Bayesian Networks I

Part 1: Probability Refresher

Notation

We will use the following notation (*same as we used for stochastic processes*):

$$P[X_1 = x_1 \wedge X_2 = x_2 \wedge \ldots \wedge X_n = x_n] = P(x_1, x_2, \ldots, x_n).$$

Joint Distributions

Given random variables $X_1, X_2, \ldots, X_n$, their joint distribution is the probability distribution on tuples $(x_1, x_2, \ldots, x_n)$ of their possible values, i.e. for us it will be given by:

$$P[X_1 = x_1 \wedge X_2 = x_2 \wedge \ldots \wedge X_n = x_n] = P(x_1, x_2, \ldots, x_n)$$

Example:

X_1 is a binary random variable which is 1 if it rains and 0 otherwise,

X_2 is a binary random variable which is 1 if it is sunny and 0 otherwise,

X_3 is a binary random variable which is 1 if there is a rainbow and 0 otherwise.

Then $P(1,0,1)$ is the probability that, at the same time: it rains, it is not sunny and there is a rainbow *(we would expect this probability to be close to 0).*

Joint Distribution (Example)

$P(x_1, x_2, x_3)$ from the previous slide, i.e. $P(\text{rains}, \text{sunny}, \text{rainbow})$, represented as a table.

rains	**sunny**	**rainbow**	**P**
0	**0**	**0**	0.4
0	**0**	**1**	0
0	**1**	**0**	0.2
0	**1**	**1**	0
1	**0**	**0**	0.2
1	**0**	**1**	0
1	**1**	**0**	0.1
1	**1**	**1**	0.1

Marginal Distributions

Given a joint distribution on random variables $X_1, X_2, \ldots, X_n$, and their subset $\mathscr{A} = \{X_{i_1}, X_{i_2}, \ldots, X_{i_k}\} \subseteq \{X_1, X_2, \ldots, X_n\}$, the marginal distribution of the variables $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ is their distribution

$$P_{\mathscr{A}}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = P[X_{i_1} = x_{i_1} \wedge \ldots \wedge X_{i_k} = x_{i_k}]$$

and it satisfies:

$$P_{\mathscr{A}}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = \sum_{x_{j_1}, x_{j_2}, \ldots, x_{j_{n-k}}} P[X_{i_1} = x_{i_1} \wedge X_{i_2} = x_{i_2} \wedge \ldots \wedge X_{i_k} = x_{i_k} \wedge X_{j_1} = x_{j_1} \wedge X_{j_2} = x_{j_2} \wedge \ldots \wedge X_{j_{n-k}} = x_{j_{n-k}}]$$

Each of these $x_{j_1}, \ldots, x_{j_{n-k}}$ is summed over its range, e.g. if it is binary then over {0,1} etc.

Marginal Distributions - Example (1/2)

Recall the table:

X_1 (rains)	X_2 (sunny)	X_3	P
0	0	0	0.4
0	0	1	0
0	1	0	0.2
0	1	1	0
1	0	0	0.2
1	0	1	0
1	1	0	0.1
1	1	1	0.1

What is the probability $P[X_2 = 1]$? That is… What is the probability that it is sunny?

In our notation, $\mathscr{A} = \{X_1\}$, $P_{\mathscr{A}}(x) = P[X_2 = x]$*. Or using the alternative notation when* $\mathscr{A}$ *is a singleton, also* $P_{X_2}(x) = P[X_2 = x]$*.*

Marginal Distributions - Example (2/2)

Recall the table:

X_1 (rains)	X_2 (sunny)	X_3	P
0	0	0	0.4
0	0	1	0
0	1	0	0.2
0	1	1	0
1	0	0	0.2
1	0	1	0
1	1	0	0.1
1	1	1	0.1

What is the probability $P[X_2 = 1]$? That is… What is the probability that it is sunny?

$$P[X_2 = 1] = P(0,1,0) + P(0,1,1) + P(1,1,0) + P(1,1,1) = 0.2 + 0 + 0.1 + 0.1 = 0.4$$

Conditional Distribution (1/2)

Special case (two random variables X and Y):

Conditional probability of X given Y is defined as:

$$P[X = x \mid Y = y] = \frac{P[X = x \wedge Y = y]}{P[Y = y]} = \frac{P(x, y)}{P_Y(y)}.$$

Undefined for y's that have zero probability, i.e. when $P[Y = y] = 0$.

We will use the notation $P_{X|Y}(x \mid y) = P[X = x \mid Y = y]$.

(To simplify many formulas, we normally use the assumption that undefined $\cdot\, 0 = 0$*, so for instance it will allow us to write* $P(x, y) = P_{X|Y}(x \mid y)P_Y(y) = P_{Y|X}(y \mid x)P_X(x)$ *for all values* x, y*.)*

Conditional Distribution (2/2)

General case

Conditional probability of $\mathbf{Y} = (X_{i_1}, X_{i_2}, \ldots, X_{i_k})$ given $\mathbf{Z} = (X_{j_1}, X_{j_2}, \ldots, X_{j_l})$ is defined as:

$$P[\mathbf{Z} = \mathbf{z} \,|\, \mathbf{Y} = \mathbf{y}] = \frac{P[\mathbf{Z} = \mathbf{z} \wedge \mathbf{Y} = \mathbf{y}]}{P[\mathbf{Y} = \mathbf{y}]} = \frac{P_{\mathbf{Z},\mathbf{Y}}(\mathbf{z}, \mathbf{y})}{P_{\mathbf{Y}}(\mathbf{y})},$$

where $\mathbf{z} = (z_1, z_2, \ldots, z_l)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_l)$.

Conditional Distribution (Example)

Recall the table:

X_1 (rains)	X_2 (sunny)	X_3	P
0	0	0	0.4
0	0	1	0
0	1	0	0.2
0	1	1	0
1	0	0	0.2
1	0	1	0
1	1	0	0.1
1	1	1	0.1

What is the probability $P[X_2 = 1 \mid X_3 = 1]$ ($P_{X_2|X_3}(1 \mid 1)$)? That is… What is the probability that it is sunny given that there is rainbow?

Conditional Distribution (Example)

Recall the table:

X_1 (rains)	X_2 (sunny)	X_3	P
0	0	0	0.4
0	0	1	0
0	1	0	0.2
0	1	1	0
1	0	0	0.2
1	0	1	0
1	1	0	0.1
1	1	1	0.1

What is the probability $P[X_2 = 1 \mid X_3 = 1]$ ($P_{X_2|X_3}(1 \mid 1)$)? That is… What is the probability that it is sunny given that there is rainbow?

$$P[X_2 = 1 \wedge X_3 = 1] = P_{\{X_2,X_3\}}(1,1) = P(0,1,1) + P(1,1,1) = 0.1 + 0.1 = 0.2,$$

$$P[X_3 = 1] = P_{X_3}(1) = P(0,0,1) + P(0,1,1) + P(1,0,1) + P(1,1,1) = 0.2,$$

$$P[X_2 = 1 \mid X_3 = 1] = \frac{P[X_2 = 1 \wedge X_3 = 1]}{P[X_3 = 1]} = \frac{0.2}{0.2} = 1.$$

Part 2: Bayesian Networks - Motivation

Curse of Dimensionality

Example: Let's consider a joint distribution on 100 binary random variables. How large does the table representing this distribution need to be?

Answer: The table will need to have 2^{100} *rows (which means* $2^{100} - 1$ *parameters to set).*

So, clearly, representing joint distributions exhaustively is not an option when we have more than a handful of examples.

Independence

Definition (special case of two random variables): Two random variables X and Y are said to be independent if

$$P[X = x \wedge Y = y] = P[X = x] \cdot P[Y = y]$$

for all possible values x and y (i.e. using the other notation, if $P_{X,Y}(x, y) = P_X(x) \cdot P_Y(y)$ for all possible values x and y).

Joint Independence

Definition: Random variables $X_1, X_2, \ldots, X_n$ are independent if

$P[X_1 = x_1 \wedge X_2 = x_2 \wedge \ldots \wedge X_n = x_n] = P[X_1 = x_1] \cdot P[X_2 = x_2] \cdot \ldots \cdot P[X_n = x_n]$

for all values $x_1, x_2, \ldots, x_n$.

Joint Independence (Events)

Note. *For independence of a collection of events (recall that an event is a subset of the sample space),the situation is a bit more complicated.*

Let $A_1, A_2, \ldots, A_n$ *be events. Then these events are independent if*

$$P[A_{i_1} \wedge A_{i_2} \wedge \ldots \wedge A_{i_k}] = P[A_{i_1}] \cdot P[A_{i_2}] \cdot \ldots \cdot P[A_{i_k}]$$

holds for every subset $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ *of the events* $A_1, A_2, \ldots, A_n$ *and for every* $k > 0$.

Independence Is Too Strict

Question: *How many parameters do we need to describe a distribution of n independent binary random variables?*

Answer: *We need only n parameters (compare this with $2^n - 1$ that we need for a general distribution of n binary random variables).*

Unfortunately, independence is a condition which is too strict for many distributions. Therefore we will need something else…

Example: Independence holds e.g. when throwing n dice or when running independent trials of some experiment…

Conditional Independence (1/4)

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 (2/4)

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

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

Conditional Independence (3/4)

Why the two definitions are equivalent?

Proof: **Def. 1 => Def. 2.**

$$P_{X|Y,Z}(x \mid y, z) = \frac{P_{X,Y,Z}(x, y, z)}{P_{Y,Z}(y, z)} = \frac{P_{X,Y|Z}(x, y \mid z)P_Z(z)}{P_{Y|Z}(y \mid z)P_Z(z)} = \frac{P_{X,Y|Z}(x, y \mid z)}{P_{Y|Z}(y \mid z)} =$$

$$= \frac{P_{X|Z}(x \mid z)P_{Y|Z}(y \mid z)}{P_{Y|Z}(y \mid z)} = P_{X|Z}(x \mid z).$$

Similarly, we of course also have $P_{Y|X,Z}(y \mid x, z) = P_{Y|Z}(y \mid z)$.

Conditional Independence (4/4)

Why the two definitions are equivalent?

Proof: **Def. 2 => Def. 1.**

$$P_{X,Y|Z}(x,y\,|\,z) = \frac{P_{X,Y,Z}(x,y,z)}{P_Z(z)} = \frac{P_{X|Y,Z}(x\,|\,y,z)P_{Y,Z}(y,z)}{P_Z(z)} =$$

$$= \frac{P_{X|Y,Z}(x\,|\,y,z)P_{Y|Z}(y\,|\,z)P_Z(z)}{P_Z(z)} = P_{X|Y,Z}(x\,|\,y,z)P_{Y|Z}(y\,|\,z) = P_{X|Z}(x\,|\,z)P_{Y|Z}(y\,|\,z)$$

Conditional Independence (Example)

Example:

Alice throws a coin with sides marked by 0 and 1 (that will be X_1*). She then sends a message over noisy channels to Bob and Eve about the result of the coin flip. Since the channel is noisy, what Bob receives (that will be* X_2*) and what Eve receives (that will be* X_3*) is not necessarily the same as what Alice sent.*

Assuming the noise in the two channels is independent, it holds

$$X_2 \perp\!\!\!\perp X_3 \,|\, X_1$$

That is, given the result of Alice's coin toss, what Bob and Eve observe is independent. However, without this conditioning, what Bob and Eve observe is not independent (imagine e.g. that the noise is small and corrupts the message only with probability 0.001…).

