X,Y,Z) = P(Y)P(X|Y)P(Z|Y)}$$

$$P(Z|X,Y) = \frac{P(X,Y,Z)}{P(X,Y)} = \frac{P(Y)P(X|Y)P(Z|Y)}{P(Y)P(X|Y)} = P(Z|Y)$$

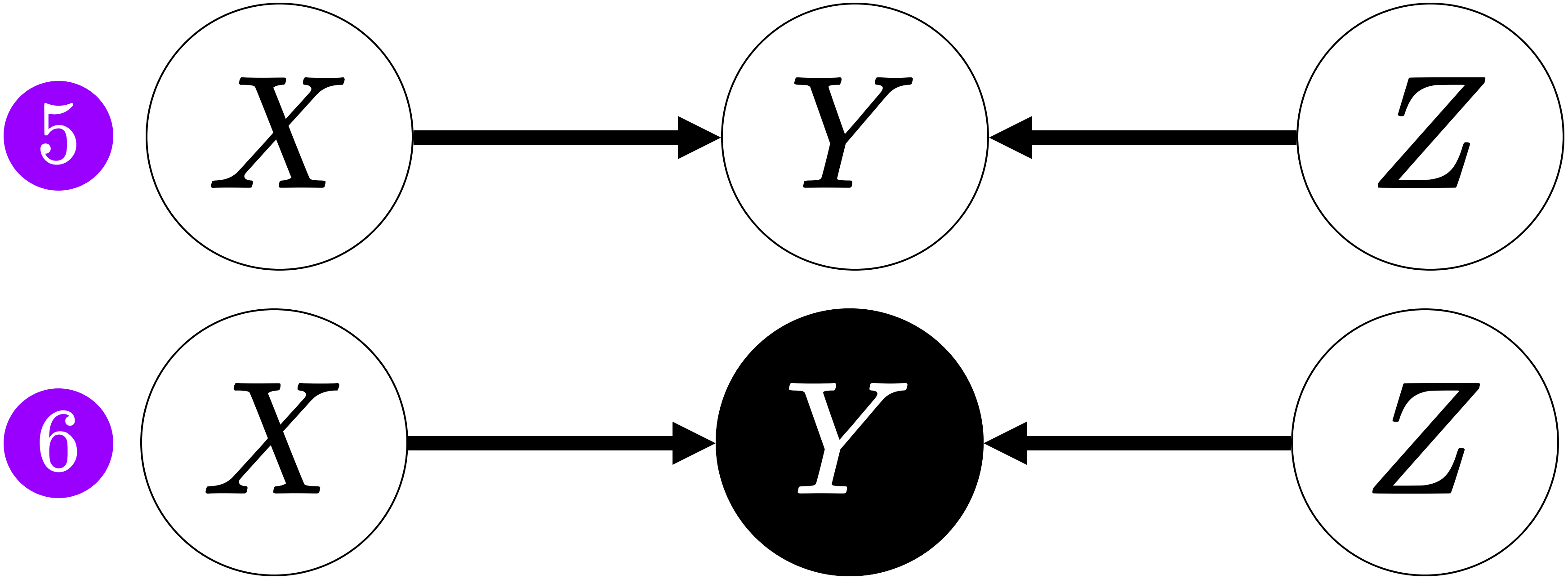
Independence in BN



$$\boxed{P(X,Y,Z) = P(Y)P(X|Y)P(Z|Y)}$$

$$P(Z|X,Y) = \frac{P(X,Y,Z)}{P(X,Y)} = \frac{P(Y)P(X|Y)P(Z|Y)}{P(Y)P(X|Y)} = P(Z|Y)$$

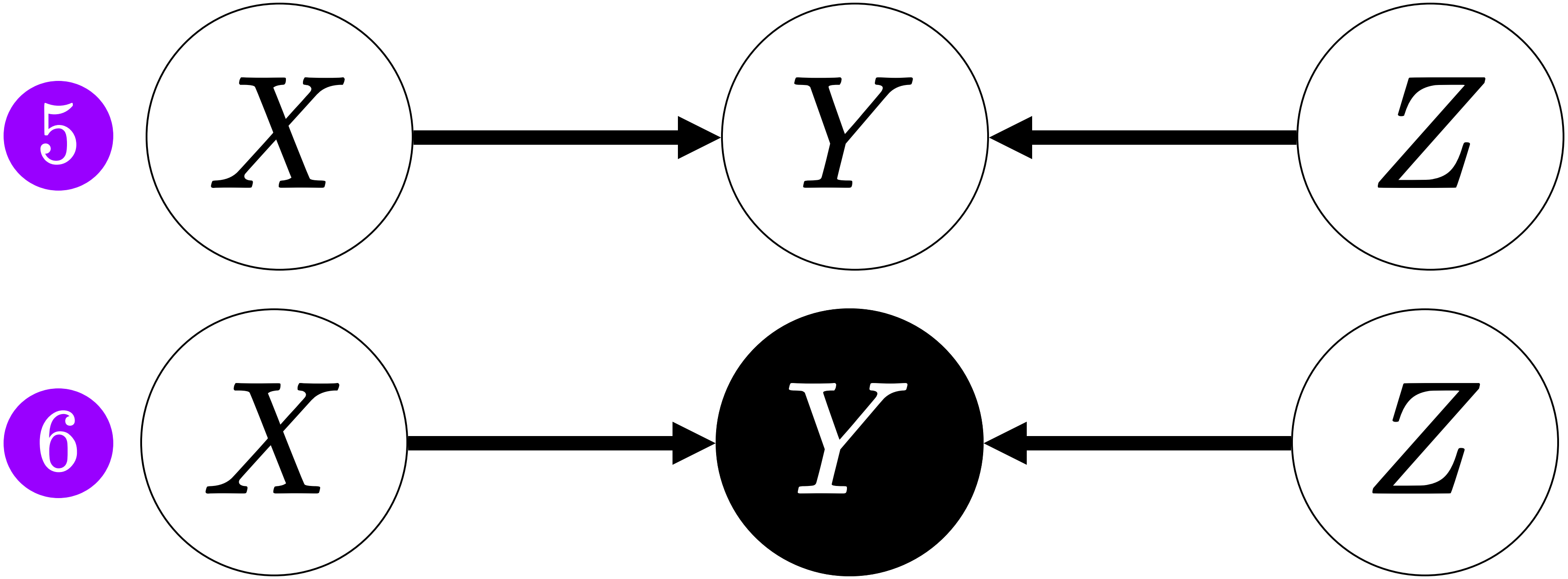
Independence in BN



Explaining Away

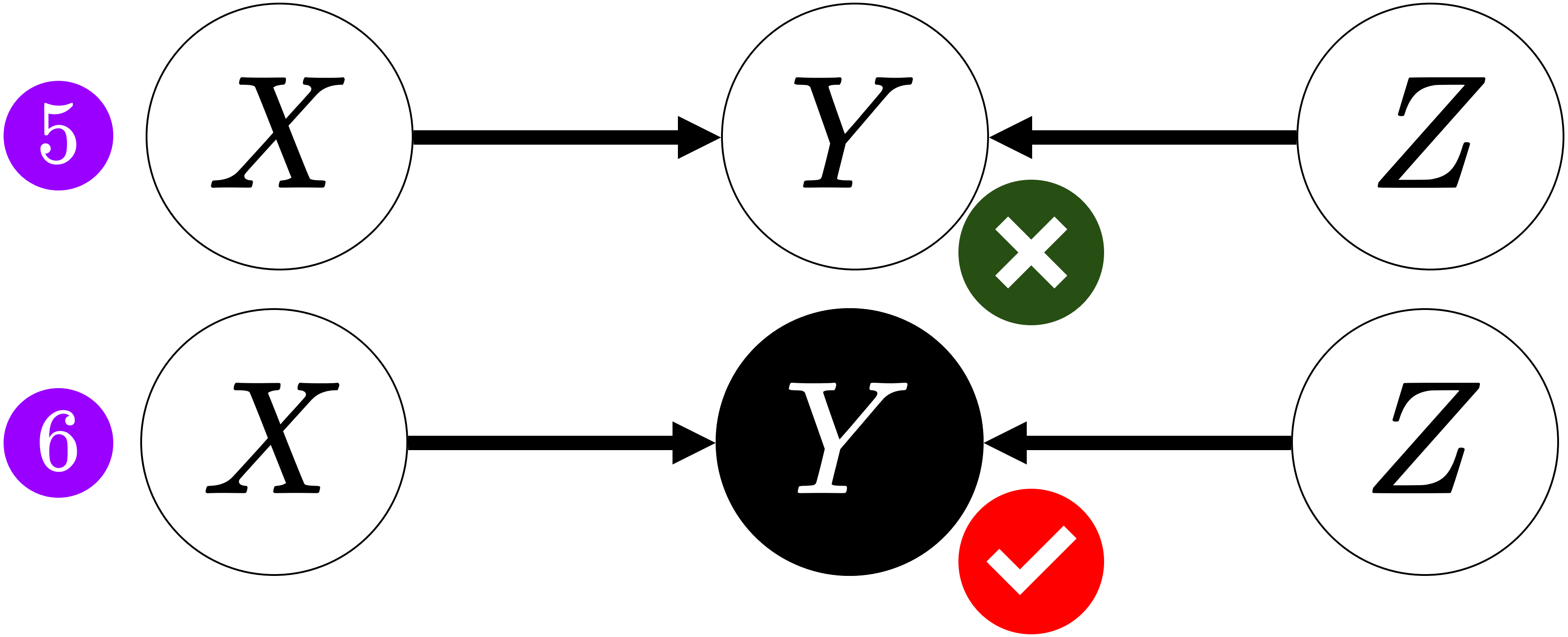
In which case X and Z are independent?

Independence in BN



$$\boxed{P(X, Y, Z) = P(X)P(Z)P(Y|X, Z)}$$

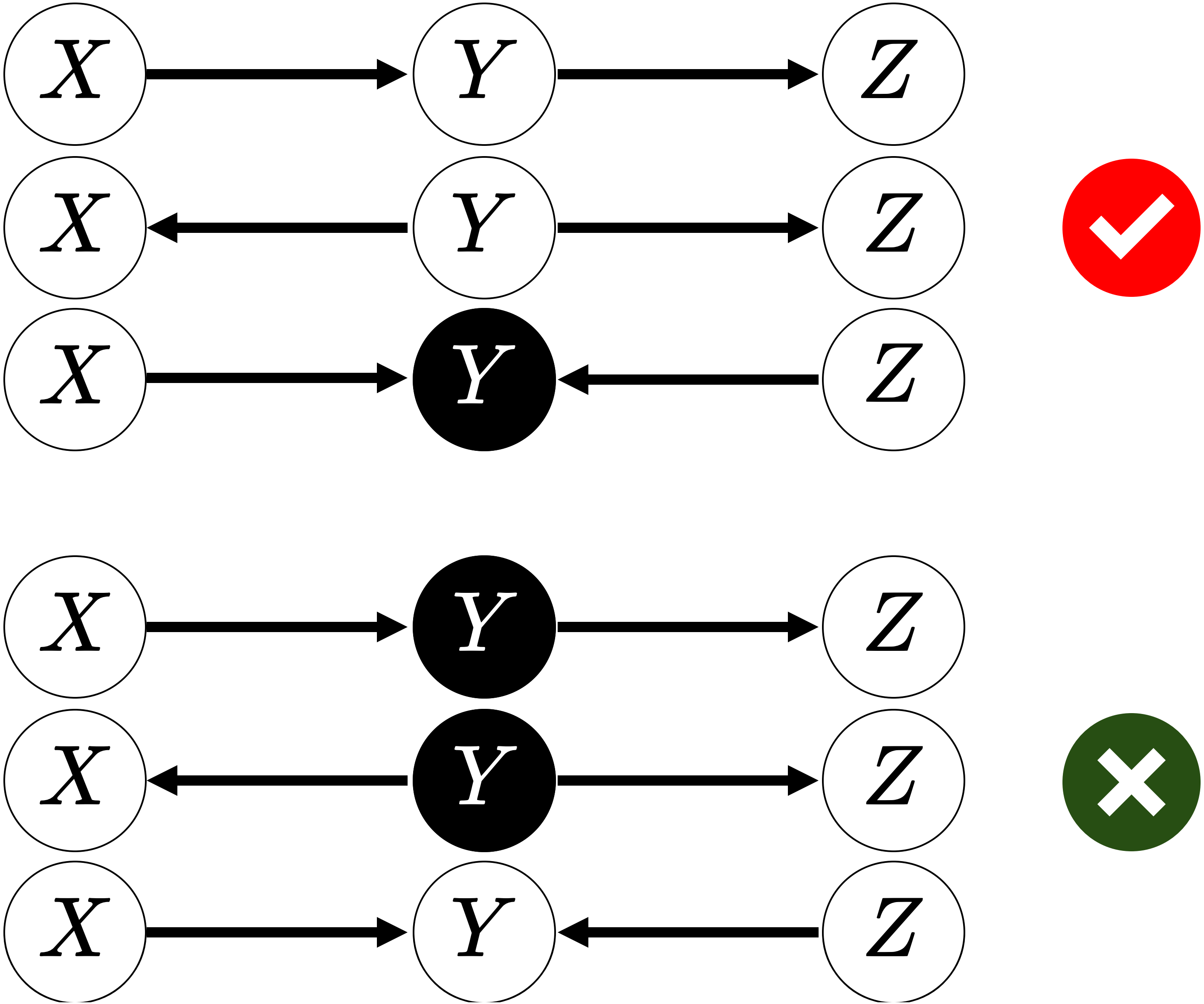
Independence in BN



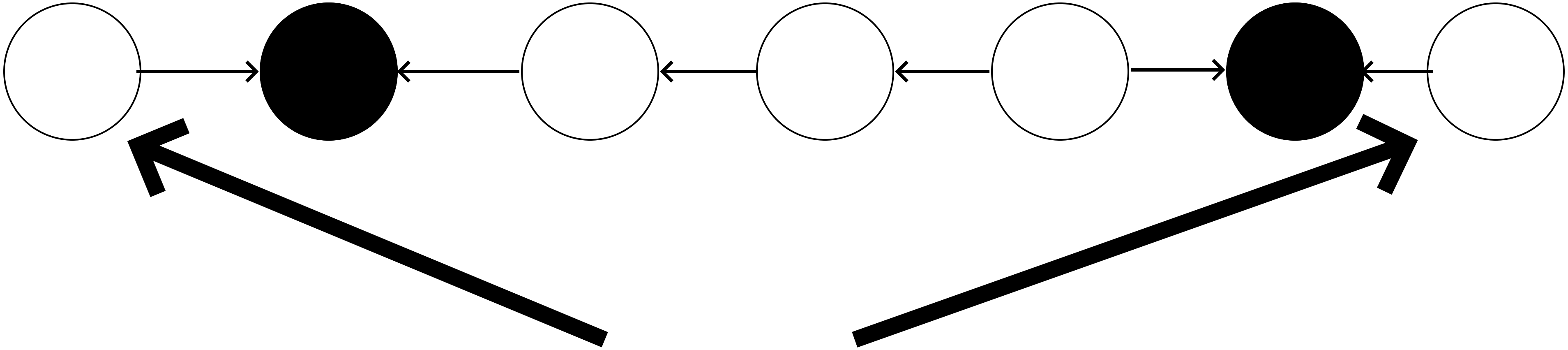
$$\boxed{P(X,Y,Z) = P(X)P(Z)P(Y|X,Z)}$$

$$P(X,Z) = \sum_Y P(X)P(Z)P(Y|X,Z) =$$
$$P(X)P(Z)\sum_Y P(Y|X,Z) = P(X)P(Z)$$

Independence in BN

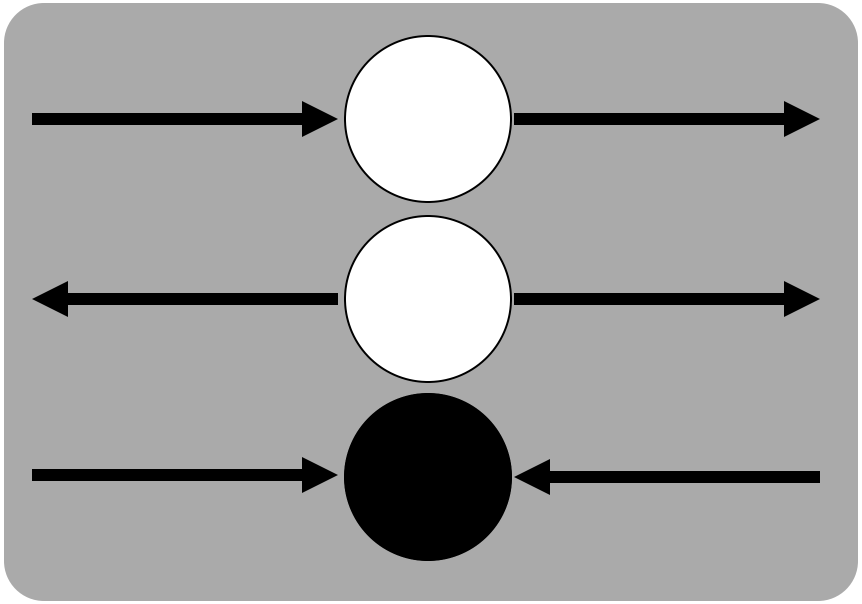


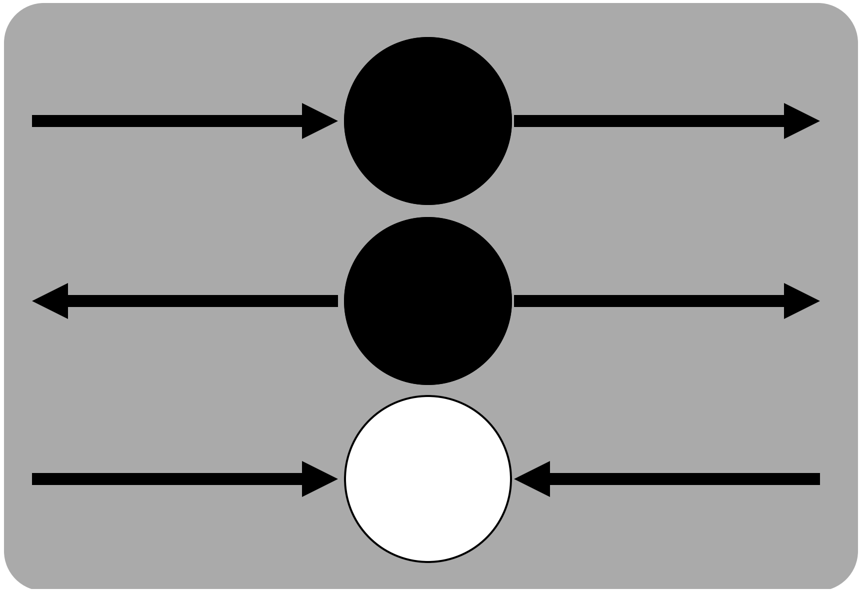
Independence in BN



Are they independent?

