Bayesian Networks II

Monday, March 28, 2022

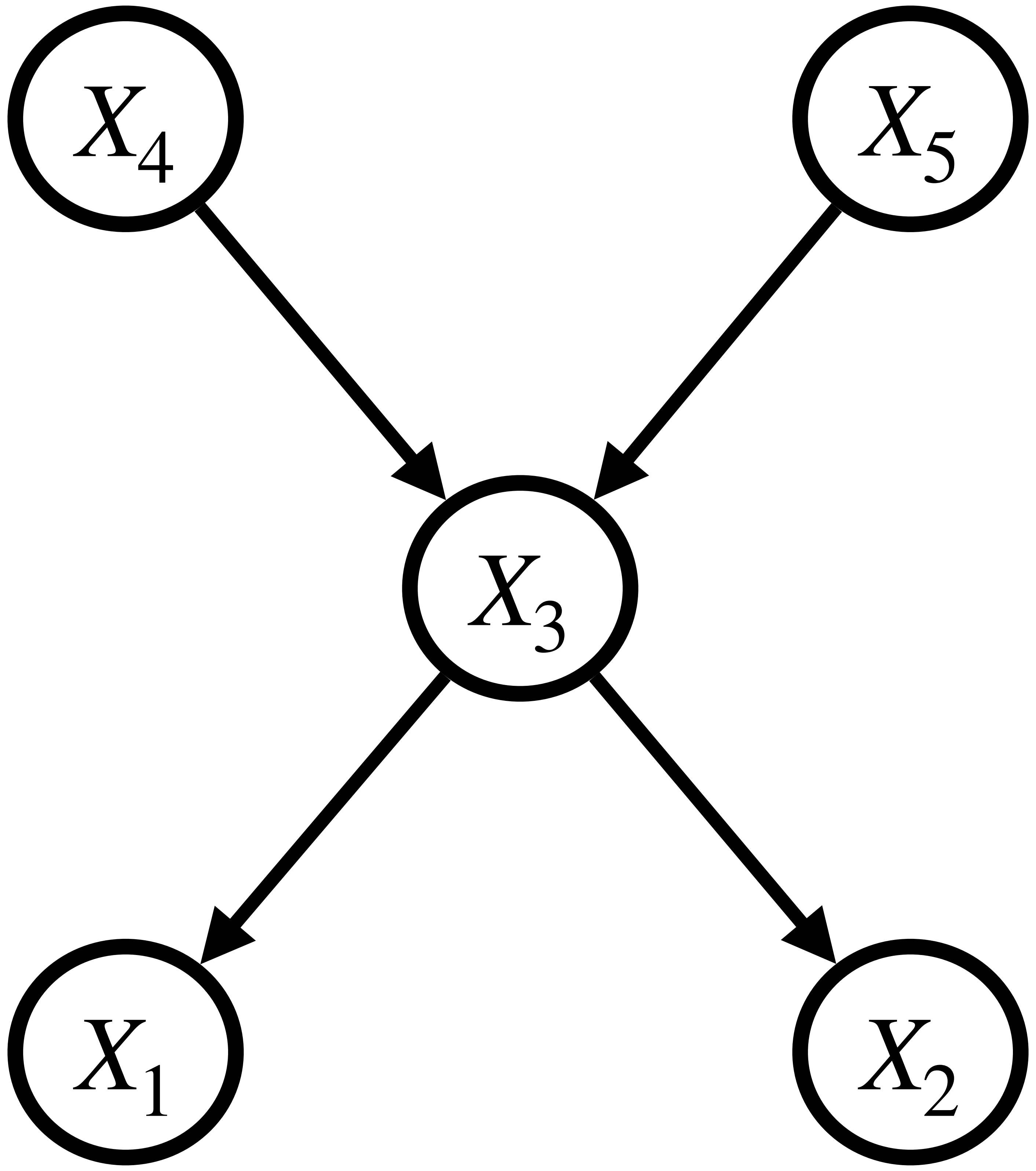
Part 1: Recap

Materials

Great materials on BNs from Volodymyr Kuleshov and Stefano Ermon from Stanford:

https://ermongroup.github.io/cs228-notes/

Bayesian Network (The Graph)



Bayesian Network Distribution

Given a BN with a graph G, the BN induces the following distribution:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P_{X_i|Par(X_i)}\left(x_i \,|\, \text{par}_{\mathbf{x}}(X_i)\right).$$

Conditional Independence

Definition (special case of 3 random variables X, Y, Z):

Definition 1: X and Y are conditionally independent given Z if

$$P[X = x \wedge Y = y \,|\, Z = z] = P[X = x \,|\, Z = z] \cdot P[Y = y \,|\, Z = z]$$

holds for all values x, y, z *(using the alternative notation:*
$P_{X,Y|Z}(x, y \,|\, z) = P_{X|Z}(x \,|\, z) \cdot P_{Y|Z}(y \,|\, z)$*)*.

Definition 2: X and Y are conditionally independent given Z if

$$P[X = x \,|\, Y = y \wedge Z = z] = P[X = x \,|\, Z = z]$$

holds for all values x, y, z *(using the alternative notation:*
$P_{X|Y,Z}(x \,|\, y, z) = P_{X|Z}(x \,|\, z)$*)*.

Conditional Independence

Notation: The notation for X and Y are conditionally independent given Z is written:

$$X \perp\!\!\!\perp Y \,|\, Z$$

D-Separation

Given a Bayesian network and a set of variables $\mathscr{E}$ that are conditioned on, we will want to detect those random variables that are conditionally independent given the values of the variables in $\mathscr{E}$.

Two variables X_1 and X_2 are conditionally independent given $\mathscr{E}$ if there is no active path connecting them.

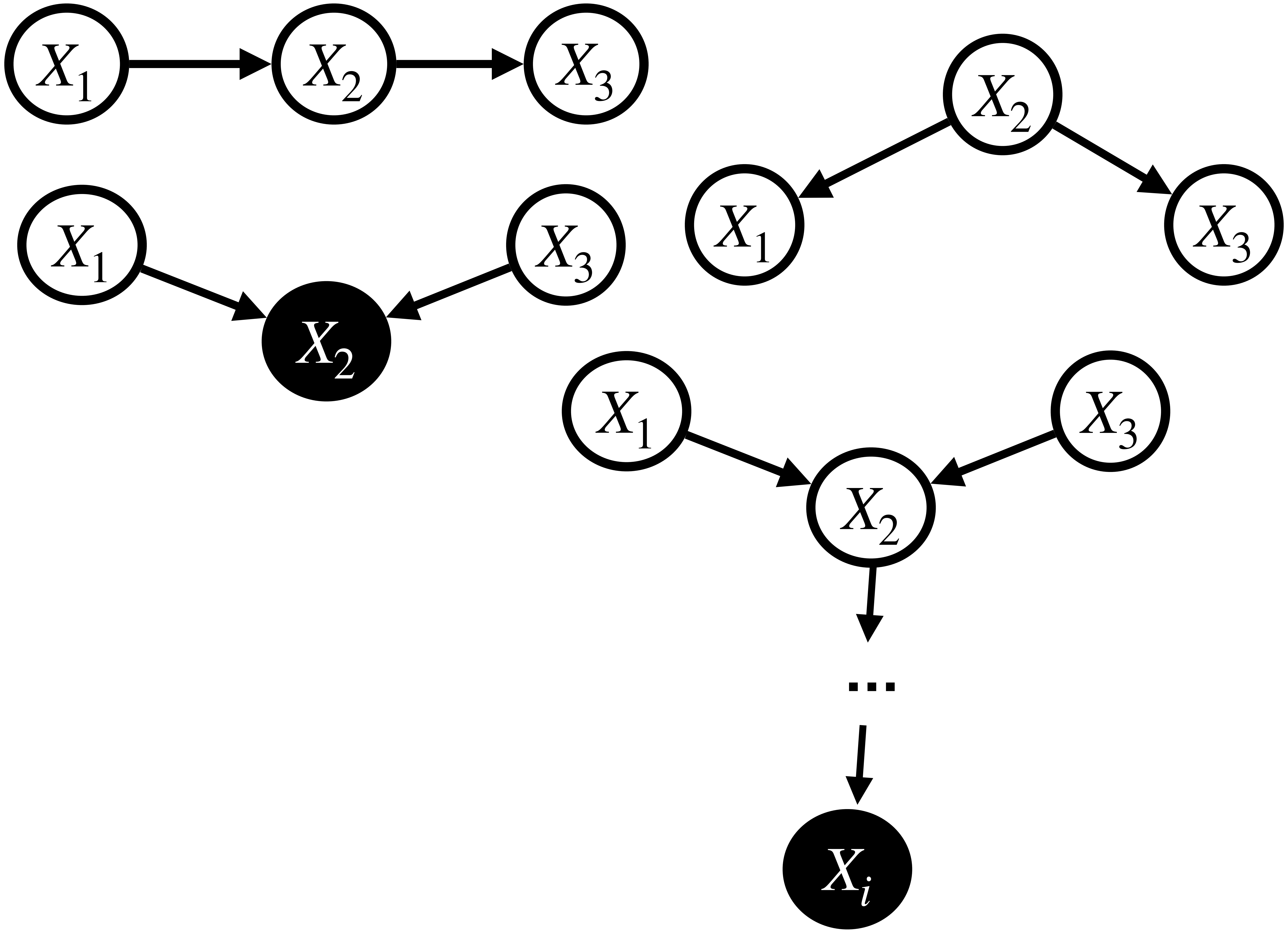
Active Path (1/3)

We will be checking all **undirected** paths between the two variables (i.e. ignoring the direction of the edges).

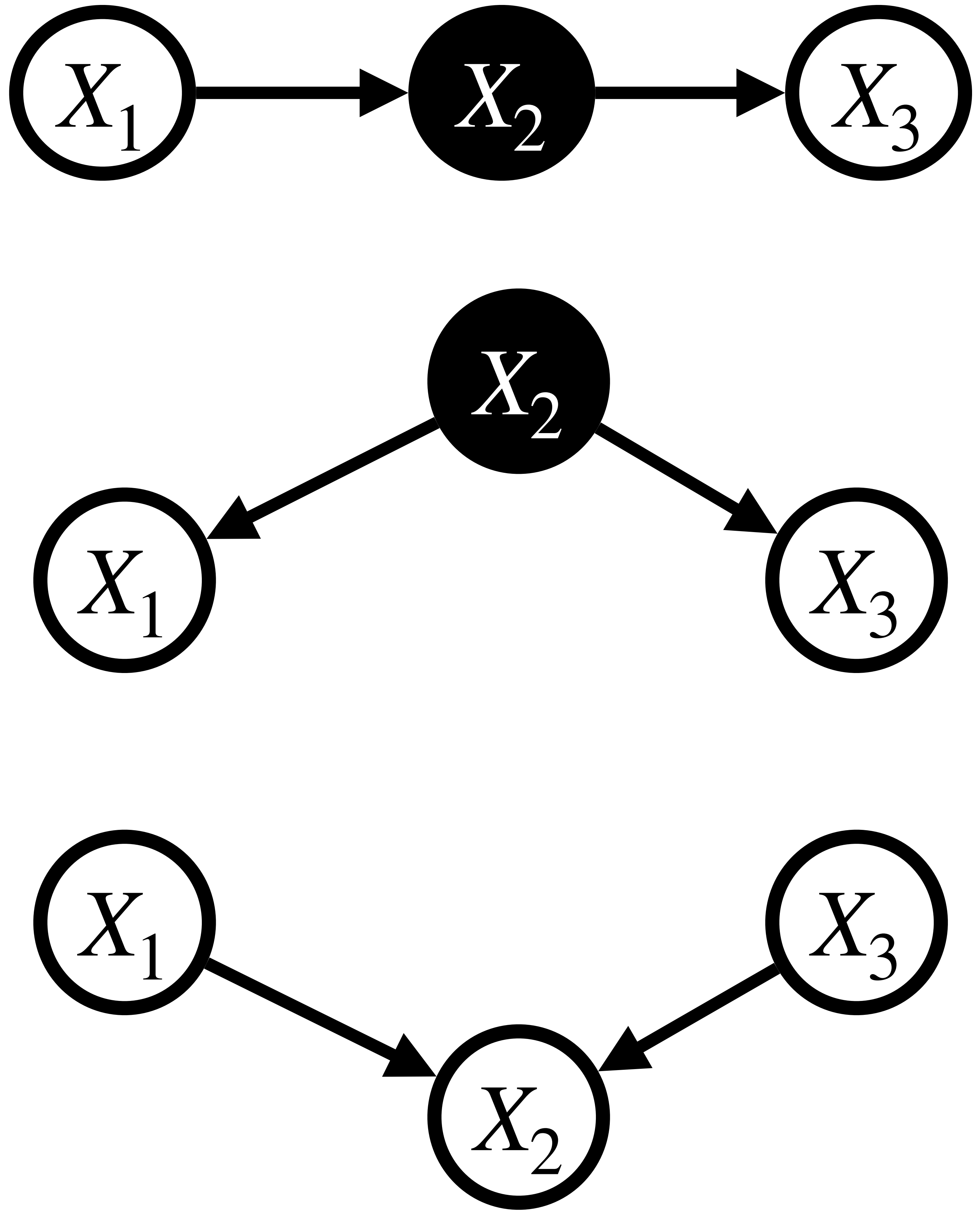
Terminology: Nodes which we condition on will be called **observed nodes** and the others **unobserved nodes**.

Active Path (2/3)

Active triples:



Blocked triples:

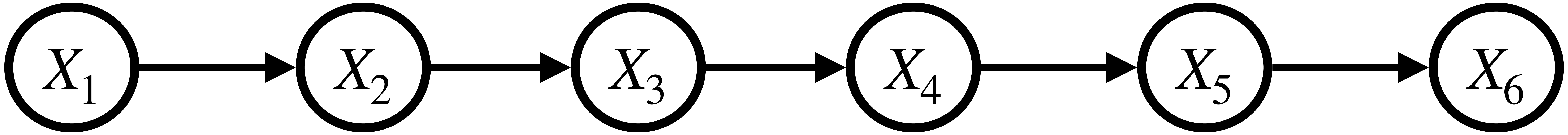


Active Path (3/3)

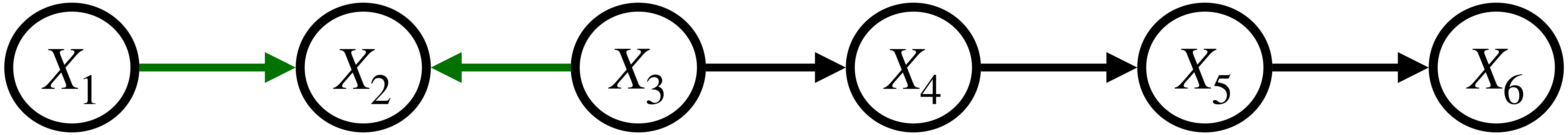
Definition: A path is active if all triples along it are active. Otherwise it is blocked.

EXAMPLES:

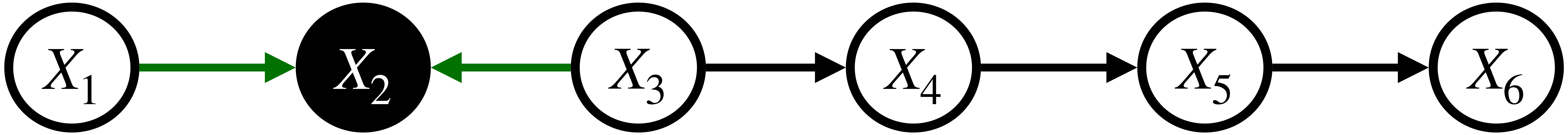
The path from X_1 to X_6 is active.



The path from X_1 to X_6 is blocked.



The path from X_1 to X_6 is active.

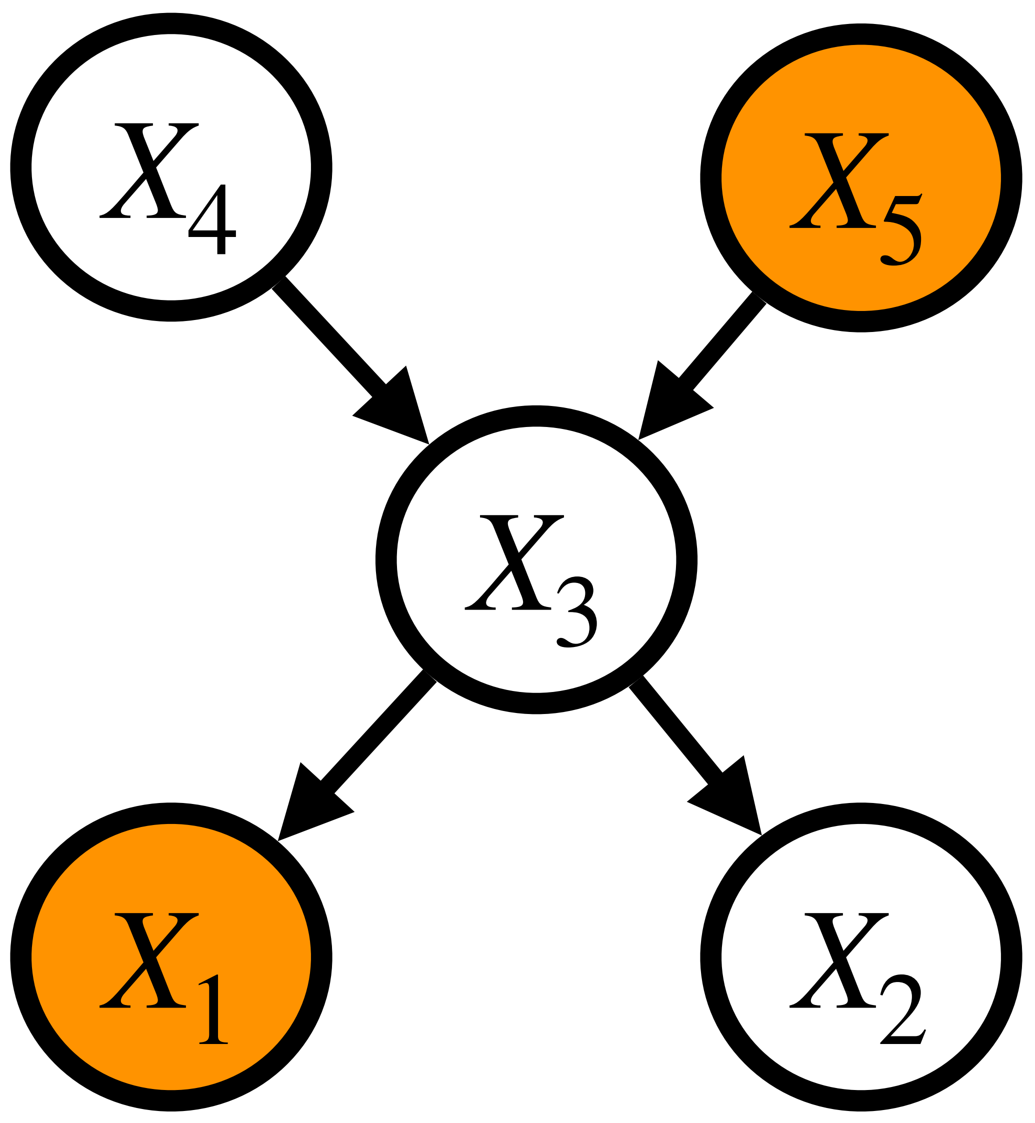


Part 2: Variable Elimination Algorithm (Intuition)

Marginal Inference

Problem: Given a BN on random variables $X_1, X_2, \ldots, X_n$, compute the probability $P_{X_{i_1}, X_{i_2}, \ldots, X_{i_k}}(x_{i_1}, x_{i_2}, \ldots, x_{i_k})$, where $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ is a subset of the random variables $X_1, X_2, \ldots, X_n$.

Example: Compute $P_{X_1, X_5}(x_1, x_5)$ from the BN shown here:



Let's Simplify Notation (1/2)

To simplify notation, we will assume that:

- We have a joint distribution, given by a BN, on random variables $Y_1, Y_2, \ldots, Y_k, Z_1, \ldots, Z_l$ *(this is the same as was before* $X_1, X_2, \ldots, X_n$*)*.
- We want to compute the marginal probability $P_{Y_1,\ldots,Y_k}(y_1, \ldots, y_k)$.
- We will call $Z_1, \ldots, Z_l$ *unobserved random variables.*

Let's Simplify Notation (2/2)

What we want to compute is now:

$$P_{Y_1,\ldots,Y_k}(y_1,\ldots,y_k) = \sum_{z_1}\sum_{z_2}\cdots\sum_{z_l} P(y_1,\ldots,y_k,z_1,\ldots,z_l)$$

where $P(y_1,\ldots,y_k,z_1,\ldots,z_l)$ is the joint probability given by the Bayesian network.

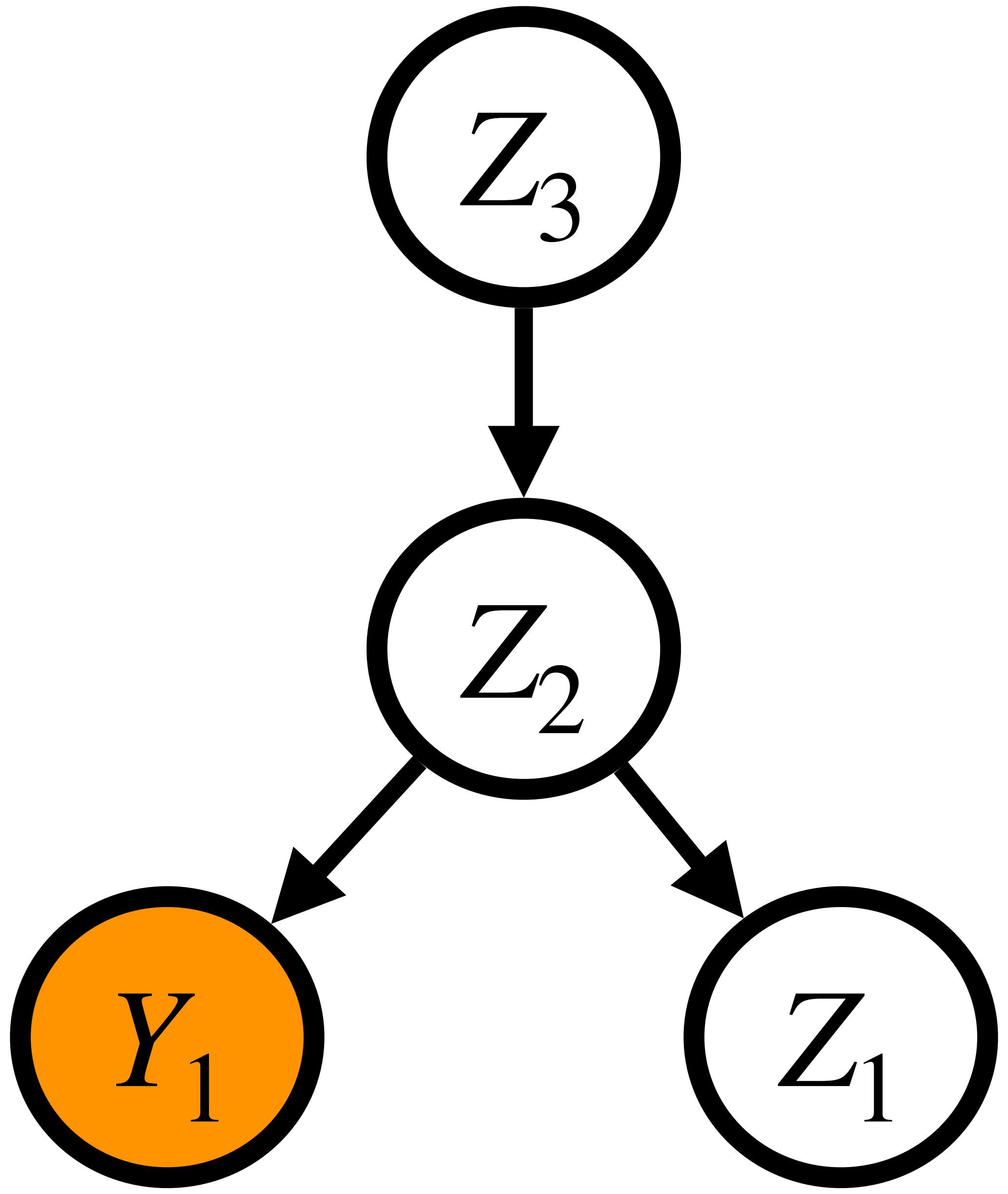
Naive Approach

Naive idea (we won't be able to do better in the worst case):

Compute the following sum explicitly:

$$P_{Y_1}(y_1) = \sum_{z_1} \sum_{z_2} \sum_{z_3} P(y_1, z_1, z_2, z_3).$$

This will have exponential complexity in the number of random variables.