How Many Parameters?

Question: How many parameters would we need in the previous example (*we always use the fact that probabilities sum up to 1*)?

We can use: $P_{X_1,X_2,X_3}(x_1, x_2, x_3) = P_{X_2|X_1}(x_2 \,|\, x_1)P_{X_3|X_1}(x_3 \,|\, x_1)P_{X_1}(x_1)$.

2 parameters for $P_{X_2|X_1}$ *(we need to determine* $P_{X_2|X_1}(0 \,|\, 1)$*, from which we can compute* $P_{X_2|X_1}(1 \,|\, 1) = 1 - P_{X_2|X_1}(0 \,|\, 1)$*, and similarly* $P_{X_3|X_1}(0 \,|\, 0)$*…).*

2 parameters for $P_{X_3|X_1}$ *(similar reasoning as above…)*

1 parameter for P_{X_1}***.***

5 parameters in total. If we did not use conditional independence, we would need $2^3 - 1 = 7$ ***parameters (this may not seem like much gain but it would be higher if we had more than three variables).***

Multi-Variate Case

Both of the equivalent definitions of conditional independence are straightforwardly generalized into the multi-variate case:

Definition 1: Random vectors $\mathbf{X}$ and $\mathbf{Y}$ are conditionally independent given $\mathbf{Z}$ if

$$P_{\mathbf{X},\mathbf{Y}|\mathbf{Z}}(\mathbf{x},\mathbf{y}\,|\,\mathbf{z}) = P_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}\,|\,\mathbf{z}) \cdot P_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}\,|\,\mathbf{z}))$$

for all possible values of the vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$.

Definition 2: Random vectors $\mathbf{X}$ and $\mathbf{Y}$ are conditionally independent given $\mathbf{Z}$ if

$$P_{\mathbf{X}|\mathbf{Y},\mathbf{Z}}(\mathbf{x}\,|\,\mathbf{y},\mathbf{z}) = P_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}\,|\,\mathbf{z})$$

for all possible values of the vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$.

Trivial Factorization (Chain Rule)

Any joint distribution of a random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ can be written as:

$$P_{\mathbf{X}}(x_1, x_2, \dots, x_n) = P_{X_1}(x_1) P_{X_2|X_1}(x_2 \mid x_1) P_{X_3|X_1,X_2}(x_3 \mid x_1, x_2) \dots P_{X_n|X_1,\dots,X_{n-1}}(x_n \mid x_1, \dots, x_{n-1})$$

Trivial Factorization (Chain Rule)

Any joint distribution of a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ can be written as:

$$P_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = P_{X_1}(x_1)P_{X_2|X_1}(x_2 \,|\, x_1)P_{X_3|X_1,X_2}(x_3 \,|\, x_1, x_2)\ldots P_{X_n|X_1,\ldots,X_{n-1}}(x_n \,|\, x_1, \ldots, x_{n-1})$$

The above can be simplified if we know that some conditional independencies hold, e.g. if X_2 and X_3 are conditionally independent given X_1 then we can replace $P_{X_3|X_1,X_2}(x_3 \,|\, x_1, x_2)$ by $P_{X_3|X_1}(x_3 \,|\, x_1)$ etc.

Part 3: Bayesian Networks

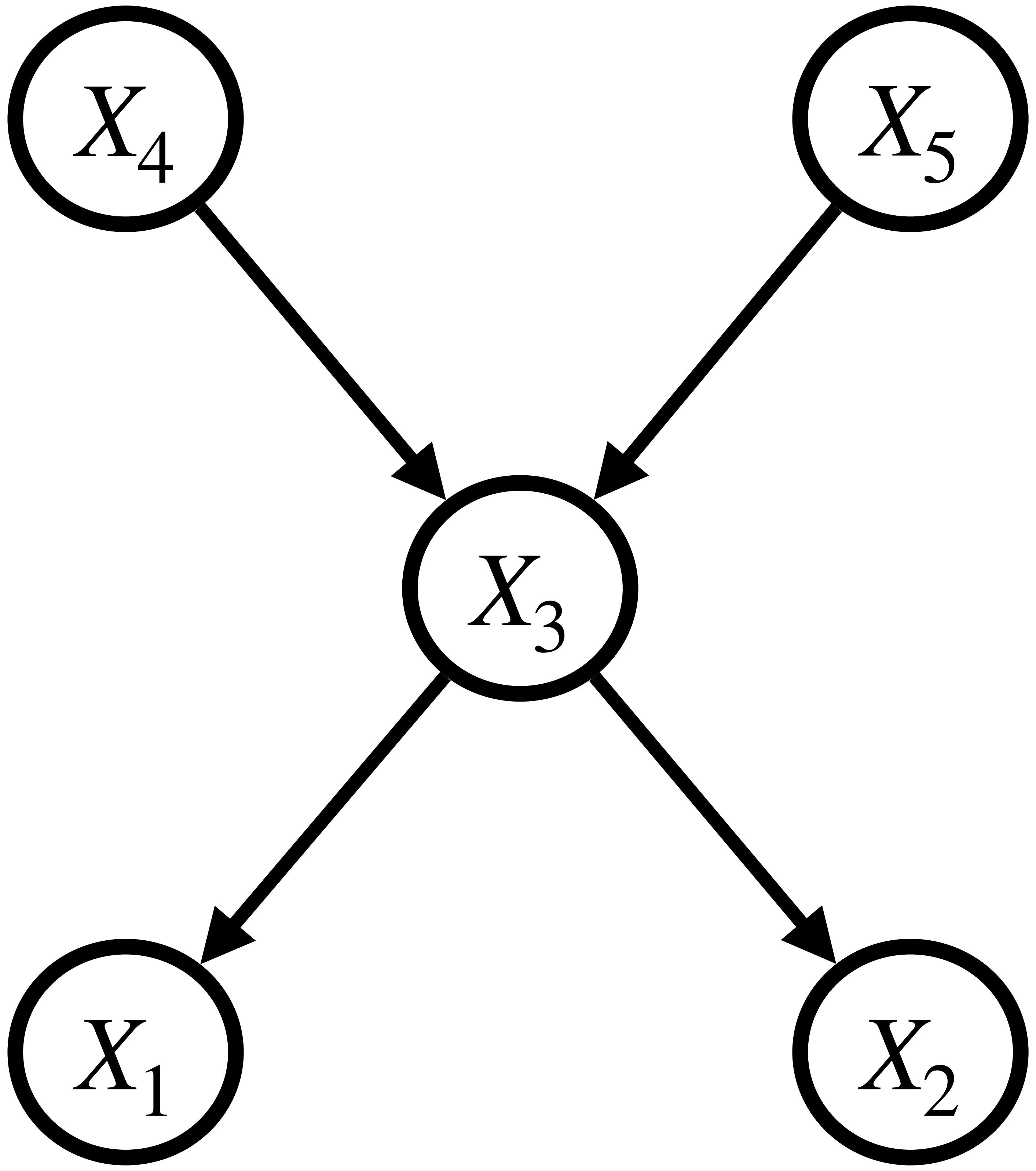
Bayesian Network

Let: $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random vector and let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denote a vector of values.

Definition: A Bayesian network for a joint distribution of the random vector $\mathbf{X}$ is given by:

- A directed acyclic graph G. The nodes of G correspond to the random variables $X_1, X_2, \ldots, X_n$.
- For every random variable X_i, a conditional distribution of X_i given its parents.

Bayesian Network (The Graph)



Bayesian Networks: Notation

Let: $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random vector and let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denote a vector of values. Let

Notation: To simplify notation in what follows, we will denote by

Par(X_i) … the vector of parents of X_i (the random variables corresponding to the parents)

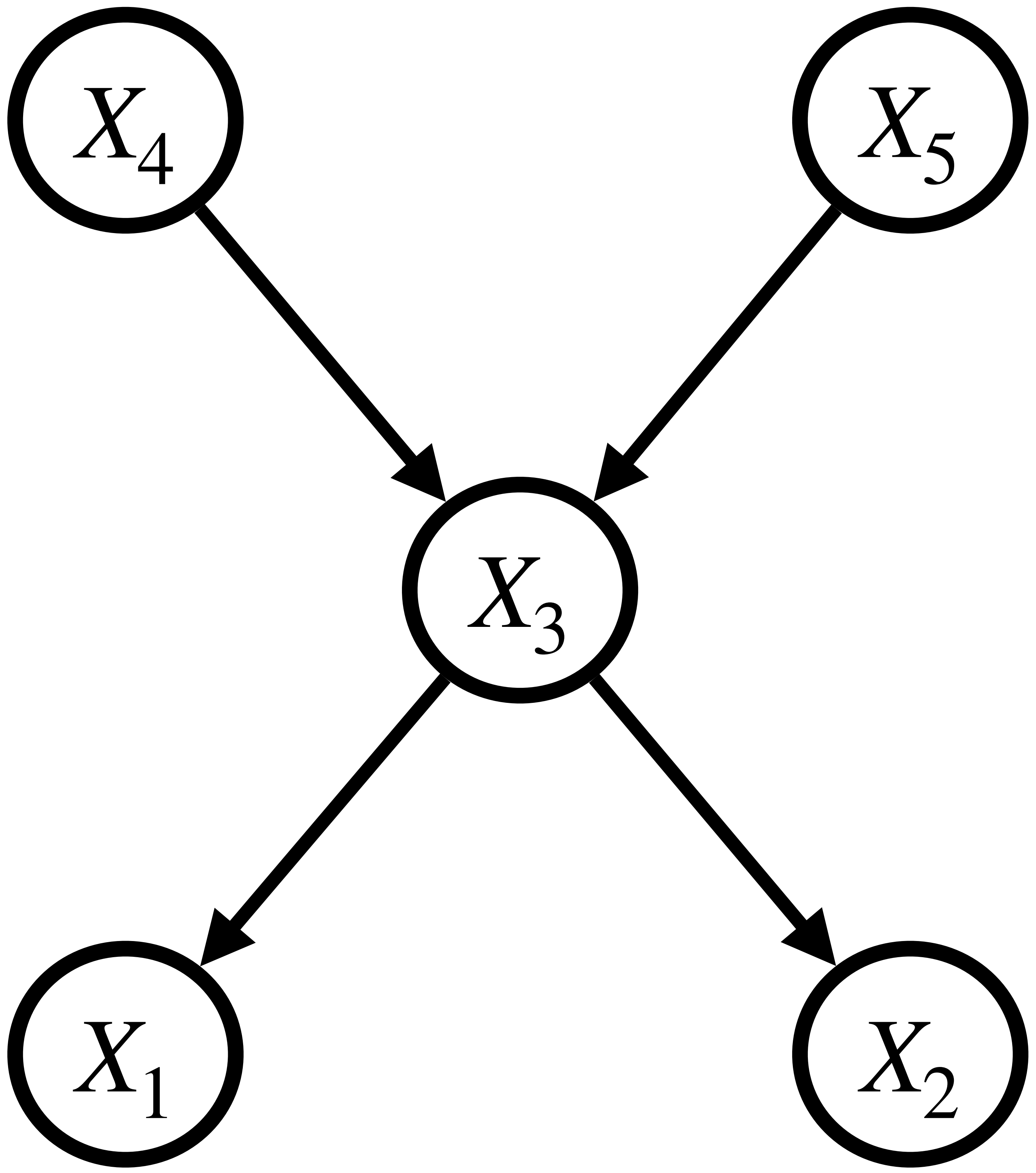
par$_{\mathbf{x}}(X_i)$ … the vector of **values** of the parents of X_i (the values are supposed to be taken from the vector of values $\mathbf{x}$).