Independence in BN

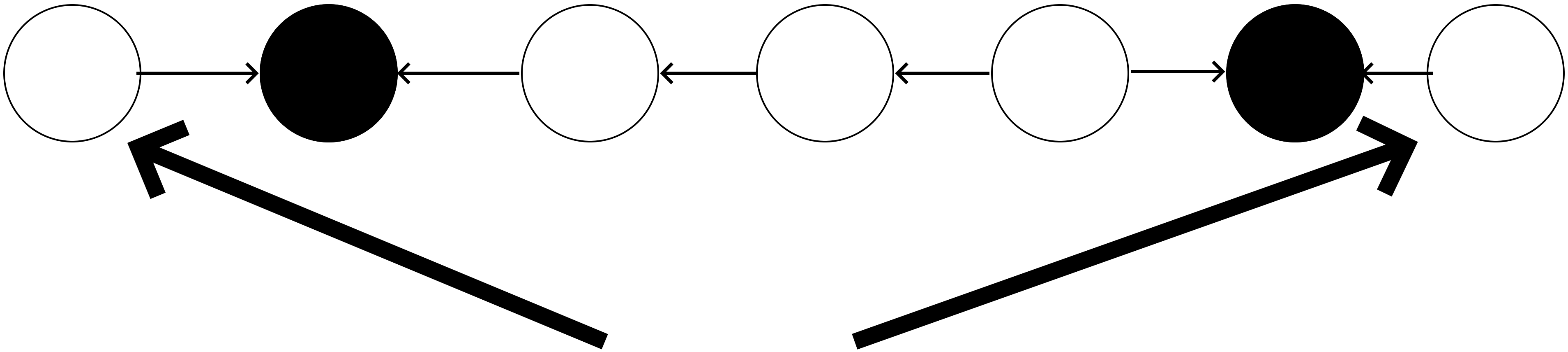




Open Gates

Closed Gates

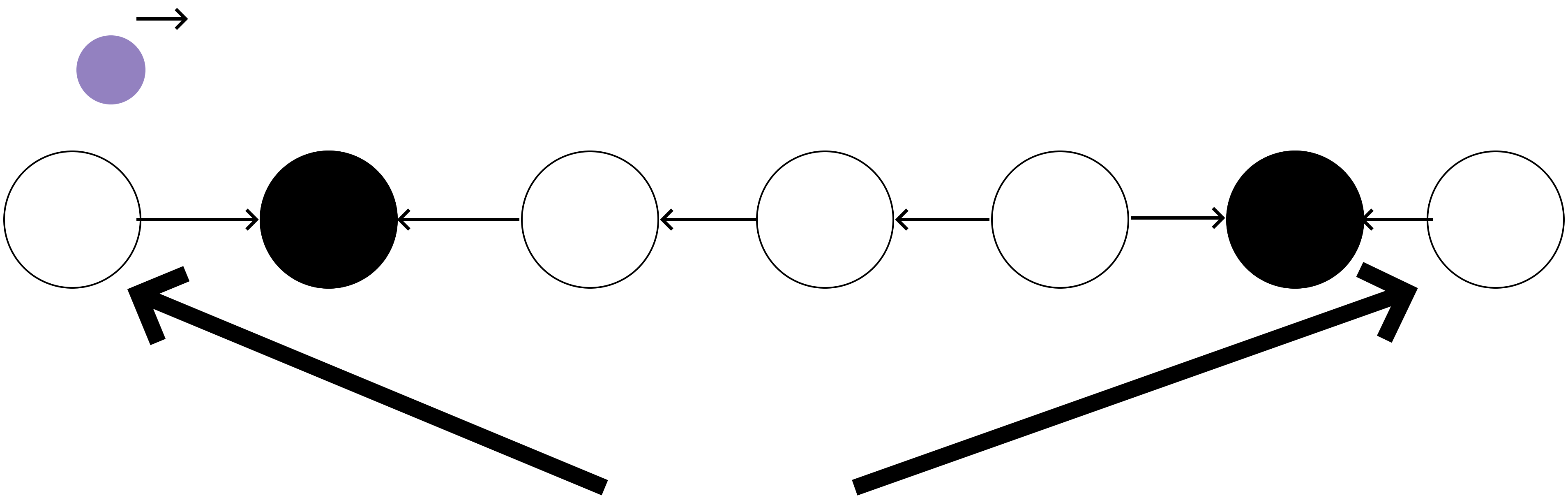
Independence in BN



Are they independent?

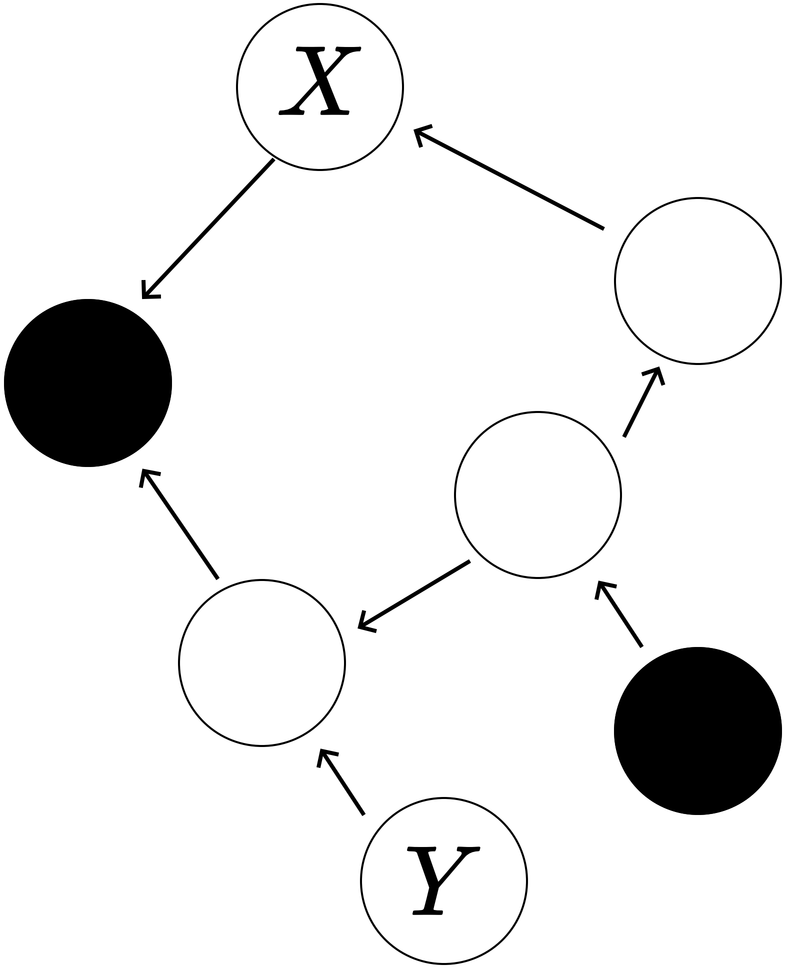
Bayes' Ball Algorithm

Can the ball reach the other node from one node (passing through various gates along the path)? If there is such a path where all gates are open, then they are not independent. Otherwise they are independent.



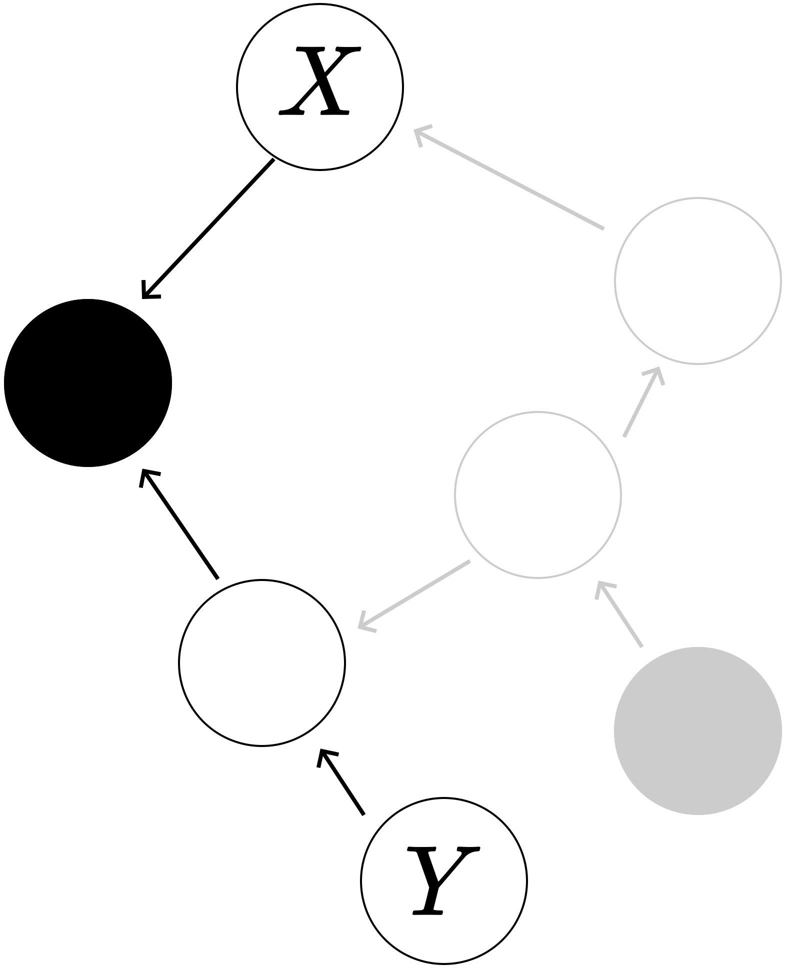
Are they independent?

Bayes' Ball Algorithm



Are X and Y independent?

Bayes' Ball Algorithm

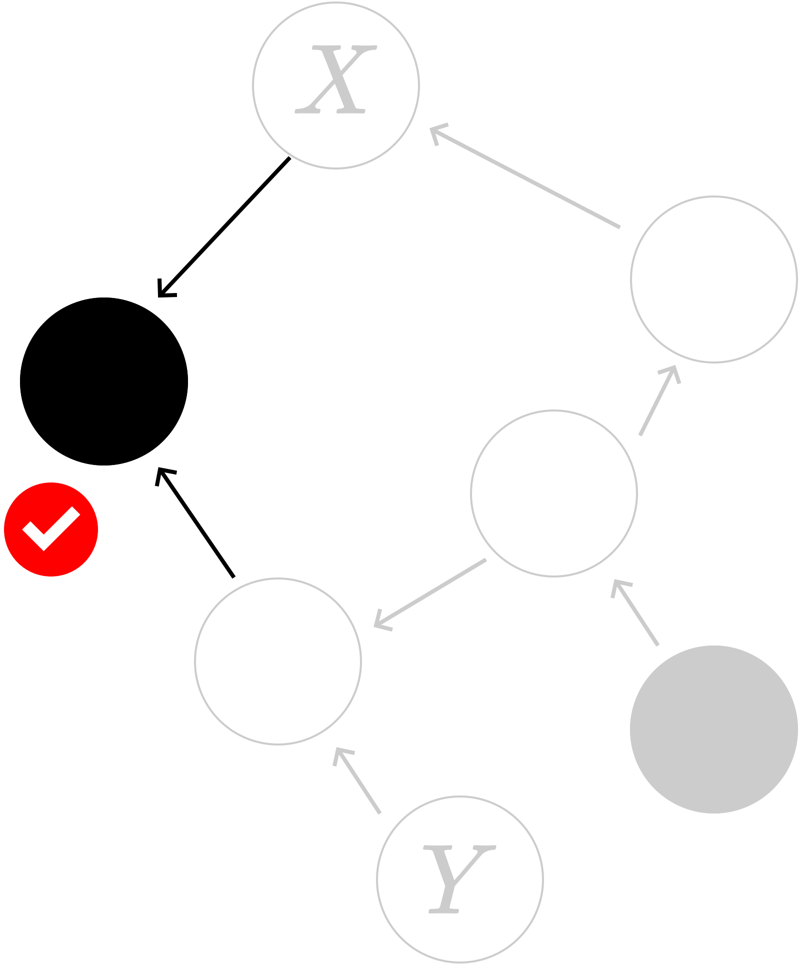


Open Path

Are X and Y independent?

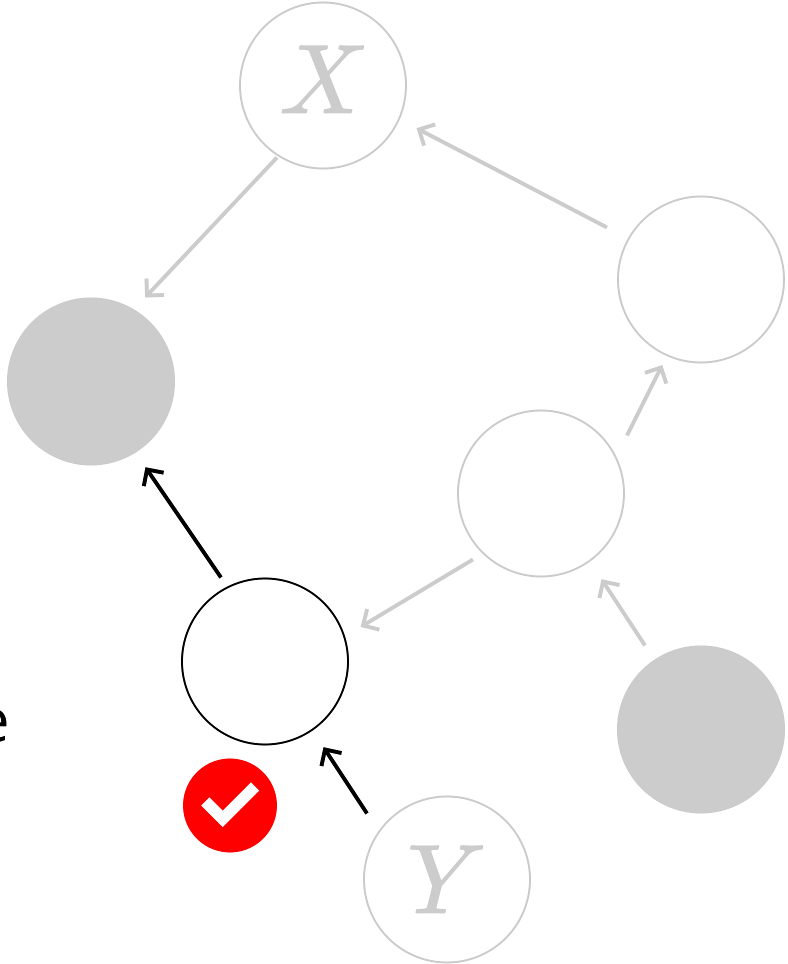
Bayes' Ball Algorithm

Open Gate



Are X and Y independent?

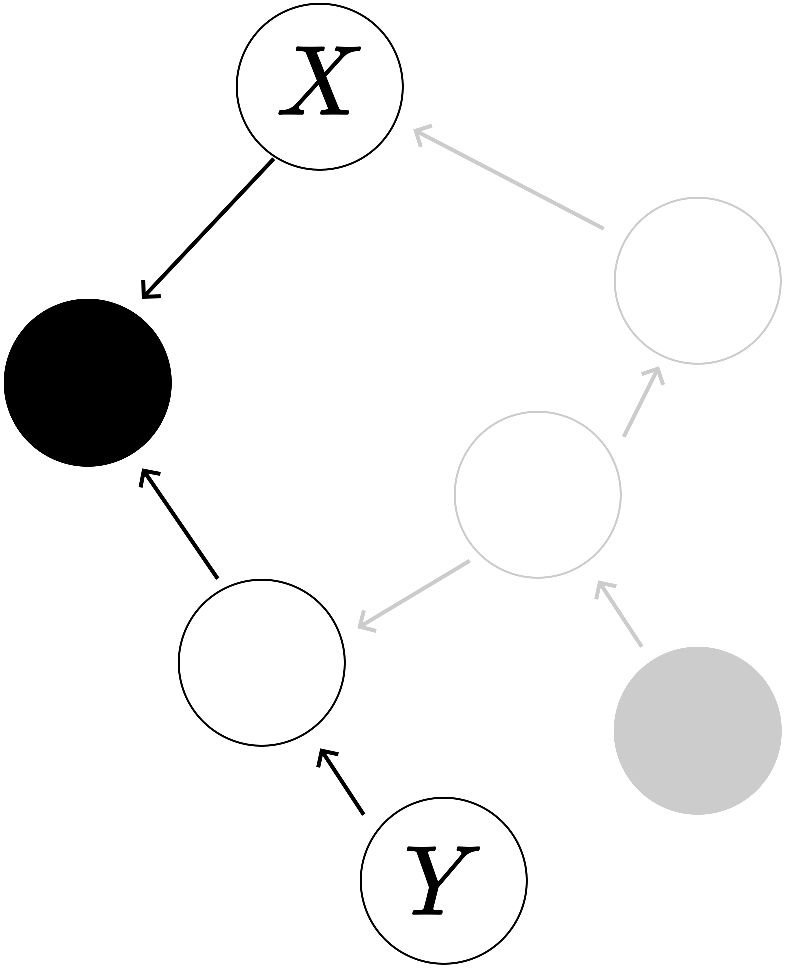
Bayes' Ball Algorithm



Open Gate

Are X and Y independent?