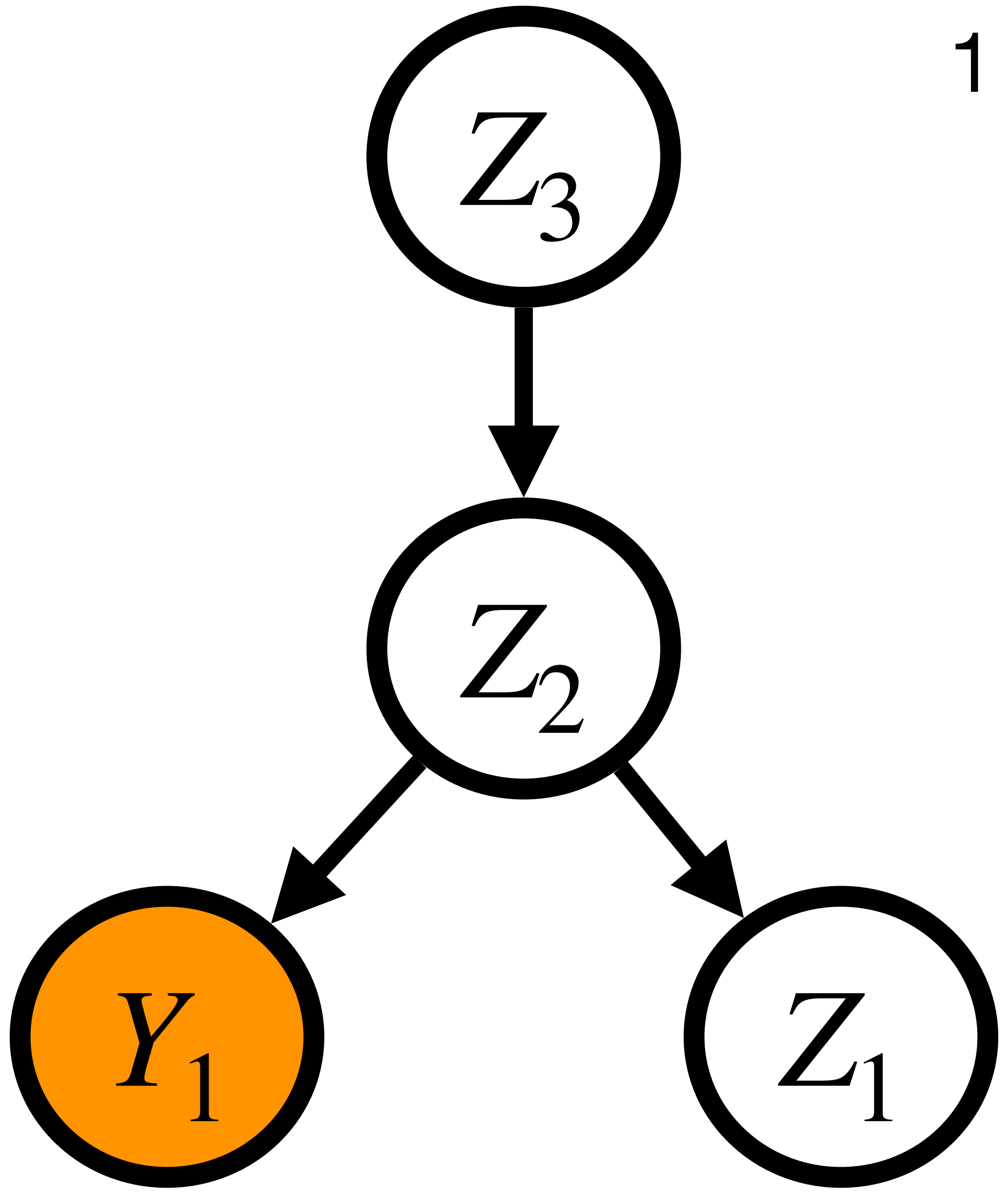
Naive Approach: Example (1/2)

Z_3	P_{Z_3}
0	0.4
1	0.6

Y_1	Z_2	$P_{Y_1\|Z_2}$
0	0	0.2
1	0	0.8
0	1	0.9
1	1	0.1

Z_1	Z_2	$P_{Z_1\|Z_2}$
0	0	0.5
1	0	0.5
0	1	0.1
1	1	0.9

Z_1	Z_2	$P_{Z_2\|Z_3}$
0	0	0.5
1	0	0.5
0	1	0.1
1	1	0.9



$$P[Y_1 = 1] = P_{Y_1}(1) = ?$$

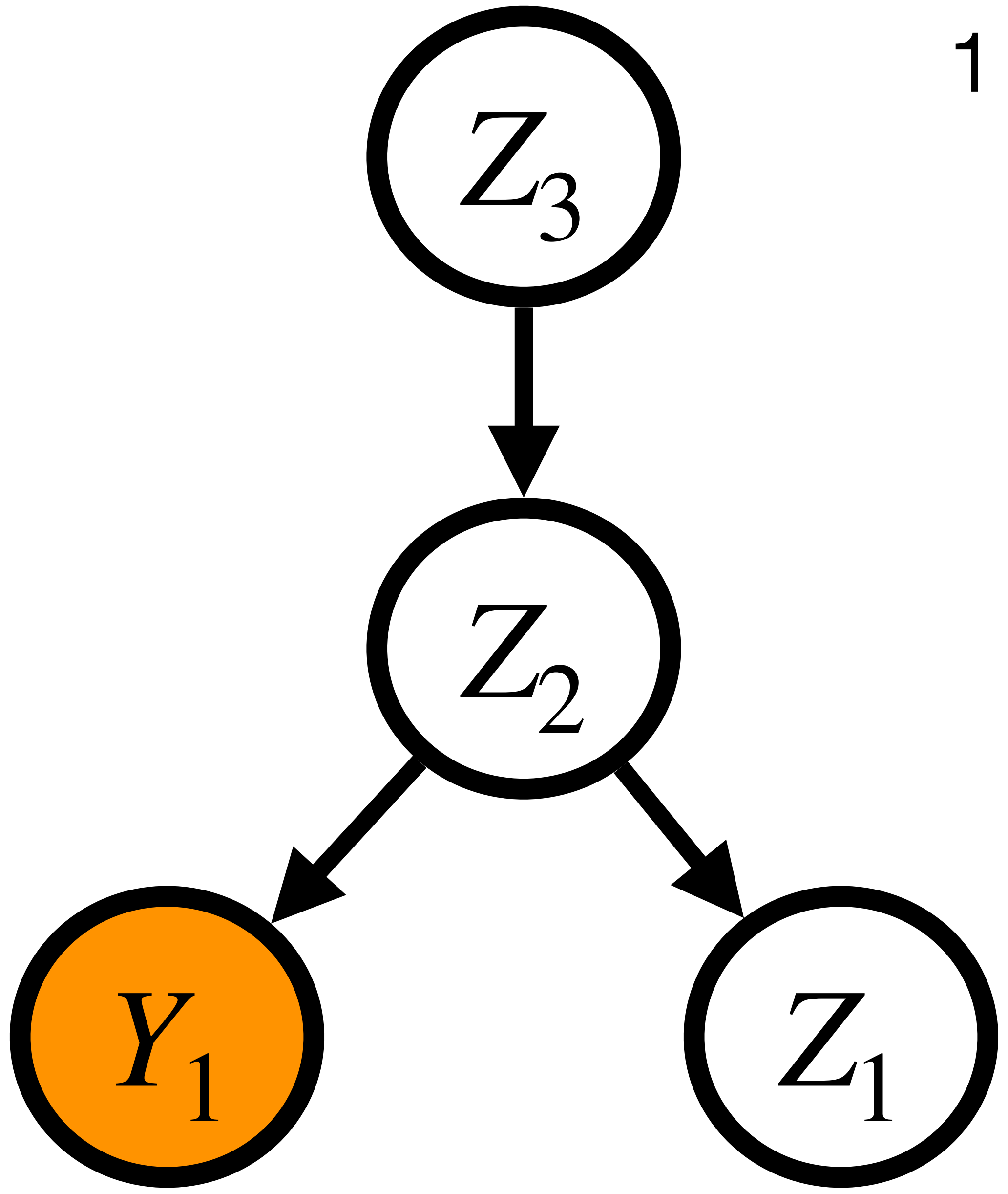
Naive Approach: Example (2/2)

Z_3	P_{Z_3}
0	0.4
1	0.6

Y_1	Z_2	$P_{Y_1\|Z_2}$
0	0	0.2
1	0	0.8
0	1	0.9
1	1	0.1

Z_1	Z_2	$P_{Z_1\|Z_2}$
0	0	0.5
1	0	0.5
0	1	0.1
1	1	0.9

Z_1	Z_2	$P_{Z_2\|Z_3}$
0	0	0.5
1	0	0.5
0	1	0.1
1	1	0.9



$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

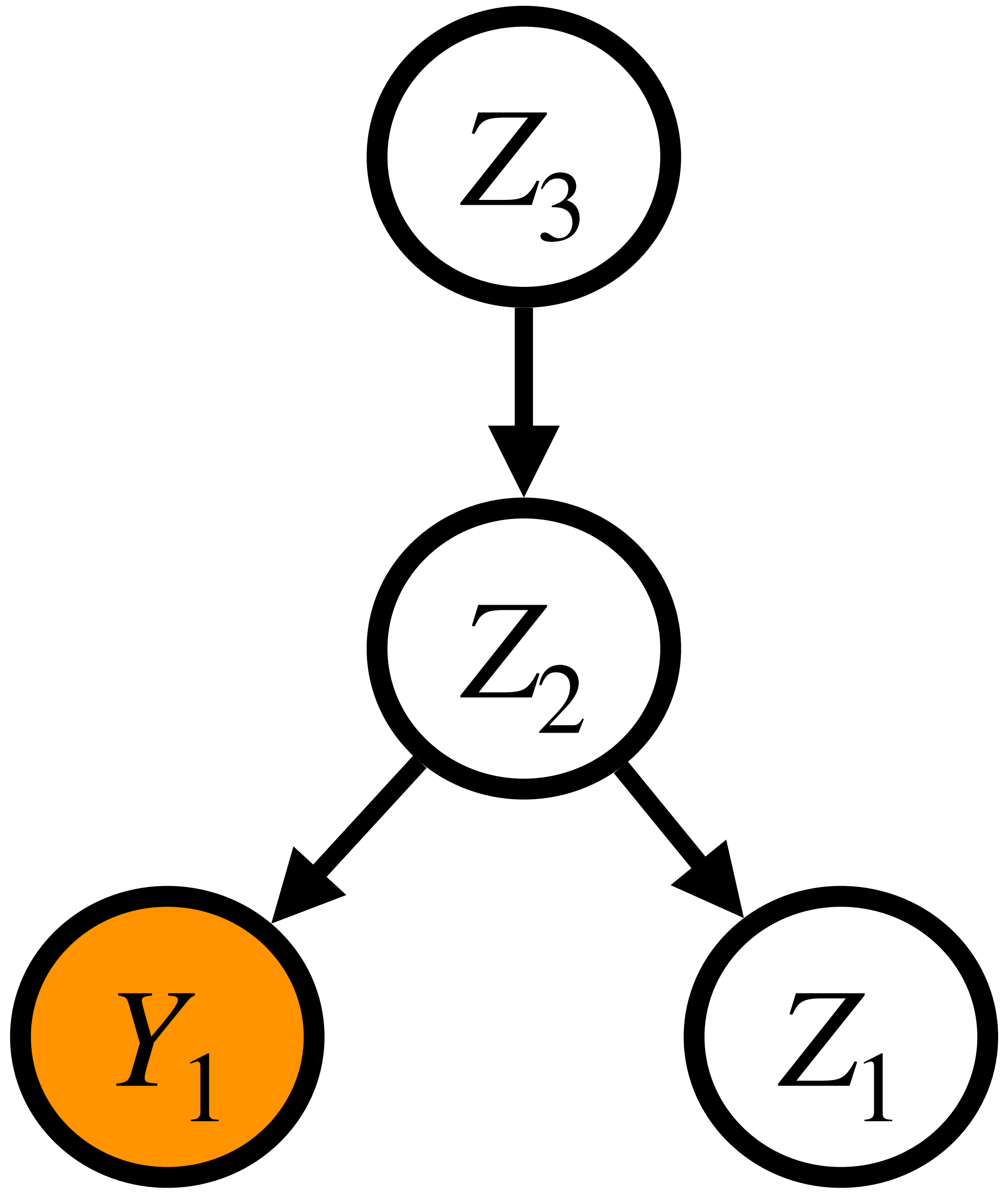
$$\sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1\,|\,z_2)P_{Z_1|Z_2}(z_1\,|\,z_2)P_{Z_2|Z_3}(z_2\,|\,z_3)P_{Z_3}(z_3)$$

$$= 0.8\cdot 0.5\cdot 0.5\cdot 0.4 + 0.8\cdot 0.5\cdot 0.1\cdot 0.6 + 0.1\cdot 0.1\cdot 0.5\cdot 0.4 +$$
$$+0.1\cdot 0.1\cdot 0.9\cdot 0.6 + 0.8\cdot 0.5\cdot 0.5\cdot 0.4 + 0.8\cdot 0.5\cdot 0.1\cdot 0.6$$
$$+0.1\cdot 0.9\cdot 0.5\cdot 0.4 + 0.1\cdot 0.9\cdot 0.9\cdot 0.6 = 0.282$$

We need $2^3 - 1 = 7$ additions and 24 multiplications...

Variable Elimination: Basic Idea (1/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1} \sum_{z_2=0}^{1} \sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$



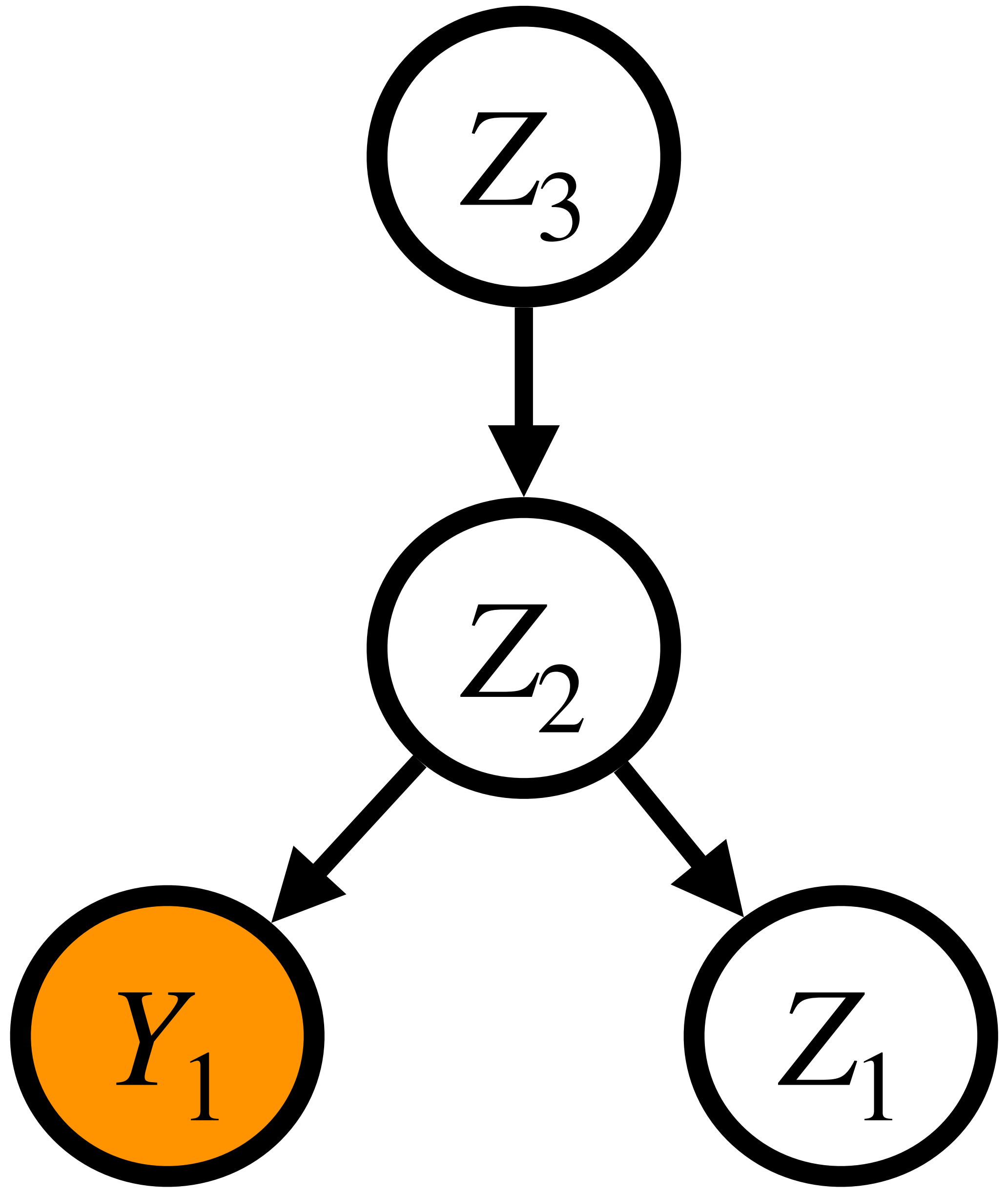
$$G_1(0) = 0.5 + 0.5 = 1$$

$$G_1(1) = 0.1 + 0.9 = 1$$

Variable Elimination: Basic Idea (1/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_1|Z_2}(z_1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) =$$



$$G_1(0) = 0.5 + 0.5 = 1$$

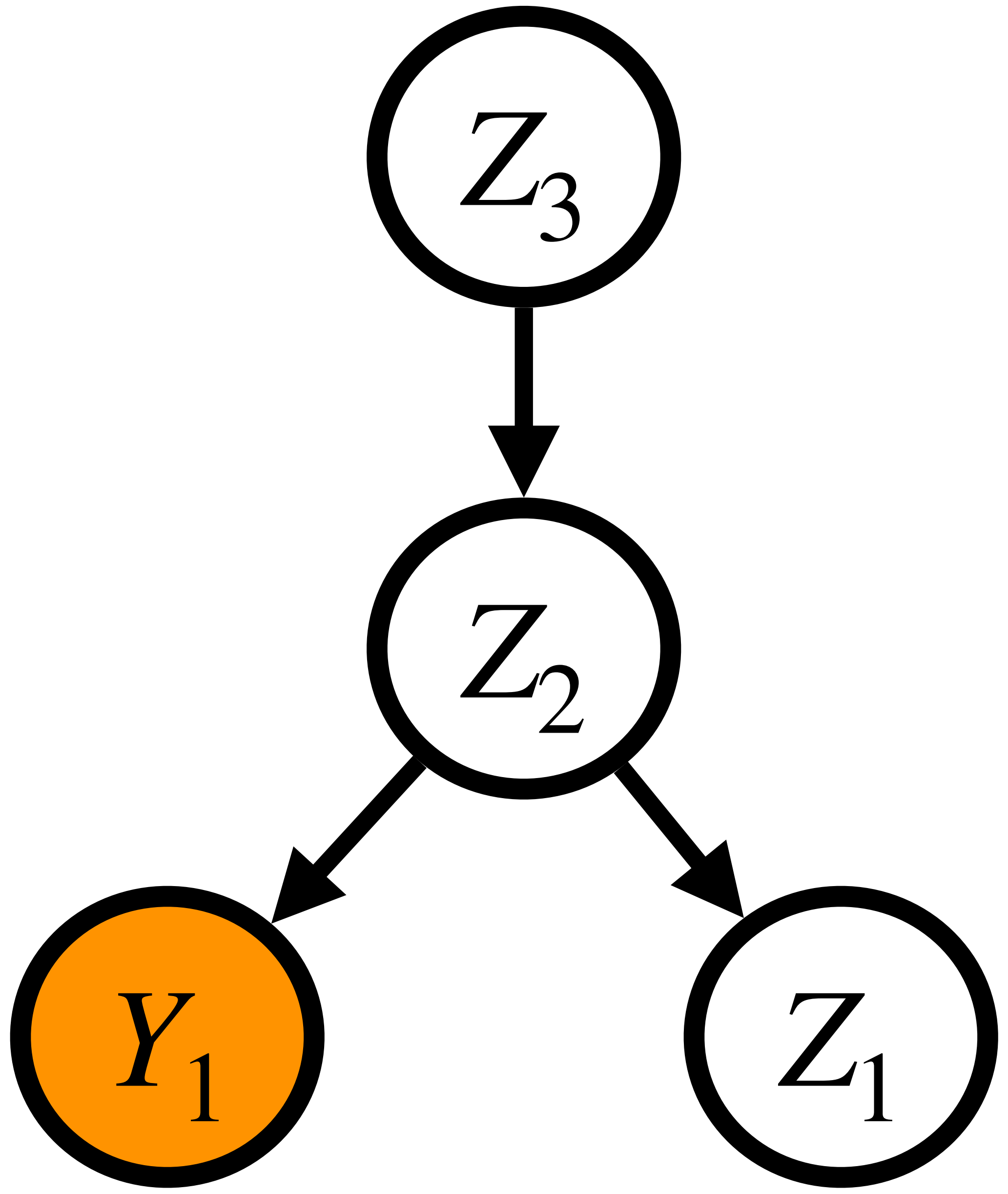
$$G_1(1) = 0.1 + 0.9 = 1$$

Variable Elimination: Basic Idea (1/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_1|Z_2}(z_1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) =$$

$$= \sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) \underbrace{\sum_{z_1=0}^{1} P_{Z_1|Z_2}(z_1 \mid z_2)}_{=G(z_2)}$$



$$G_1(0) = 0.5 + 0.5 = 1$$

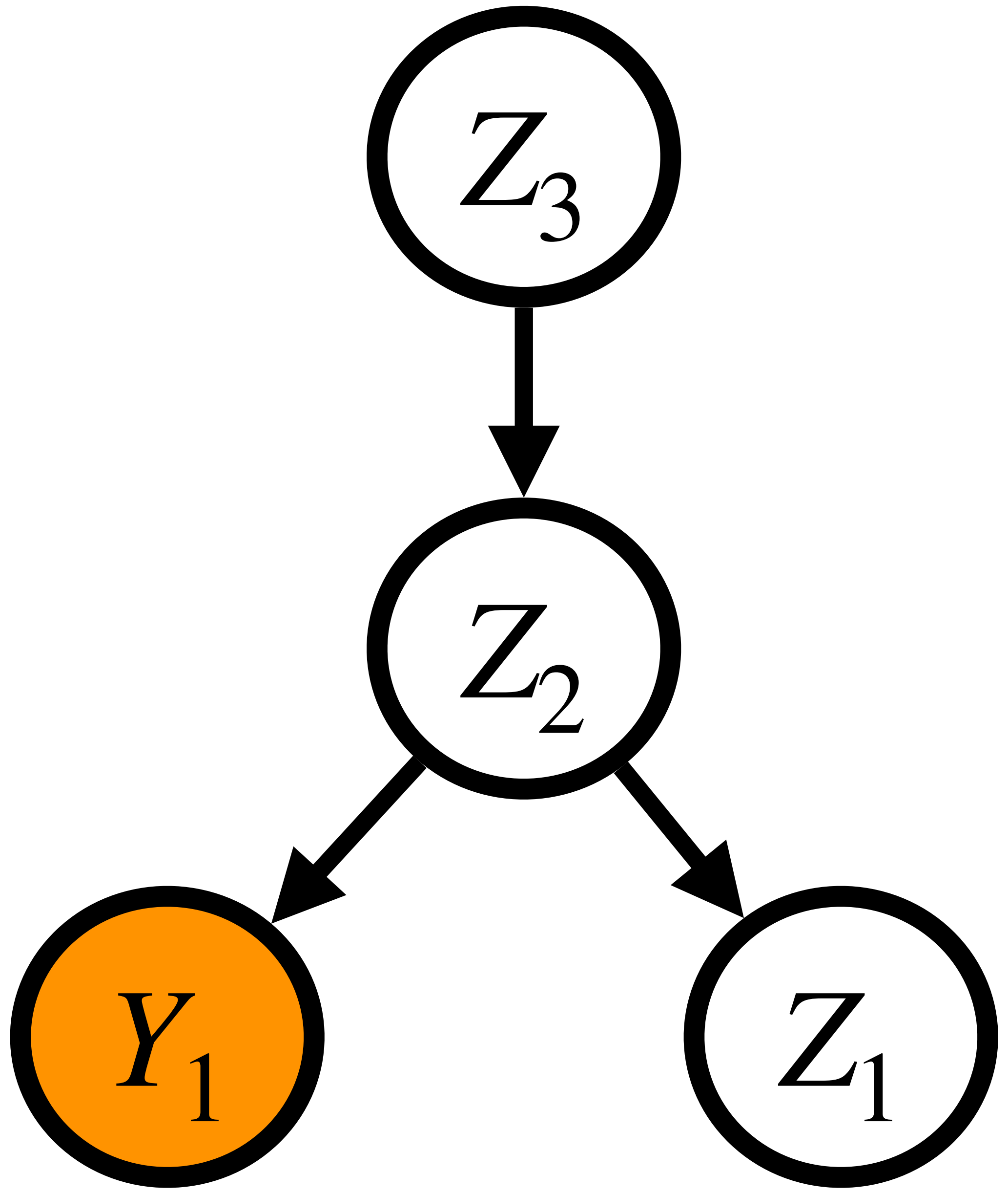
$$G_1(1) = 0.1 + 0.9 = 1$$

Variable Elimination: Basic Idea (2/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1} \sum_{z_2=0}^{1} \sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1} \sum_{z_2=0}^{1} \sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_1|Z_2}(z_1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) =$$