Bayesian Network Distribution

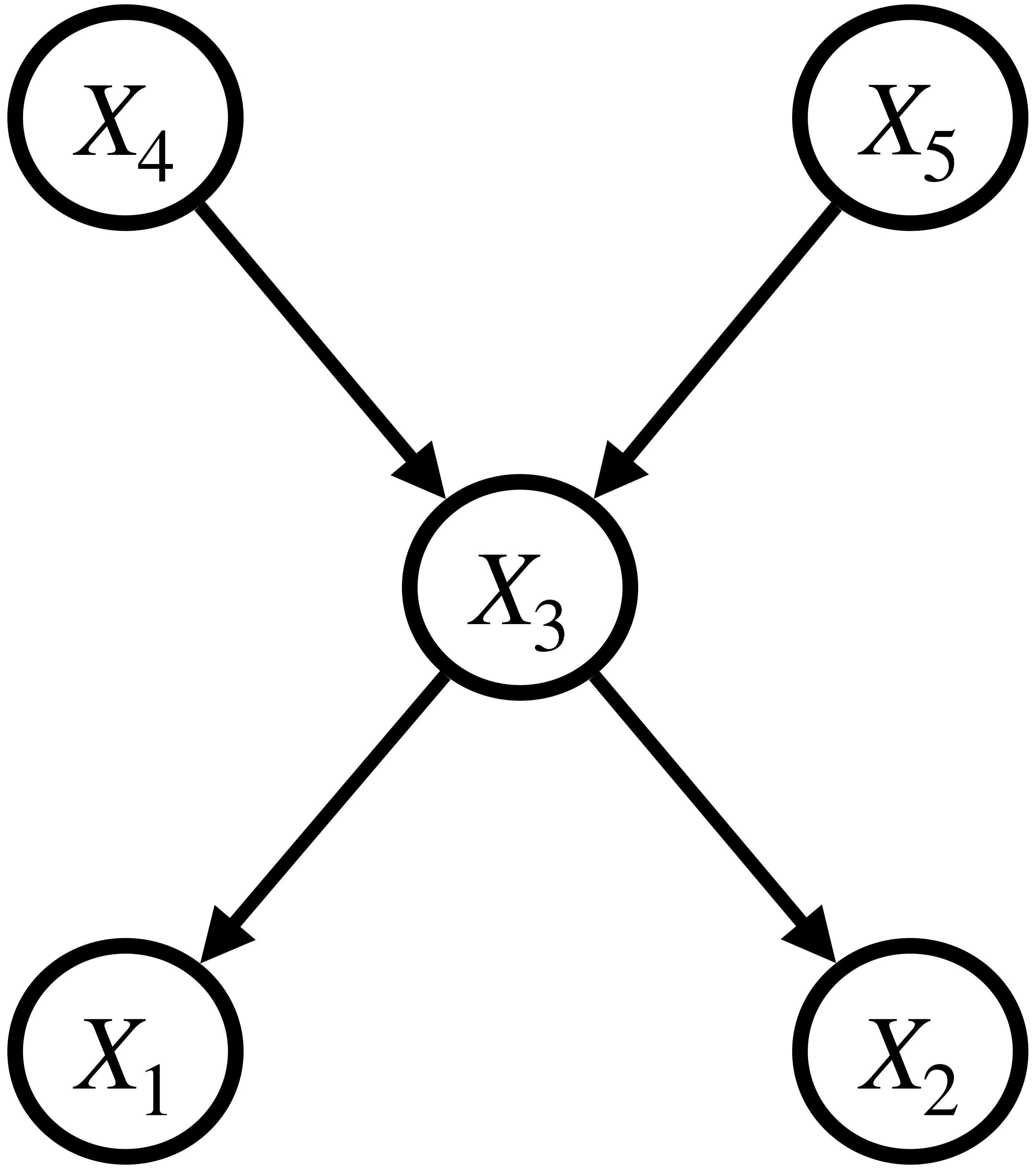
Given a BN with a graph G, the BN induces the following distribution:

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

Bayesian Network: Example (1/3)



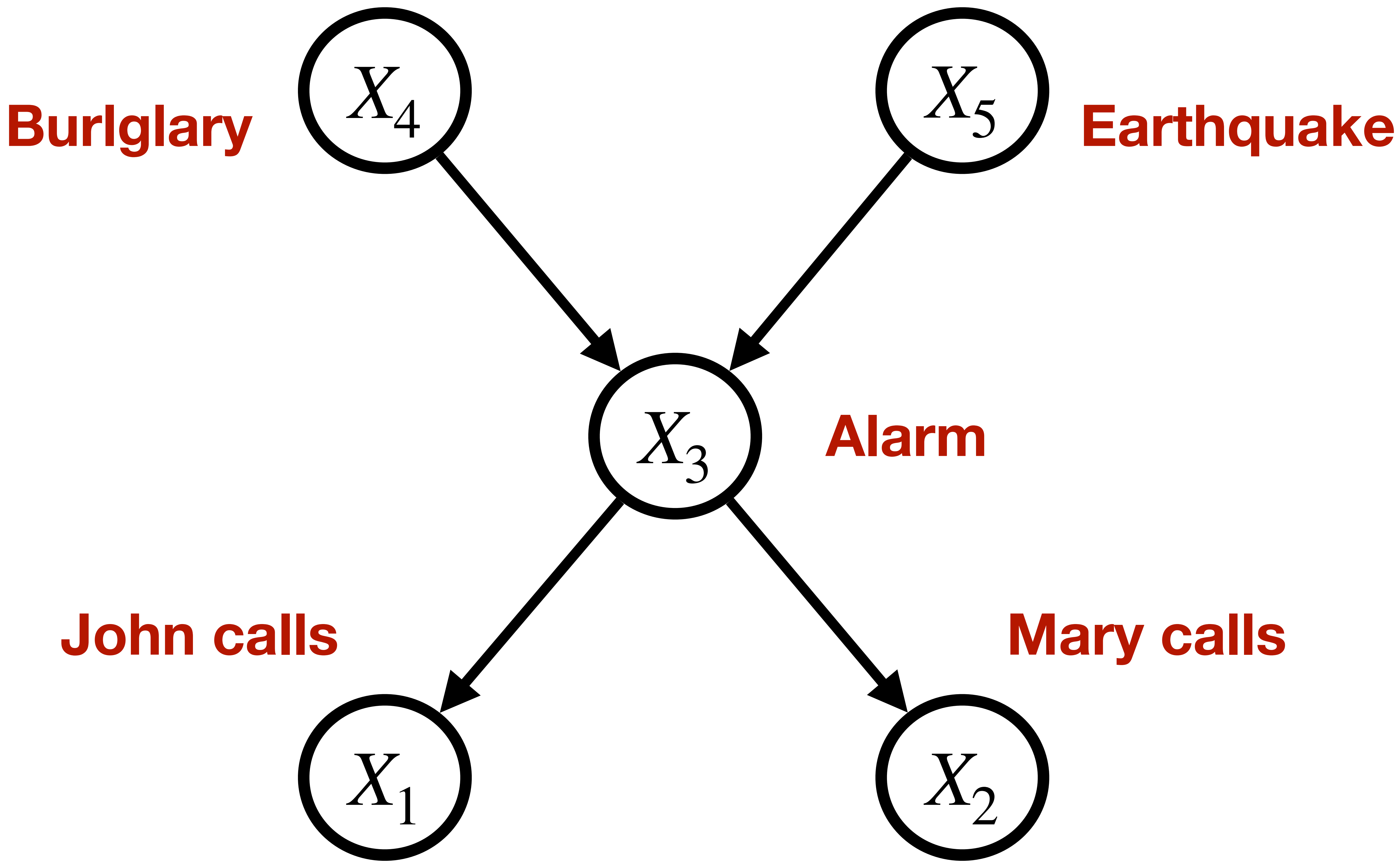
Bayesian Network: Example (2/3)



$$P(x_1, \ldots, x_5) = P_{X_4}(x_4) P_{X_5}(x_5) P_{X_3|X_4,X_5}(x_3 \,|\, x_4, x_5) P_{X_1|X_3}(x_1 \,|\, x_3) P_{X_2|X_3}(x_2 \,|\, x_3)$$

Bayesian Network: Example (3/3)

Let's make the example concrete



$$P(x_1, \ldots, x_5) = P_{X_4}(x_4)P_{X_5}(x_5)P_{X_3|X_4,X_5}(x_3 \mid x_4, x_5)P_{X_1|X_3}(x_1 \mid x_3)P_{X_2|X_3}(x_2 \mid x_3)$$

Conditional Independence in BNs

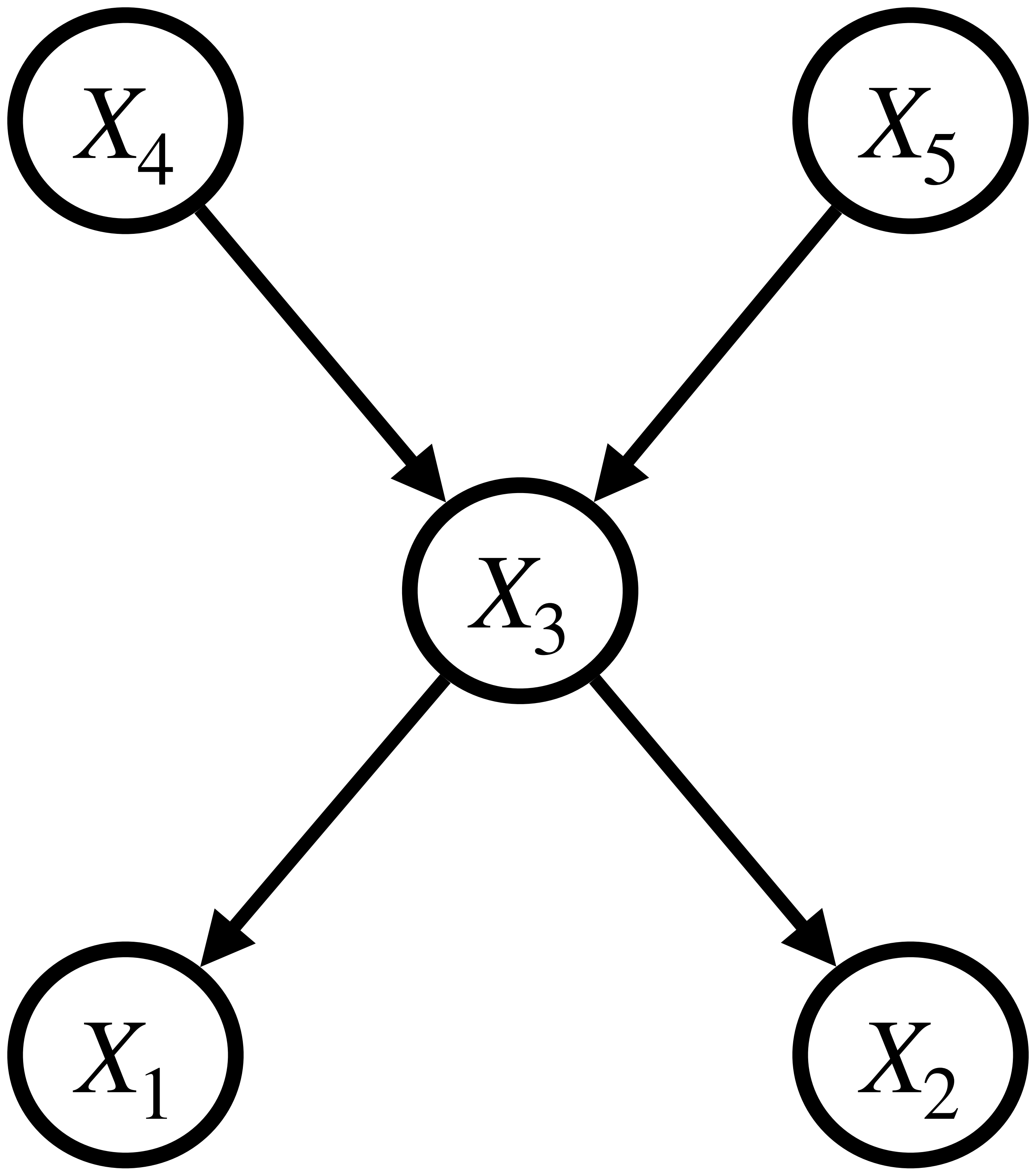
What are the conditional independence assumptions behind BNs?