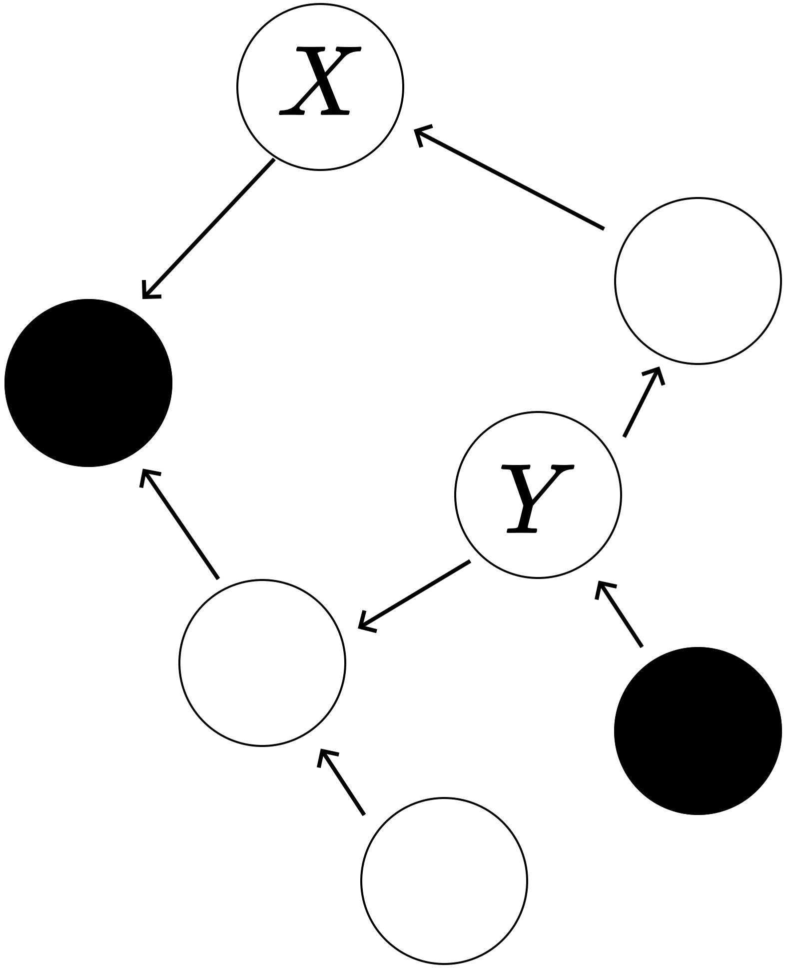
Bayes' Ball Algorithm



Open Path

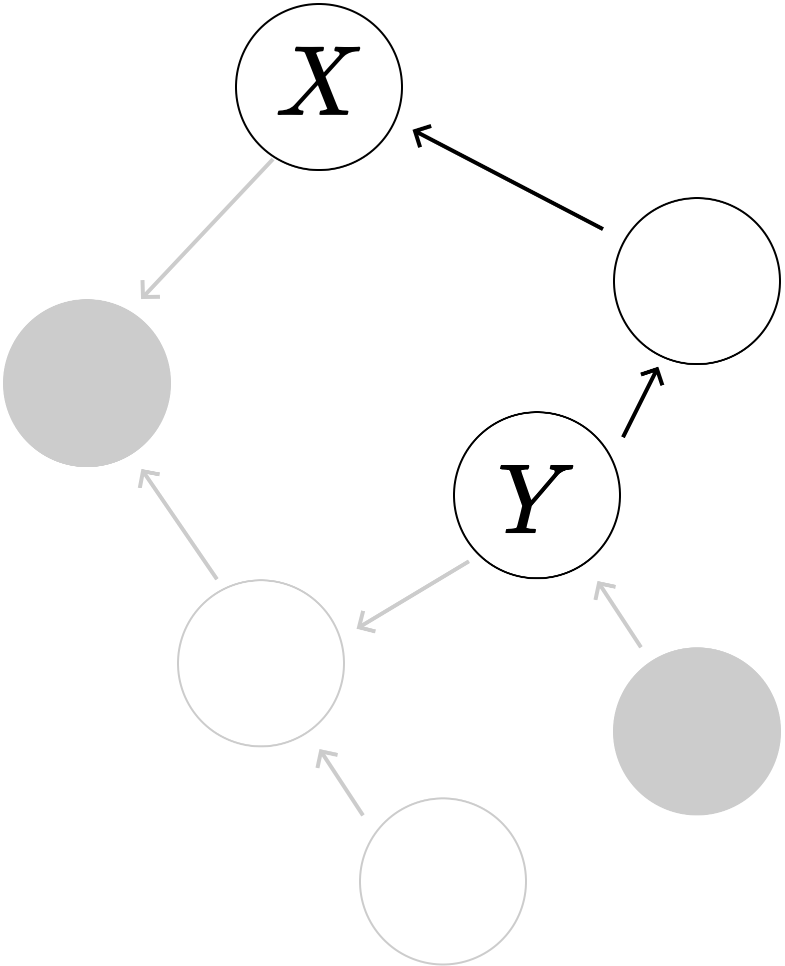
Are X and Y independent?

Bayes' Ball Algorithm



Are X and Y independent?

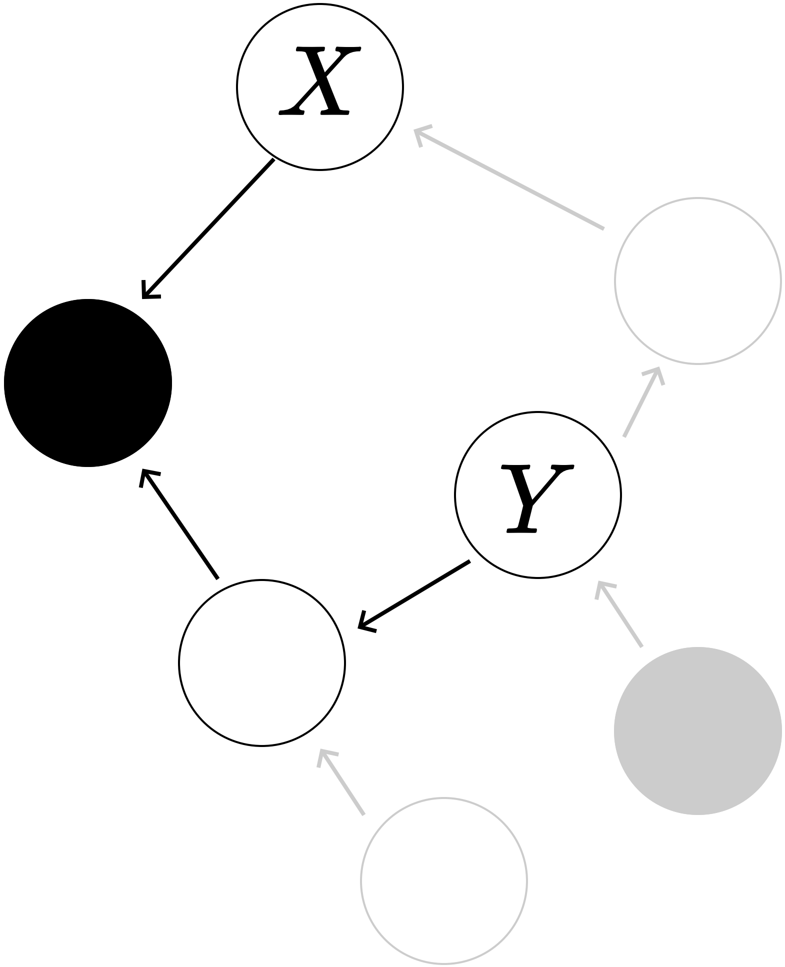
Bayes' Ball Algorithm



Open Path

Are X and Y independent?

Bayes' Ball Algorithm

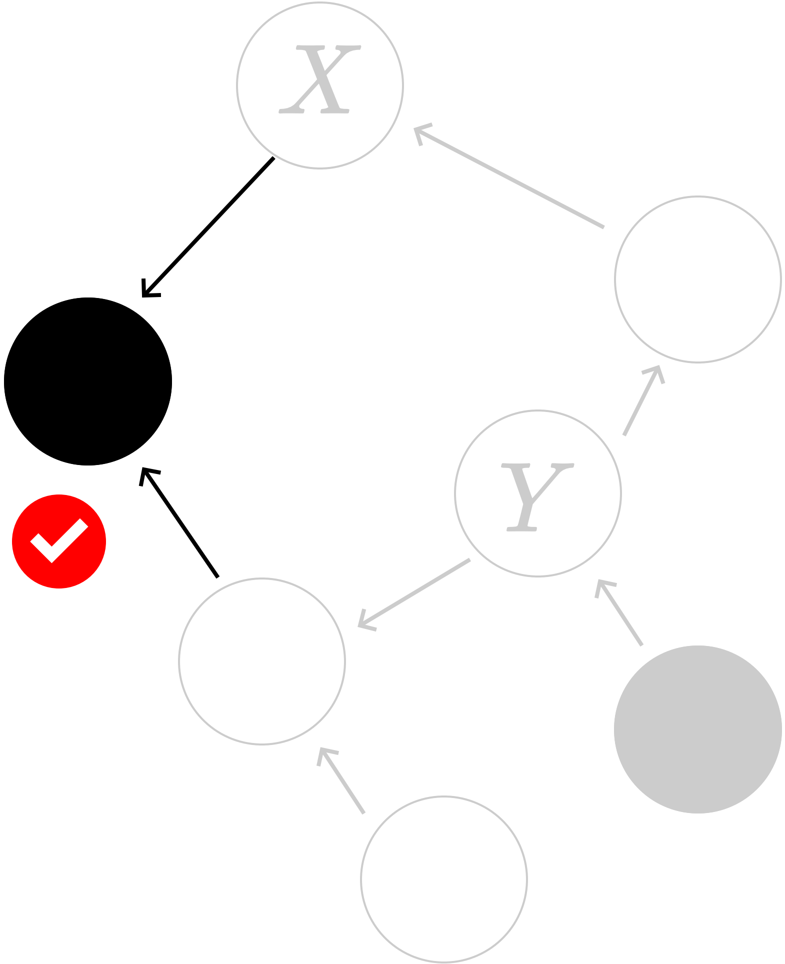


Open Path

Are X and Y independent?

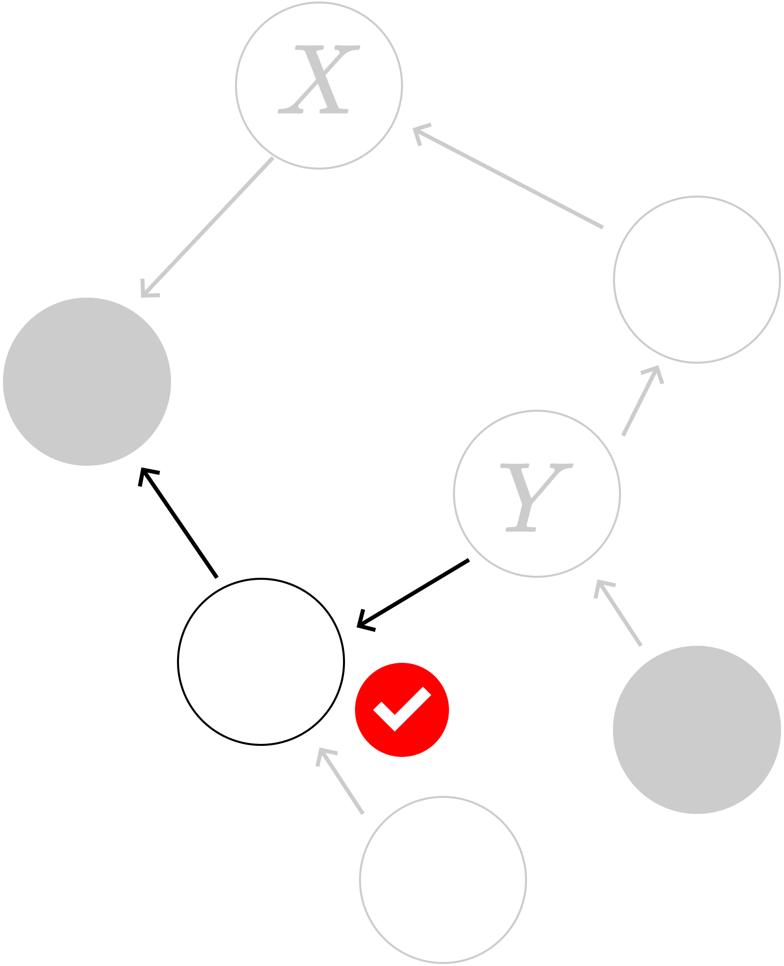
Bayes' Ball Algorithm

Open Gate



Are X and Y independent?

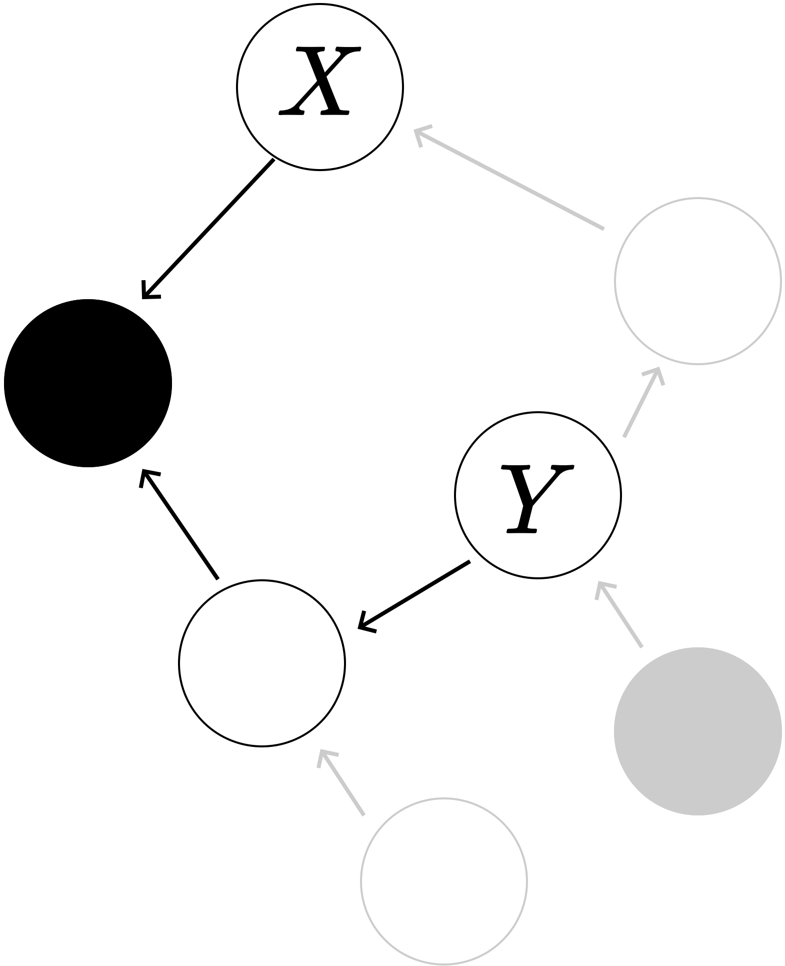
Bayes' Ball Algorithm



Open Gate

Are X and Y independent?