$$= \sum_{z_2=0}^{1} \sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) \underbrace{\sum_{z_1=0}^{1} P_{Z_1|Z_2}(z_1 \mid z_2)}_{=G_1(z_2)}$$



$$G_1(0) = 0.5 + 0.5 = 1$$

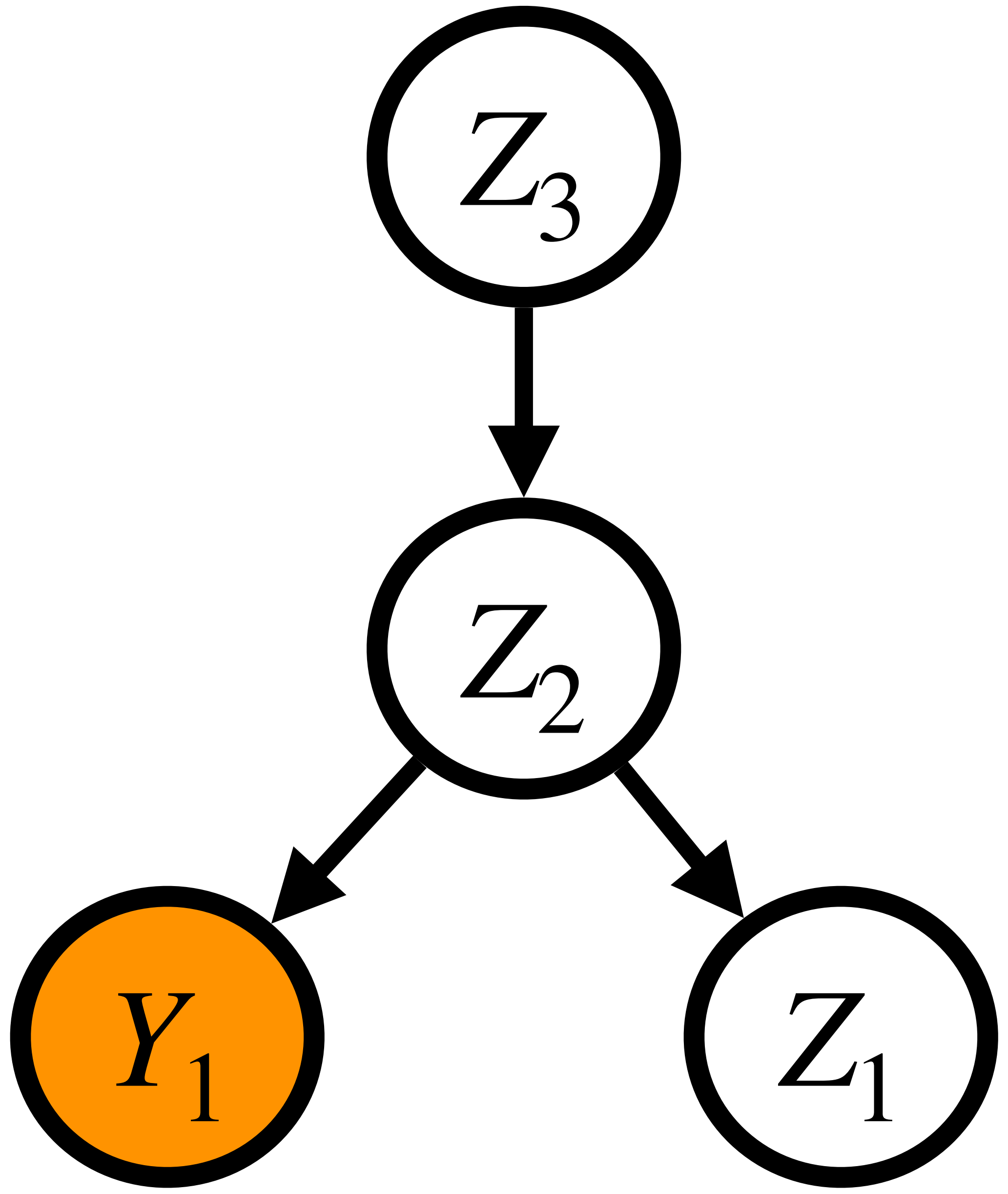
$$G_1(1) = 0.1 + 0.9 = 1$$

Variable Elimination: Basic Idea (3/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_1|Z_2}(z_1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) =$$

$$= \sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) G_1(z_2)$$



$$G_1(0) = 0.5 + 0.5 = 1$$

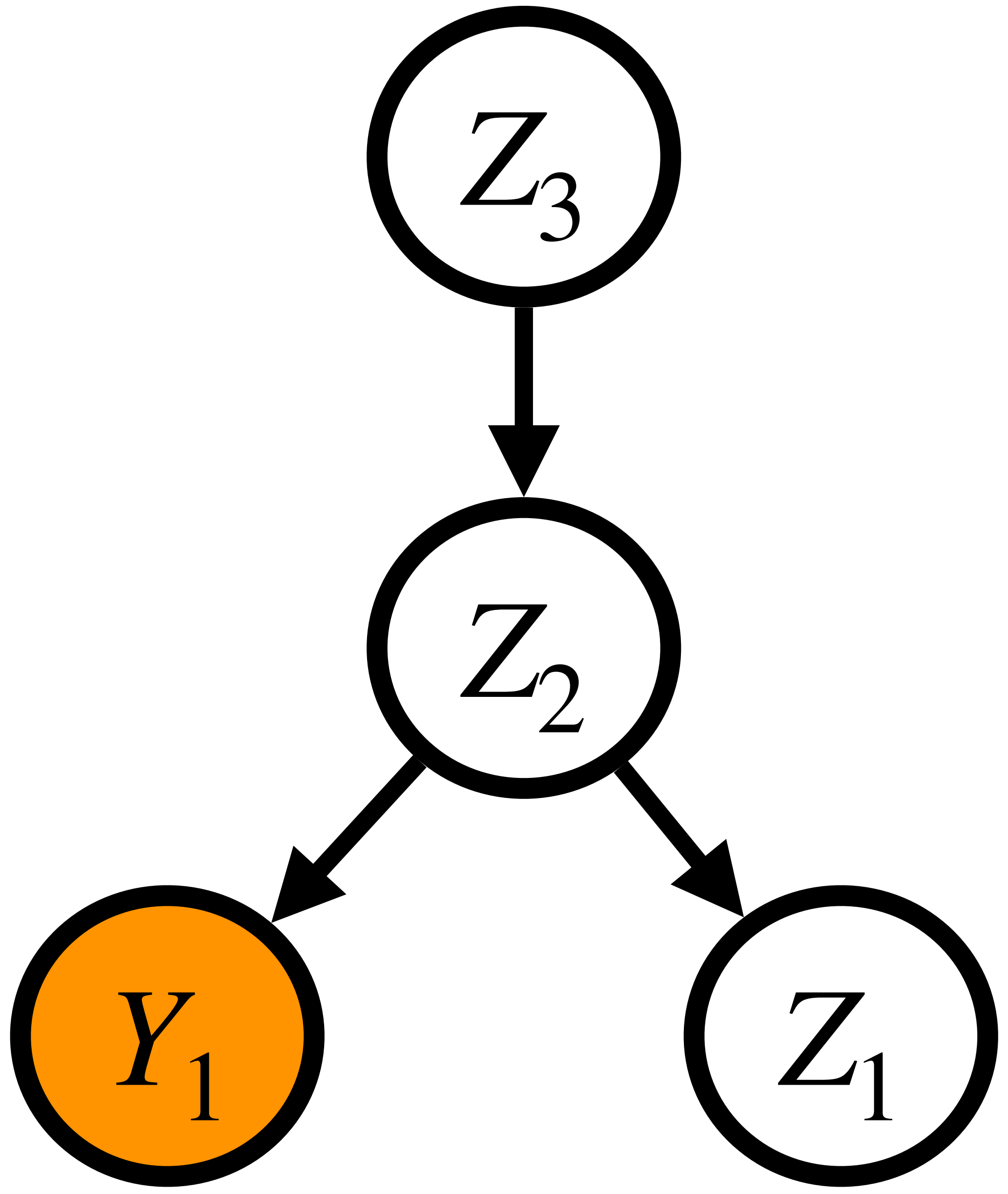
$$G_1(1) = 0.1 + 0.9 = 1$$

Variable Elimination: Basic Idea (4/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_1|Z_2}(z_1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) =$$

$$= \sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) G_1(z_2)$$



$$G_1(0) = 0.5 + 0.5 = 1$$

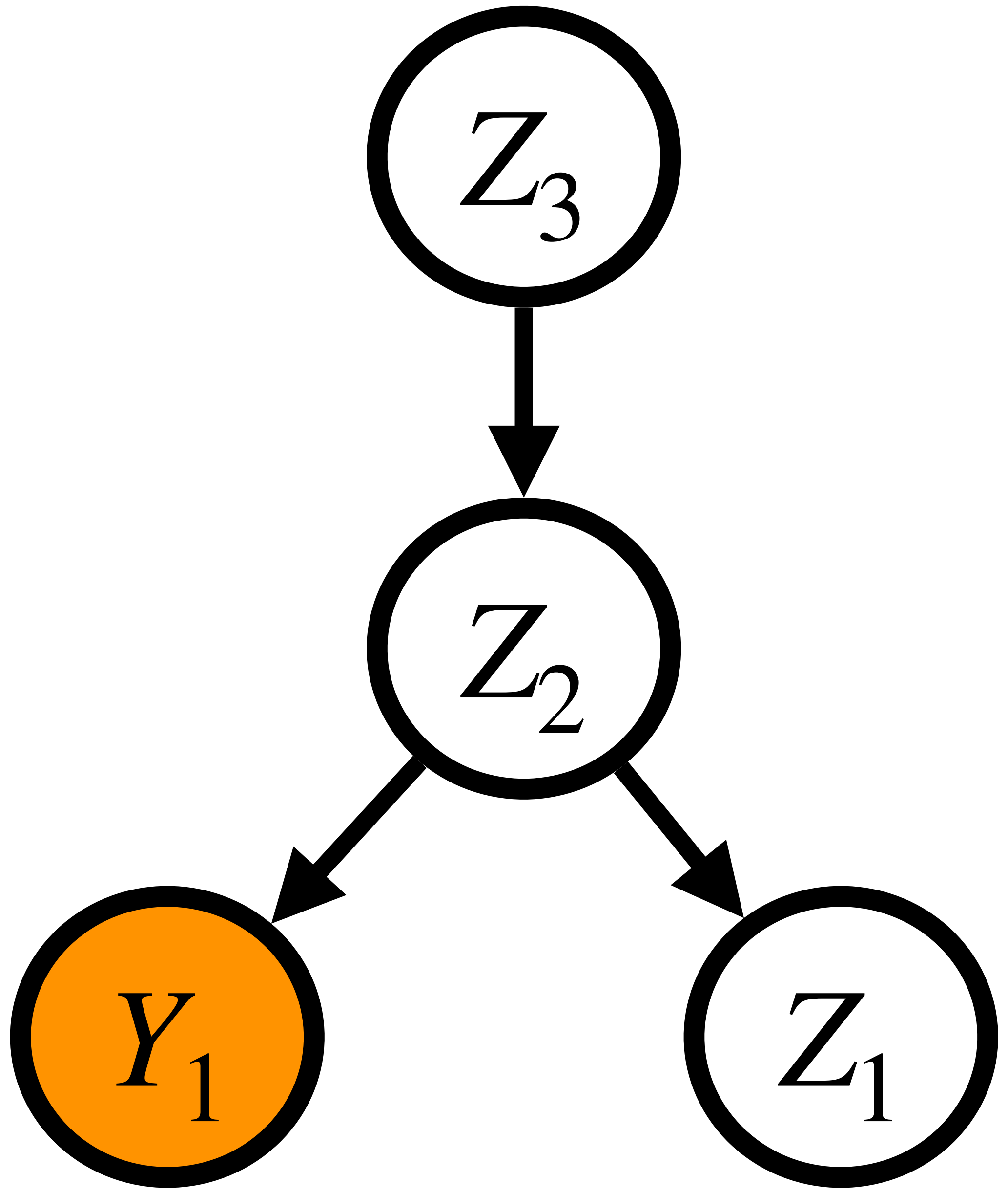
$$G_1(1) = 0.1 + 0.9 = 1$$

Variable Elimination: Basic Idea (5/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_1|Z_2}(z_1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) =$$

$$= \sum_{z_3=0}^{1} P_{Z_3}(z_3) \sum_{z_2=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) G_1(z_2)$$



$$G_1(0) = 0.5 + 0.5 = 1$$

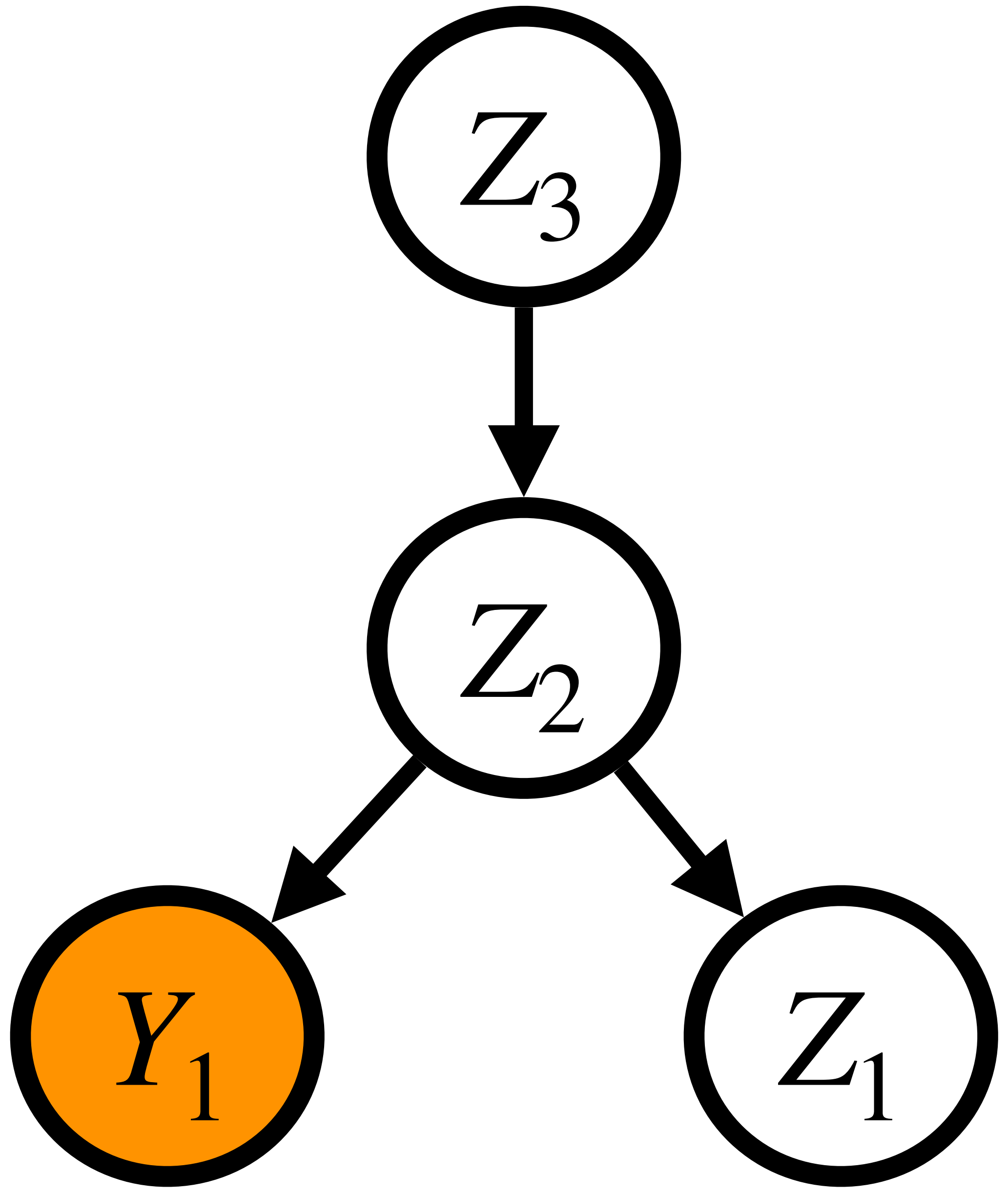
$$G_1(1) = 0.1 + 0.9 = 1$$

Variable Elimination: Basic Idea (6/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_1|Z_2}(z_1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) =$$

$$= \sum_{z_3=0}^{1} P_{Z_3}(z_3) \underbrace{\sum_{z_2=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) G_1(z_2)}_{G_2(z_3)}$$



$$G_2(0) = 0.8 \cdot 0.5 \cdot 1 + 0.1 \cdot 0.5 \cdot 1 = 0.45$$

$$G_2(1) = 0.8 \cdot 0.1 \cdot 1 + 0.1 \cdot 0.9 \cdot 1 = 0.17$$

$$G_1(0) = 0.5 + 0.5 = 1$$

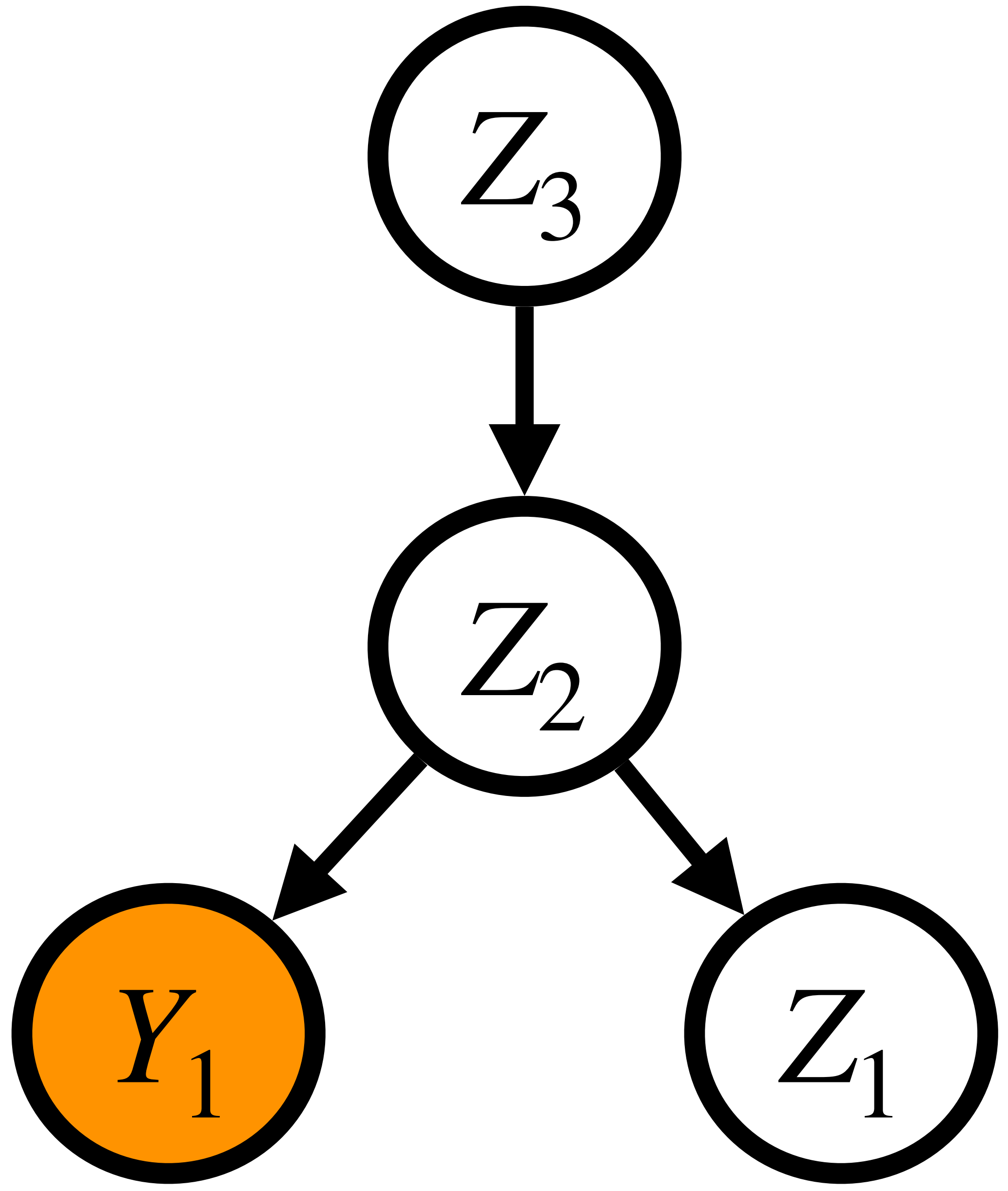
$$G_1(1) = 0.1 + 0.9 = 1$$

Variable Elimination: Basic Idea (7/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_1|Z_2}(z_1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) =$$

$$= \sum_{z_3=0}^{1} P_{Z_3}(z_3) G_2(z_3)$$



$$G_2(0) = 0.8 \cdot 0.5 \cdot 1 + 0.1 \cdot 0.5 \cdot 1 = 0.45 \qquad G_1(0) = 0.5 + 0.5 = 1$$

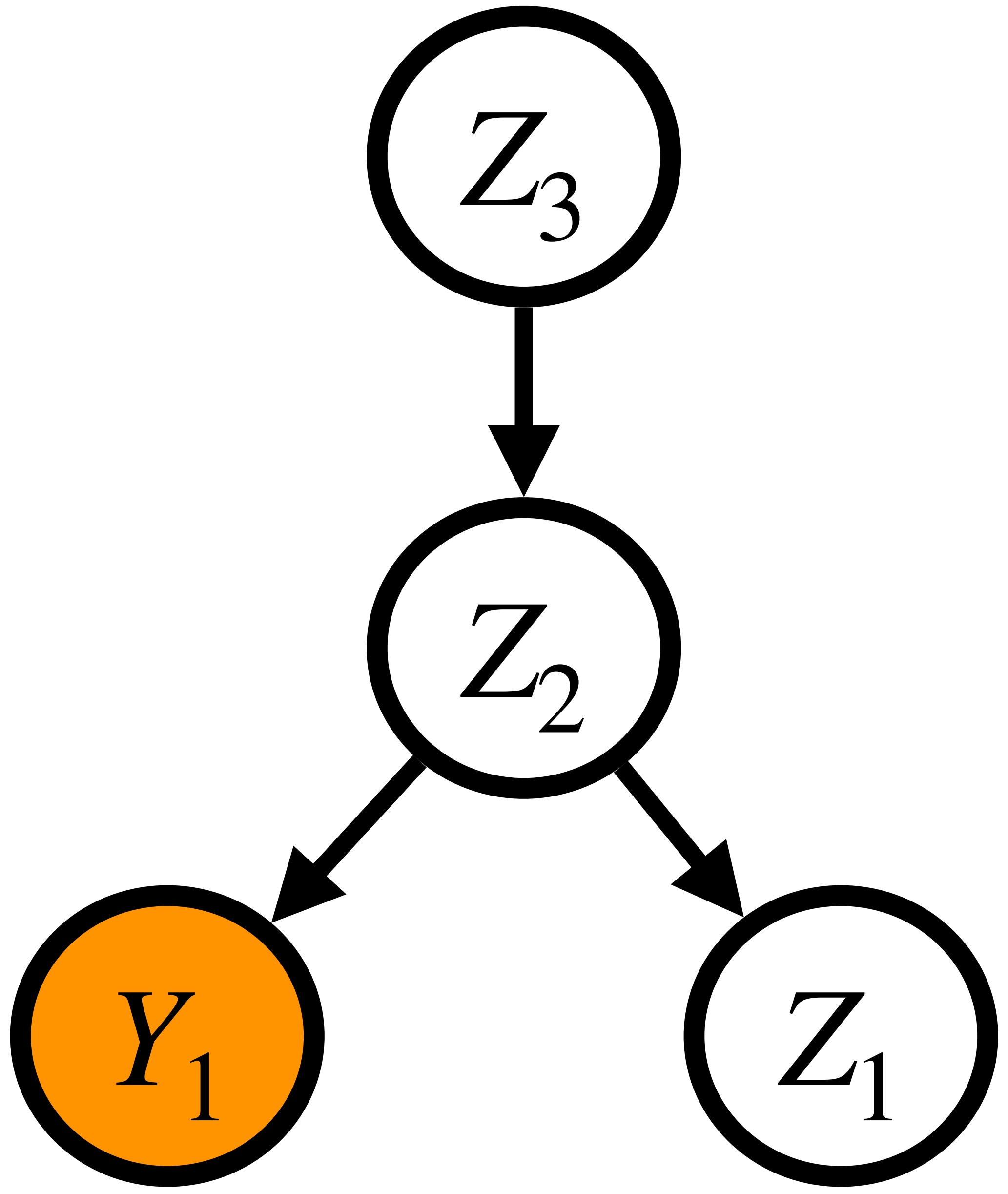
$$G_2(1) = 0.8 \cdot 0.1 \cdot 1 + 0.1 \cdot 0.9 \cdot 1 = 0.17 \qquad G_1(1) = 0.1 + 0.9 = 1$$