The main one is that, for every X_i: X_i is conditionally independent of its ancestors given its parents.

This can be equivalently stated as follows:

Let Anc(X_i) be the ancestors of X_i, i.e. nodes in the BN from which X_i can be reached.

$$P_{X_i|Par(X_i)}(x_i \,|\, \text{par}_{\mathbf{x}}(X_i)) = P_{X_i|Anc(X_i)}(x_i \,|\, \text{anc}_{\mathbf{x}}(X_i)).$$



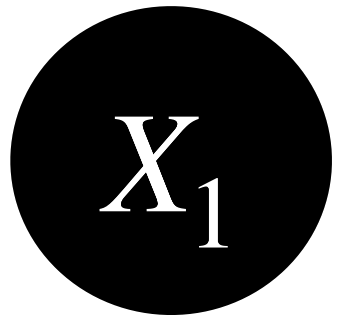
$$P_{X_1|X_3}(x_1\,|\,x_3) = P_{X_1|X_3,X_4,X_5}(x_1\,|\,x_3, x_4, x_5)$$
$$P_{X_2|X_3}(x_2\,|\,x_3) = P_{X_2|X_3,X_4,X_5}(x_2\,|\,x_3, x_4, x_5)$$

Part 4: More on Conditional Independence (D-Separation)

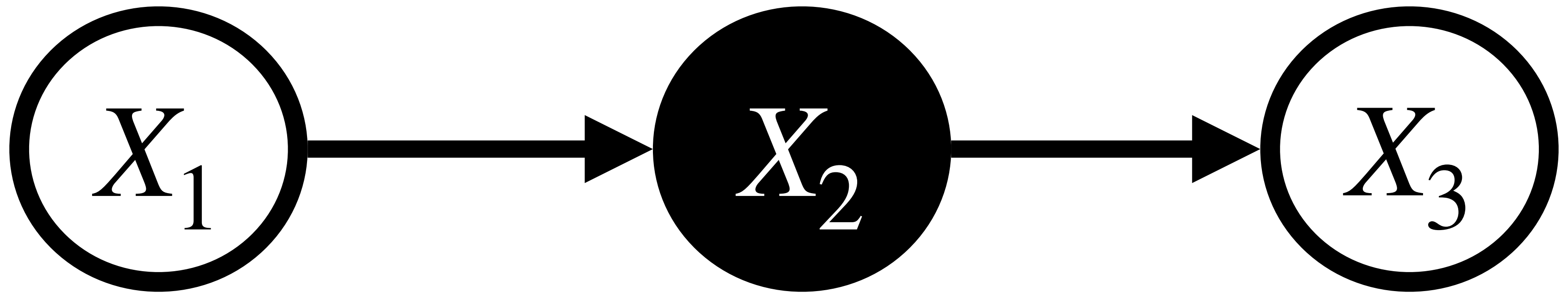
More Conditional Independencies?

In general, a BN encodes many conditional independencies. We will now learn to recognize them.

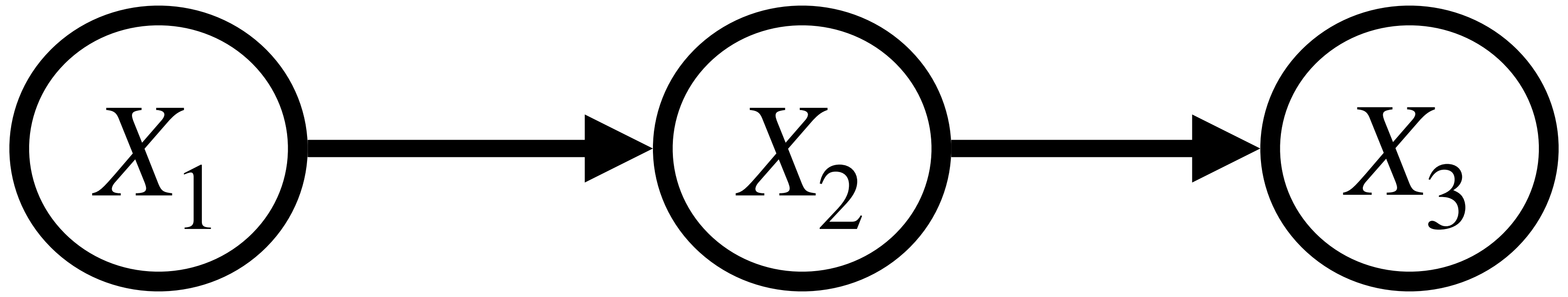
In what follows, nodes on which we condition will be shown as full black ovals, e.g.:



Causal Chain (1/2)



$$X_1 \perp\!\!\!\perp X_3 \mid X_2$$



X_1 **and** X_3 **not independent (unconditionally)**

Causal Chain (2/2)

The conditional independence part can be shown as follows:

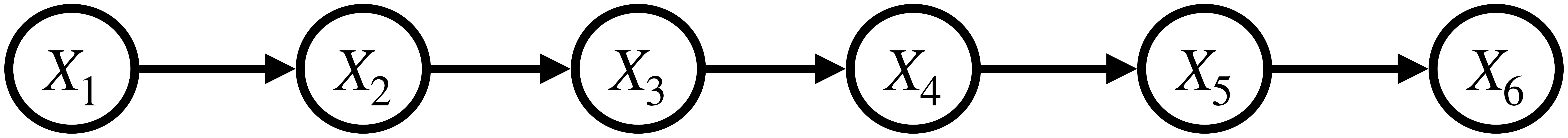
$$P_{X_1,X_3|X_2}(x_1, x_3 \,|\, x_2) = \frac{P_{X_1,X_2,X_3}(x_1, x_2, x_3)}{P_{X_2}(x_2)} =$$

$$= \frac{P_{X_1}(x_1)P_{X_2|X_1}(x_2 \,|\, x_1)P_{X_3|X_2}(x_3 \,|\, x_2)}{P_{X_2}(x_2)} = \underbrace{\frac{P_{X_1}(x_1)P_{X_2|X_1}(x_2 \,|\, x_1)}{P_{X_2}(x_2)}}_{=P_{X_1|X_2}(x_1|x_2)} P_{X_3|X_2}(x_3 \,|\, x_2) =$$

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

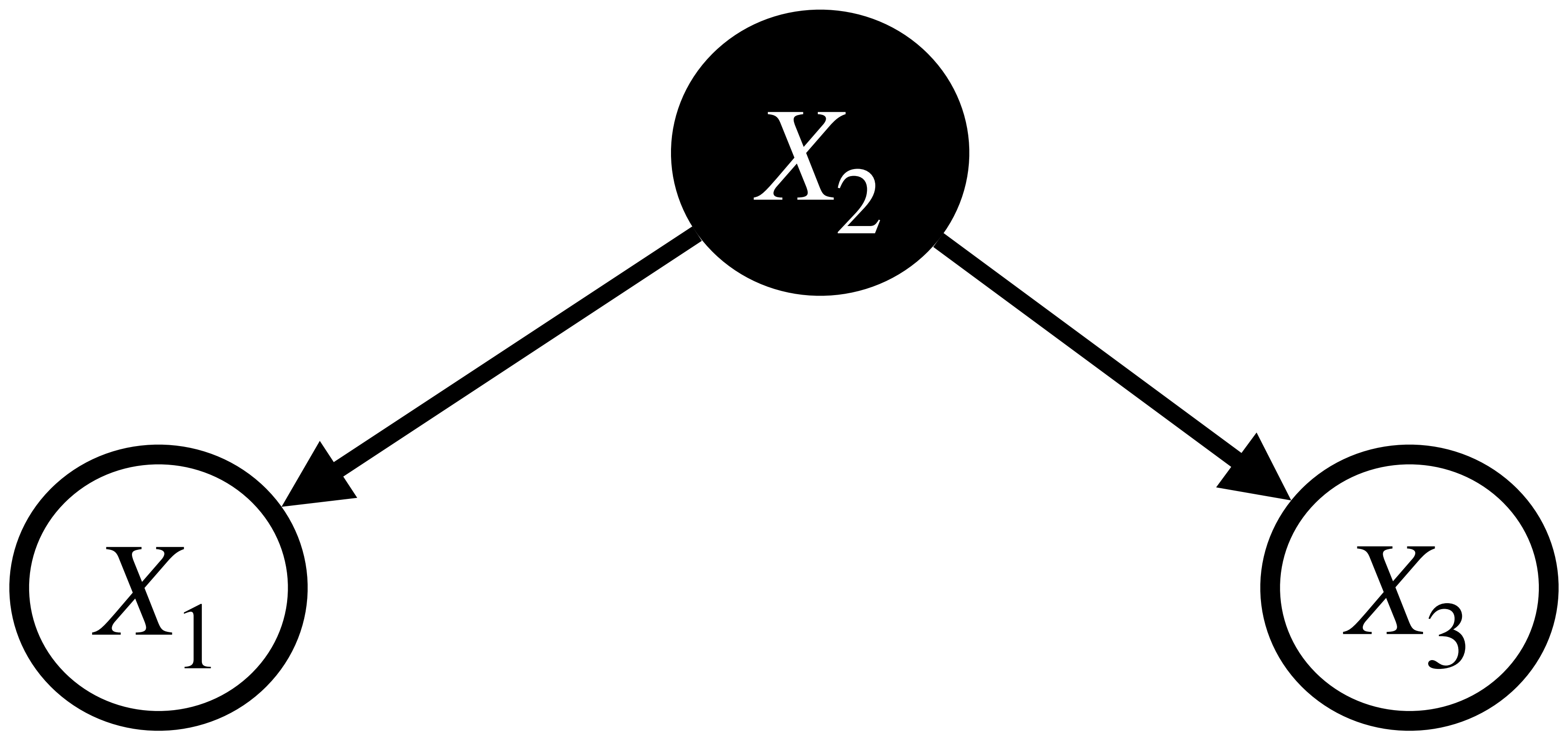
Causal Chain: Example

You know one example already… Markov process.