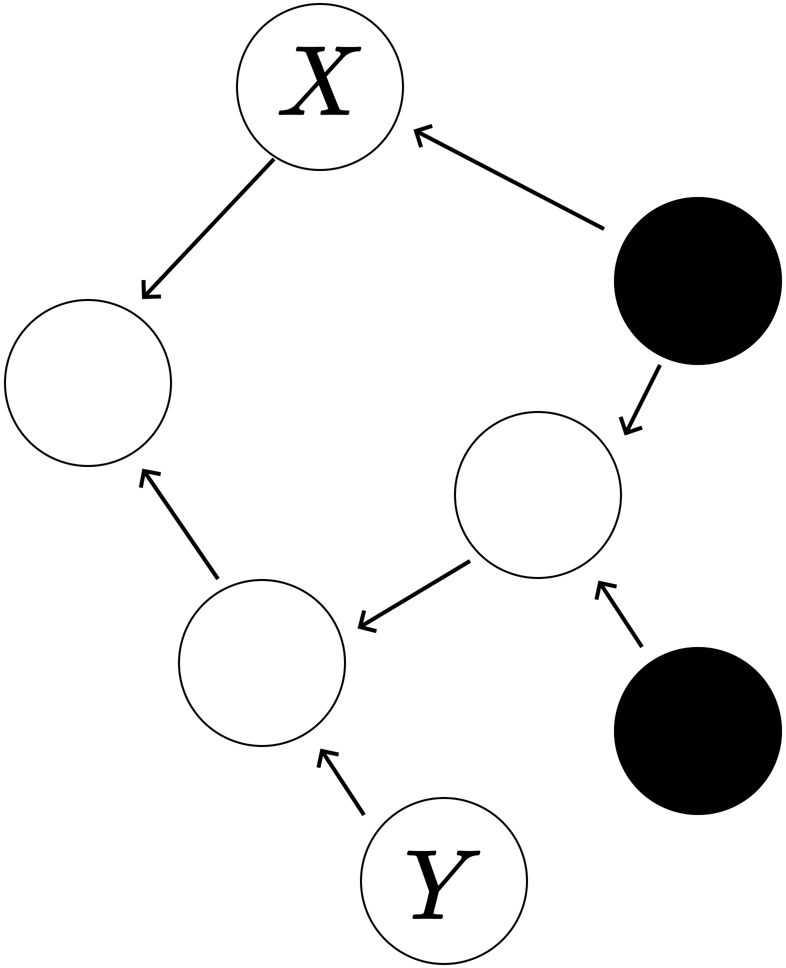
Bayes' Ball Algorithm



Open Path

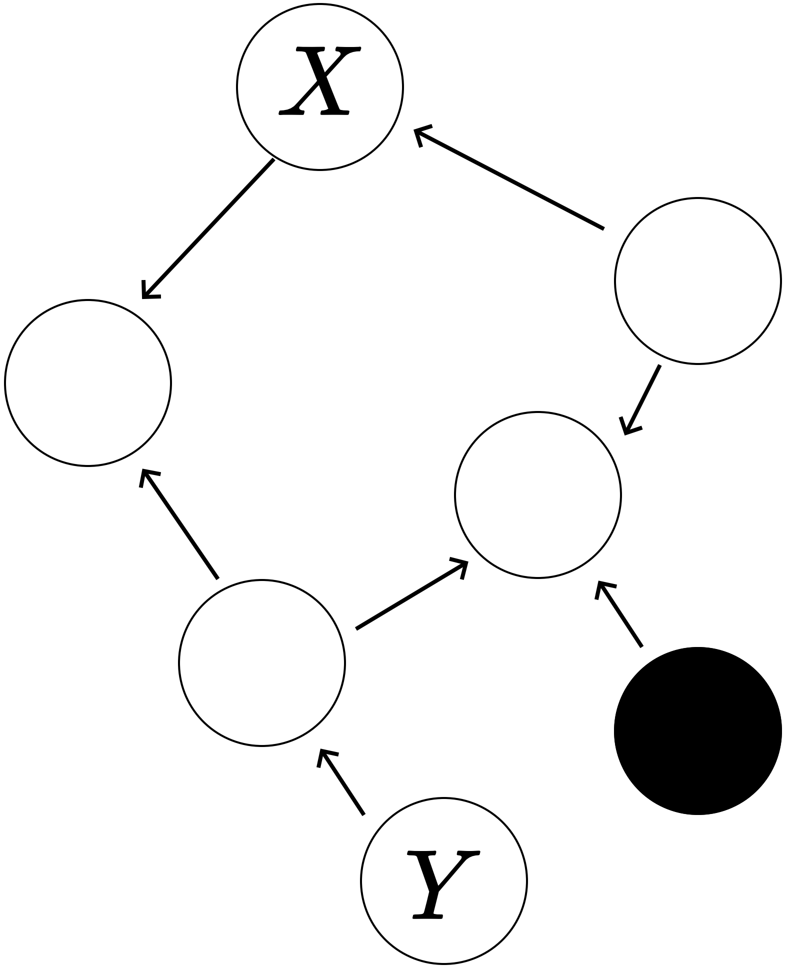
Are X and Y independent?

Bayes' Ball Algorithm



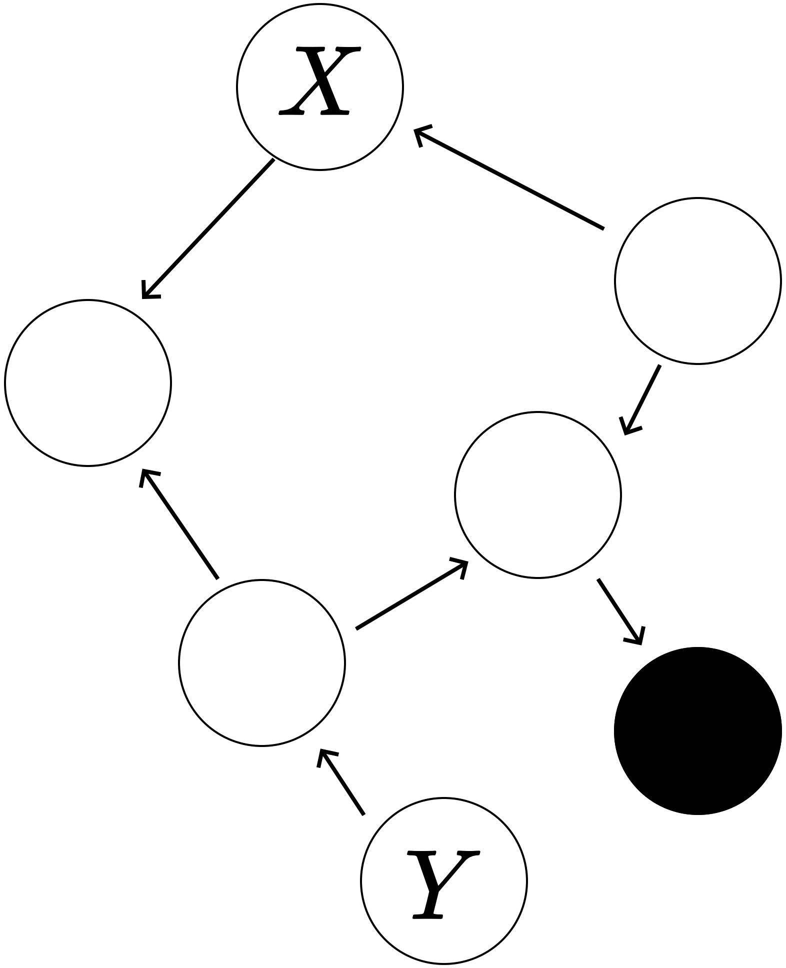
Are X and Y independent?

Bayes' Ball Algorithm



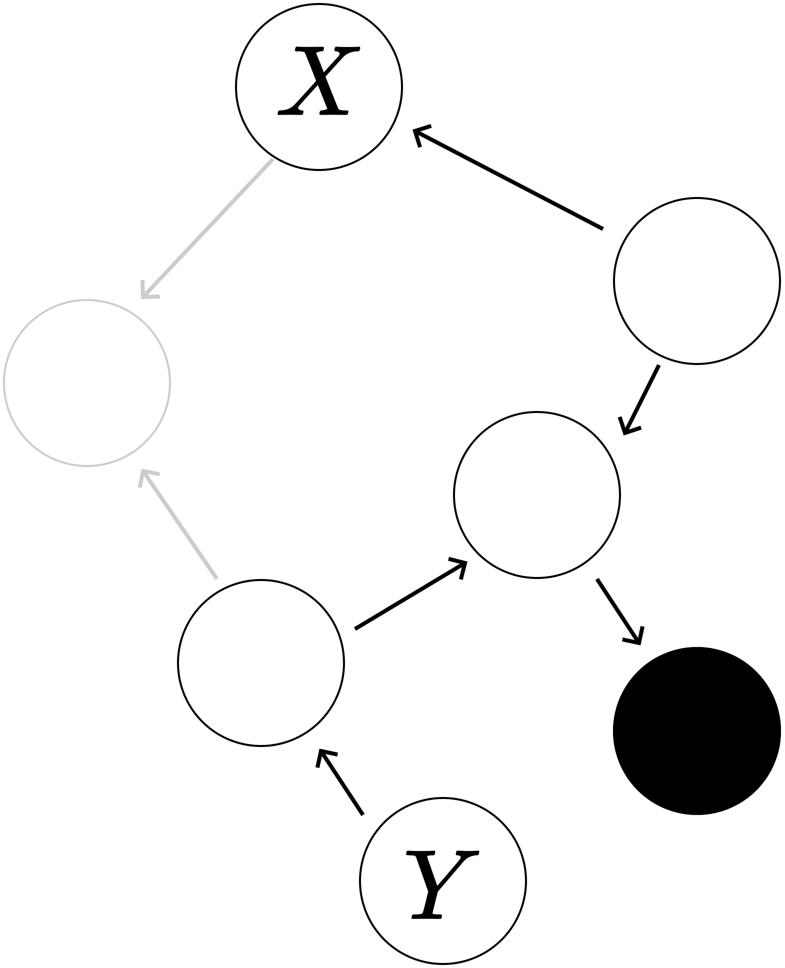
Are X and Y independent?

Bayes' Ball Algorithm



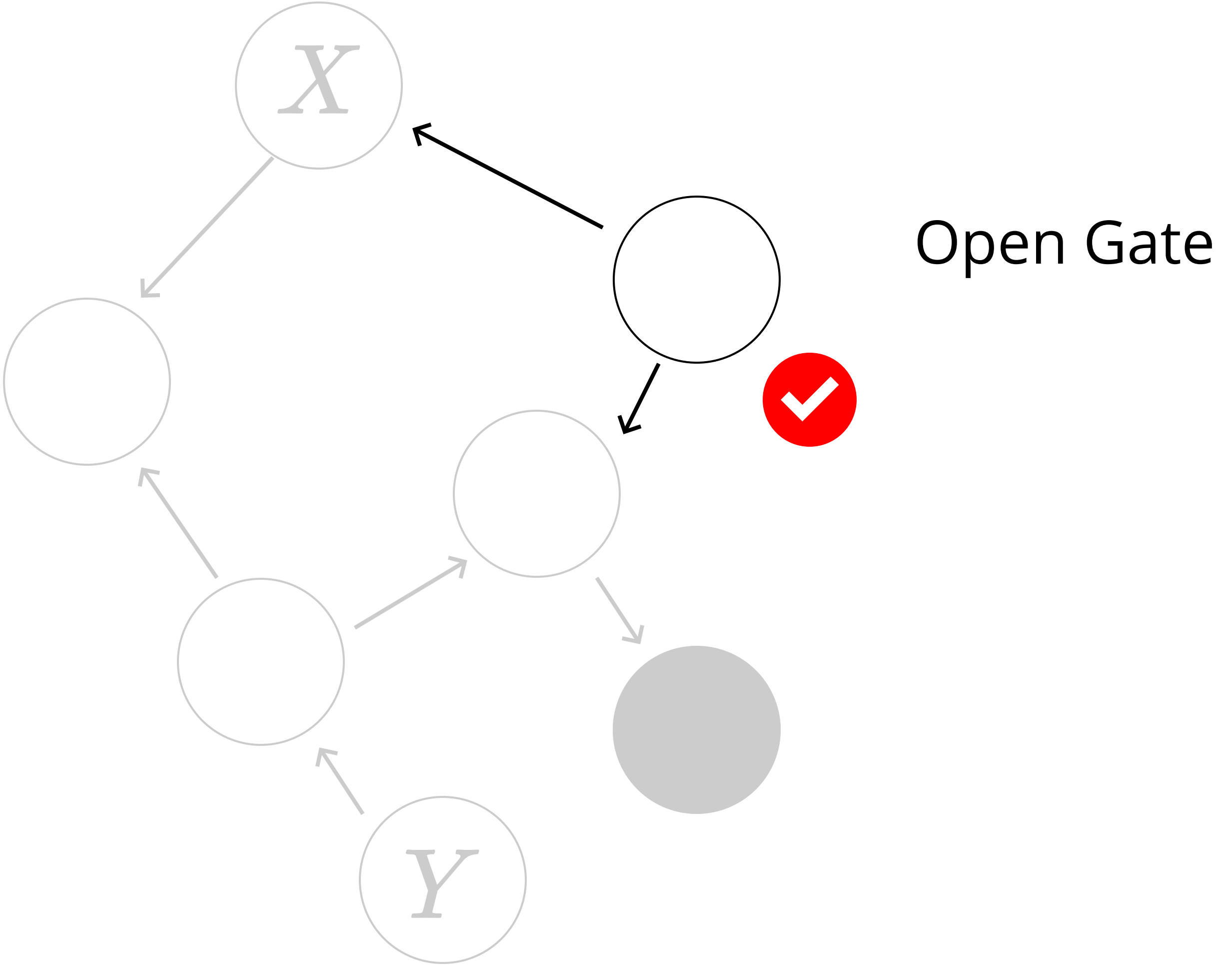
Are X and Y independent?

Bayes' Ball Algorithm



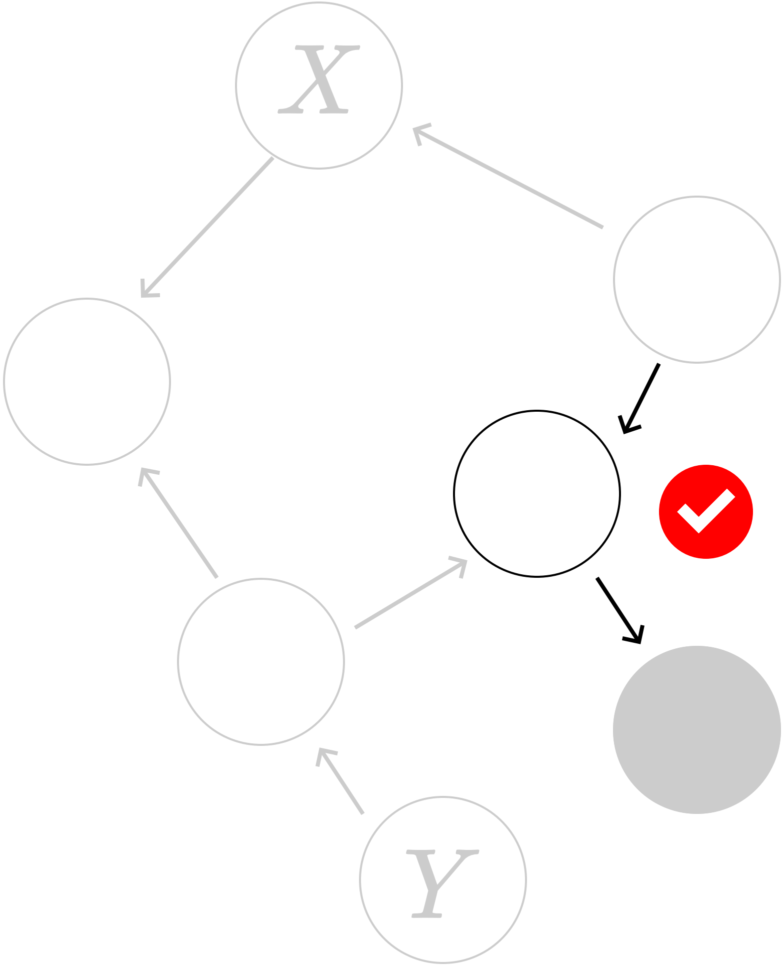
Are X and Y independent?

Bayes' Ball Algorithm



Are X and Y independent?