Variable Elimination: Basic Idea (8/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_1|Z_2}(z_1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) =$$

$$= \sum_{z_3=0}^{1} P_{Z_3}(z_3) G_2(z_3)$$



$$G_2(0) = 0.8 \cdot 0.5 \cdot 1 + 0.1 \cdot 0.5 \cdot 1 = 0.45$$

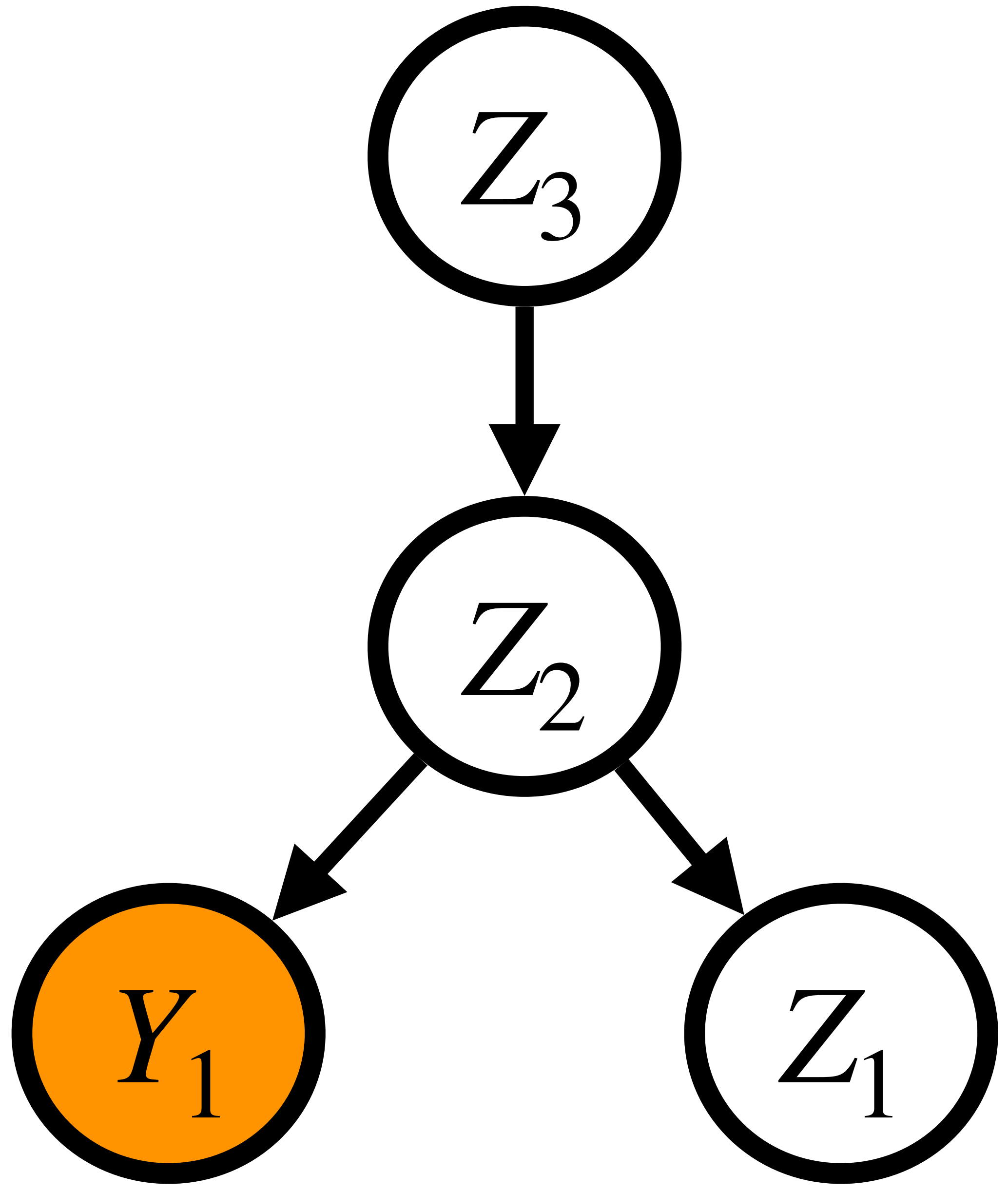
$$G_2(1) = 0.8 \cdot 0.1 \cdot 1 + 0.1 \cdot 0.9 \cdot 1 = 0.17$$

Variable Elimination: Basic Idea (9/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1} \sum_{z_2=0}^{1} \sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1} \sum_{z_2=0}^{1} \sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_1|Z_2}(z_1 \mid z_2) P_{Z_2|Z_3}(z_2 \mid z_3) P_{Z_3}(z_3) =$$

$$= \sum_{z_3=0}^{1} P_{Z_3}(z_3) G_2(z_3)$$



$$G_2(0) = 0.8 \cdot 0.5 \cdot 1 + 0.1 \cdot 0.5 \cdot 1 = 0.45$$

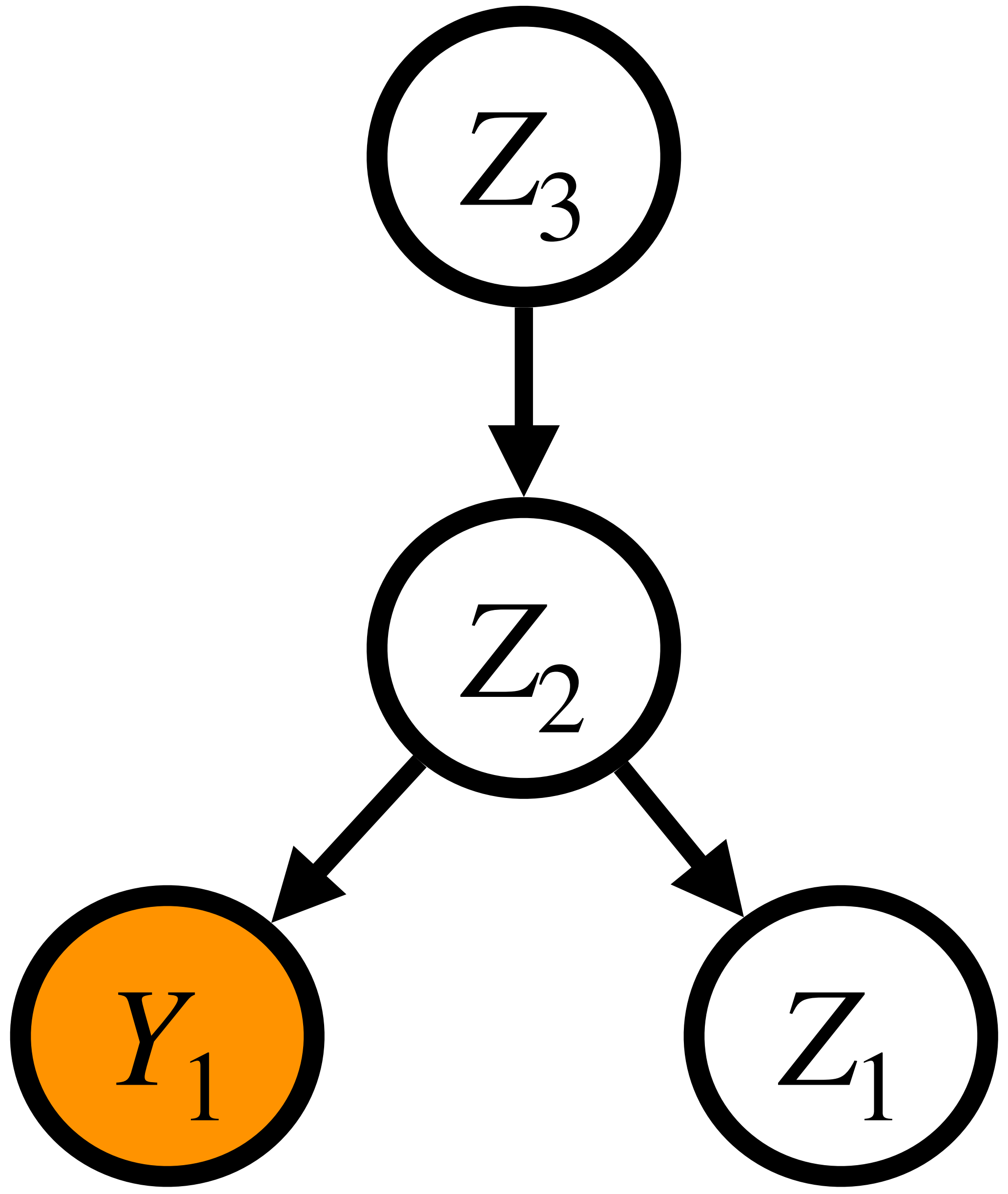
$$G_2(1) = 0.8 \cdot 0.1 \cdot 1 + 0.1 \cdot 0.9 \cdot 1 = 0.17$$

Variable Elimination: Basic Idea (10/10)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2)P_{Z_1|Z_2}(z_1 \mid z_2)P_{Z_2|Z_3}(z_2 \mid z_3)P_{Z_3}(z_3) =$$

$$= \sum_{z_3=0}^{1} P_{Z_3}(z_3)G_2(z_3)$$



$$P_{Y_1}(1) = 0.4 \cdot 0.45 + 0.6 \cdot 0.17 = 0.282$$

$$G_2(0) = 0.8 \cdot 0.5 \cdot 1 + 0.1 \cdot 0.5 \cdot 1 = 0.45$$

$$G_2(1) = 0.8 \cdot 0.1 \cdot 1 + 0.1 \cdot 0.9 \cdot 1 = 0.17$$

How Many Operations Did We Need?

$$G_2(0) = 0.8 \cdot 0.5 \cdot 1 + 0.1 \cdot 0.5 \cdot 1 = 0.45$$
$$G_2(1) = 0.8 \cdot 0.1 \cdot 1 + 0.1 \cdot 0.9 \cdot 1 = 0.17$$

$$G_1(0) = 0.5 + 0.5 = 1$$
$$G_1(1) = 0.1 + 0.9 = 1$$

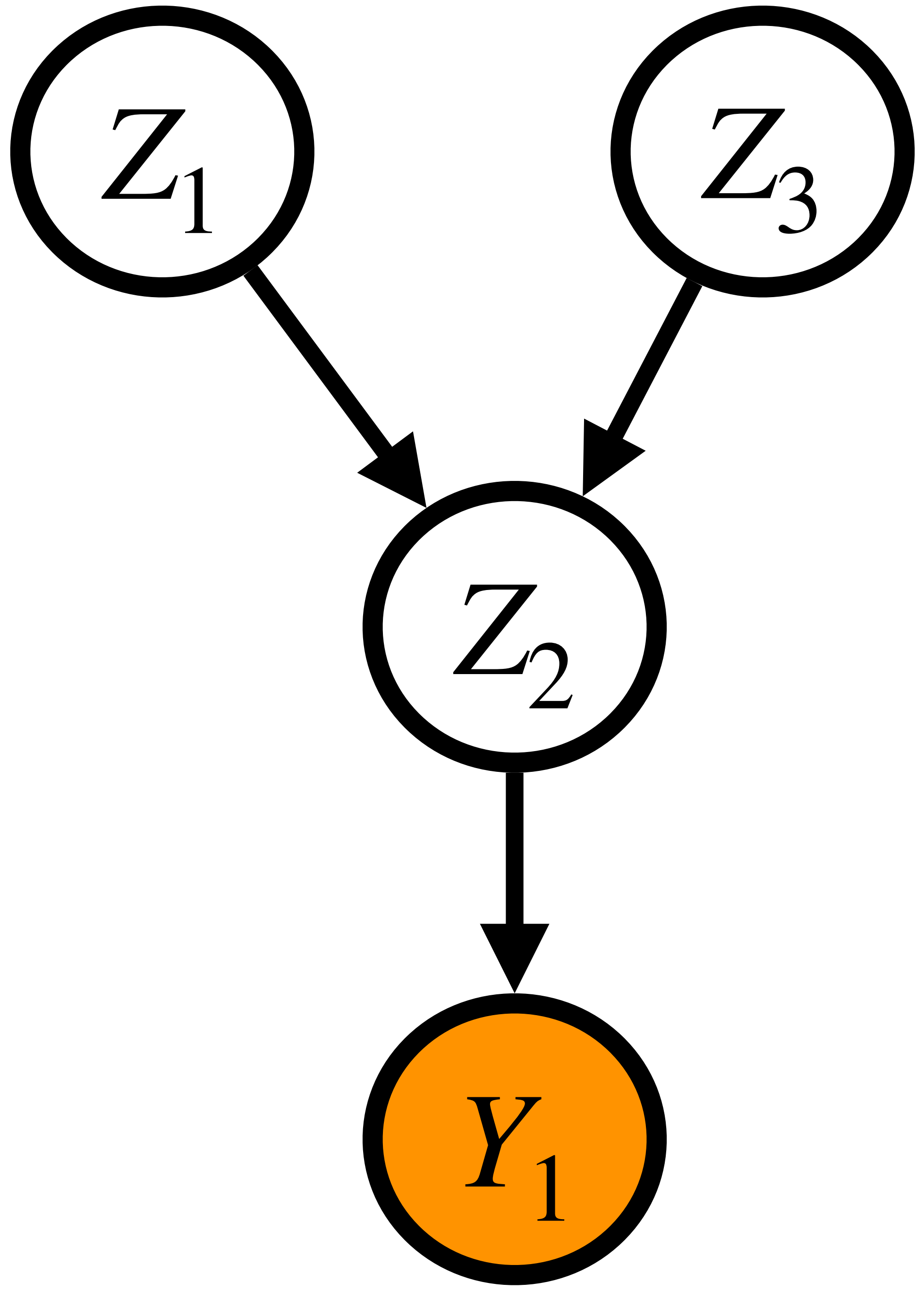
$$P_{Y_1}(1) = 0.4 \cdot 0.45 + 0.6 \cdot 0.17 = 0.282$$

10 multiplications and 5 additions (we needed 24 multiplications and 7 additions fot the naive approach!)

Variable Elimination: Example II (1/8)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_1,Z_3}(z_2 \mid z_1, z_3) P_{Z_1}(z_1) P_{Z_3}(z_3)$$

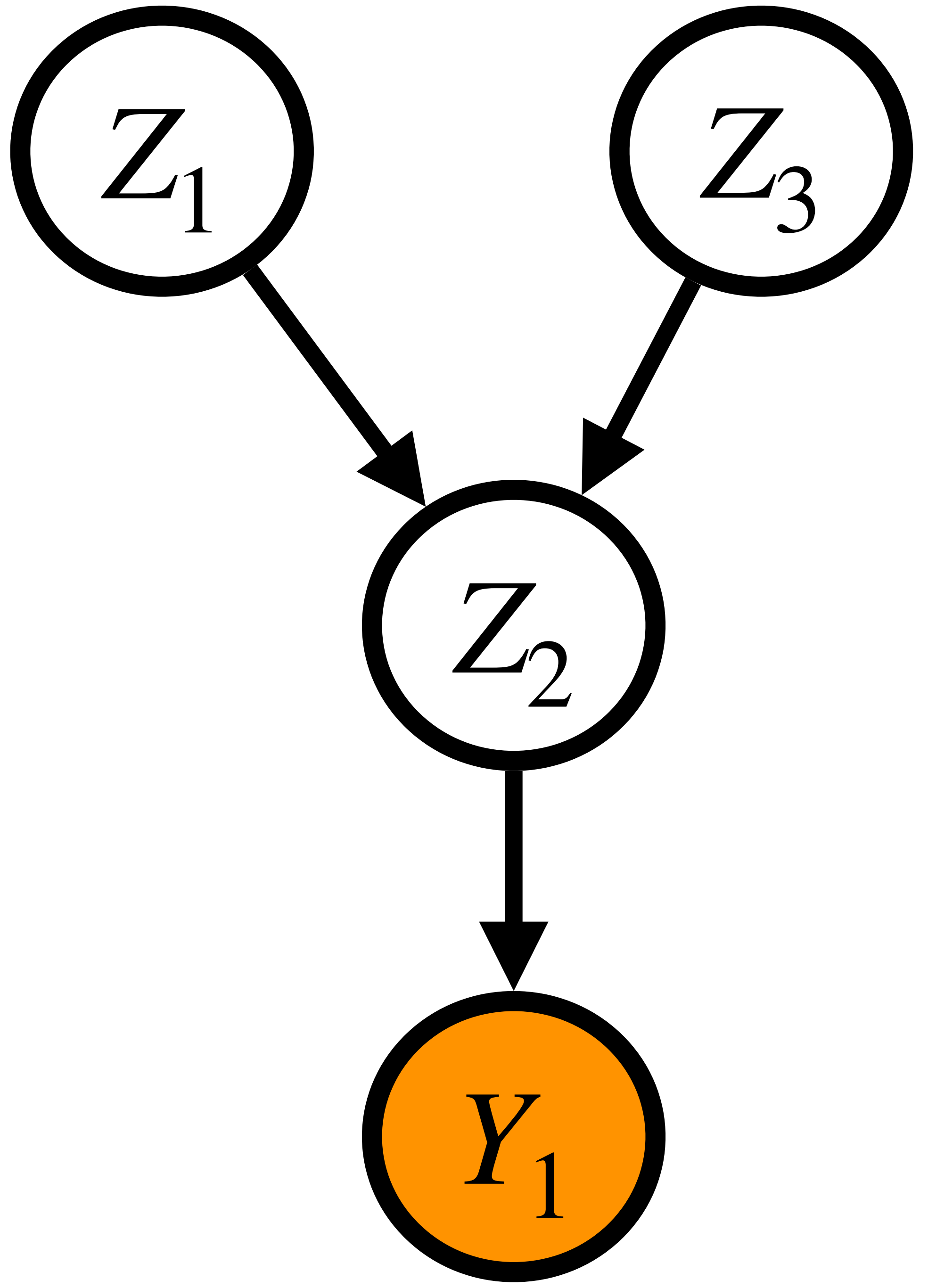


Variable Elimination: Example II (2/8)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_1,Z_3}(z_2 \mid z_1, z_3) P_{Z_1}(z_1) P_{Z_3}(z_3) =$$

$$= \sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_3}(z_3) \sum_{z_1=0}^{1} P_{Z_1}(z_1) P_{Z_2|Z_1,Z_3}(z_2 \mid z_1, z_3)$$

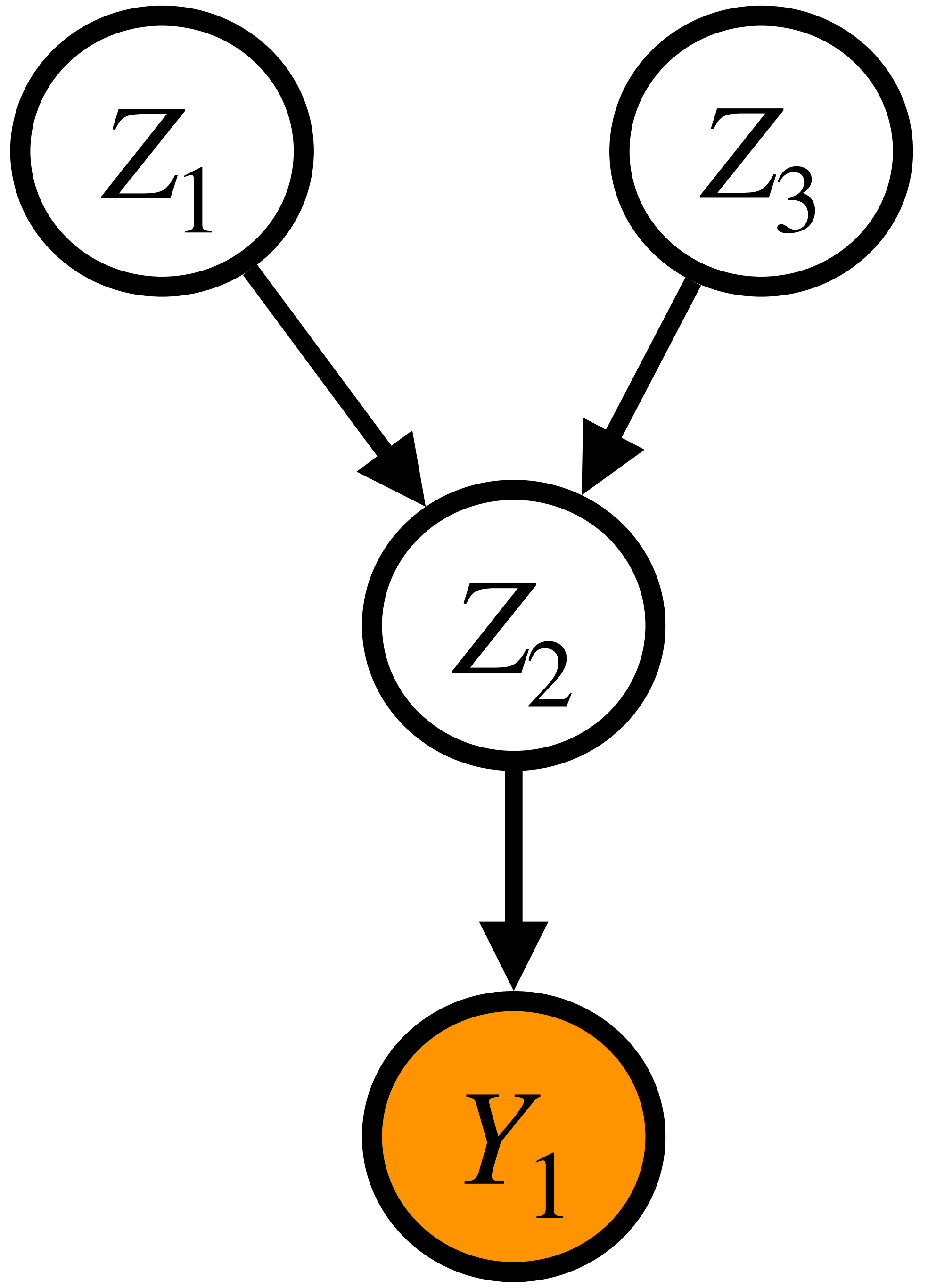


Variable Elimination: Example II (3/8)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_1,Z_3}(z_2 \mid z_1, z_3) P_{Z_1}(z_1) P_{Z_3}(z_3) =$$

$$= \sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_3}(z_3) \underbrace{\sum_{z_1=0}^{1} P_{Z_1}(z_1) P_{Z_2|Z_1,Z_3}(z_2 \mid z_1, z_3)}_{=G_1(z_2, z_3)}$$