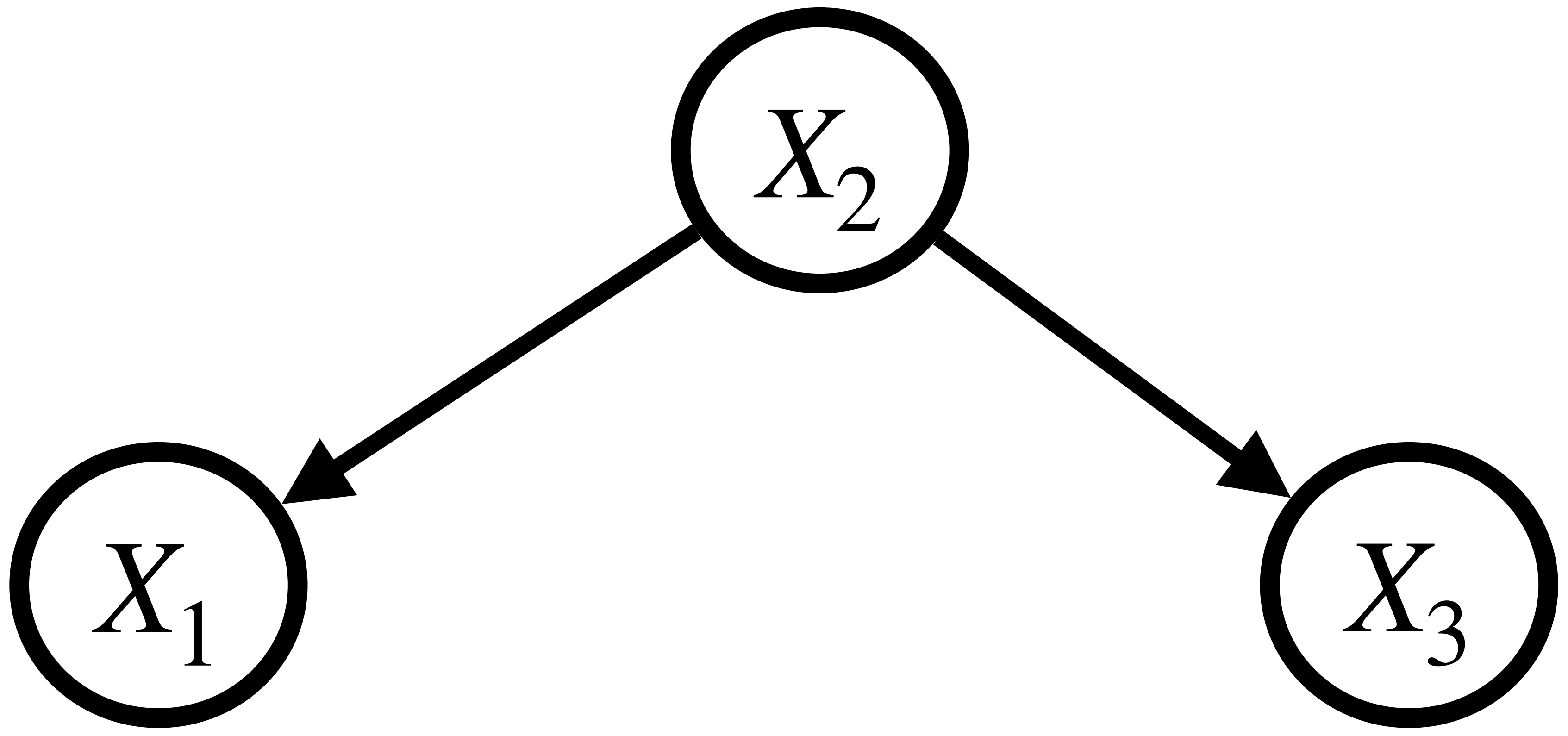


The possible values of each X_i are the states from the state space S.

Common Cause (1/2)



$$X_1 \perp\!\!\!\perp X_3 \mid X_2$$



X_1 and X_3 not independent (unconditionally)

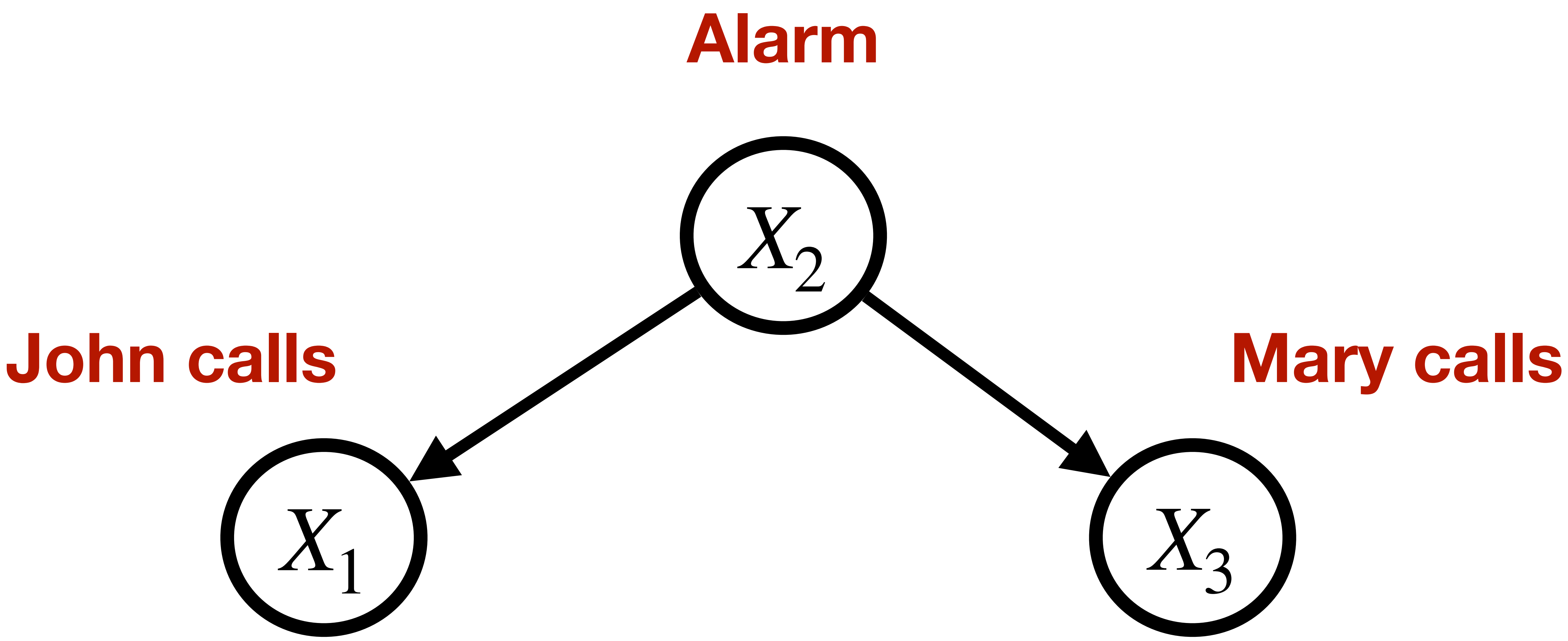
Common Cause (2/2)

The conditional independence part can be shown as follows:

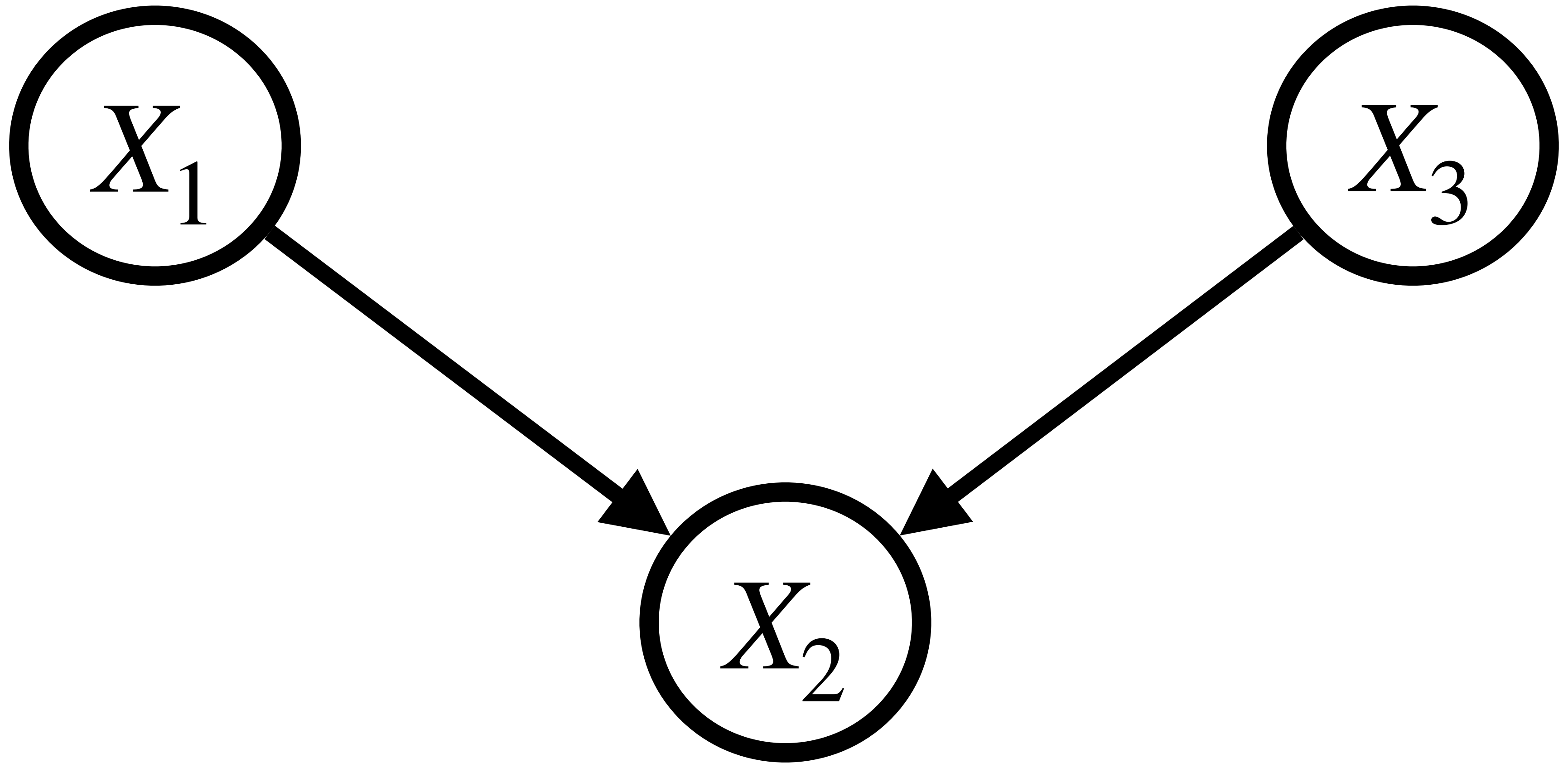
$$P_{X_1,X_3|X_2}(x_1, x_3 \,|\, x_2) = \frac{P_{X_1,X_2,X_3}(x_1, x_2, x_3)}{P_{X_2}(x_2)} =$$