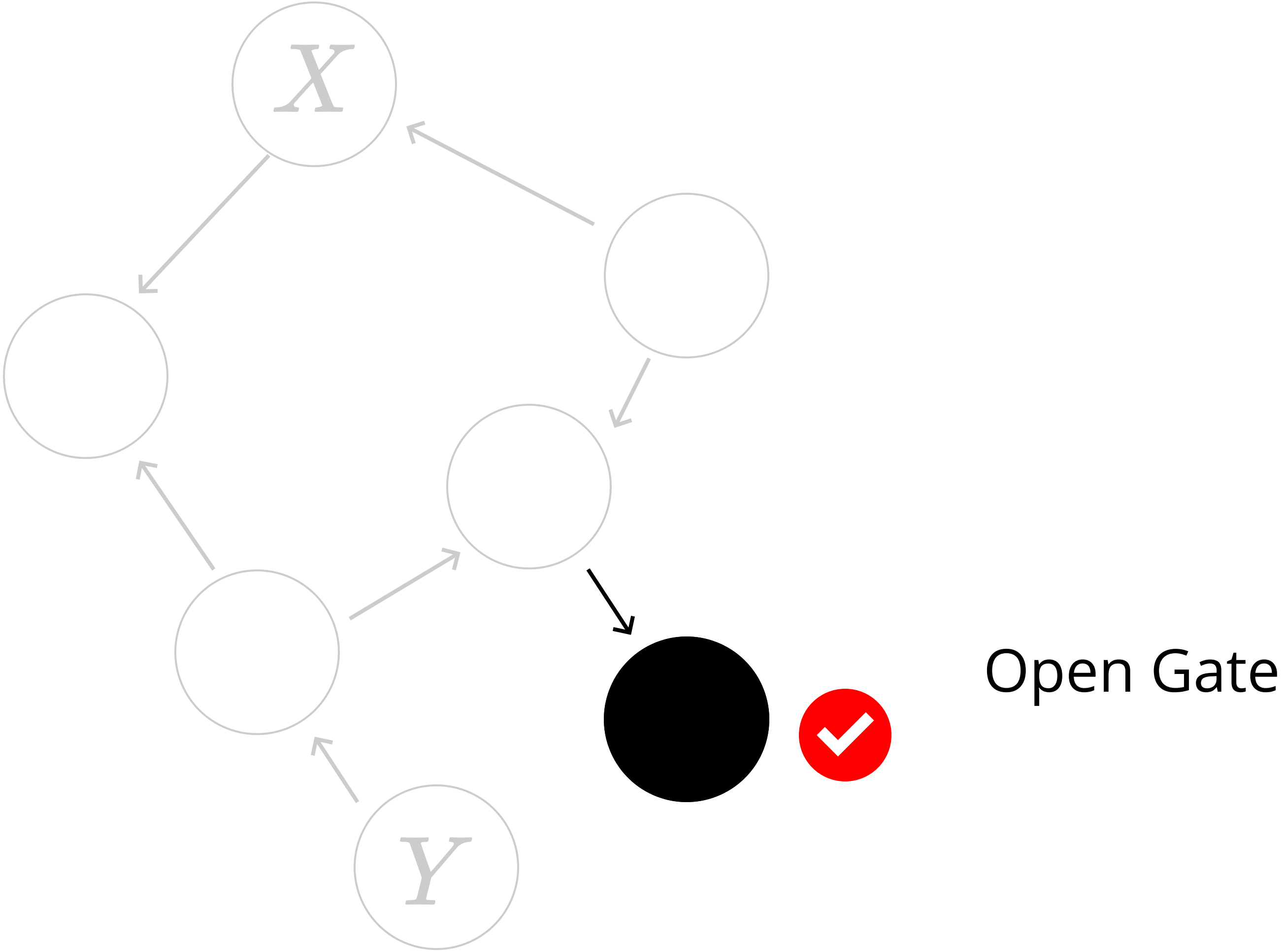
Bayes' Ball Algorithm



Open Gate

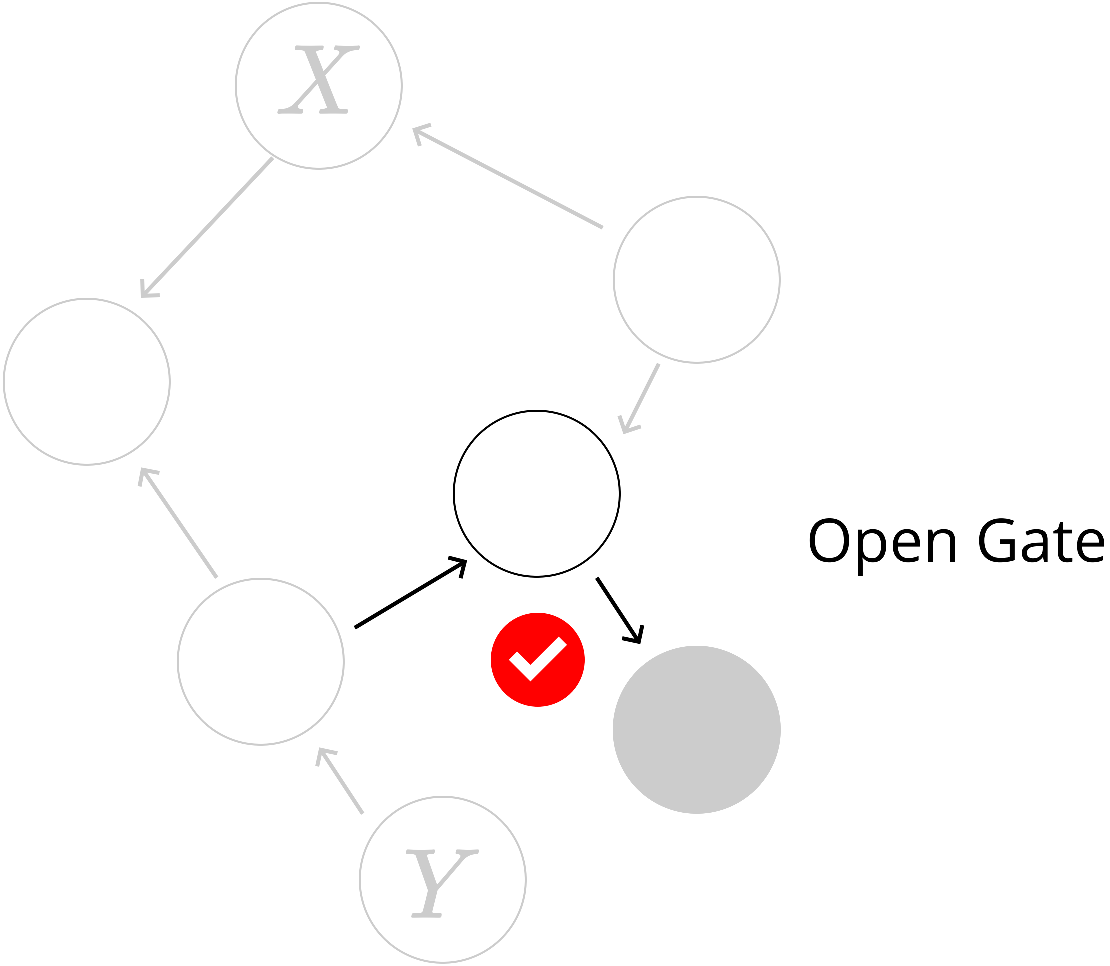
Are X and Y independent?

Bayes' Ball Algorithm



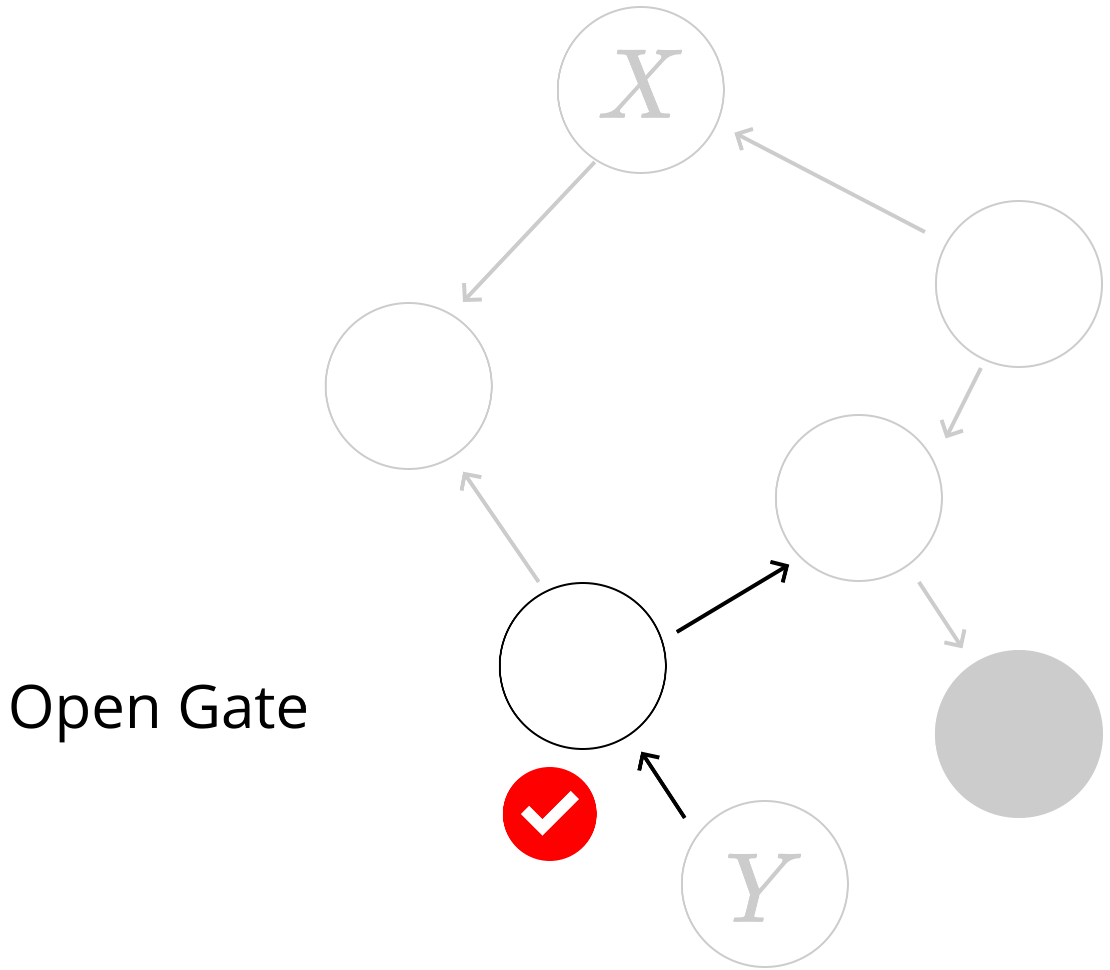
Are X and Y independent?

Bayes' Ball Algorithm



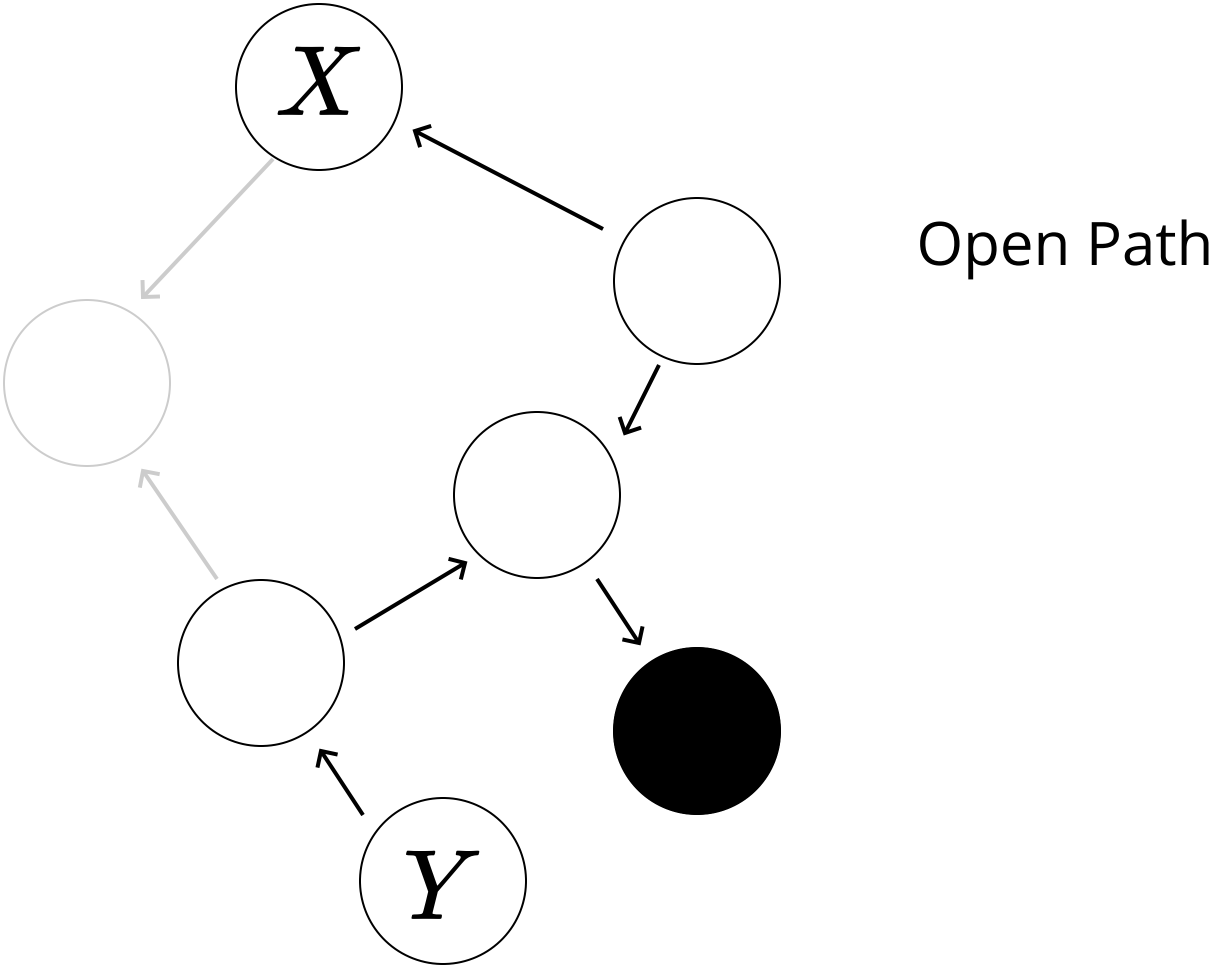
Are X and Y independent?

Bayes' Ball Algorithm



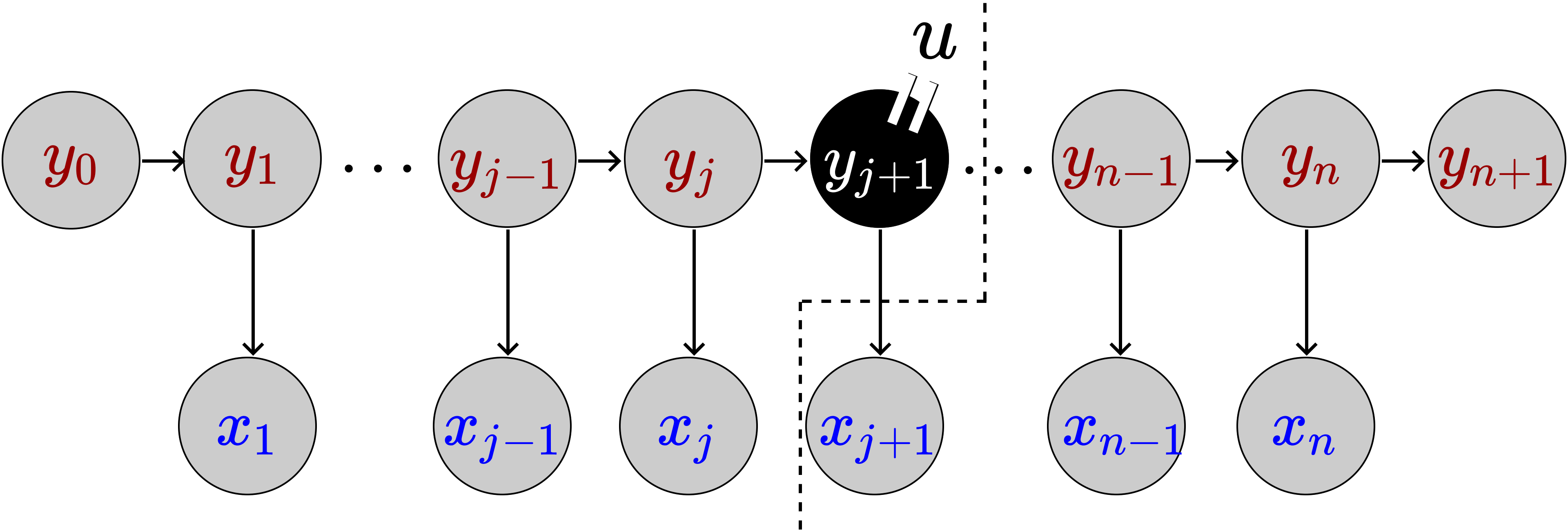
Are X and Y independent?

Bayes' Ball Algorithm



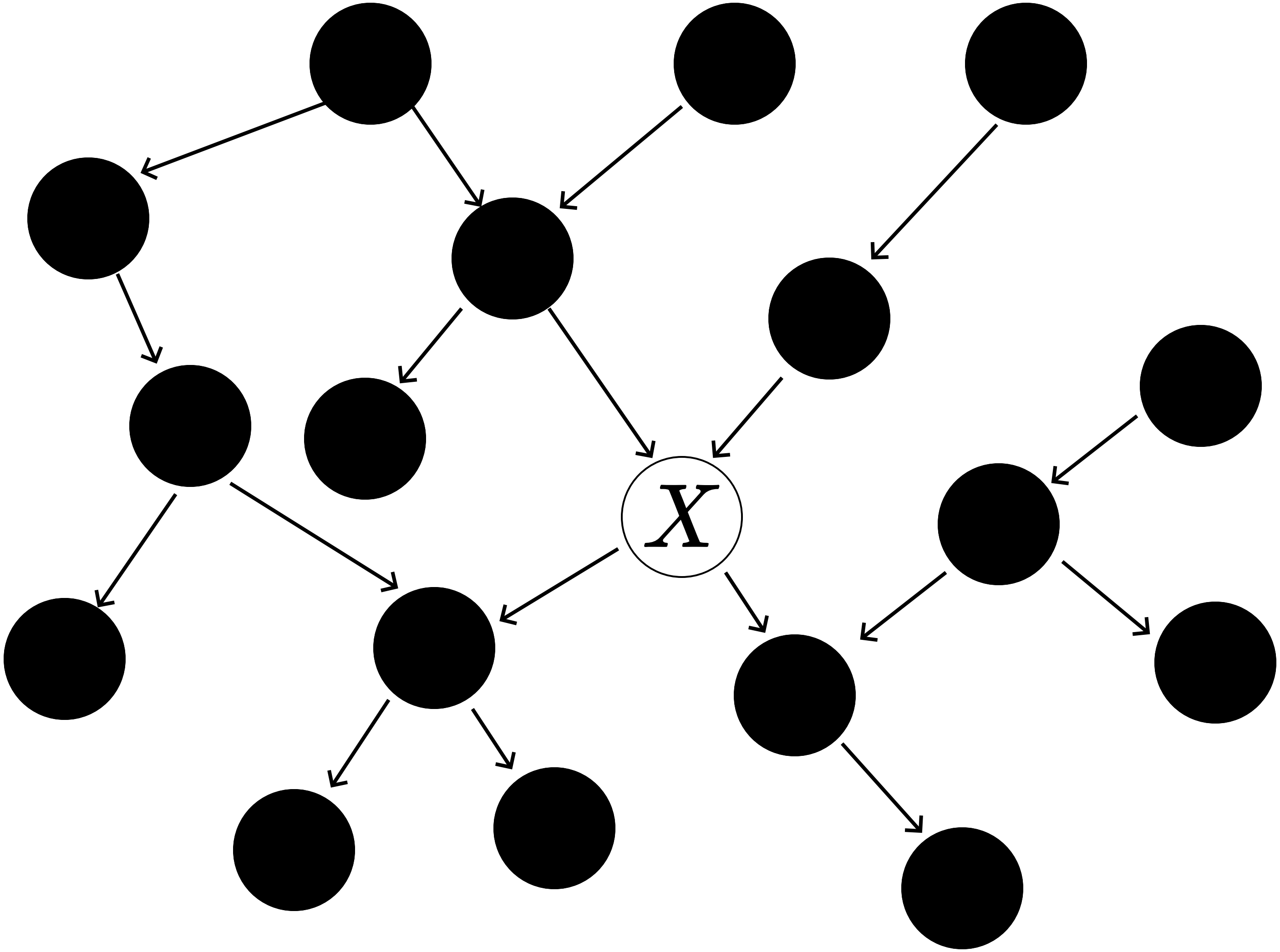
$$\text{Are } X \text{ and } Y \text{ independent?}$$