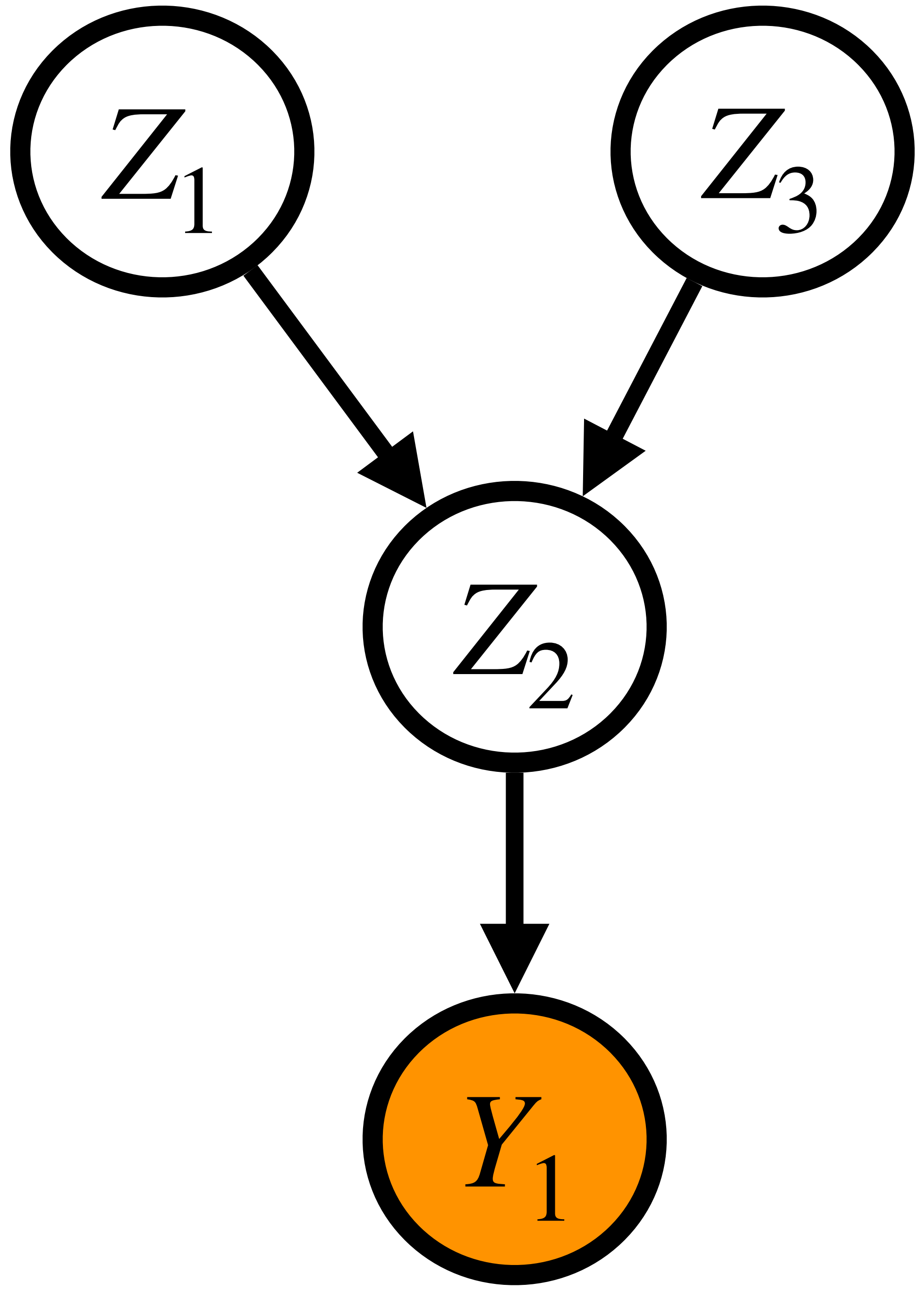
$$G_1(0,0) = \ldots, \quad G_1(0,1) = \ldots$$
$$G_1(1,0) = \ldots, \quad G_1(1,1) = \ldots$$

Variable Elimination: Example II (4/8)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_1,Z_3}(z_2 \mid z_1, z_3) P_{Z_1}(z_1) P_{Z_3}(z_3) =$$

$$= \sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_3}(z_3) G_1(z_2, z_3)$$

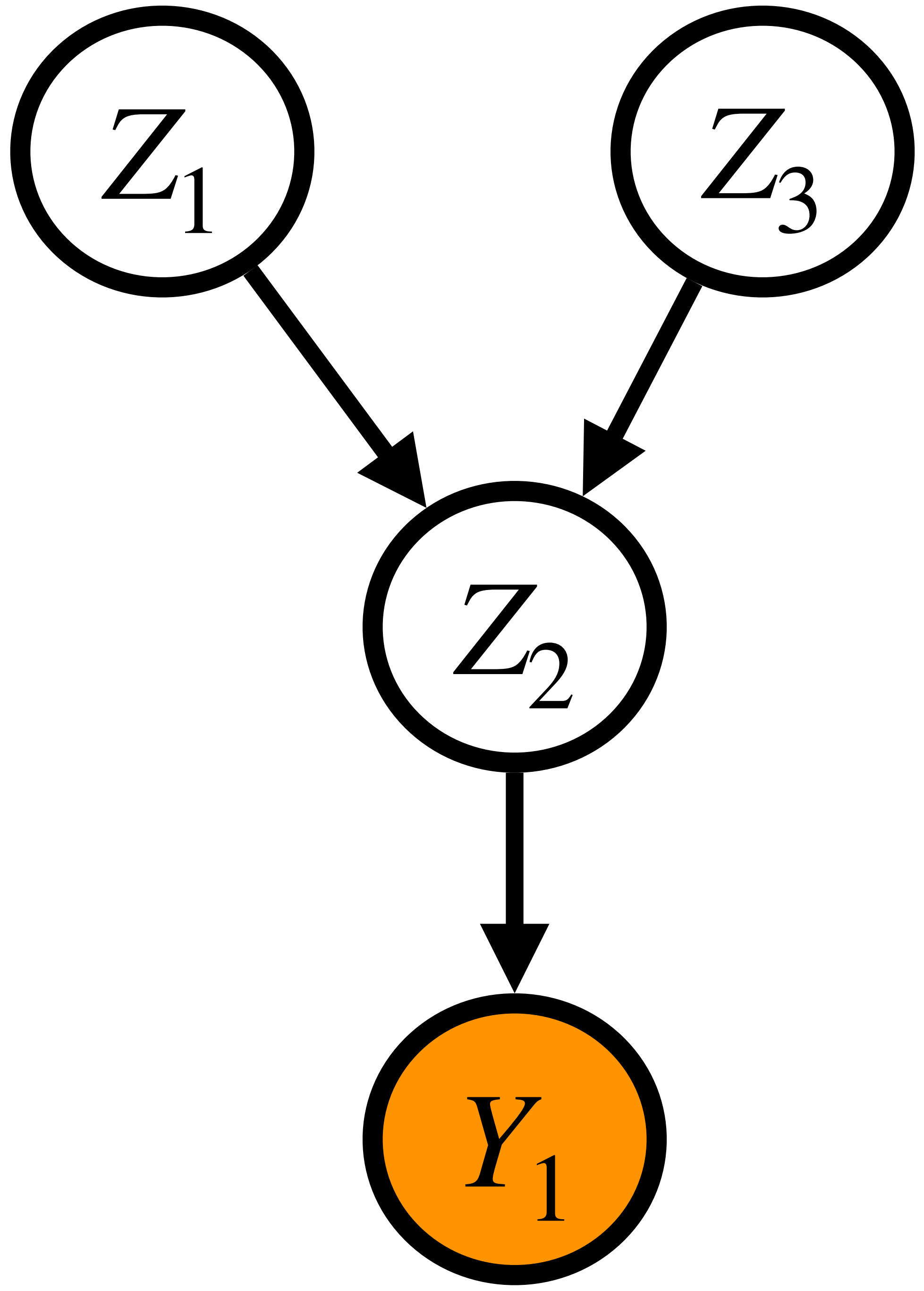


Variable Elimination: Example II (5/8)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_1,Z_3}(z_2 \mid z_1, z_3) P_{Z_1}(z_1) P_{Z_3}(z_3) =$$

$$= \sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_3}(z_3) G_1(z_2, z_3)$$

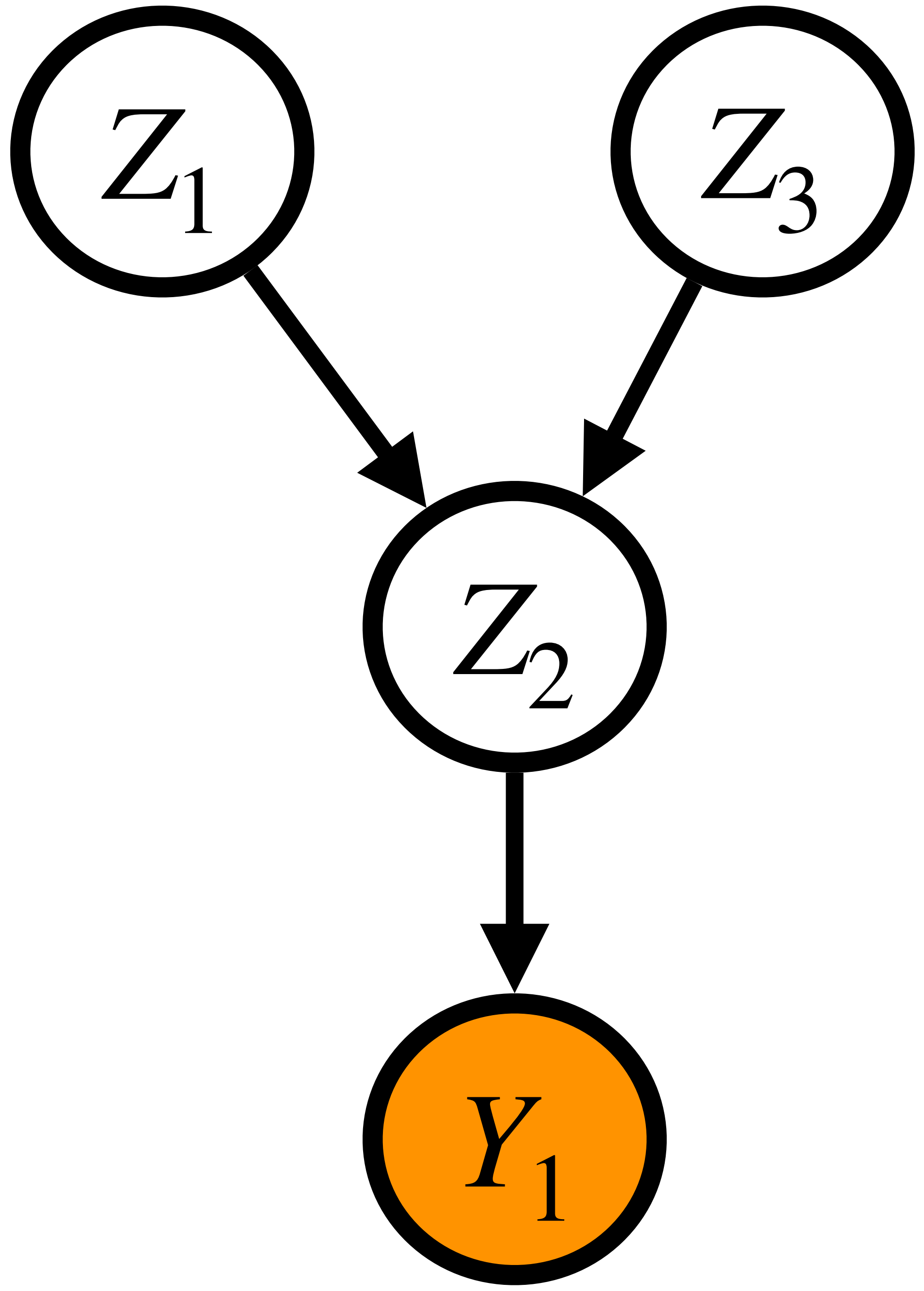


Variable Elimination: Example II (6/8)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_1,Z_3}(z_2 \mid z_1, z_3) P_{Z_1}(z_1) P_{Z_3}(z_3) =$$

$$= \sum_{z_3=0}^{1} P_{Z_3}(z_3) \sum_{z_2=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) G_1(z_2, z_3)$$



Variable Elimination: Example II (7/8)

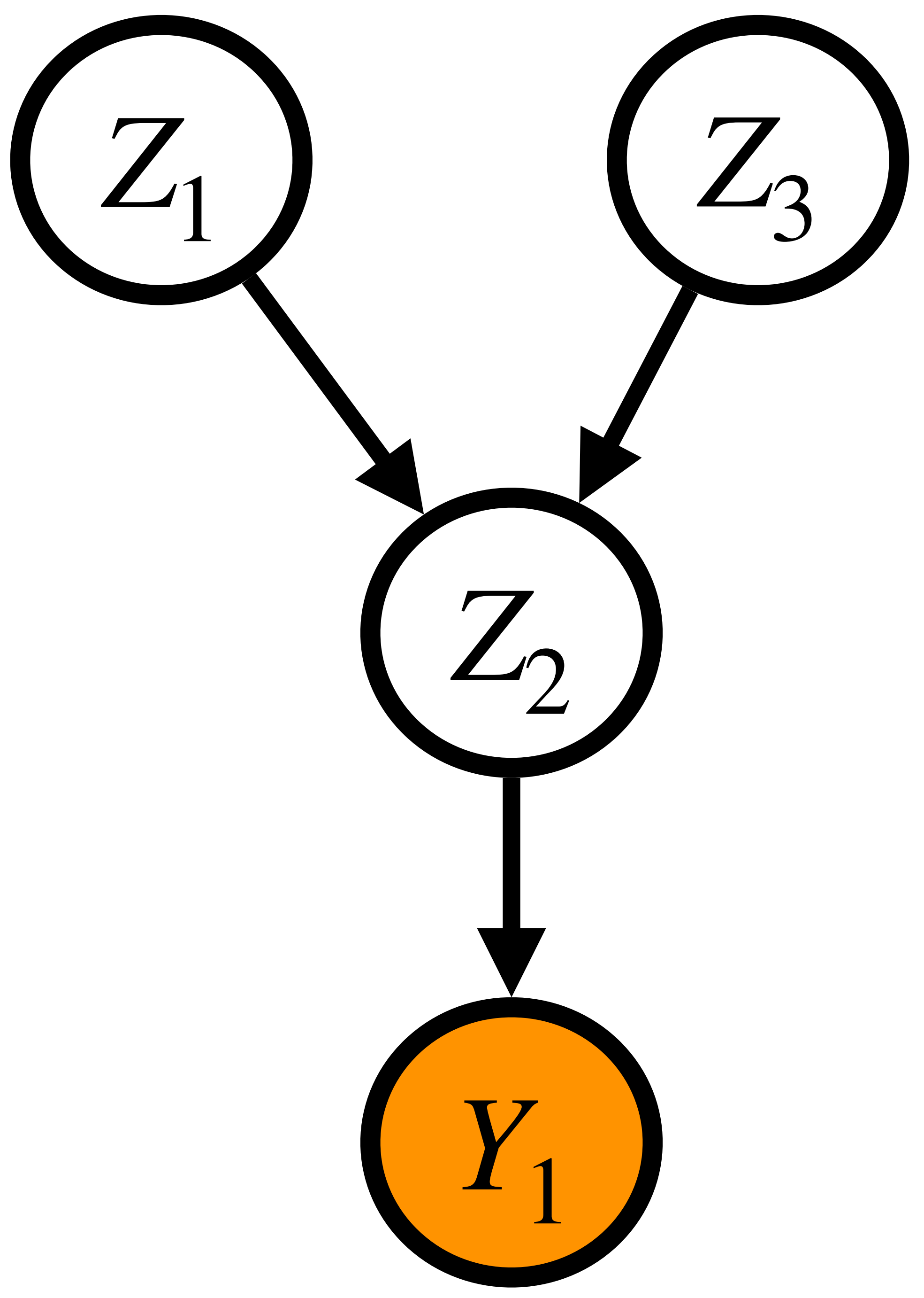
$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_1,Z_3}(z_2 \mid z_1, z_3) P_{Z_1}(z_1) P_{Z_3}(z_3) =$$

$$= \sum_{z_3=0}^{1} P_{Z_3}(z_3) \underbrace{\sum_{z_2=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) G_1(z_2, z_3)}_{=G_2(z_3)}$$

$$G_2(0) = \ldots,$$

$$G_2(1) = \ldots,$$

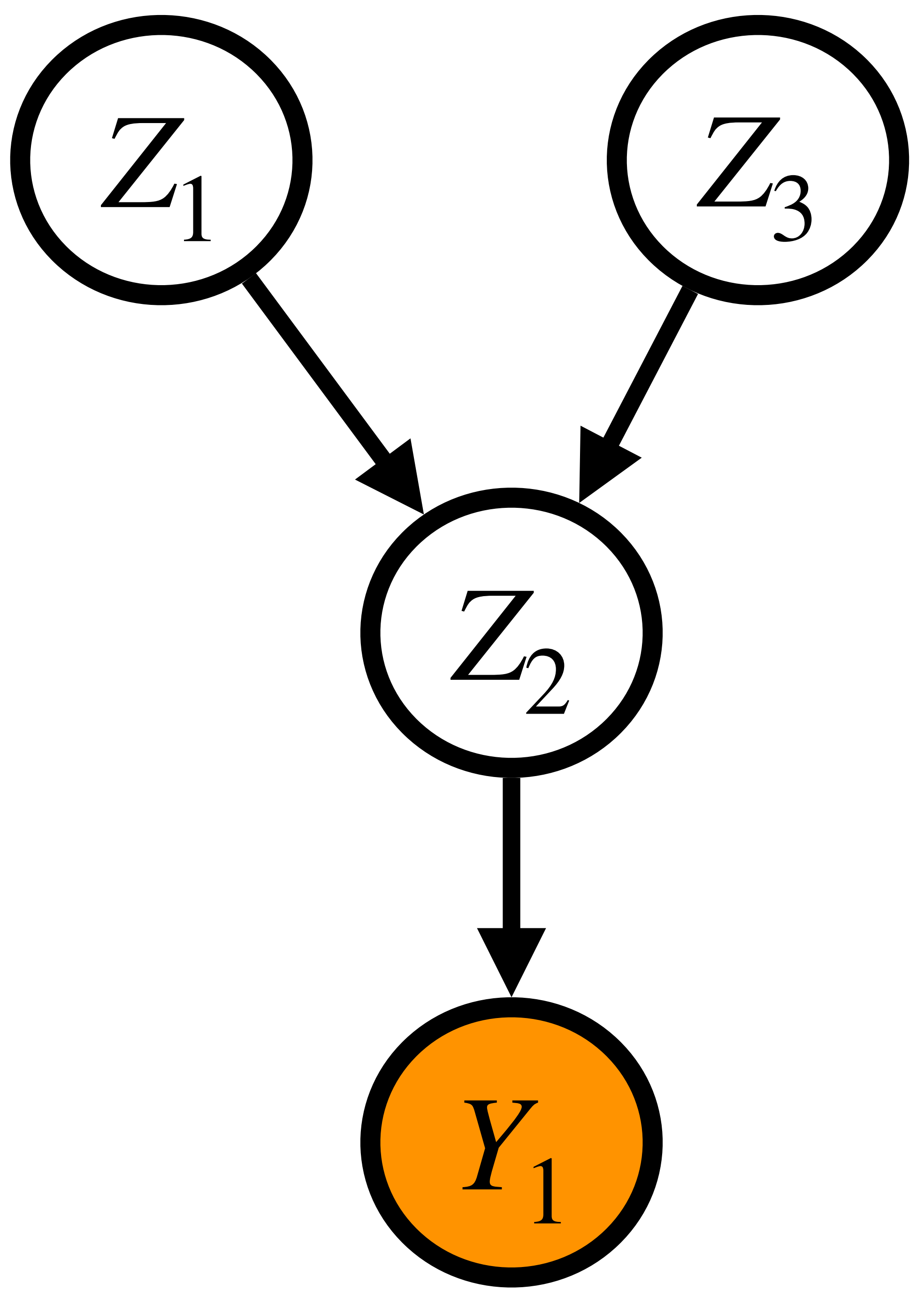


Variable Elimination: Example II (8/8)

$$P_{Y_1}(1) = \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P(1, z_1, z_2, z_3) =$$

$$= \sum_{z_1=0}^{1}\sum_{z_2=0}^{1}\sum_{z_3=0}^{1} P_{Y_1|Z_2}(1 \mid z_2) P_{Z_2|Z_1,Z_3}(z_2 \mid z_1, z_3) P_{Z_1}(z_1) P_{Z_3}(z_3) =$$

$$= \sum_{z_3=0}^{1} P_{Z_3}(z_3) G_2(z_3)$$



Factor Representation (1/2)

We will now abstract a bit… We will replace conditional probabilities in a BN by factors $\psi_i(x_i, x_{p_1}, \ldots, x_{p_k})$. This is for now just a change of notation, which will simplify things.

So instead of

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P_{X_i|Par(X_i)}\left(x_i \,|\, \mathrm{par}_{\mathbf{x}}(X_i)\right)$$

we will write

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} \psi_i(\mathbf{v}_i)$$

where each $\mathbf{v}_i$ is a tuple consisting of a subset of $\{x_1, x_2, \ldots, x_n\}$.

Factor Representation (2/2)

We will now abstract a bit… We will replace conditional probabilities in a BN by factors $\psi_i(x_i, x_{p_1}, \ldots, x_{p_k})$. This is for now just a change of notation, which will simplify things.

So instead of

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P_{X_i|Par(X_i)}\left(x_i \,|\, \mathrm{par}_{\mathbf{x}}(X_i)\right)$$

we will write

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} \psi_i(\mathbf{v}_i)$$

where each $\mathbf{v}_i$ is a tuple consisting of a subset of $\{x_1, x_2, \ldots, x_n\}$.

This can simply be done by setting $\psi_i(\mathbf{v}_i) = P_{X_i|Par(X_i)}\left(x_i \,|\, \mathrm{par}_{\mathbf{x}}(X_i)\right)$ and $\mathbf{v}_i = (x_i, x'_1, \ldots, x'_{k_i})$ where $x'_1, x'_2, \ldots, x'_{k_i}$ are the parents of X_i (more precisely, their values).

Two Operations: Product

Product:

Given factors $\psi_1(x_{i_1}, \ldots, x_{i_k})$ and $\psi_2(x_{j_1}, \ldots, x_{j_k})$, their product is defined simply as:

$$\psi_{1\times 2}(x_{h_1}, \ldots, x_{h_k}) = \psi_1(x_{i_1}, \ldots, x_{i_k}) \cdot \psi_2(x_{j_1}, \ldots, x_{j_k}),$$

where $\{x_{h_1}, \ldots, x_{h_k}\} = \{x_{i_1}, \ldots, x_{i_k}\} \cup \{x_{j_1}, \ldots, x_{j_k}\}$. *(Note that this will be represented as a table that will have a row for every possible combination of the values of the variables in it.)*

Example:

$\psi_1(x_1, x_2) = P_{X_1|X_2}(x_1 \,|\, x_2)$, $\psi_2(x_2, x_3) = P_{X_2|X_3}(x_2 \,|\, x_3)$. Then

$$\psi_{1\times 2}(x_1, x_2, x_3) = P_{X_1|X_2}(x_1 \,|\, x_2) \cdot P_{X_2|X_3}(x_2 \,|\, x_3).$$

Two Operations: Marginalization

Product:

Given a factor $\psi(x_{i_1}, \ldots, x_{i*}, \ldots, x_{i_k})$ and a variable x_{i*}, the marginalization is:

$$\tau_{i*}(x_{i_1}, \ldots, x_{i*}, \ldots, x_{i_k}) = \sum_{x_{i*}} \psi(x_{i_1}, \ldots, x_{i*}, \ldots, x_{i_k}).$$

Example:

Let us have a factor $\psi(x_1, x_2, x_3)$. Then

$$\tau_2(x_1, x_3) = \sum_{x_2} \psi(x_1, x_2, x_3).$$

Note: *The operations product and marginalization are similar in spirit to join and projection operations from relational databases.*

Variable Elimination: Algorithm

Assume that the unobserved random variables of the BN are ordered as $Z_1, Z_2, \ldots, Z_n$.

For $i = 1, \ldots, n$**:**

Collect all factors containing Z_i and compute their product $\psi^{(i)}_{prod}$.

Marginalize out Z_i from $\psi^{(i)}_{prod}$ and call the result τ.

Remove all factors containing Z_i and add τ instead.

The Elimination Order Matters

The number of operations (~runtime) depends on the ordering of the unobserved variables that we use when eliminating them.

The runtime depends on the size of the intermediate factors and that depends on the order.

Finding the best ordering is an NP-hard problem (but there are heuristics).

Part 3: Approximate Methods

Approximate Methods

Variable elimination is an **exact method.**

Exact inference is computationally hard. Computing marginal probabilities in Bayesian networks is **#P-hard**.

Now we will take a look at approximate methods based on sampling.