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

Common Cause: Example

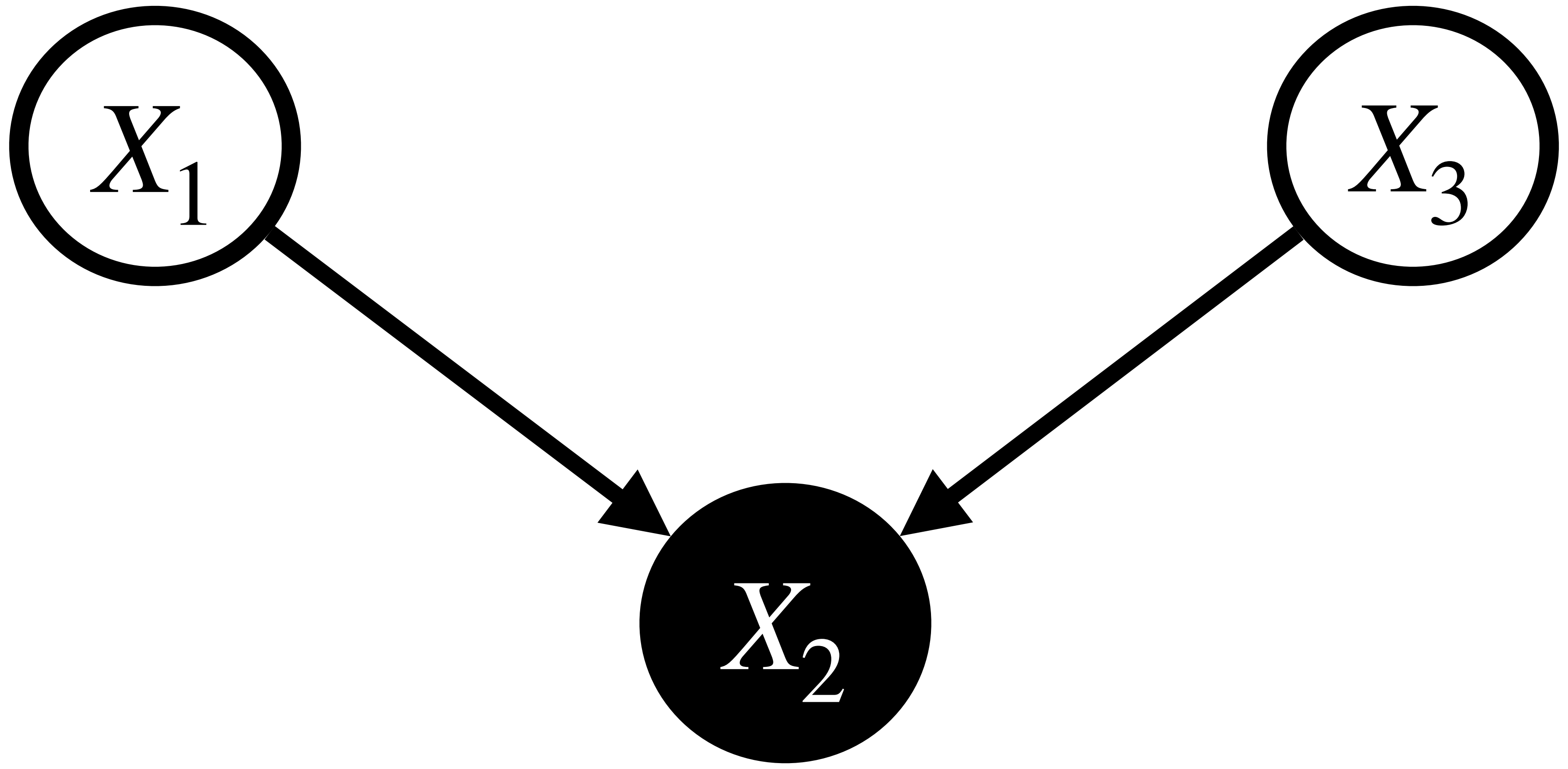


Common Effect (1/2)



Independent unconditionally

$$X_1 \perp\!\!\!\perp X_3$$



But X_1 and X_3 are NOT independent given the value of X_2!!!!

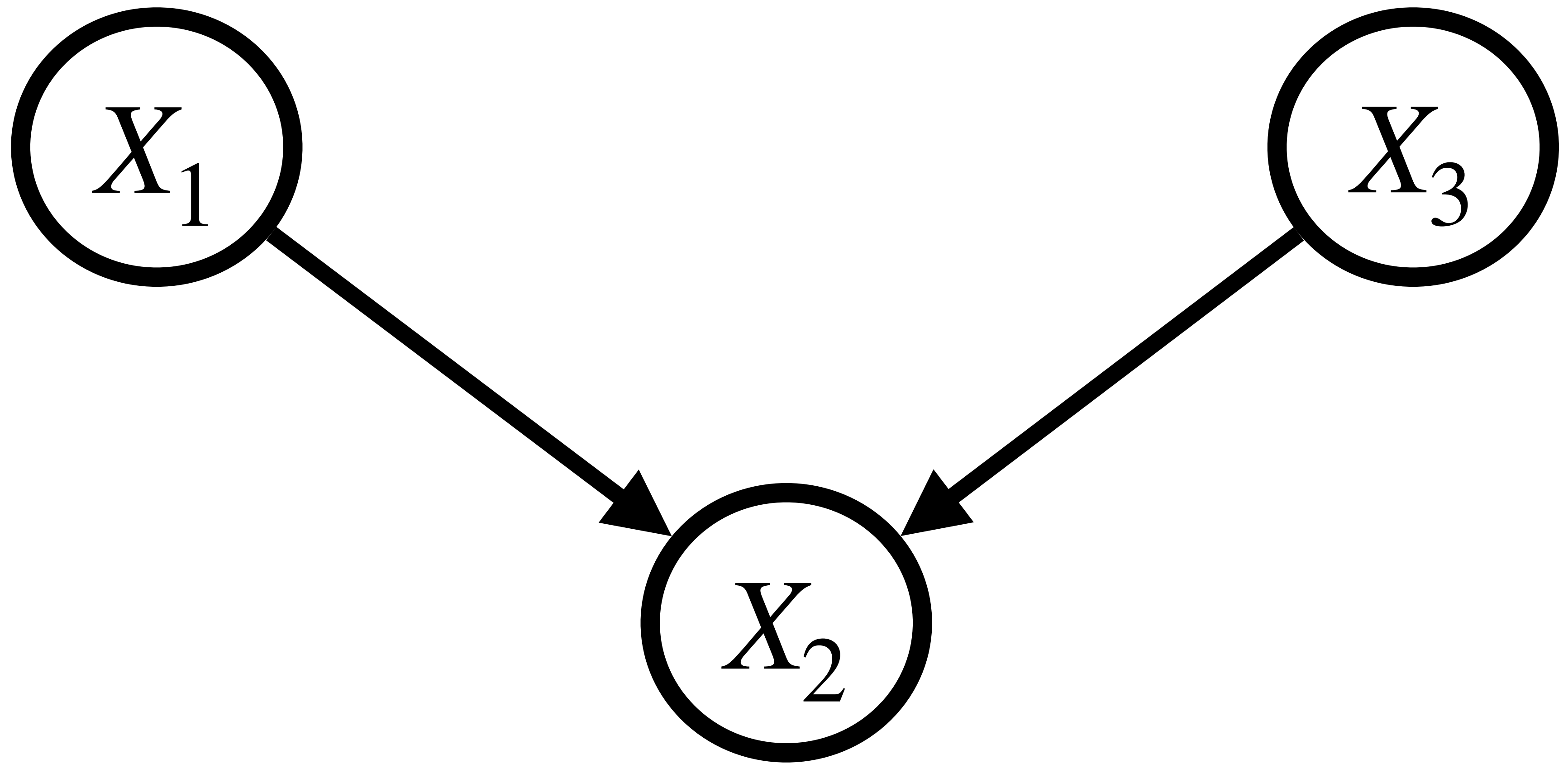
Common Effect (2/2)

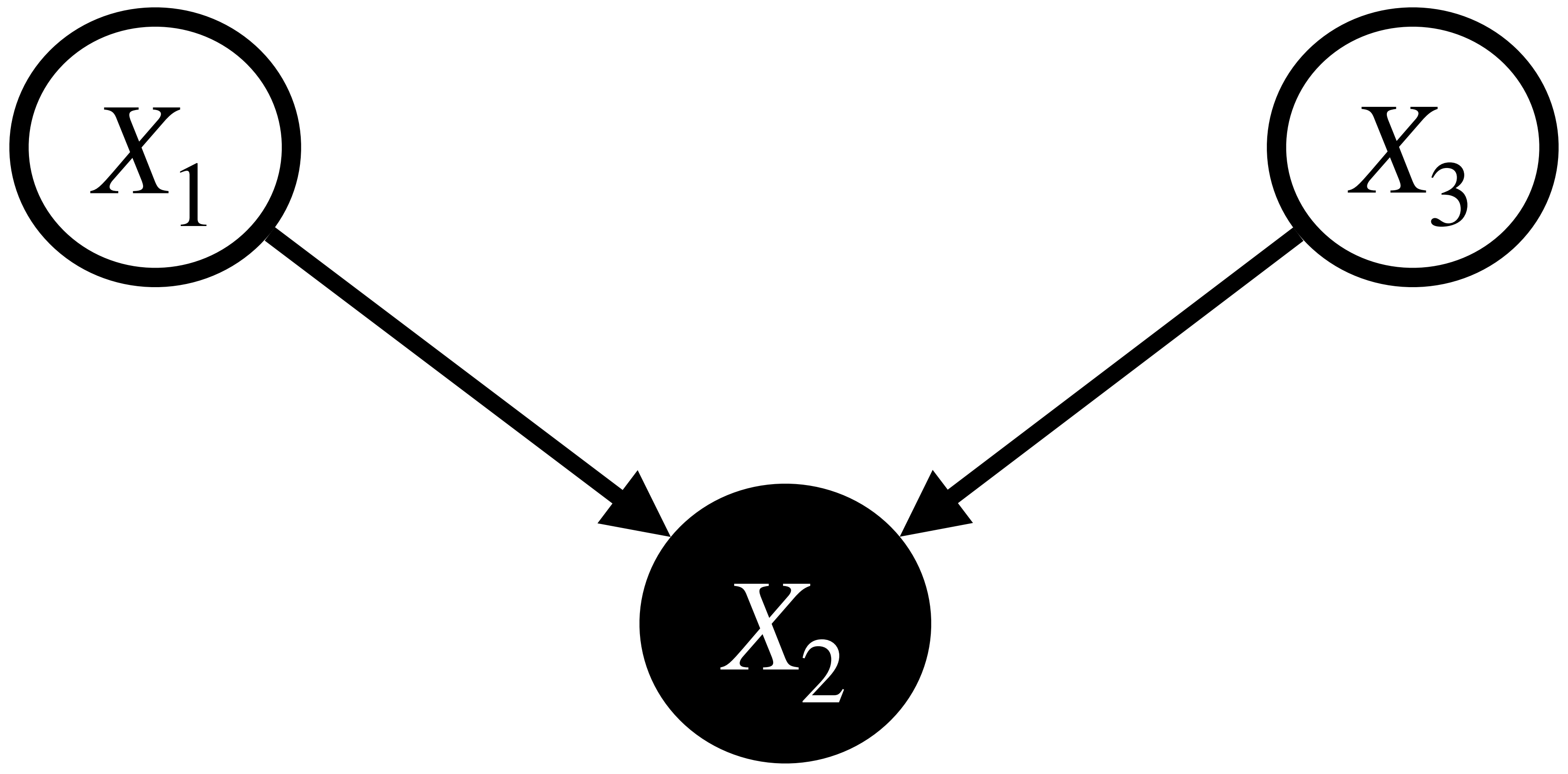
The independence part can be shown as follows:

$$P_{X_1,X_3}(x_1, x_3) = \sum_{x_2} P_{X_1,X_2,X_3}(x_1, x_2, x_3) = \sum_{x_2} P_{X_2|X_1,X_3}(x_2 \,|\, x_1, x_3) P_{X_1}(x_1) P_{X_3}(x_3) =$$

$$= P_{X_1}(x_1) P_{X_3}(x_3) \underbrace{\sum_{x_2} P_{X_2|X_1,X_3}(x_2 \,|\, x_1, x_3)}_{=1} = P_{X_1}(x_1) P_{X_3}(x_3).$$