Bayes Net - HMM



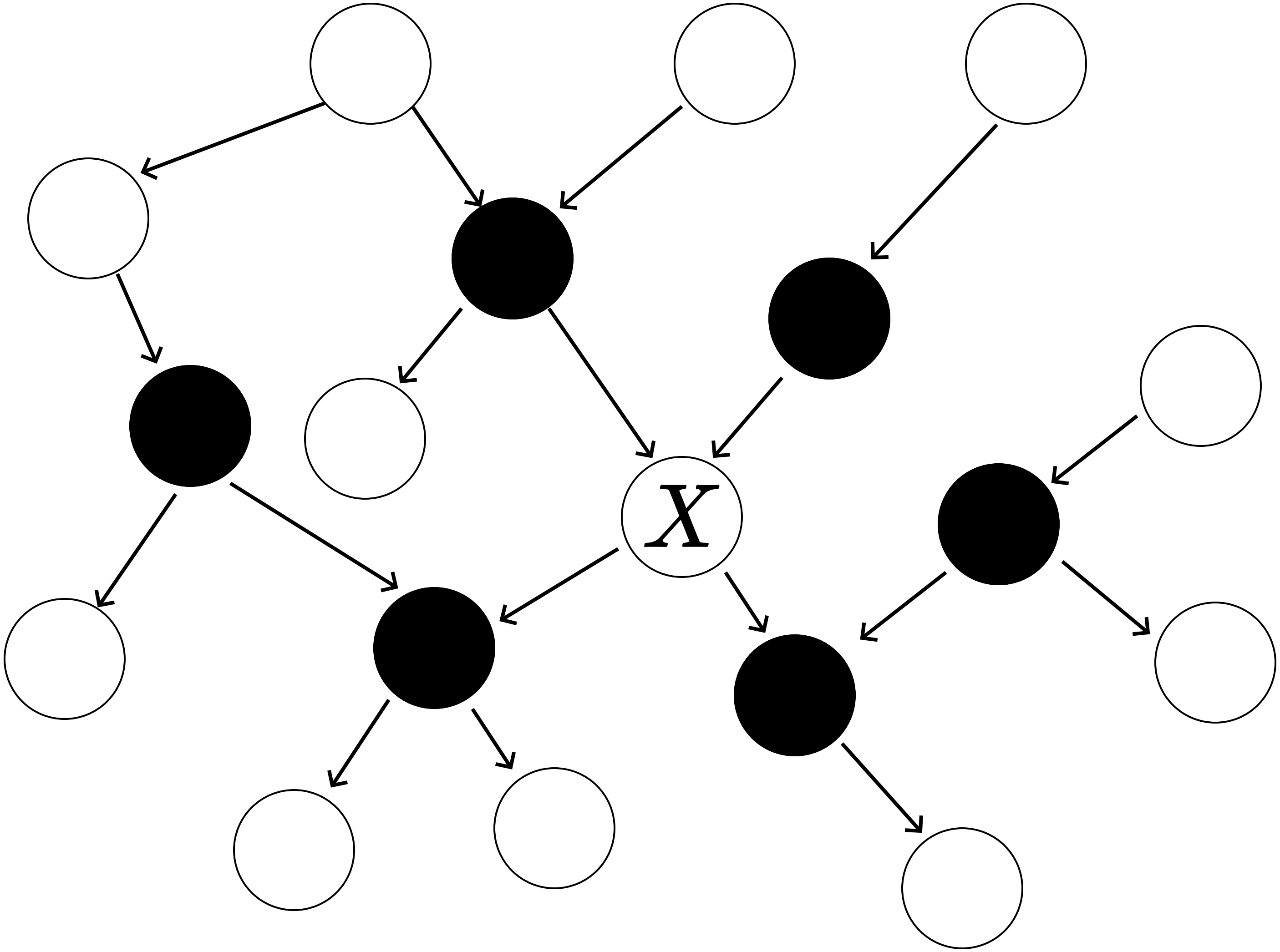
Given $y_{j+1} = u$, now we can see the two portions of the network are independent of each other.

Markov Blanket



$$p(X|\mathbf{V}_{-X})$$

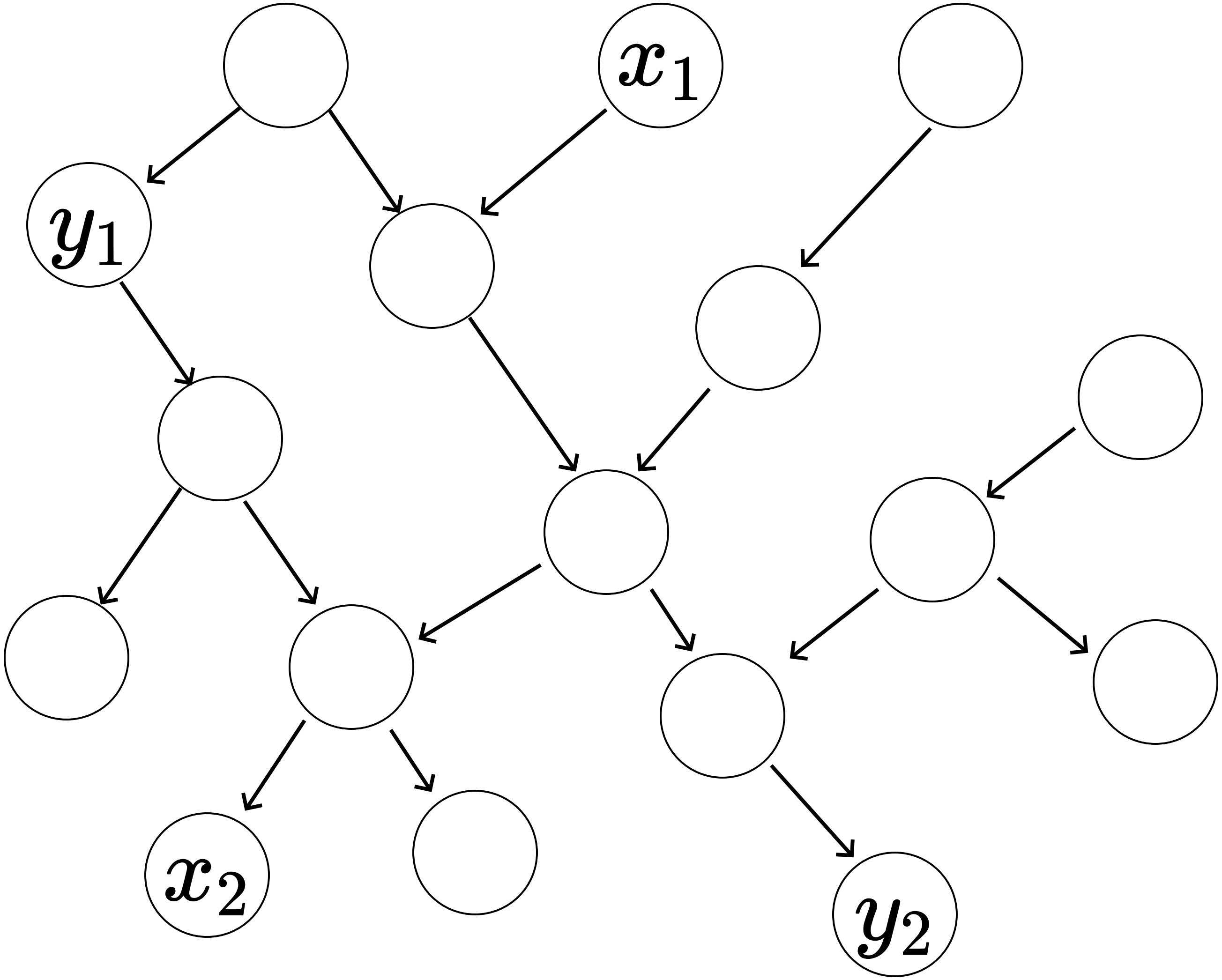
Markov Blanket



$$p(X|\mathbf{V}_{-X})$$

Approximate Inference

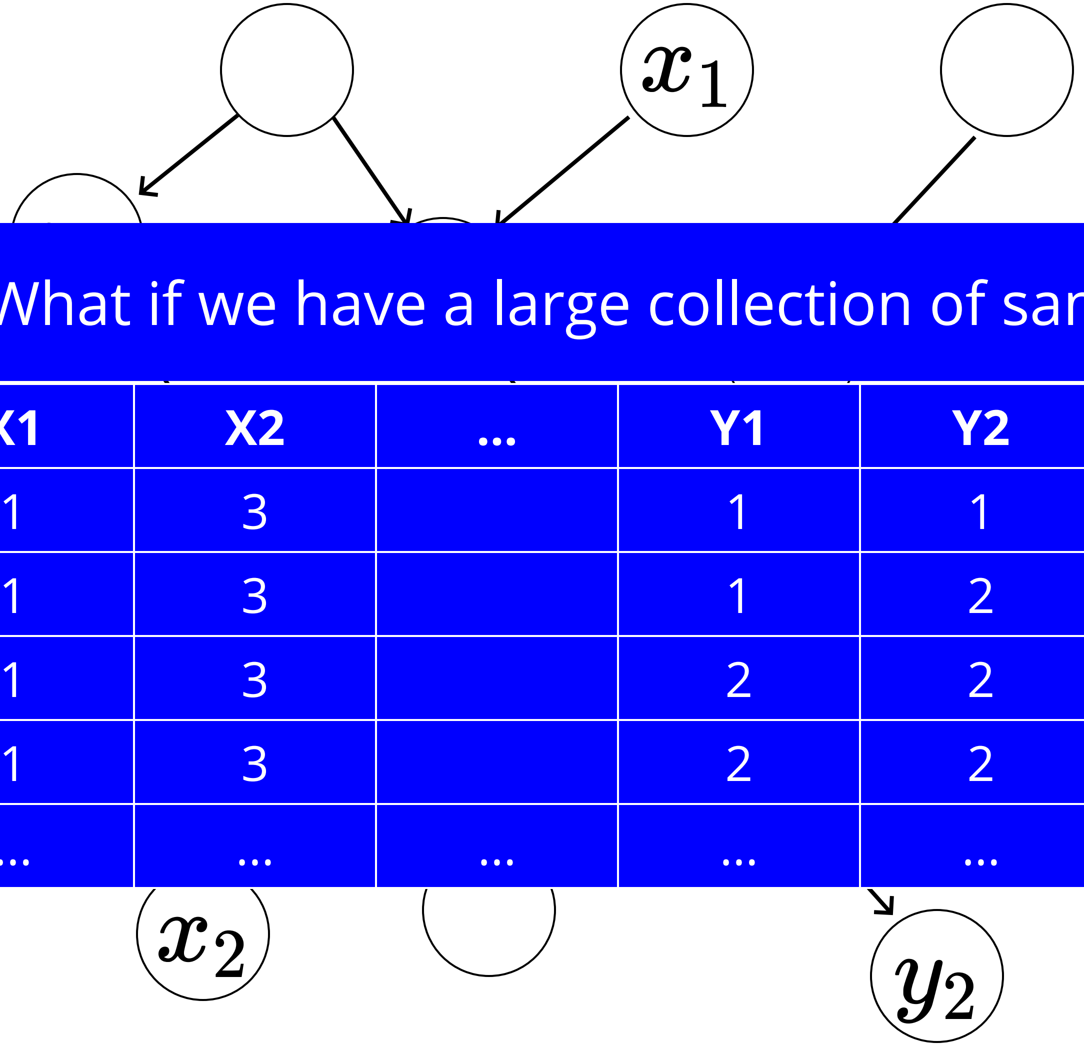
Optional



$$p(y_1 = 1, y_2 = 2 | x_1 = 1, x_2 = 3) =?$$

Optional

Approximate Inference



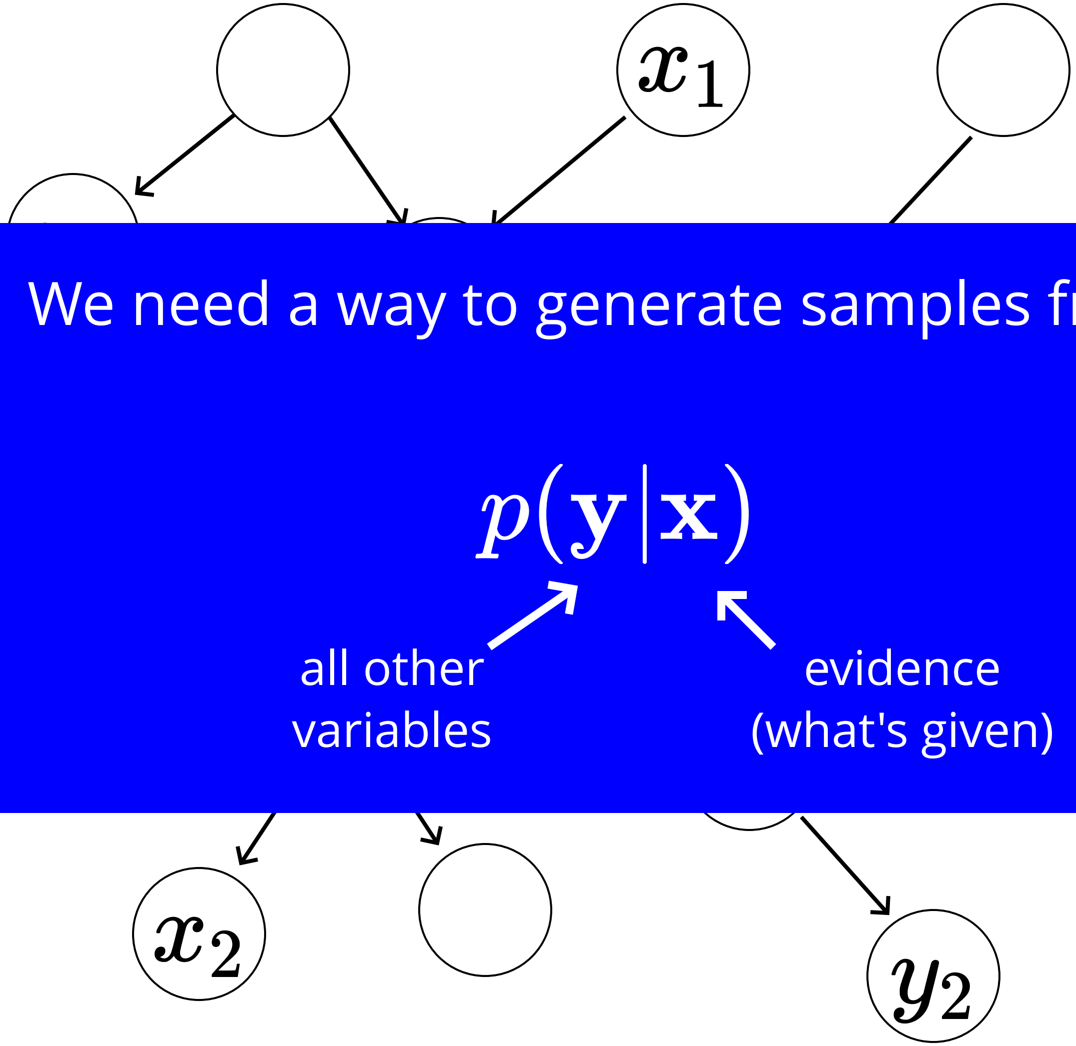
What if we have a large collection of samples:

X1	**X2**	**...**	**Y1**	**Y2**	**...**
1	3		1	1	
1	3		1	2	
1	3		2	2	
1	3		2	2	
...	...	...	...	...	...

$$p(y_1 = 1, y_2 = 2 | x_1 = 1, x_2 = 3) =?$$

Optional

Approximate Inference



We need a way to generate samples from:

$$p(\mathbf{y}|\mathbf{x})$$

all other variables

evidence (what's given)

$$p(y_1 = 1, y_2 = 2 | x_1 = 1, x_2 = 3) = ?$$

Optional

Gibbs Sampling

1. Randomly initialize $\mathbf{y}^{(0)} = \langle y_1^{(0)}, y_2^{(0)}, \ldots, y_n^{(0)} \rangle$

2. For $t = 1, \ldots, T$ do

   (a) For $k = 1, \ldots, n$ do

   $$y_k^{(t)} \sim P(y_k | y_1^{(t)}, \ldots, y_{k-1}^{(t)}, y_{k+1}^{(t-1)}, y_{k+2}^{(t-1)}, \ldots, y_n^{(t-1)}, \mathbf{x})$$

   (b) Collect the t-th sample as $\langle y_1^{(t)}, y_2^{(t)}, \ldots, y_n^{(t)} \rangle$