Forward Sampling

Generating samples from a distribution given by a Bayesian network (without evidence) is easy…

Algorithm:

Let the nodes of $X_1, X_2, \ldots, X_n$ be ordered topologically. (*That is, for all i, all $X_i's$ parents must preceed X_i in this ordering.)*

Initialize: Sampled $= []$

For $i = 1,\ldots,n$**:**

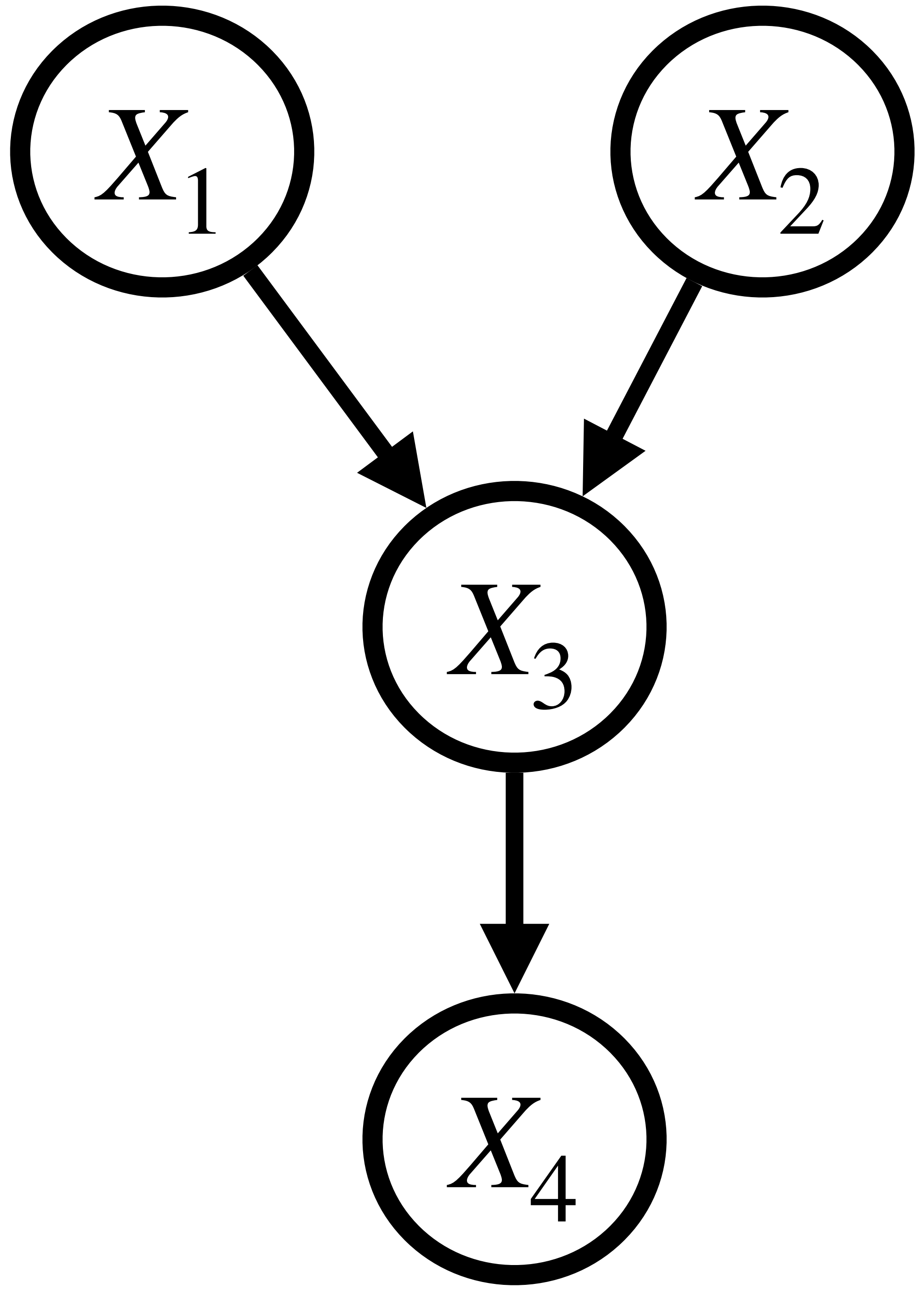
Let $X_{i_1}, X_{i_2}, \ldots, X_{i_l}$ be the parents of X_i and $P_{X_i, Par(X_i)}(x \,|\, , x_{i_1}, \ldots, x_{i_l})$ be the conditional distribution of X_i.

Sample $x_i \sim P_{X_i|Par(X_i)}(\,.\,|\,$Sampled$[i_1]$, Sampled$[i_2], \ldots,$ Sampled$[i_l]$).

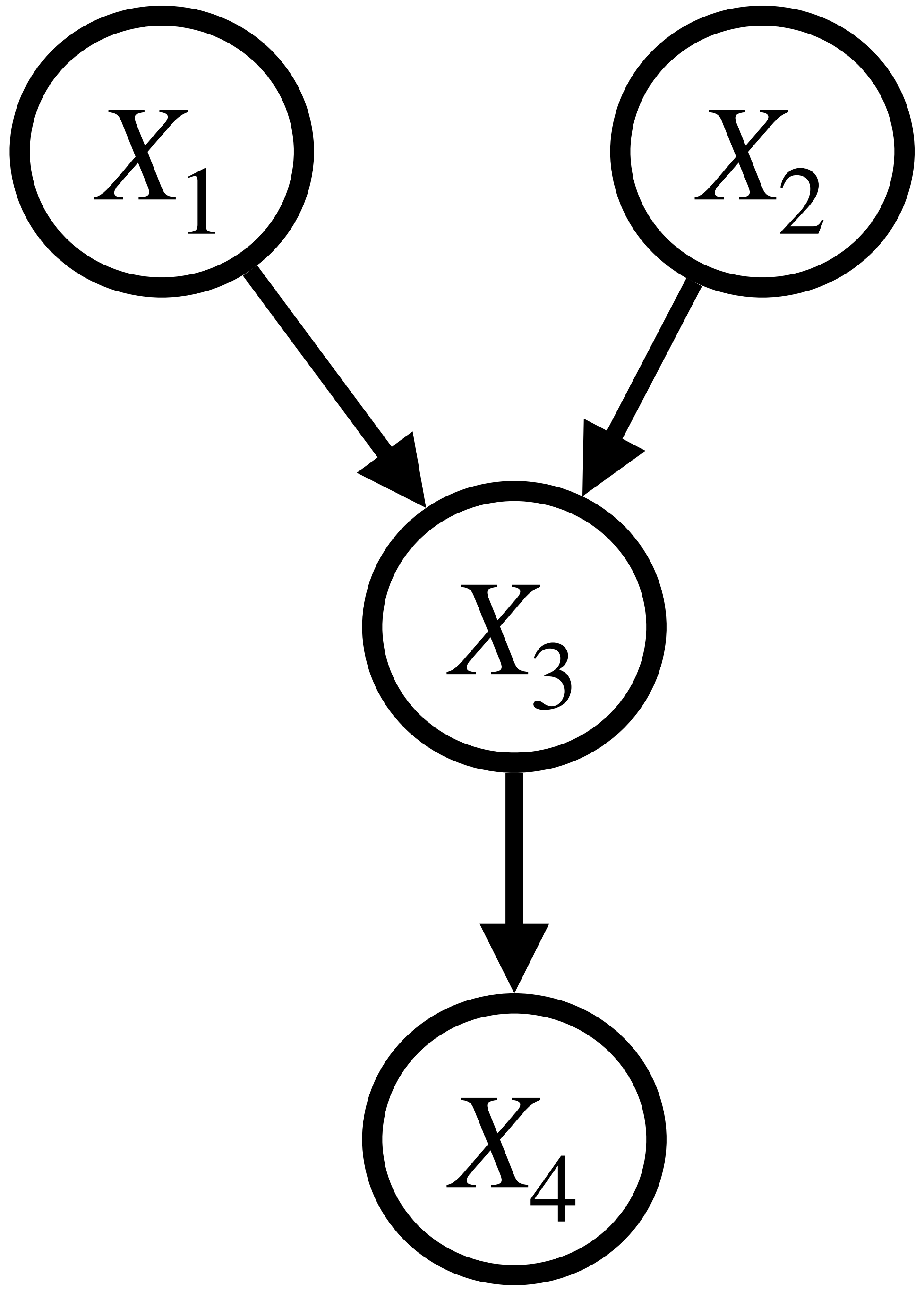
Set Sampled$[i] = x_i$**.**

Forward Sampling: Example

Topological ordering? … X_1, X_2, X_3, X_4 ✅



Forward Sampling: Example

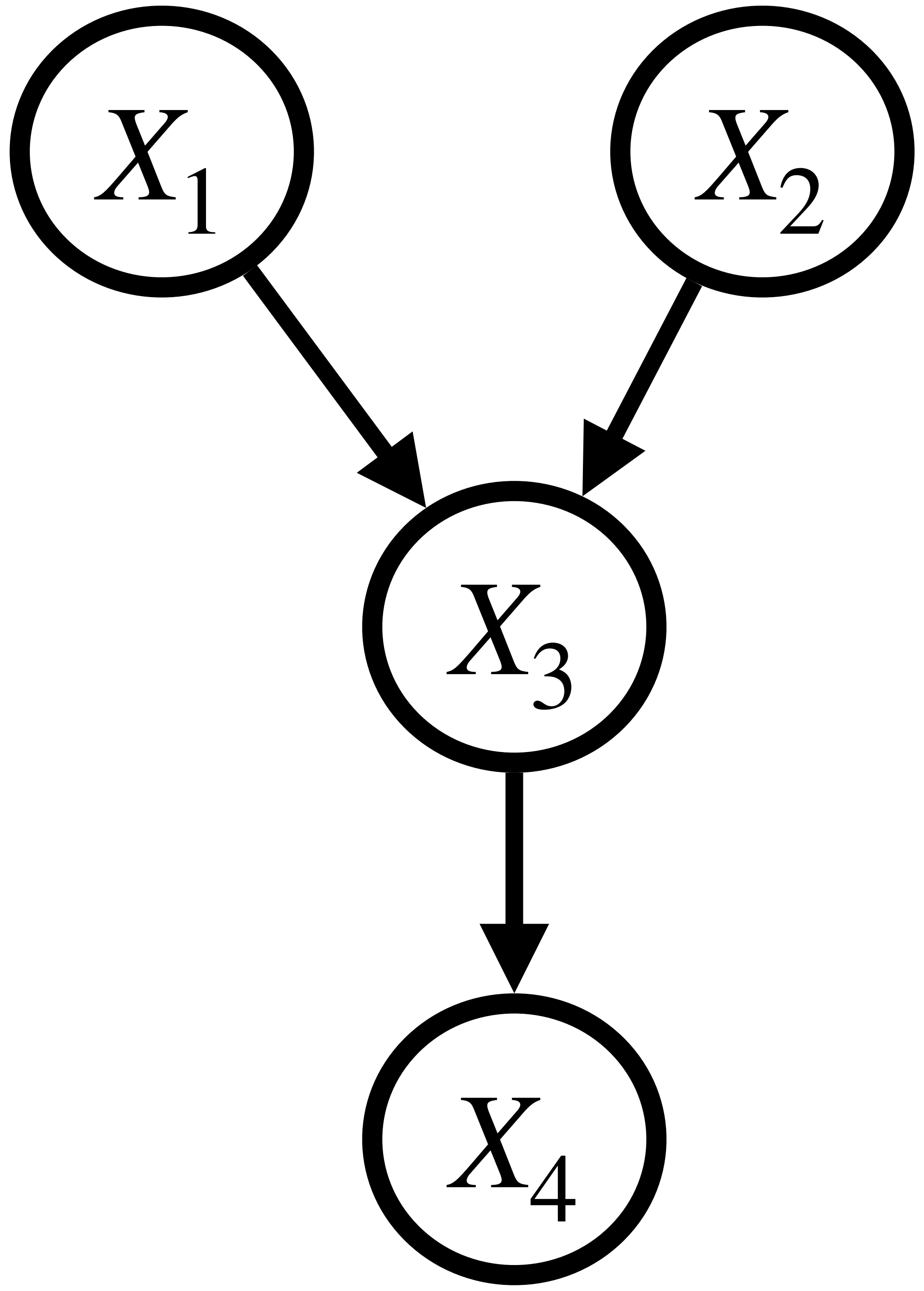


Topological ordering? … X_1, X_2, X_3, X_4 ✅

Sample $x_1 \sim P_{X_1}(\,.\,)$, e.g. $x_1 = 0$

Sampled $= [0]$

Forward Sampling: Example



Topological ordering? … X_1, X_2, X_3, X_4 ✅

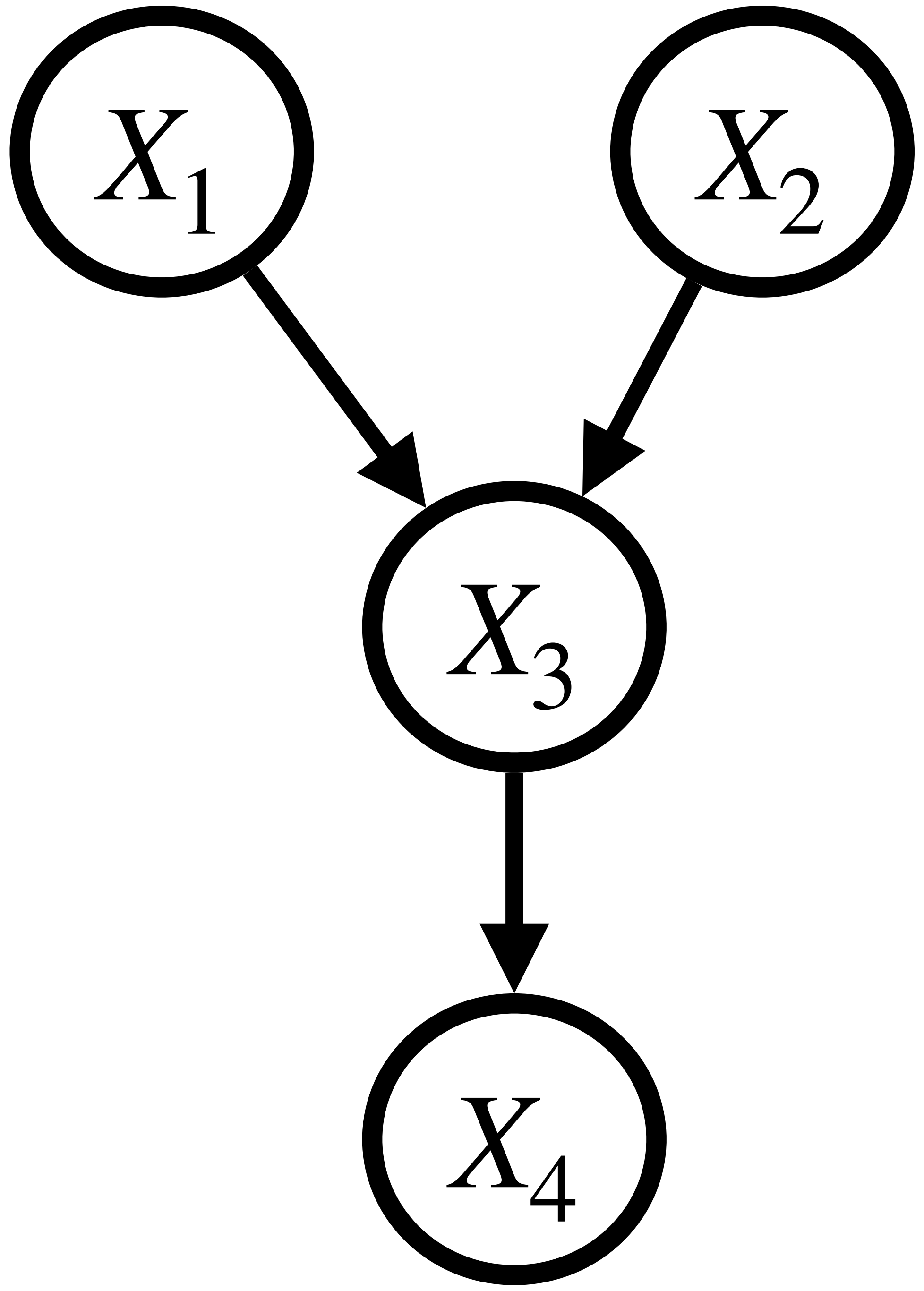
Sample $x_1 \sim P_{X_1}(\,.\,)$, e.g. $x_1 = 0$

Sampled $= [0]$

Sample $x_2 \sim P_{X_2}(\,.\,)$, e.g. $x_2 = 1$

Sampled $= [0,1]$

Forward Sampling: Example



Topological ordering? … X_1, X_2, X_3, X_4 ✅

Sample $x_1 \sim P_{X_1}(\,.\,)$, e.g. $x_1 = 0$

Sampled $= [0]$

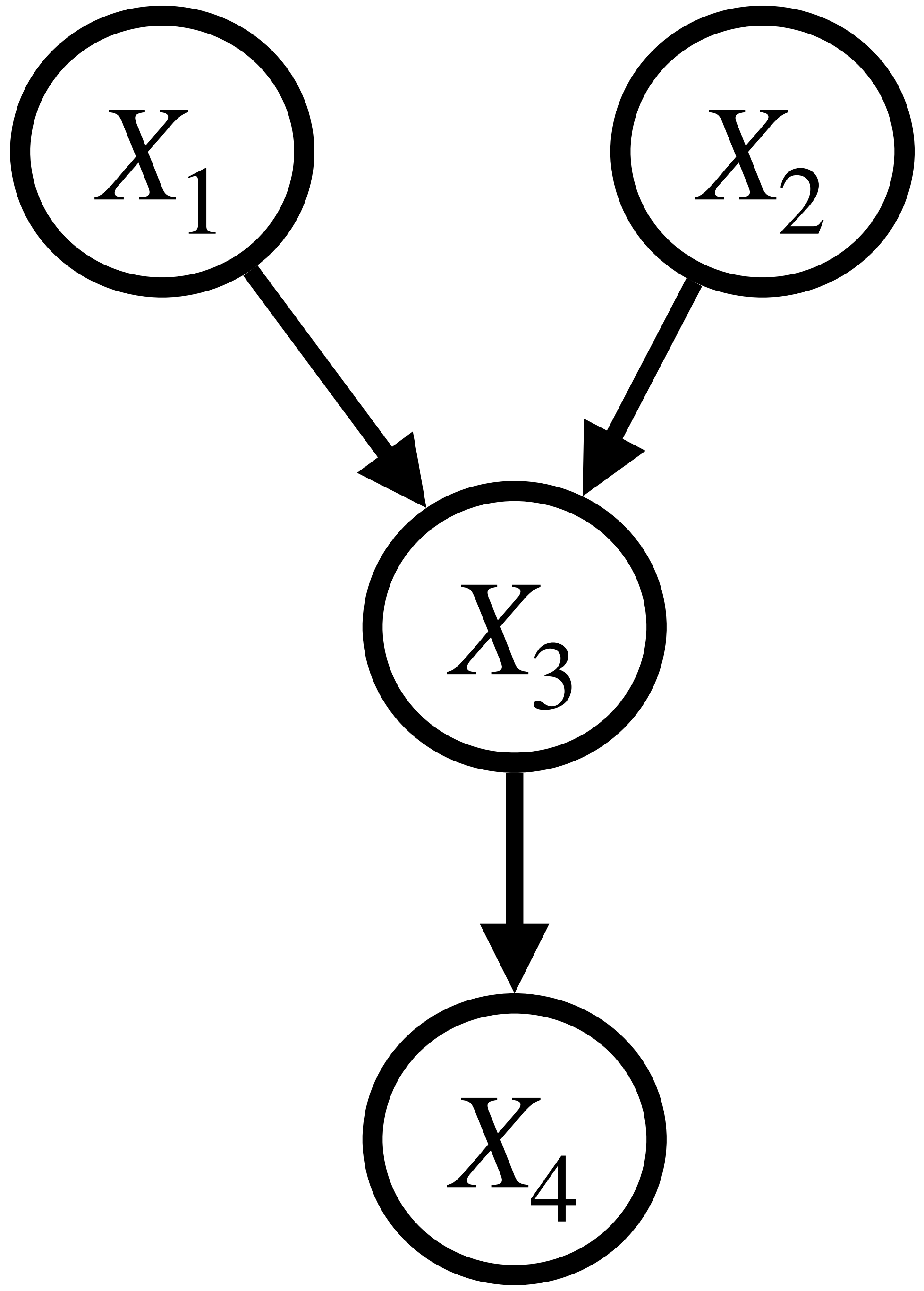
Sample $x_2 \sim P_{X_2}(\,.\,)$, e.g. $x_2 = 1$

Sampled $= [0,1]$

Sample $x_3 \sim P_{X_3|X_1,X_2}(\,.\,|\,0,1)$, e.g. $x_3 = 1$

Sampled $= [0,1,1]$

Forward Sampling: Example



Topological ordering? … X_1, X_2, X_3, X_4 ✅

Sample $x_1 \sim P_{X_1}(\,.\,)$, e.g. $x_1 = 0$

Sampled $= [0]$

Sample $x_2 \sim P_{X_2}(\,.\,)$, e.g. $x_2 = 1$

Sampled $= [0,1]$

Sample $x_3 \sim P_{X_3|X_1,X_2}(\,.\,|\,0,1)$, e.g. $x_3 = 1$

Sampled $= [0,1,1]$

Sample $x_4 \sim P_{X_4|X_3}(\,.\,|\,1)$, e.g. $x_4 = 1$

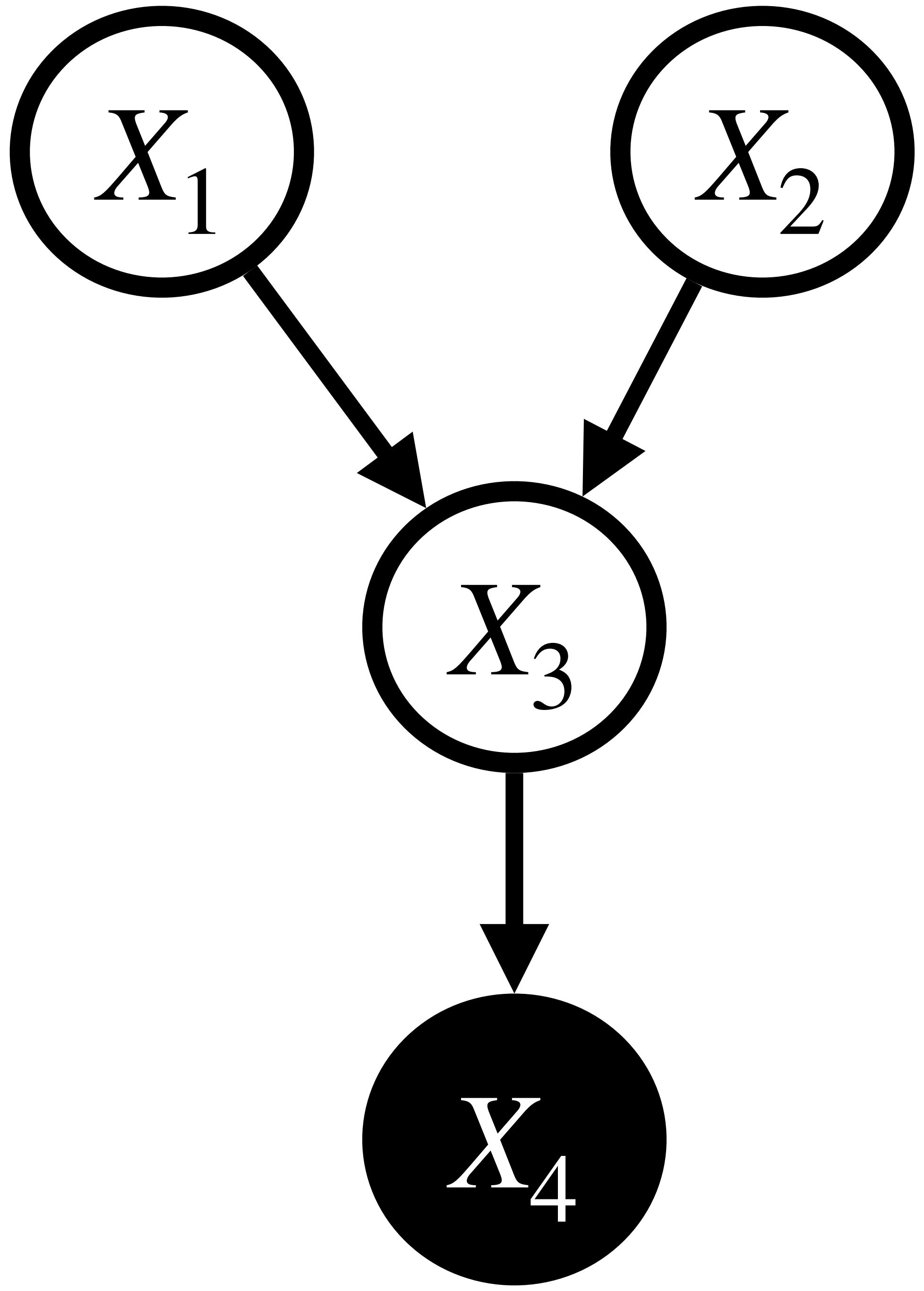
Sampled $= [0,1,1,1]$

Monte Carlo Estimation

We can use **forward sampling to estimate marginal probabilities** (without conditioning):

$$P_Y(\mathbf{y}) \approx \frac{1}{N} \sum_{(\mathbf{y}^{(i)}, \mathbf{z}^{(i)})} \mathbb{I}(\mathbf{y}^{(i)} = \mathbf{y}).$$

Sampling with Evidence



How can we sample from a BN if there is evidence on some random variables?