Common Effect: Example





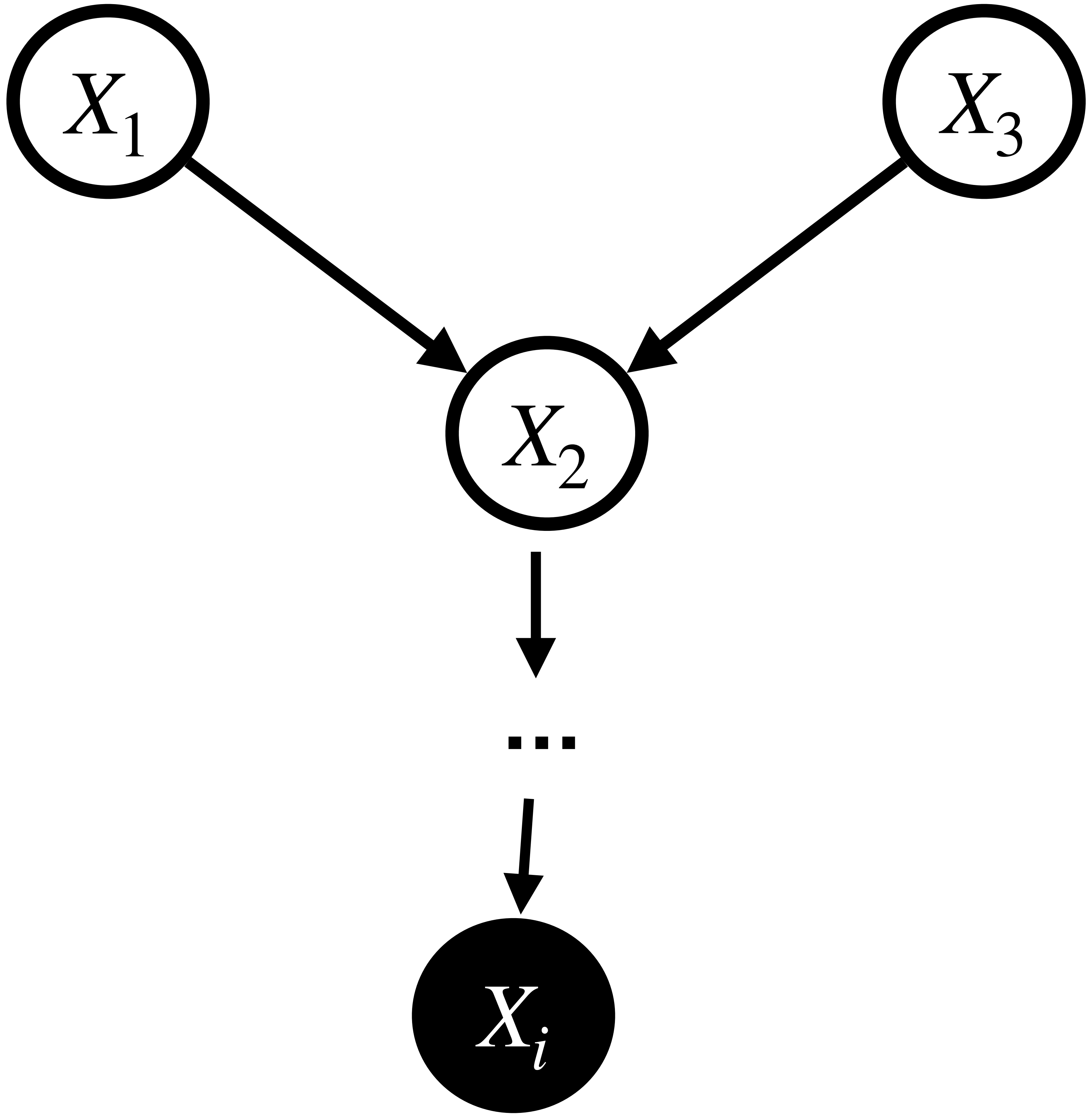
X_1 … flip a coin (the result is 0 or 1)

X_3 … flip a coin (the result is 0 or 1)

$X_2 = X_1 \oplus X_3$.

Then if we do not condition on X_2, X_1 and X_3 are independent, but if we do condition on X_2 then just fixing the value of X_1 determines the value of X_3, so they are not conditionally independent given X_2!

Common Effect - Descendants



X_1 **and** X_3 **are NOT independent given the value of** X_i**!!!!**

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.

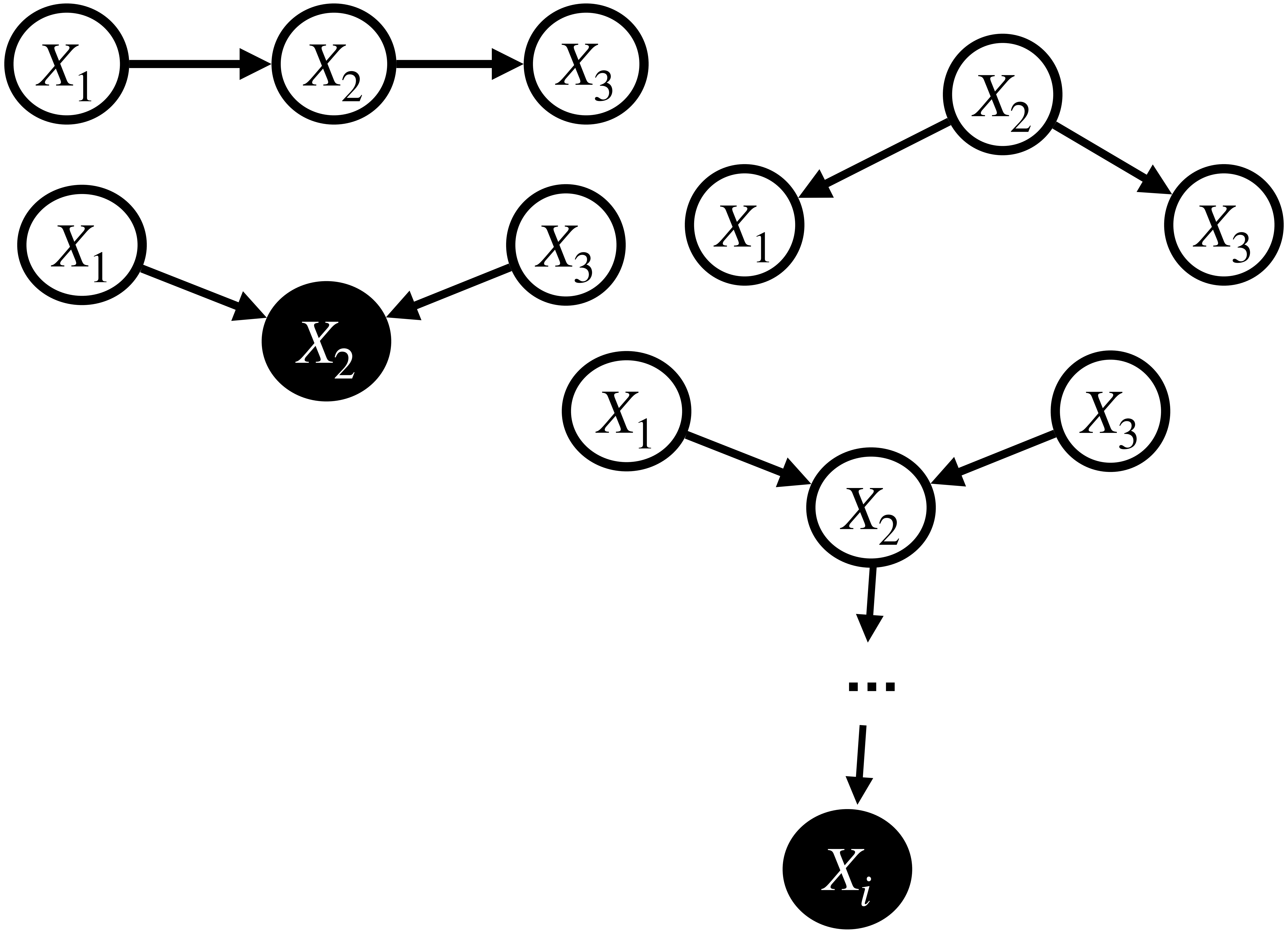
Active Path (1/3)

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

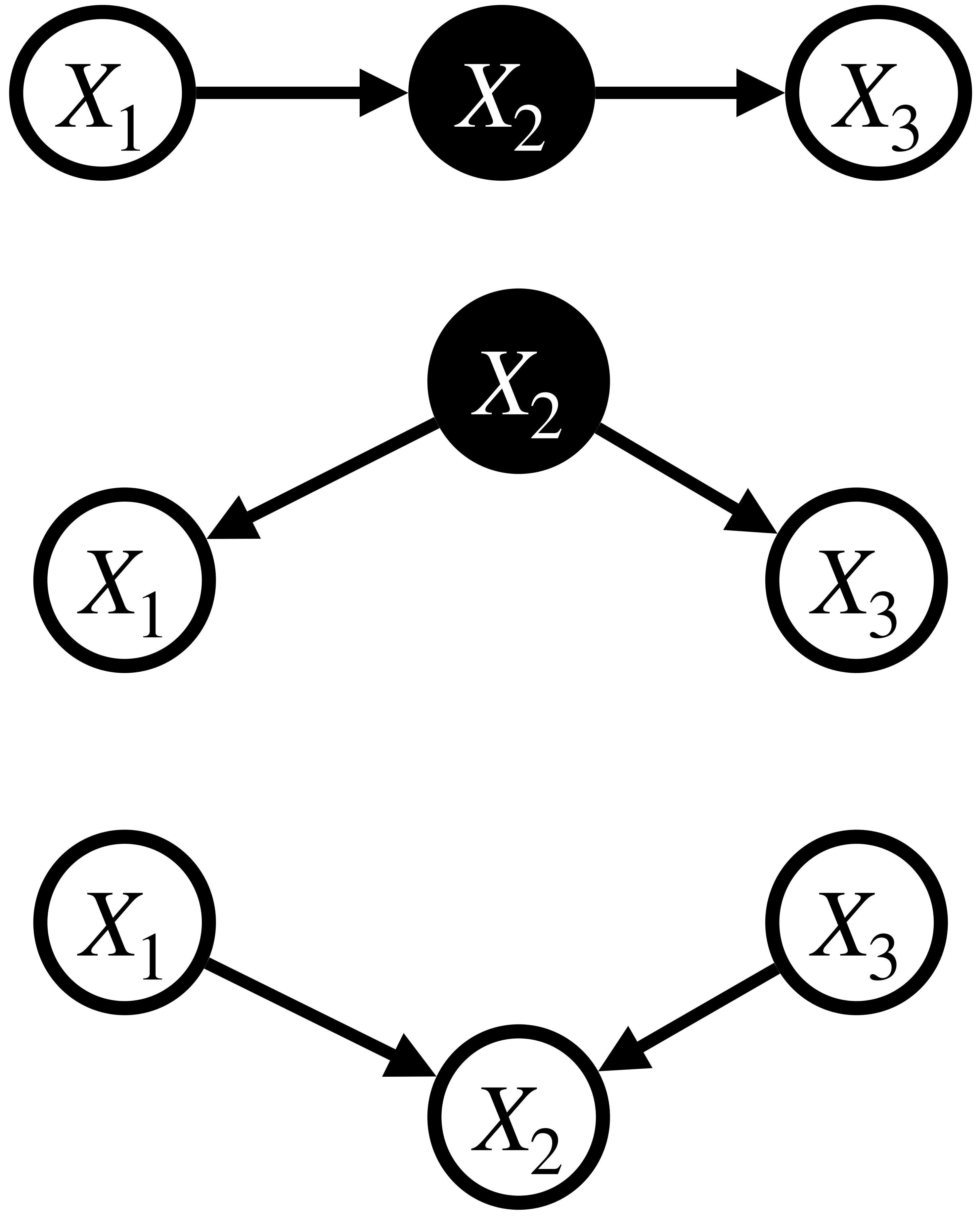
Terminology: Nodes which we condition on will be called **observed nodes** and the others **unobserved nodes**.