3. Return the collection of samples

$\downarrow$

These samples are from $p(\mathbf{y}|\mathbf{x})$

Gibbs Sampling

Optional

1. Randomly initialize $\mathbf{y}^{(0)} = \langle y_1$

You may simplify this conditional probability with Markov blanket

2. For $t = 1, \ldots, T$ do

   (a) For $k = 1, \ldots, n$ do

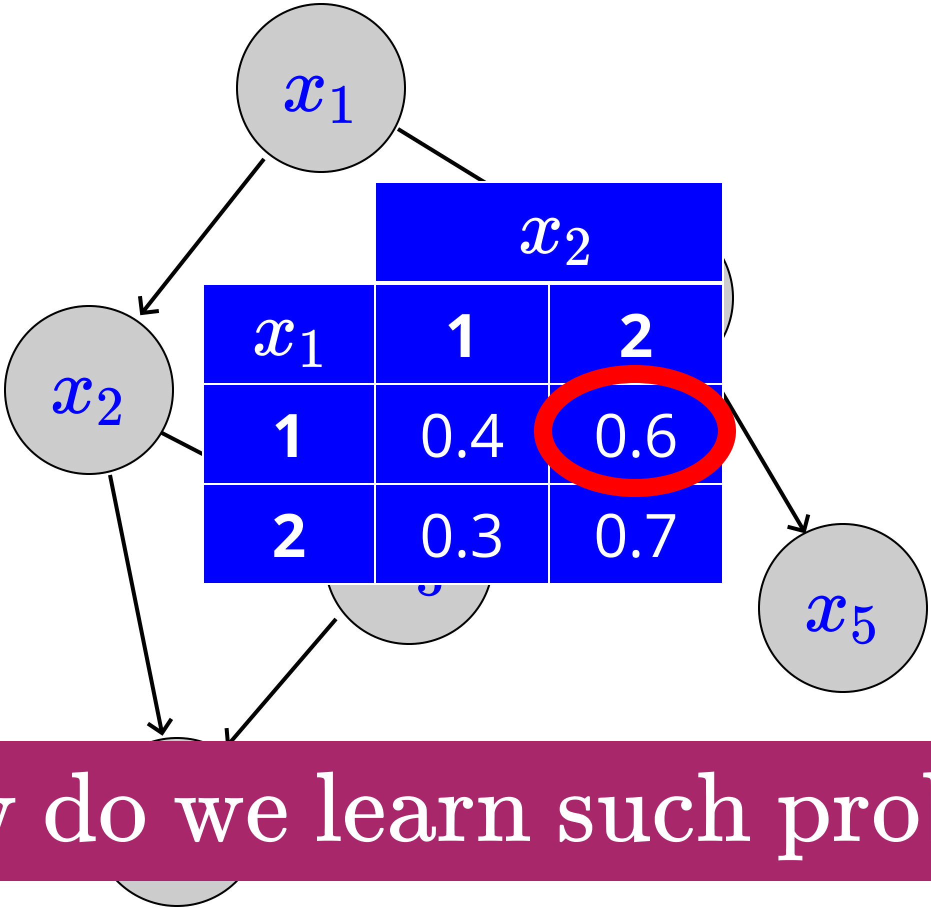
   $$y_k^{(t)} \sim P(y_k | y_1^{(t)}, \ldots, y_{k-1}^{(t)}, y_{k+1}^{(t-1)}, y_{k+2}^{(t-1)}, \ldots, y_n^{(t-1)}, \mathbf{x})$$

   (b) Collect the t-th sample as $\langle y_1^{(t)}, y_2^{(t)}, \ldots, y_n^{(t)} \rangle$

3. Return the collection of samples

You may use then to perform approximate inference

Bayesian Networks

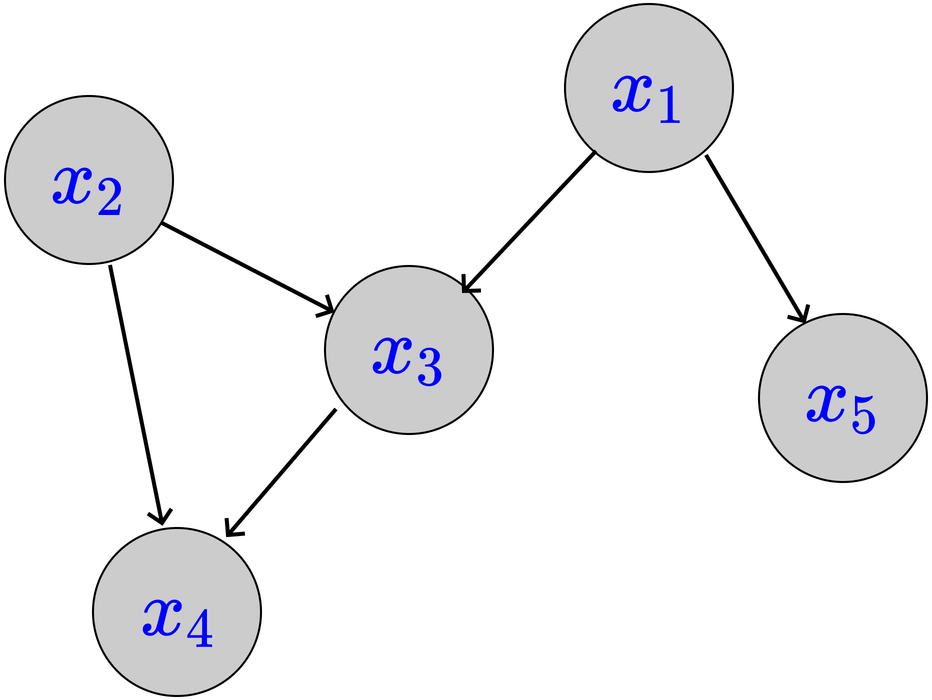


	x_2	
x_1	**1**	**2**
1	0.4	0.6
2	0.3	0.7

How do we learn such probability values?

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

Learning in Bayes Net



$x_1 \in \{1, 2, \ldots, r_1\}$

$x_2 \in \{1, 2, \ldots, r_2\}$

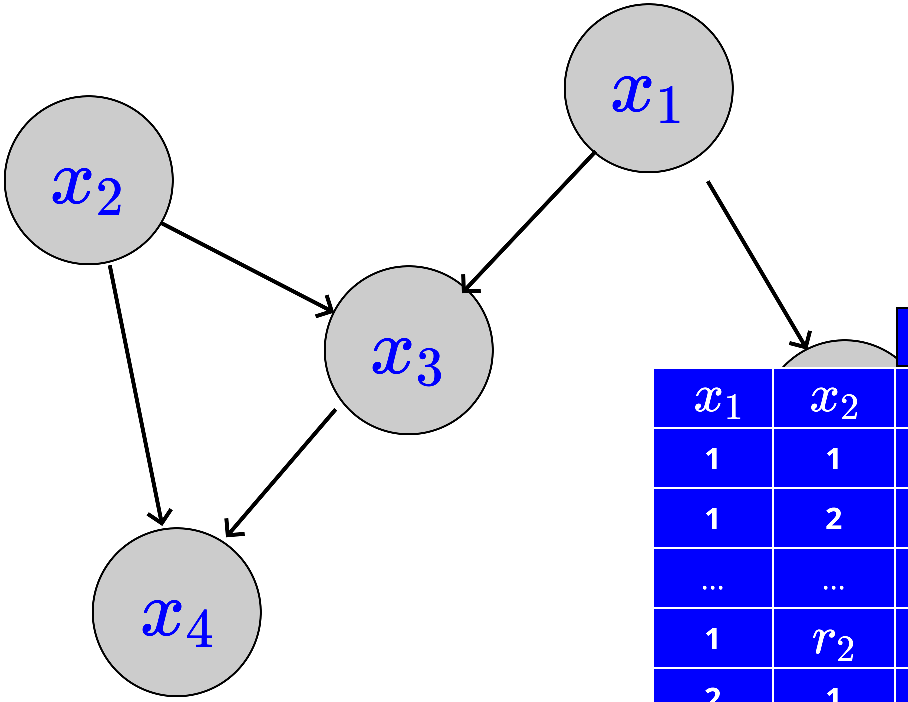
$x_3 \in \{1, 2, \ldots, r_3\}$

$x_4 \in \{1, 2, \ldots, r_4\}$

$x_5 \in \{1, 2, \ldots, r_5\}$

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

Bayesian Networks

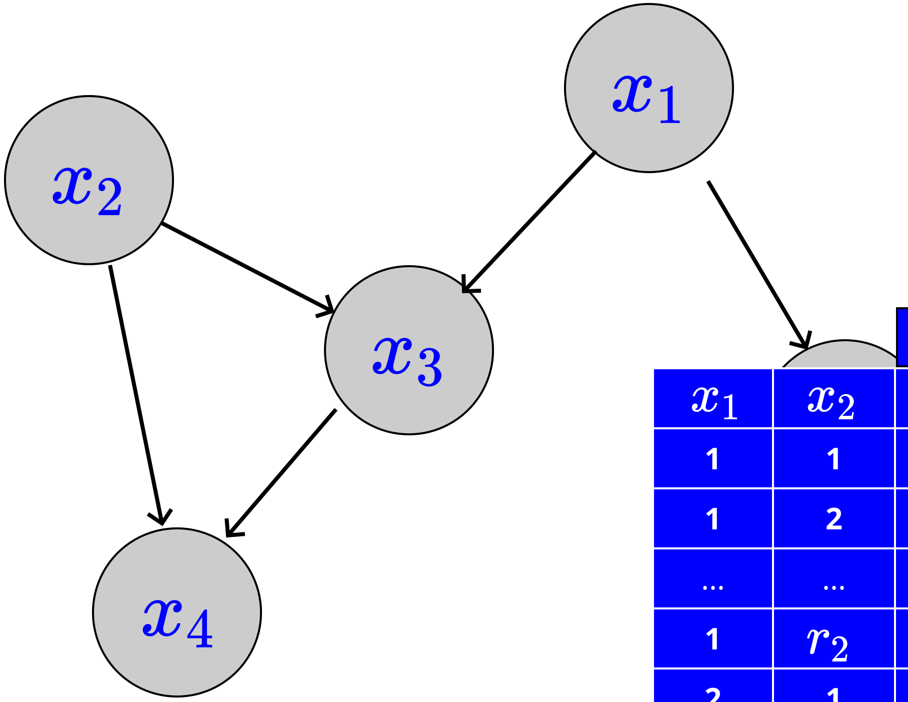


$$x_1 \in \{1, 2, \ldots, r_1\}$$
$$x_2 \in \{1, 2, \ldots, r_2\}$$
$$x_3 \in \{1, 2, \ldots, r_3\}$$
$$\ldots, r_4\}$$
$$\ldots, r_5\}$$

		x_3			
x_1	x_2	1	2	…	r_3
1	1				
1	2				
…	…				
1	r_2				
2	1				
2	2				
…	…				
2	r_2				
…	…				
r_1	1				
r_1	2				
…	…				
r_1	r_2				

$$p(x_1, x_2$$
$$= p(x_1)p(x_2)p(x_3|x_1 \quad \ldots \quad x_1)$$

Bayesian Networks



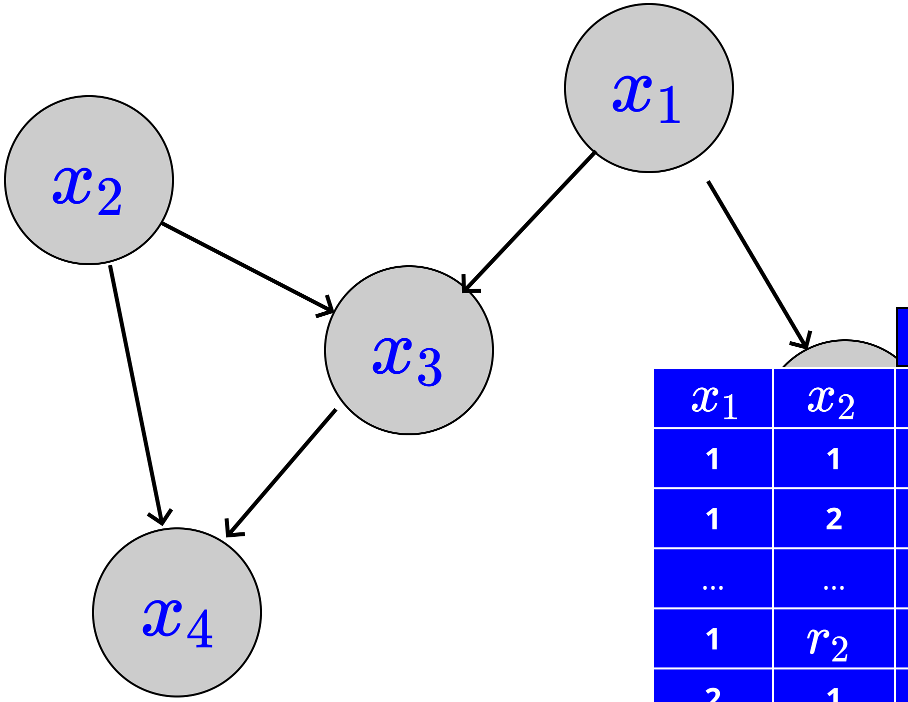
$$x_1 \in \{1, 2, \dots, r_1\}$$
$$x_2 \in \{1, 2, \dots, r_2\}$$
$$x_3 \in \{1, 2, \dots, r_3\}$$
$$\dots, r_4\}$$
$$\dots, r_5\}$$

		x_3			
x_1	x_2	1	2	...	r_3
1	1				
1	2				
...	...				
1	r_2				
2	1				
2	2				
...	...				
2	r_2				
...	...				
r_1	1				
r_1	2				
...	...				
r_1	r_2				

$\theta_3(1|1,1)$

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

Bayesian Networks



$x_1 \in \{1, 2, \ldots, r_1\}$

$x_2 \in \{1, 2, \ldots, r_2\}$

$x_3 \in \{1, 2, \ldots, r_3\}$

$\ldots, r_4\}$

$\ldots, r_5\}$

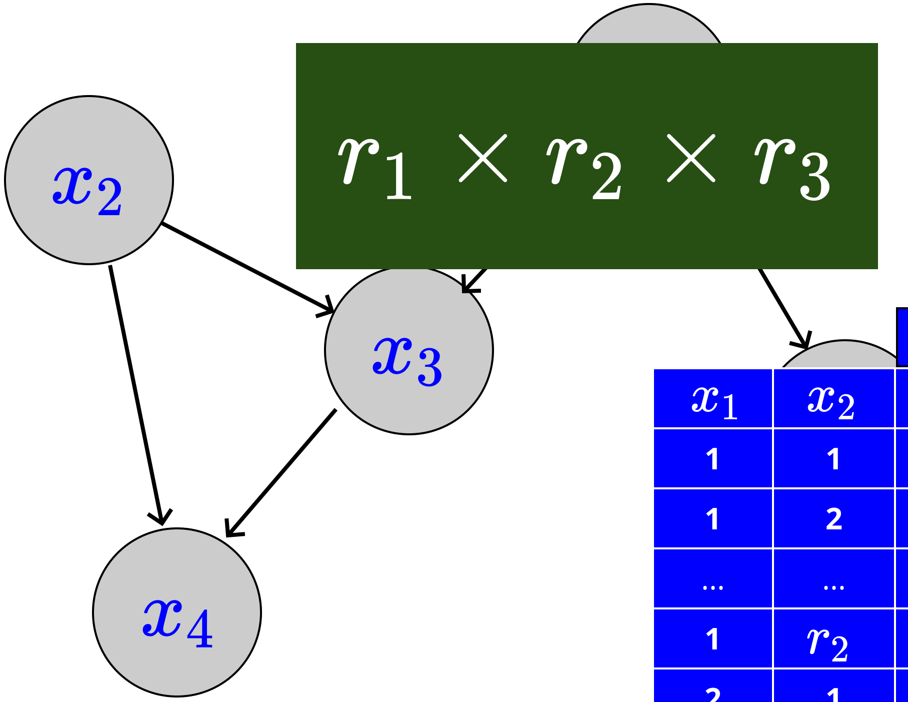
		x_3			
x_1	x_2	1	2	...	r_3
1	1				
1	2				
...	...				
1	r_2				
2	1				
2	2				
...	...				
2					
...					
r_1	1				
r_1	2				
...	...				
r_1	r_2				

$\theta_3(r_3|r_1, r_2)$

$$p(x_1, x_2$$

$$= p(x_1)p(x_2)p(x_3|x_1 \ldots x_1)$$

Bayesian Networks



$$r_1 \times r_2 \times r_3$$

$$x_1 \in \{1, 2, \ldots, r_1\}$$
$$x_2 \in \{1, 2, \ldots, r_2\}$$
$$x_3 \in \{1, 2, \ldots, r_3\}$$
$$\ldots, r_4\}$$
$$\ldots, r_5\}$$

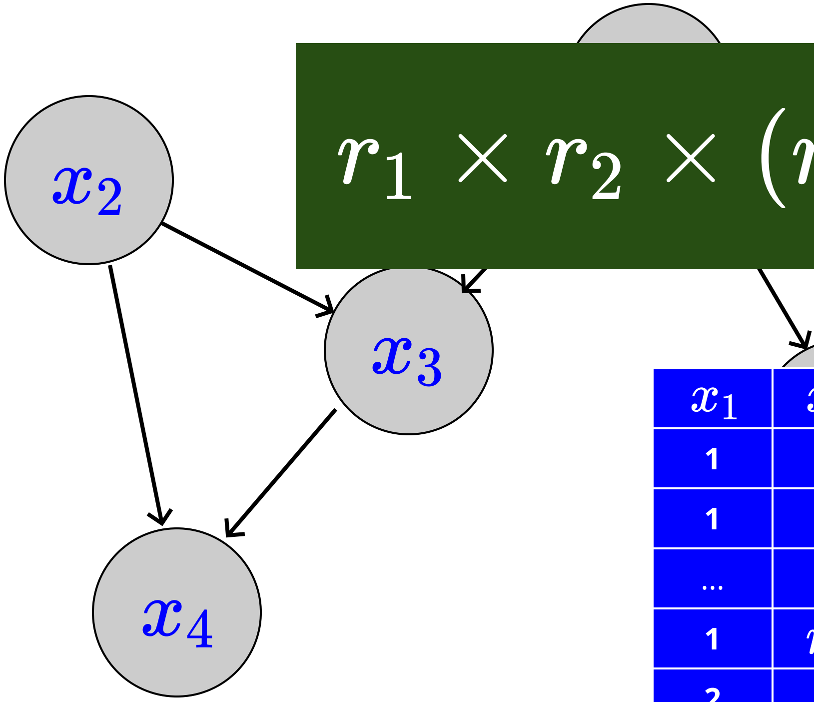
		x_3			
x_1	x_2	1	2	...	r_3
1	1				
1	2				
...	...				
1	r_2				
2	1				
2	2				
...	...				
2	r_2				
...	...				
r_1	1				
r_1	2				
...	...				
r_1	r_2				

What is the number of parameters involved in this table?