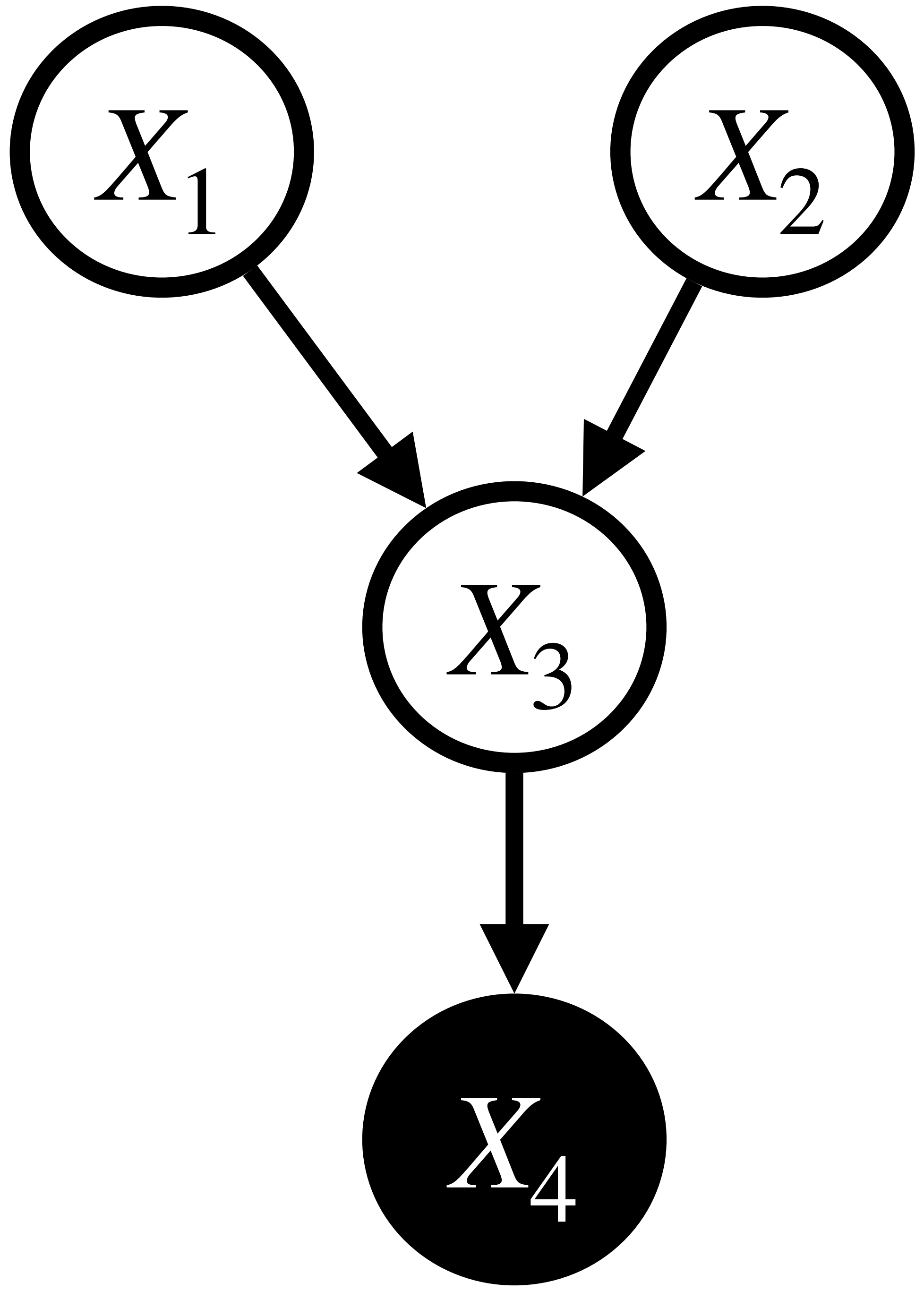
Example: How can we sample $(x_1, x_2, x_3) \sim P_{X_1,X_2,X_3|X_4}(\,.\,,\,.\,,\,.\,|\,x_4)$?

Note: *One possibility is to use marginal inference and to compute $P_{X_1,X_2,X_3|X_4}(x_1, x_2, x_3\,|\,x_4)$ for all tuples $(x_1, x_2, x_3) \in \{0,1\}^3$ as $P(x_1, x_2, x_3, 1)/P_{X_4}(1)$. However, that is something we want to avoid… after all we sample to do approximate marginal inference.*

Evidence:

e.g. $X_4 = 1$

Rejection Sampling (1/2)



Evidence:

e.g. $X_4 = 1$

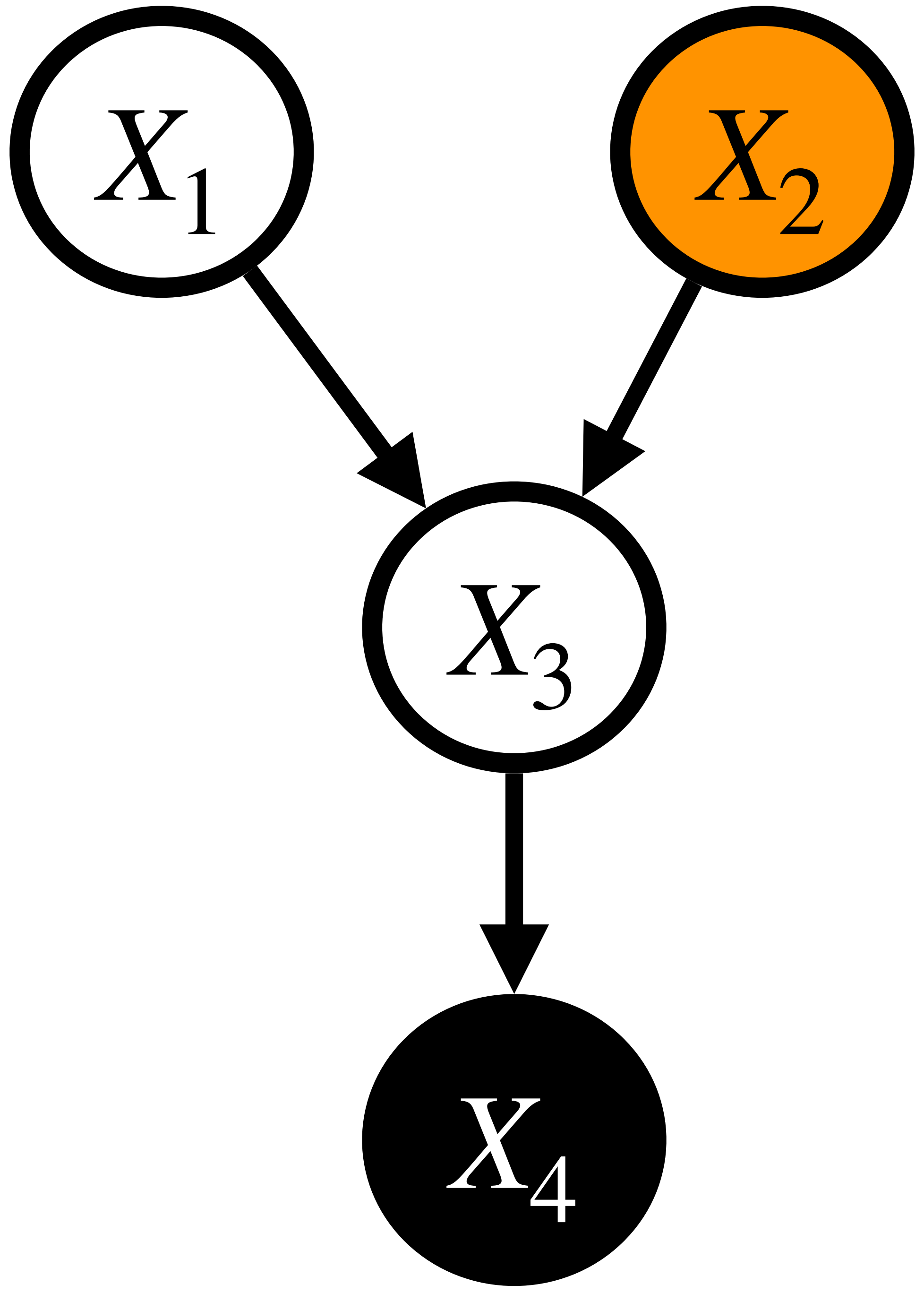
Basic idea:

We know how to sample without evidence (forward sampling). Let $(x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, x_4^{(1)}), (x_1^{(2)}, x_2^{(2)}, x_3^{(2)}, x_4^{(2)}), \ldots, (x_1^{(N)}, x_2^{(N)}, x_3^{(N)}, x_4^{(N)})$ be samples from the BN without taking evidence into account.

To get samples from the distribution conditioned on $X_4 = 1$, we just need to filter out those samples where $x_4^{(i)} \neq 1$.

Downside: What if $P[X_4 = 1]$ is small? Then we will need many samples from the unconditional distribution to get enough samples from the conditional one…

Rejection Sampling (2/2)



Monte Carlo Estimation with rejection sampling:

We are interested in estimating $P[\mathbf{Y} = \mathbf{y} \mid \mathbf{E} = \mathbf{e}] = P_{\mathbf{Y}|\mathbf{E}}(\mathbf{y} \mid \mathbf{e})$, e.g., $P[X_2 = 1 \mid X_4 = 1] = P_{X_2|X_4}(1 \mid 1)$.

Algorithm RejectionMC($BN, \mathbf{Y}, \mathbf{E}, \mathbf{y}, \mathbf{e}$)**:**

$YES := 0, NO := 0$

For $i = 1, \ldots, N$**:**

$(x_1^{(i)}, \ldots, x_n^{(i)})$:= ForwardSampling(BN)

if $(x_1^{(i)}, \ldots, x_n^{(i)})$ is consistent with evidence $\mathbf{e}$ on $\mathbf{E}$:

if $(x_1^{(i)}, \ldots, x_n^{(i)})$ is consistent with $\mathbf{Y} = \mathbf{y}$

$YES := YES + 1$

else $NO = NO + 1$

return $YES/(YES + NO)$

Importance Sampling: Basic Idea (1/3)

Problem: Estimate

$$P_{\mathbf{Y}|\mathbf{E}}(\mathbf{y} \,|\, \mathbf{e}) = \sum_{\mathbf{z}} \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{P_{\mathbf{E}}(\mathbf{e})}.$$

Importance Sampling: Basic Idea (2/3)

We have

$$P_{\mathbf{Y}|\mathbf{E}}(\mathbf{y} \mid \mathbf{e}) = \sum_{\mathbf{z}} \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{P_{\mathbf{E}}(\mathbf{e})}$$

$$P(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0 \Rightarrow Q(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0$$

$$W_{\mathbf{e}}(\mathbf{y}, \mathbf{z}) = \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}$$

Importance Sampling: Basic Idea (2/3)

We have

$$P_{\mathbf{Y}|\mathbf{E}}(\mathbf{y} \mid \mathbf{e}) = \sum_{\mathbf{z}} \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{P_{\mathbf{E}}(\mathbf{e})} = \frac{\sum_{\mathbf{z}} P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{\sum_{\mathbf{y}', \mathbf{z}} P(\mathbf{y}', \mathbf{z}, \mathbf{e})} =$$

$$P(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0 \Rightarrow Q(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0$$

$$W_{\mathbf{e}}(\mathbf{y}, \mathbf{z}) = \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}$$

Importance Sampling: Basic Idea (2/3)

We have

$$P_{\mathbf{Y}|\mathbf{E}}(\mathbf{y} \mid \mathbf{e}) = \sum_{\mathbf{z}} \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{P_{\mathbf{E}}(\mathbf{e})} = \frac{\sum_{\mathbf{z}} P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{\sum_{\mathbf{y}', \mathbf{z}} P(\mathbf{y}', \mathbf{z}, \mathbf{e})} =$$

$$= \frac{\sum_{\mathbf{z}} P(\mathbf{y}, \mathbf{z}, \mathbf{e}) \frac{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}}{\sum_{\mathbf{y}', \mathbf{z}} P(\mathbf{y}', \mathbf{z}, \mathbf{e}) \frac{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}}$$

$$P(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0 \Rightarrow Q(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0$$

$$W_{\mathbf{e}}(\mathbf{y}, \mathbf{z}) = \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}$$

Importance Sampling: Basic Idea (2/3)

We have

$$P_{\mathbf{Y}|\mathbf{E}}(\mathbf{y} \mid \mathbf{e}) = \sum_{\mathbf{z}} \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{P_{\mathbf{E}}(\mathbf{e})} = \frac{\sum_{\mathbf{z}} P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{\sum_{\mathbf{y}', \mathbf{z}} P(\mathbf{y}', \mathbf{z}, \mathbf{e})} =$$

$$= \frac{\sum_{\mathbf{z}} P(\mathbf{y}, \mathbf{z}, \mathbf{e}) \frac{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}}{\sum_{\mathbf{y}', \mathbf{z}} P(\mathbf{y}', \mathbf{z}, \mathbf{e}) \frac{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}} = \frac{\sum_{\mathbf{z}} Q(\mathbf{y}, \mathbf{z}, \mathbf{e}) \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}}{\sum_{\mathbf{y}', \mathbf{z}} Q(\mathbf{y}', \mathbf{z}, \mathbf{e}) \frac{P(\mathbf{y}', \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}}$$

$P(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0 \Rightarrow Q(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0$

$$W_{\mathbf{e}}(\mathbf{y}, \mathbf{z}) = \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}$$

Importance Sampling: Basic Idea (2/3)

We have

$$P_{\mathbf{Y}|\mathbf{E}}(\mathbf{y} \mid \mathbf{e}) = \sum_{\mathbf{z}} \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{P_{\mathbf{E}}(\mathbf{e})} = \frac{\sum_{\mathbf{z}} P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{\sum_{\mathbf{y}', \mathbf{z}} P(\mathbf{y}', \mathbf{z}, \mathbf{e})} =$$

$$= \frac{\sum_{\mathbf{z}} P(\mathbf{y}, \mathbf{z}, \mathbf{e}) \frac{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}}{\sum_{\mathbf{y}', \mathbf{z}} P(\mathbf{y}', \mathbf{z}, \mathbf{e}) \frac{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}} = \frac{\sum_{\mathbf{z}} Q(\mathbf{y}, \mathbf{z}, \mathbf{e}) \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}}{\sum_{\mathbf{y}', \mathbf{z}} Q(\mathbf{y}', \mathbf{z}, \mathbf{e}) \frac{P(\mathbf{y}', \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}} =$$

$$= \frac{\sum_{\mathbf{y}', \mathbf{z}} Q(\mathbf{y}', \mathbf{z}, \mathbf{e}) \cdot \mathbb{I}(\mathbf{y} = \mathbf{y}') \cdot \frac{P(\mathbf{y}', \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}}{\sum_{\mathbf{y}', \mathbf{z}} Q(\mathbf{y}', \mathbf{z}, \mathbf{e}) \frac{P(\mathbf{y}', \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}}$$

$P(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0 \Rightarrow Q(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0$

$$W_{\mathbf{e}}(\mathbf{y}, \mathbf{z}) = \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}$$

Importance Sampling: Basic Idea (2/3)

We have

$$P_{\mathbf{Y}|\mathbf{E}}(\mathbf{y} \mid \mathbf{e}) = \sum_{\mathbf{z}} \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{P_{\mathbf{E}}(\mathbf{e})} = \frac{\sum_{\mathbf{z}} P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{\sum_{\mathbf{y}', \mathbf{z}} P(\mathbf{y}', \mathbf{z}, \mathbf{e})} =$$

$$= \frac{\sum_{\mathbf{z}} P(\mathbf{y}, \mathbf{z}, \mathbf{e}) \frac{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}}{\sum_{\mathbf{y}', \mathbf{z}} P(\mathbf{y}', \mathbf{z}, \mathbf{e}) \frac{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}} = \frac{\sum_{\mathbf{z}} Q(\mathbf{y}, \mathbf{z}, \mathbf{e}) \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}}{\sum_{\mathbf{y}', \mathbf{z}} Q(\mathbf{y}', \mathbf{z}, \mathbf{e}) \frac{P(\mathbf{y}', \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}} =$$

$$= \frac{\sum_{\mathbf{y}', \mathbf{z}} Q(\mathbf{y}', \mathbf{z}, \mathbf{e}) \cdot \mathbb{I}(\mathbf{y} = \mathbf{y}') \cdot \frac{P(\mathbf{y}', \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}}{\sum_{\mathbf{y}', \mathbf{z}} Q(\mathbf{y}', \mathbf{z}, \mathbf{e}) \frac{P(\mathbf{y}', \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}', \mathbf{z}, \mathbf{e})}} = \frac{\mathbb{E}_{\mathbf{Y}, \mathbf{Z} \sim Q} \left[\mathbb{I}(\mathbf{y} = \mathbf{y}') \cdot W_{\mathbf{e}}(\mathbf{Y}, \mathbf{Z}) \right]}{\mathbb{E}_{\mathbf{Y}, \mathbf{Z} \sim Q} \left[W_{\mathbf{e}}(\mathbf{Y}, \mathbf{Z}) \right]}$$

$P(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0 \Rightarrow Q(\mathbf{y}, \mathbf{z}, \mathbf{e}) > 0$

$$W_{\mathbf{e}}(\mathbf{y}, \mathbf{z}) = \frac{P(\mathbf{y}, \mathbf{z}, \mathbf{e})}{Q(\mathbf{y}, \mathbf{z}, \mathbf{e})}$$

Importance Sampling: Basic Idea (3/3)

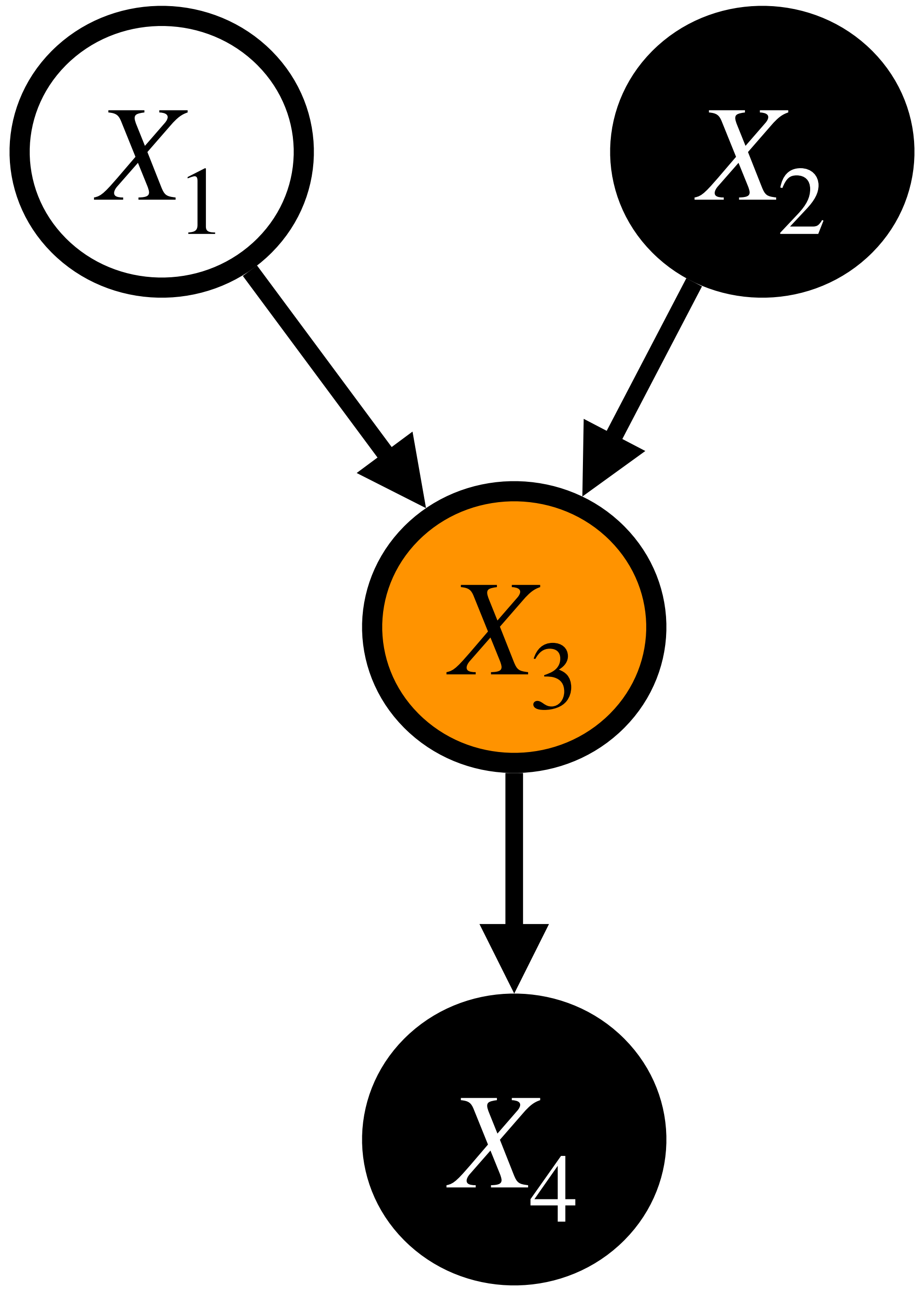
We have

$$P_{\mathbf{Y}|\mathbf{E}}(\mathbf{y}\,|\,\mathbf{e}) = \frac{\mathbb{E}_{\mathbf{Y},\mathbf{Z}\sim Q}\left[\mathbb{I}(\mathbf{y}=\mathbf{y}')\cdot W_{\mathbf{e}}(\mathbf{Y},\mathbf{Z})\right]}{\mathbb{E}_{\mathbf{Y},\mathbf{Z}\sim Q}\left[W_{\mathbf{e}}(\mathbf{Y},\mathbf{Z})\right]} \approx \frac{\sum_{(\mathbf{y}^{(i)},\mathbf{z}^{(i)})\in\mathscr{D}} \mathbb{I}(\mathbf{y}=\mathbf{y}^{(i)})\cdot W_{\mathbf{e}}(\mathbf{y}^{(i)},\mathbf{z}^{(i)})}{\sum_{(\mathbf{y}^{(i)},\mathbf{z}^{(i)})\in\mathscr{D}} W_{\mathbf{e}}(\mathbf{y}^{(i)},\mathbf{z}^{(i)})}$$

where $\mathscr{D}$ is a collection of samples which were sampled according to the distribution Q.