$(x_1, x_2$

$(x_3|x_1$ … $x_1)$

Bayesian Networks



$$r_1 \times r_2 \times (r_3 - 1)$$

$\in \{1, 2, \ldots, r_1\}$

$\in \{1, 2, \ldots, r_2\}$

$x_3 \in \{1, 2, \ldots, r_3\}$

$\ldots, r_4\}$

$\ldots, r_5\}$

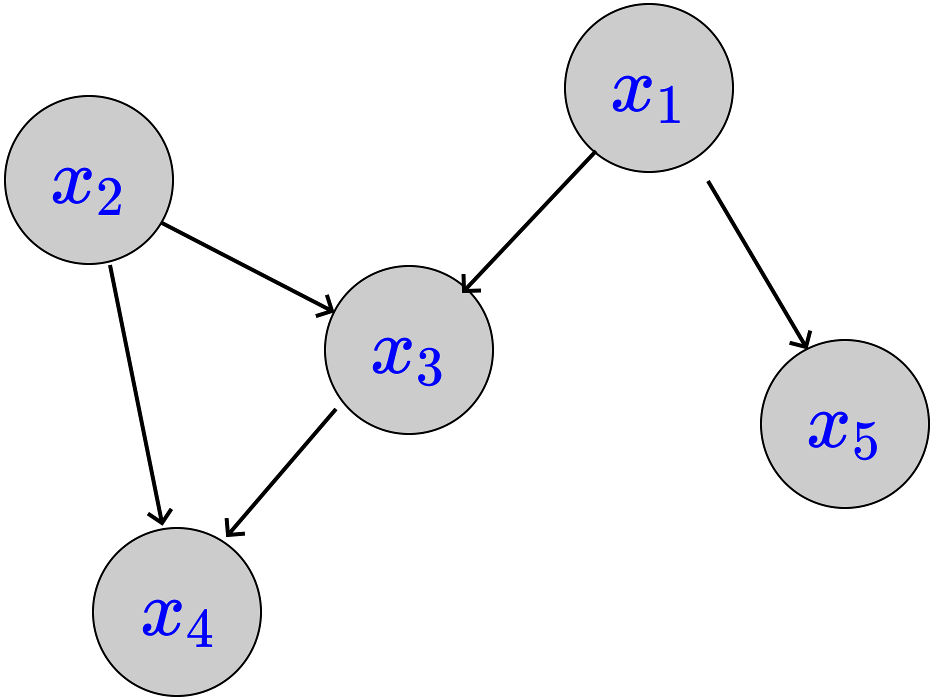
		x_3			
x_1	x_2	1	2	...	r_3
1	1				
1	2				
...	...				
1	r_2				
2	1				
2	2				
...	...				
2	r_2				
...	...				
r_1	1				
r_1	2				
...	...				
r_1	r_2				

What is the number of free parameters involved in this table?

$(x_1, x_2$

$(x_3|x_1$ $x_1)$

Learning in Bayes Nets

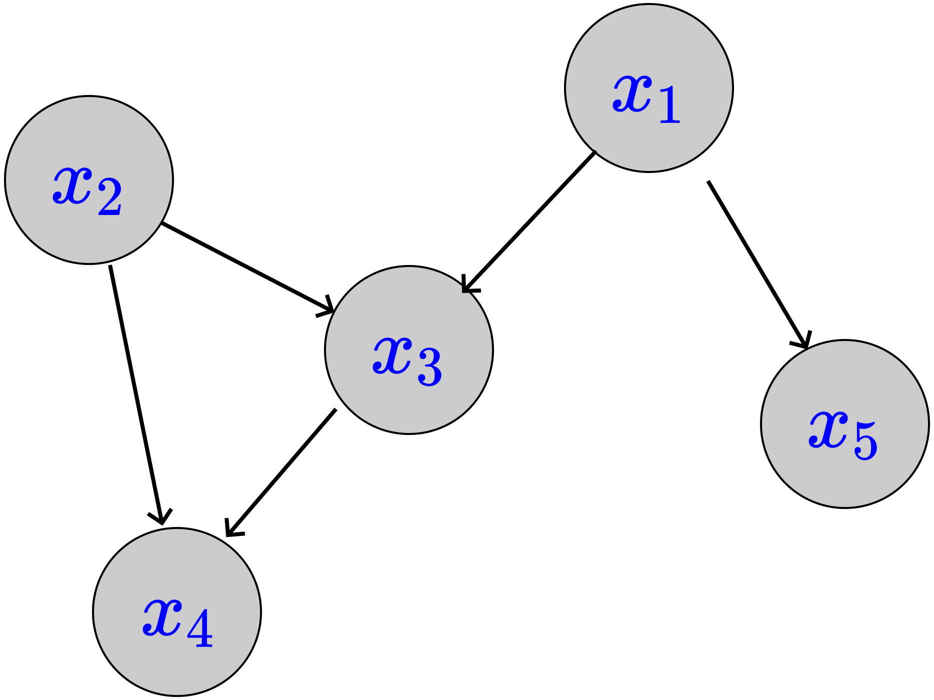


$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$
1	1	2	2	2
1	2	1	1	2
2	2	2	1	2
2	2	1	2	1
2	1	2	2	1
1	1	2	1	2
1	2	1	1	1
2	2	2	1	2
1	1	1	1	2

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

How do we learn the values for each of these model parameters?

Learning in Bayes Nets

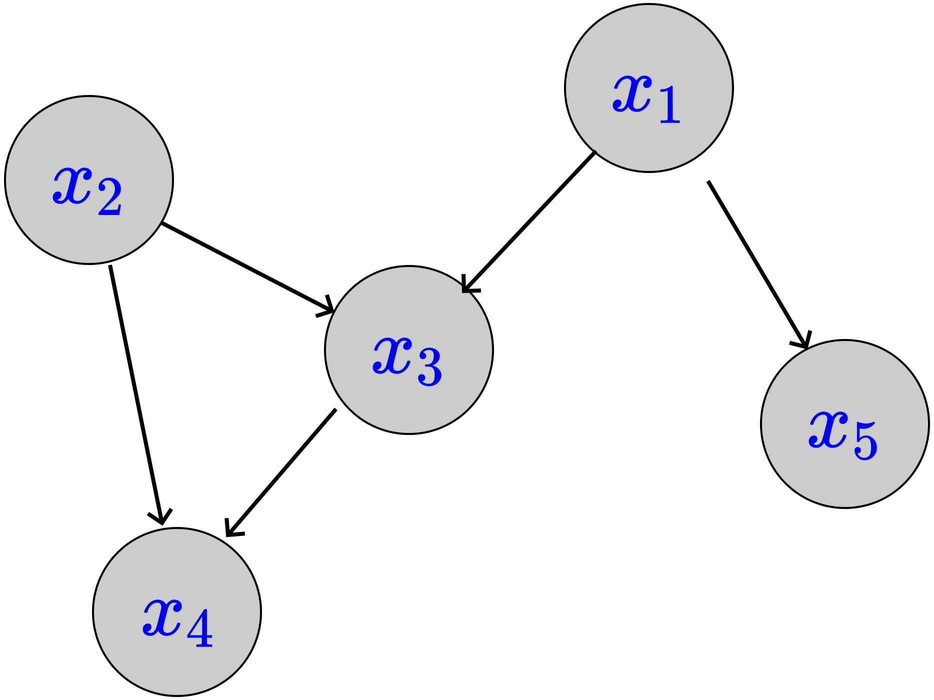


$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$
1	1	2	2	2
1	2	1	1	2
2	2	2	1	2
2	2	1	2	1
2	1	2	2	1
1	1	2	1	2
1	2	1	1	1
2	2	2	1	2
1	1	1	1	2

$$l(D;\theta;G) = \sum_{t=1}^{m} \log \left[\prod_{i=1}^{d} \theta_i (x_i^{(t)} | \mathbf{x}_{pa_i}^{(t)}) \right]$$

This is what we would like to maximize!

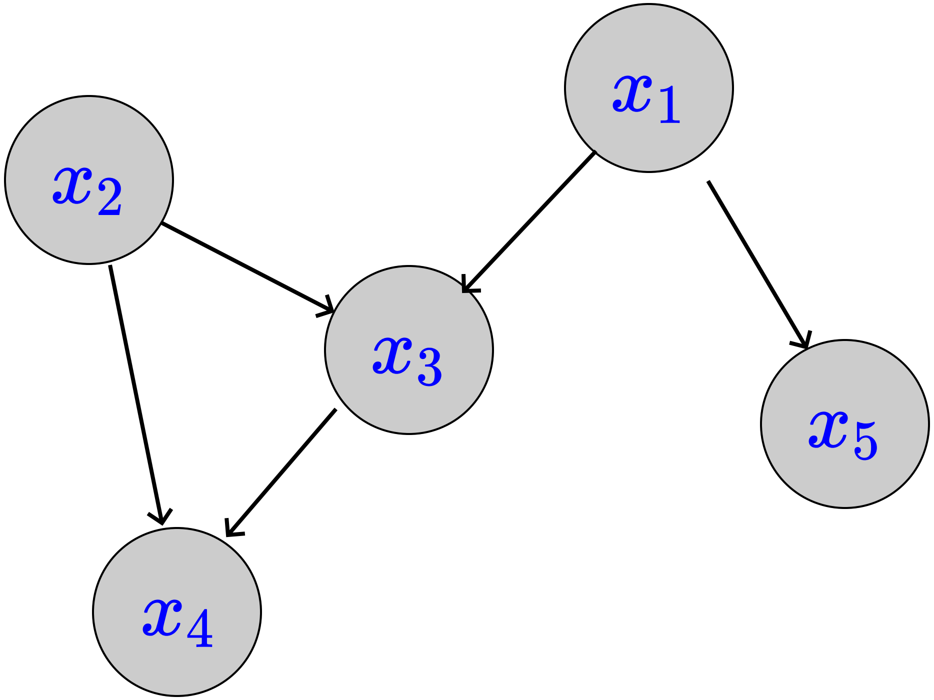
Learning in Bayes Nets



$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$
1	1	2	2	2
1	2	1	1	2
2	2	2	1	2
2	2	1	2	1
2	1	2	2	1
1	1	2	1	2
1	2	1	1	1
2	2	2	1	2
1	1	1	1	2

$$l(D;\theta;G) = \sum_{t=1}^{m} \log\left[\prod_{i=1}^{d} \theta_i(x_i^{(t)} | \mathbf{x}_{pa_i}^{(t)})\right]$$
$$= \sum_{t=1}^{m} \sum_{i=1}^{d} \log \theta_i(x_i^{(t)} | \mathbf{x}_{pa_i}^{(t)})$$

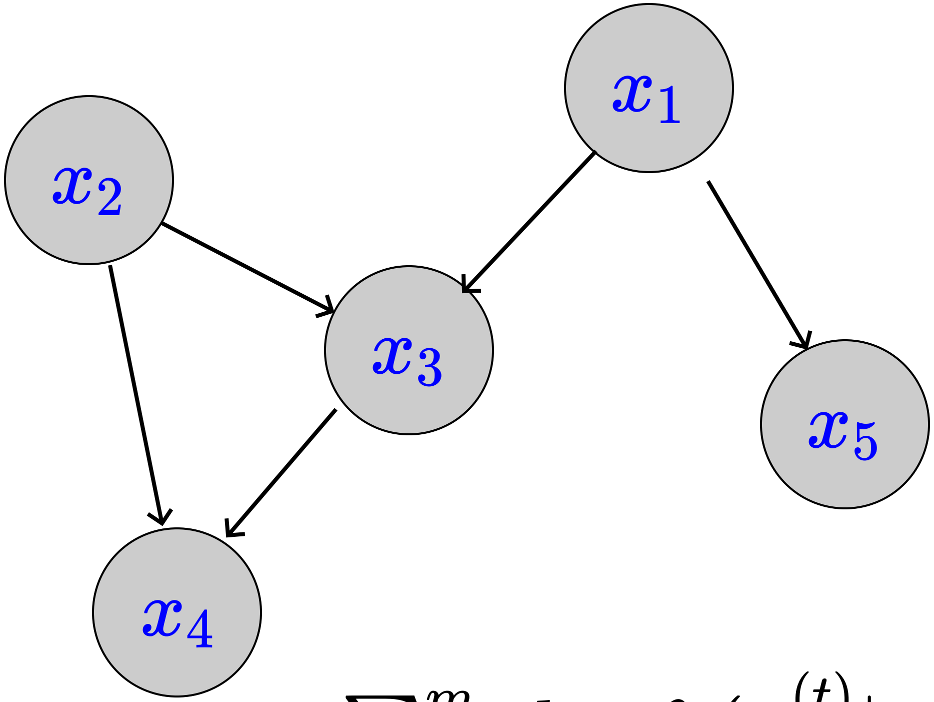
Learning in Bayes Nets



$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$
1	1	2	2	2
1	2	1	1	2
2	2	2	1	2
2	2	1	2	1
2	1	2	2	1
1	1	2	1	2
1	2	1	1	1
2	2	2	1	2
1	1	1	1	2

$$
\begin{aligned}
l(D;\theta;G) &= \sum_{t=1}^{m} \log\left[\prod_{i=1}^{d} \theta_i(x_i^{(t)}|\mathbf{x}_{pa_i}^{(t)})\right] \\
&= \sum_{t=1}^{m} \sum_{i=1}^{d} \log \theta_i(x_i^{(t)}|\mathbf{x}_{pa_i}^{(t)}) \\
&= \sum_{i=1}^{d} \left[\sum_{t=1}^{m} \log \theta_i(x_i^{(t)}|\mathbf{x}_{pa_i}^{(t)})\right]
\end{aligned}
$$

Learning in Bayes Nets



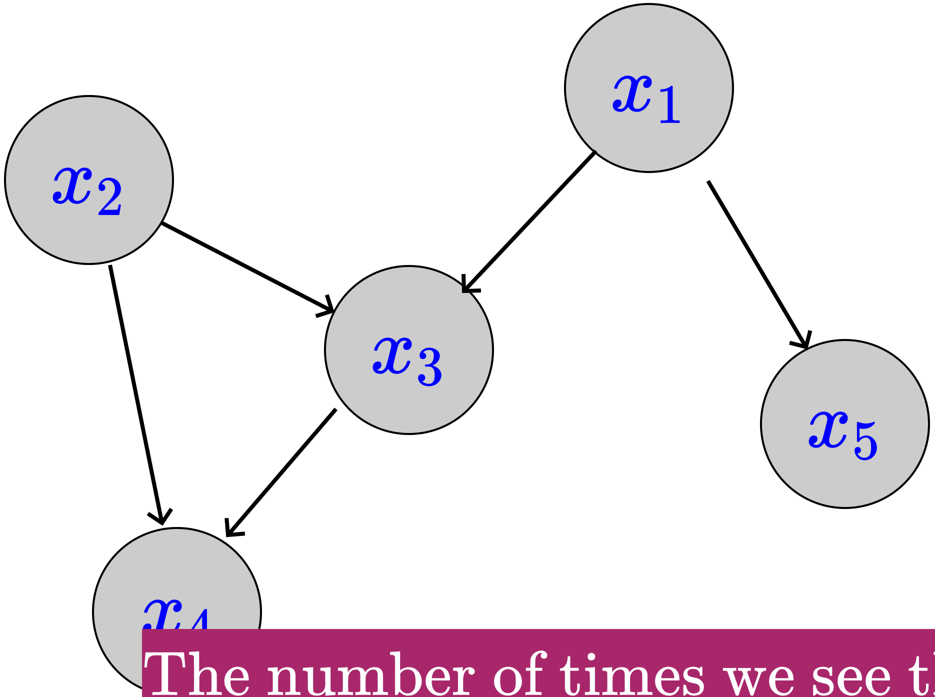
$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$
1	1	2	2	2
1	2	1	1	2
2	2	2	1	2
2	2	1	2	1
2	1	2	2	1
1	1	2	1	2
1	2	1	1	1
2	2	2	1	2
1	1	1	1	2

$$\sum_{t=1}^{m} \log \theta_i(x_i^{(t)} | \mathbf{x}_{pa_i}^{(t)})$$

$$= \sum_{x_i, \mathbf{x}_{pa_i}} \underline{\text{Count}\big((x_i, \mathbf{x}_{pa_i}) \text{ in } D\big)} \log \theta_i(x_i | \mathbf{x}_{pa_i})$$

The number of observations in data D for which $X_i = x_i, \mathbf{X}_{pa_i} = \mathbf{x}_{pa_i}$

Learning in Bayes Nets



$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$
1	1	2	2	2
1	2	1	1	2
2	2	2	1	2
2	2	1	2	1
2	1	2	2	1
1	1	2	1	2
1	2	1	1	1
2	2	2	1	2
1	1	1	1	2

The number of times we see the pattern $\mathbf{x}_{pa_i} \to x_i$ in D

$$\hat{\theta}_i(x_i|\mathbf{x}_{pa_i}) = \frac{\text{Count}\Big((x_i, \mathbf{x}_{pa_i}) \text{ in } D\Big)}{\text{Count}\Big((\mathbf{x}_{pa_i}) \text{ in } D\Big)}, x_i \in \{1, \dots, r_1\}$$

The number of times we see the pattern $\mathbf{x}_{pa_i}$ in D