The trick: Pick Q for which sampling is easy! *(We will see one particular choice of Q next, which will lead to a method called likelihood weighting.)*

Likelihood Weighting (1/4)

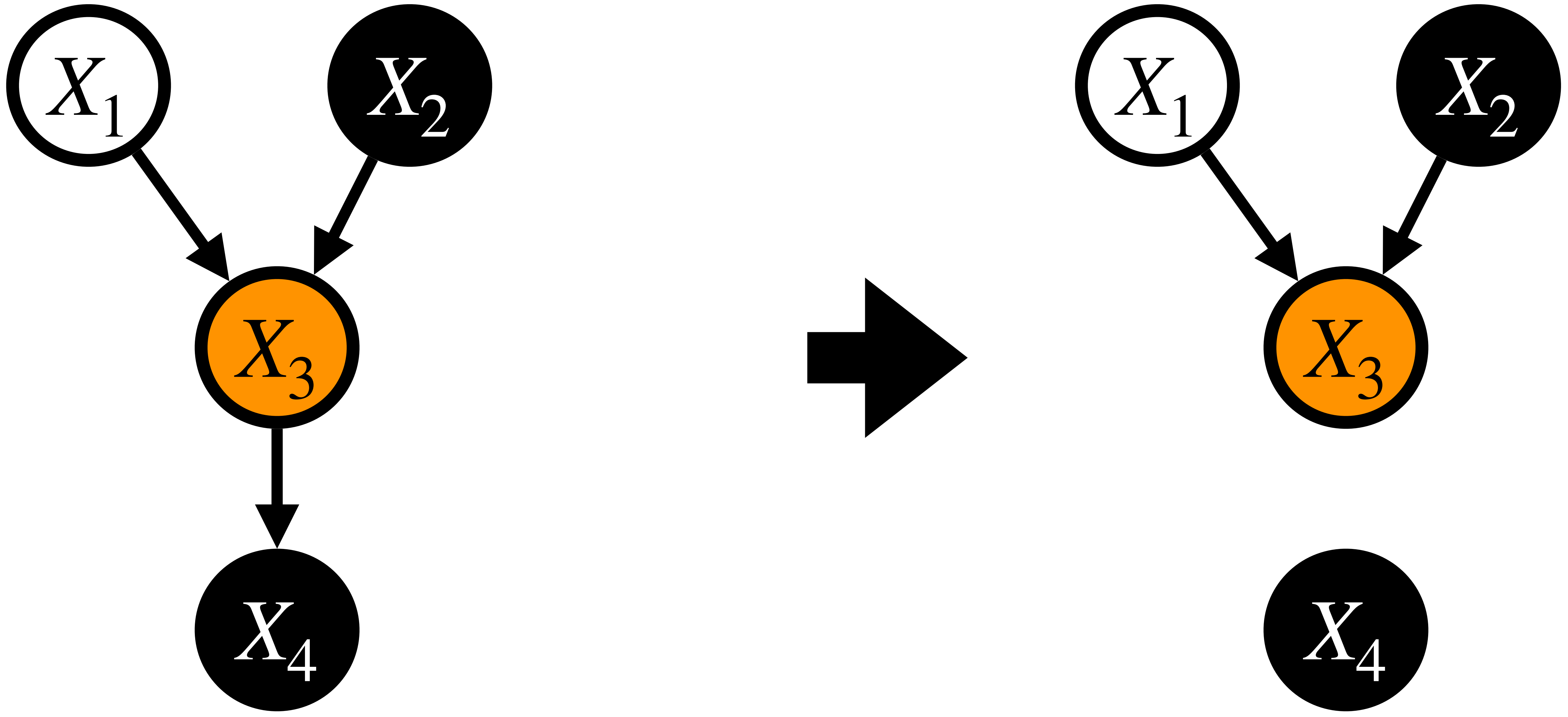


We want to find a Q **from which it is easy to sample** (we want to use forward sampling) **and such that $P(\mathbf{y}, \mathbf{z}, \mathbf{e})/Q(\mathbf{y}, \mathbf{z}, \mathbf{e})$ is easy to compute.**

Likelihood Weighting (2/4)

We want to find a Q **from which it is easy to sample** (we want to use forward sampling) **and such that $P(\mathbf{y}, \mathbf{z}, \mathbf{e})/Q(\mathbf{y}, \mathbf{z}, \mathbf{e})$ is easy to compute.**

Method: For every node from $\mathbf{E}$ (i.e. for every node on which we are conditioning), remove all edges that end in it.



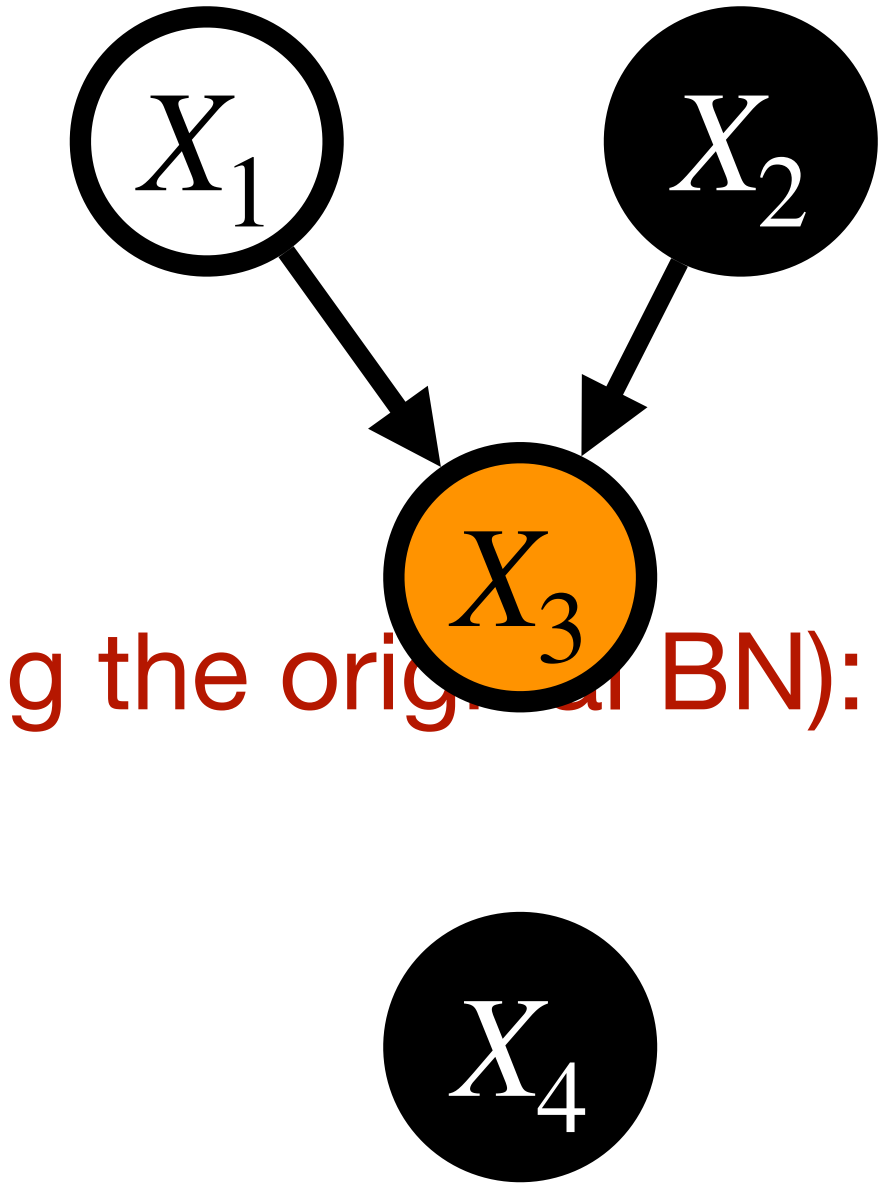
Likelihood Weighting (3/4)

We want to find a Q **from which it is easy to sample** (we want to use forward sampling) **and such that $W_{\mathbf{e}}(\mathbf{y}, \mathbf{z}) = P(\mathbf{y}, \mathbf{z}, \mathbf{e})/Q(\mathbf{y}, \mathbf{z}, \mathbf{e})$ is easy to compute.**

Method:

1. For every node from $\mathbf{E}$ (i.e. for every node on which we are conditioning), remove all edges that end in it.

2. Use forward sampling (keeping the values of nodes which are in $\mathbf{E}$ fixed to construct N samples from the modified BN. Store them in $\mathscr{D}$.



3. For every sample $(\mathbf{y}^{(i)}, \mathbf{z}^{(i)})$, compute $W_e(\mathbf{y}^{(i)}, \mathbf{z}^{(i)})$ as follows (here we are using the original BN):

$$W_e(\mathbf{y}^{(i)}, \mathbf{z}^{(i)}) = \prod_{E_i \in \mathbf{E}} P_{E_i|Par(E_i)}\left(e_i \,|\, \text{par}_{\mathbf{y}^{(i)},\mathbf{z}^{(i)}}(E_i)\right).$$

Likelihood Weighting (4/4)

Method:

1. For every node from $\mathbf{E}$ (i.e. for every node on which we are conditioning), remove all edges that end in it.

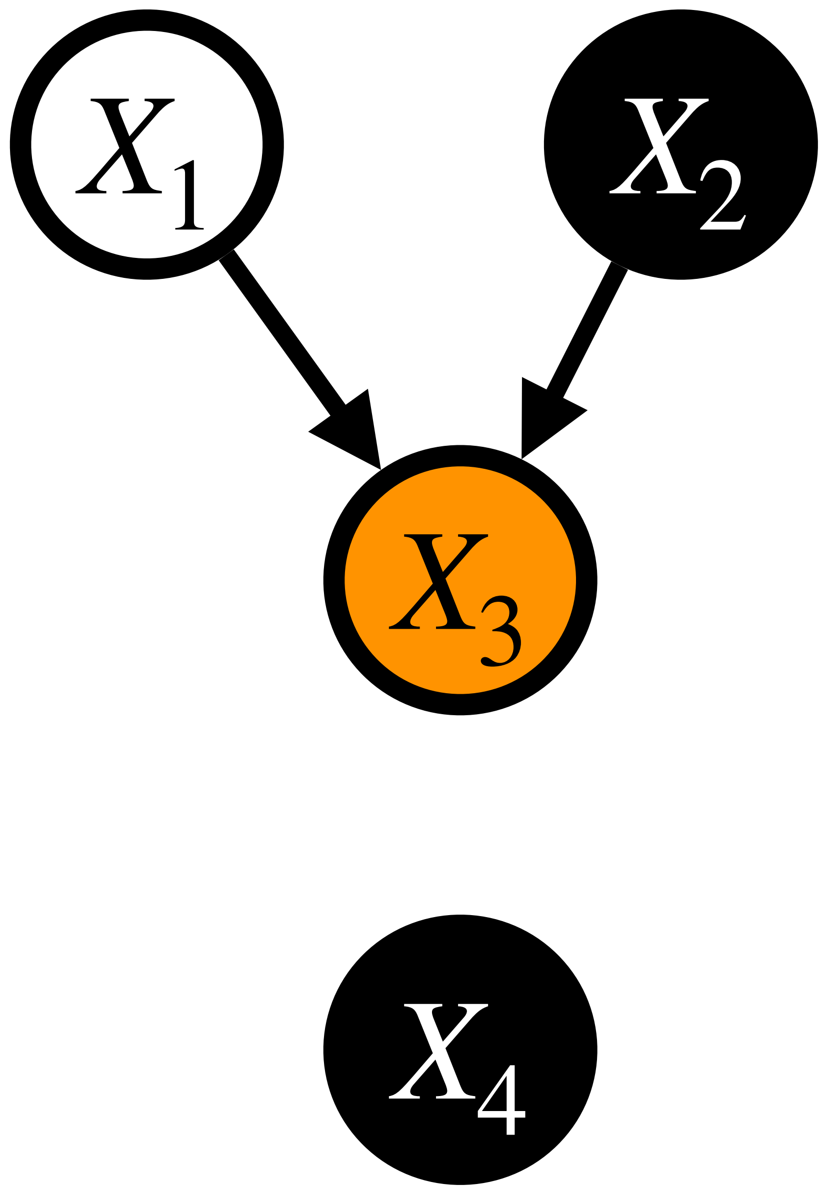
2. Use forward sampling (keeping the values of nodes which are in $\mathbf{E}$ fixed to construct N samples from the modified BN. Store them in $\mathscr{D}$.

3. For every sample $(\mathbf{y}^{(i)}, \mathbf{z}^{(i)})$, compute $W_e(\mathbf{y}^{(i)}, \mathbf{z}^{(i)})$ as follows:

$$W_e(\mathbf{y}^{(i)}, \mathbf{z}^{(i)}) = \prod_{E_i \in \mathbf{E}} P_{E_i|Par(E_i)}\left(e_i \,|\, \text{par}_{\mathbf{y}^{(i)},\mathbf{z}^{(i)}}(E_i)\right).$$

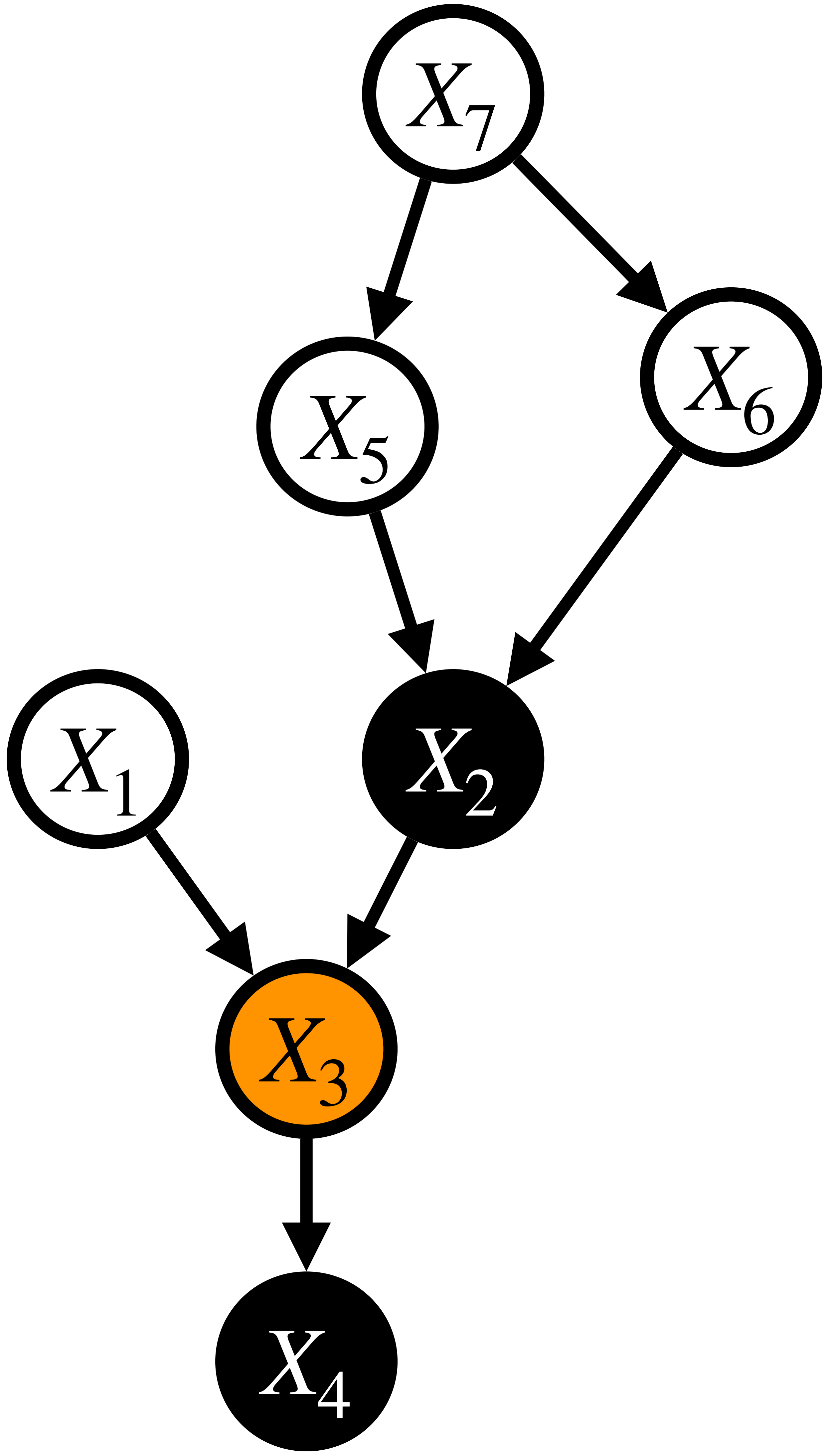
4. Compute estimate of $P_{\mathbf{Y}|\mathbf{E}}(\mathbf{y}\,|\,\mathbf{e})$ as

$$P_{\mathbf{Y}|\mathbf{E}}(\mathbf{y}\,|\,\mathbf{e}) \approx \frac{\sum_{(\mathbf{y}^{(i)},\mathbf{z}^{(i)}) \in \mathscr{D}} \mathbb{I}(\mathbf{y} = \mathbf{y}^{(i)}) \cdot W_{\mathbf{e}}(\mathbf{y}^{(i)}, \mathbf{z}^{(i)})}{\sum_{(\mathbf{y}^{(i)},\mathbf{z}^{(i)}) \in \mathscr{D}} W_{\mathbf{e}}(\mathbf{y}^{(i)}, \mathbf{z}^{(i)})}.$$



Part 4: Requisite Network

Exploiting Conditional Independence



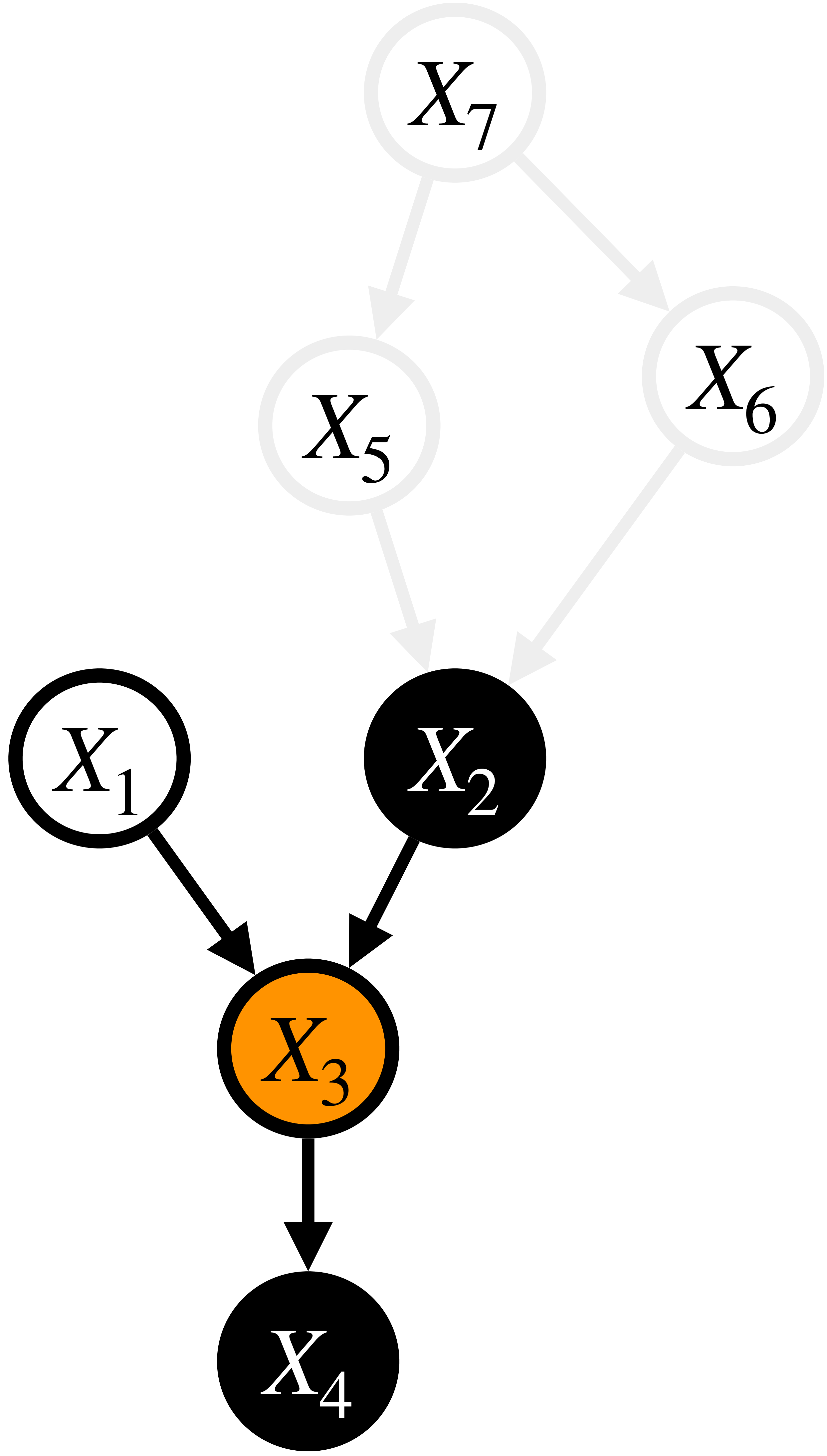
Do we need to perform inference on the whole network in order to compute $P[X_3 = 1 \mid X_2 = 0, X_1 = 1]$?

No!

We can check that $X_3 \perp\!\!\!\perp \{X_5, X_6, X_7\} \mid X_1, X_2$.

Therefore we can do inference on a smaller network…

Exploiting Conditional Independence



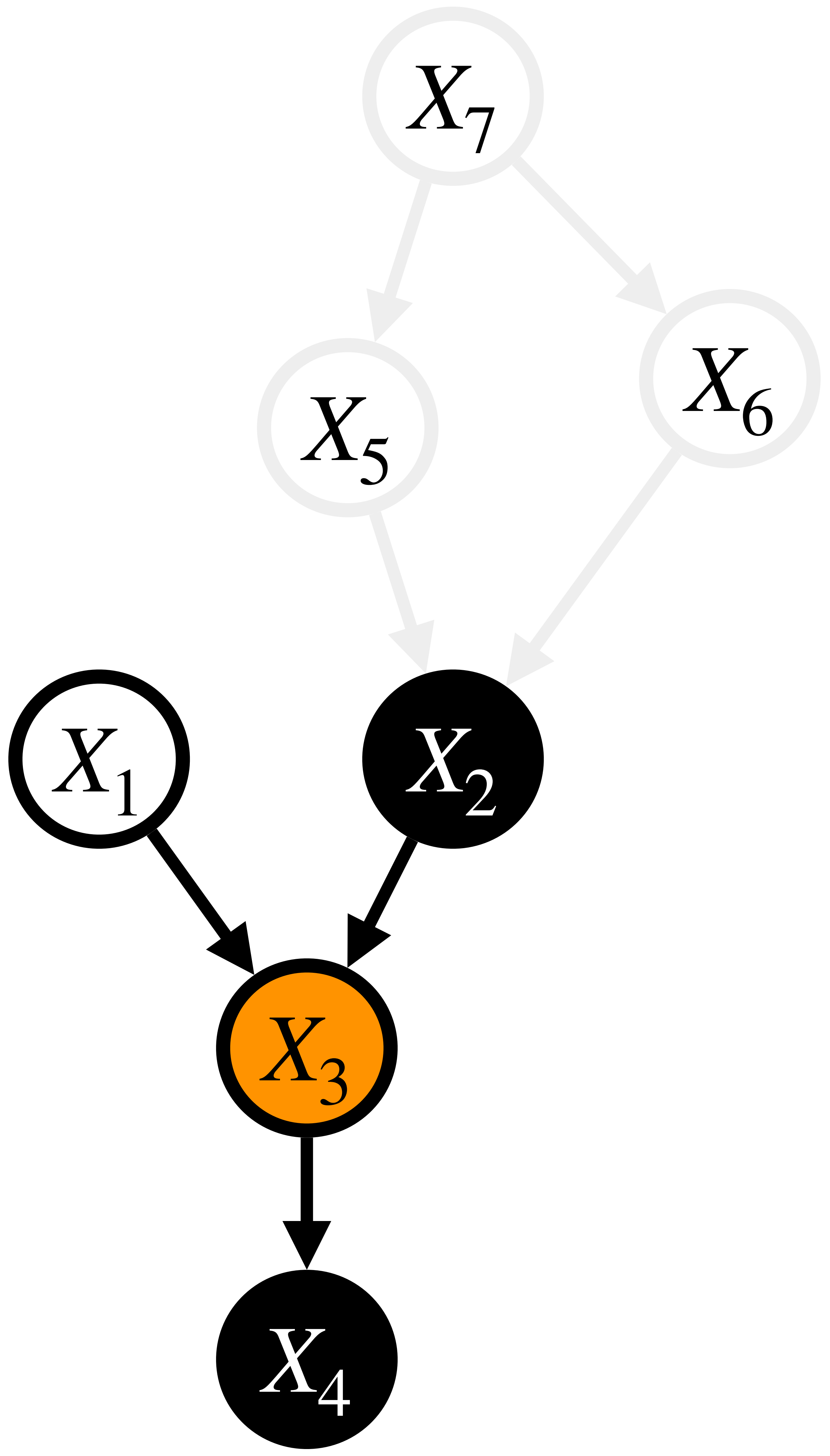
Do we need to perform inference on the whole network in order to compute $P[X_3 = 1 \mid X_2 = 0, X_1 = 1]$**?**

No!

We can check that $X_3 \perp\!\!\!\perp \{X_5, X_6, X_7\} \mid X_1, X_2$.

Therefore we can do inference on a smaller network…

Exploiting Conditional Independence



Do we need to perform inference on the whole network in order to compute $P[X_3 = 1 \mid X_2 = 0, X_1 = 1]$?

No!

We can check that $X_3 \perp\!\!\!\perp \{X_5, X_6, X_7\} \mid X_1, X_2$.

Therefore we can do inference on a smaller network…

This can be done more efficiently than by checking all paths *(whether they are active)* **by so-called Bayes-Ball Algorithm.**

Part 5: What we did not cover…

What we did not cover…

Inference: Join-tree algorithm, Gibbs sampling, variational inference…

Other inference problem: MAP-inference, MPE-inference

Learning: Maximum-likelihood learning (*for fully-observable data, this is just estimation of conditional probability tables from frequencies)*, EM algorithm *(when we have missing data or latent random variables)*, structure learning…