Active Path (2/3)

Active triples:



Blocked triples:

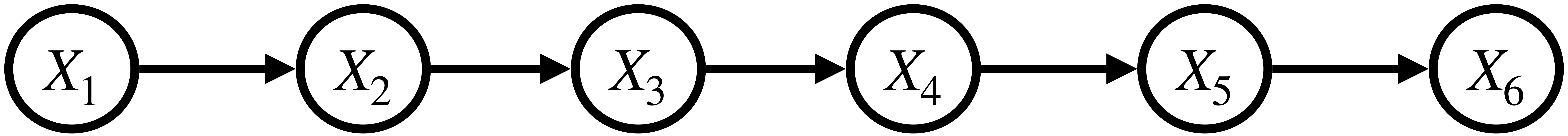


Active Path (3/3)

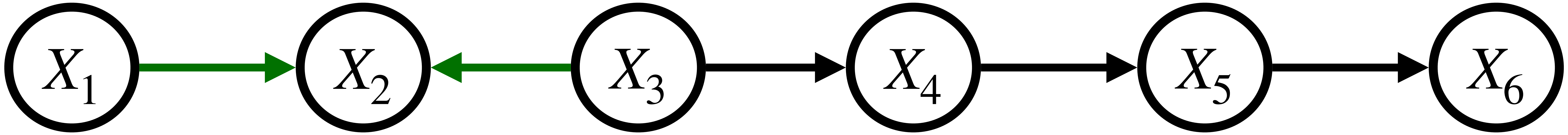
Definition: A path is active if all triples along it are active. Otherwise it is blocked.

EXAMPLES:

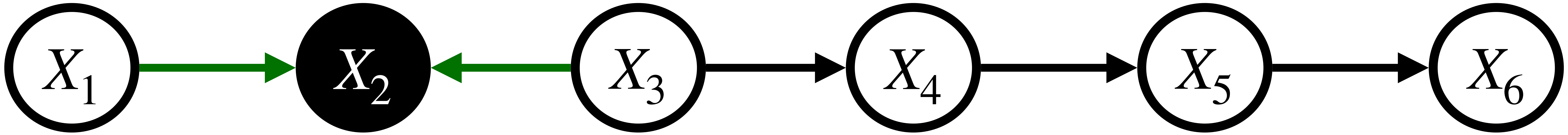
The path from X_1 to X_6 is active.



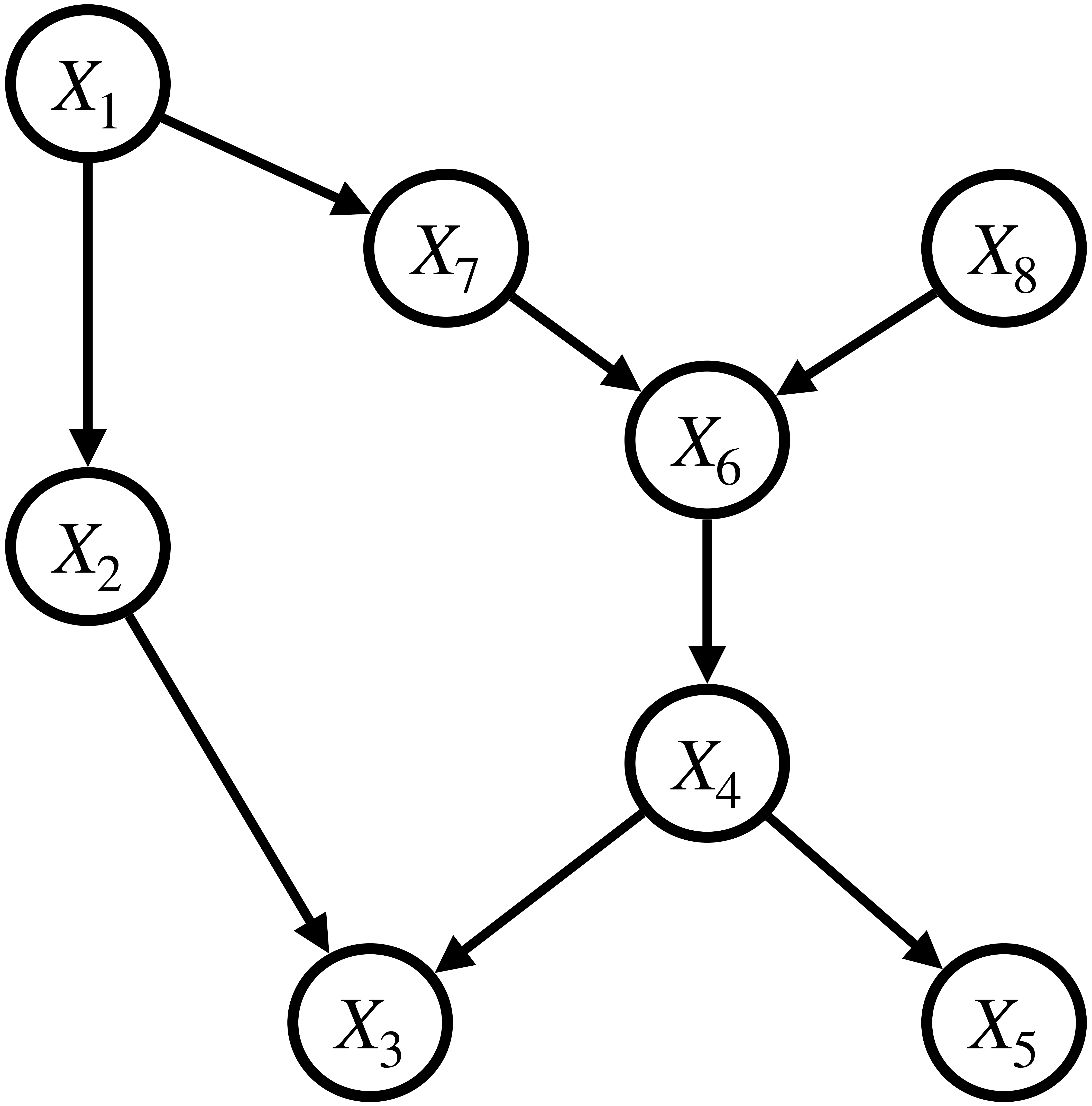
The path from X_1 to X_6 is blocked.



The path from X_1 to X_6 is active.

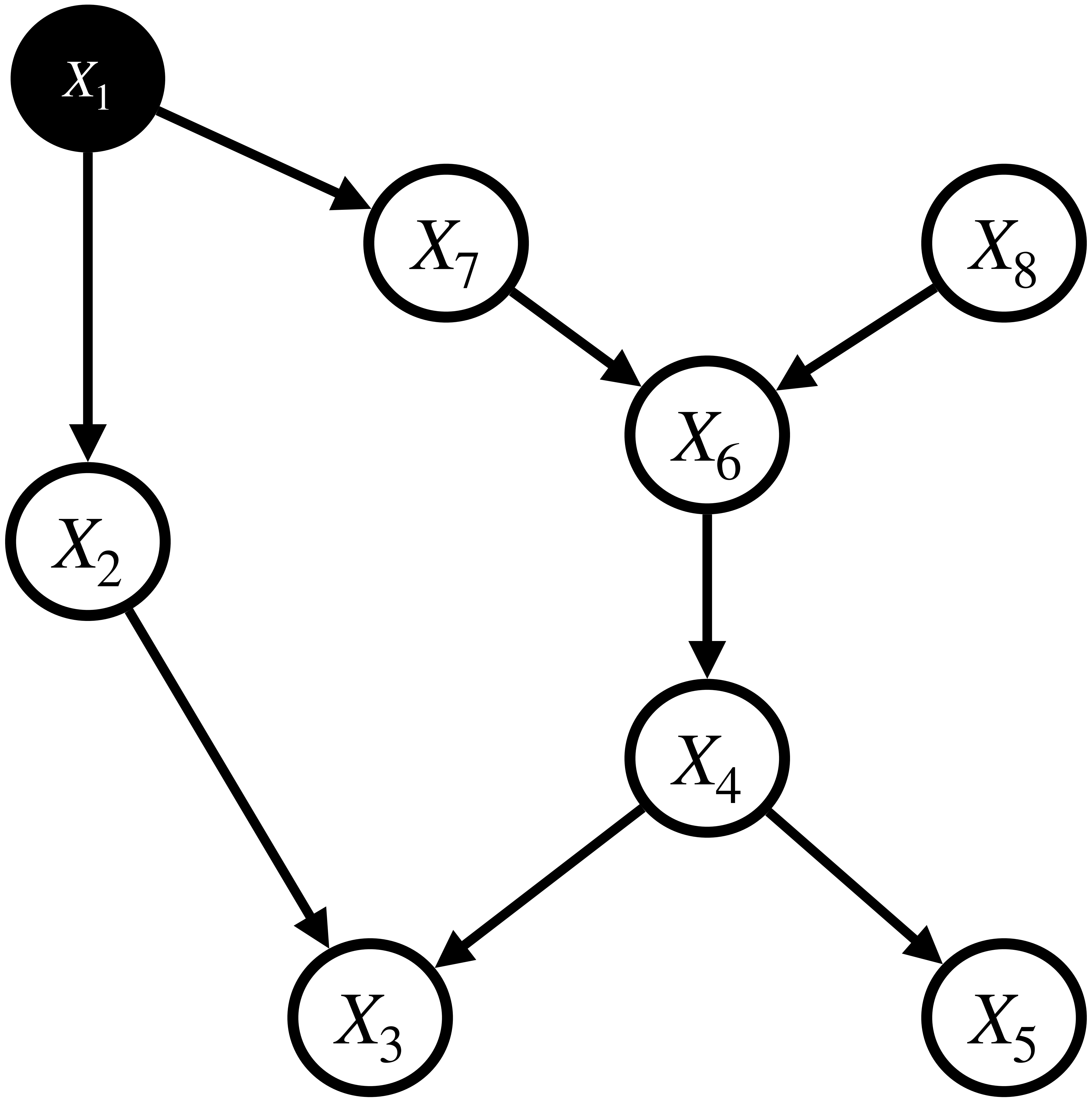


D-Separation (Examples)



Credit: Petr Pošík

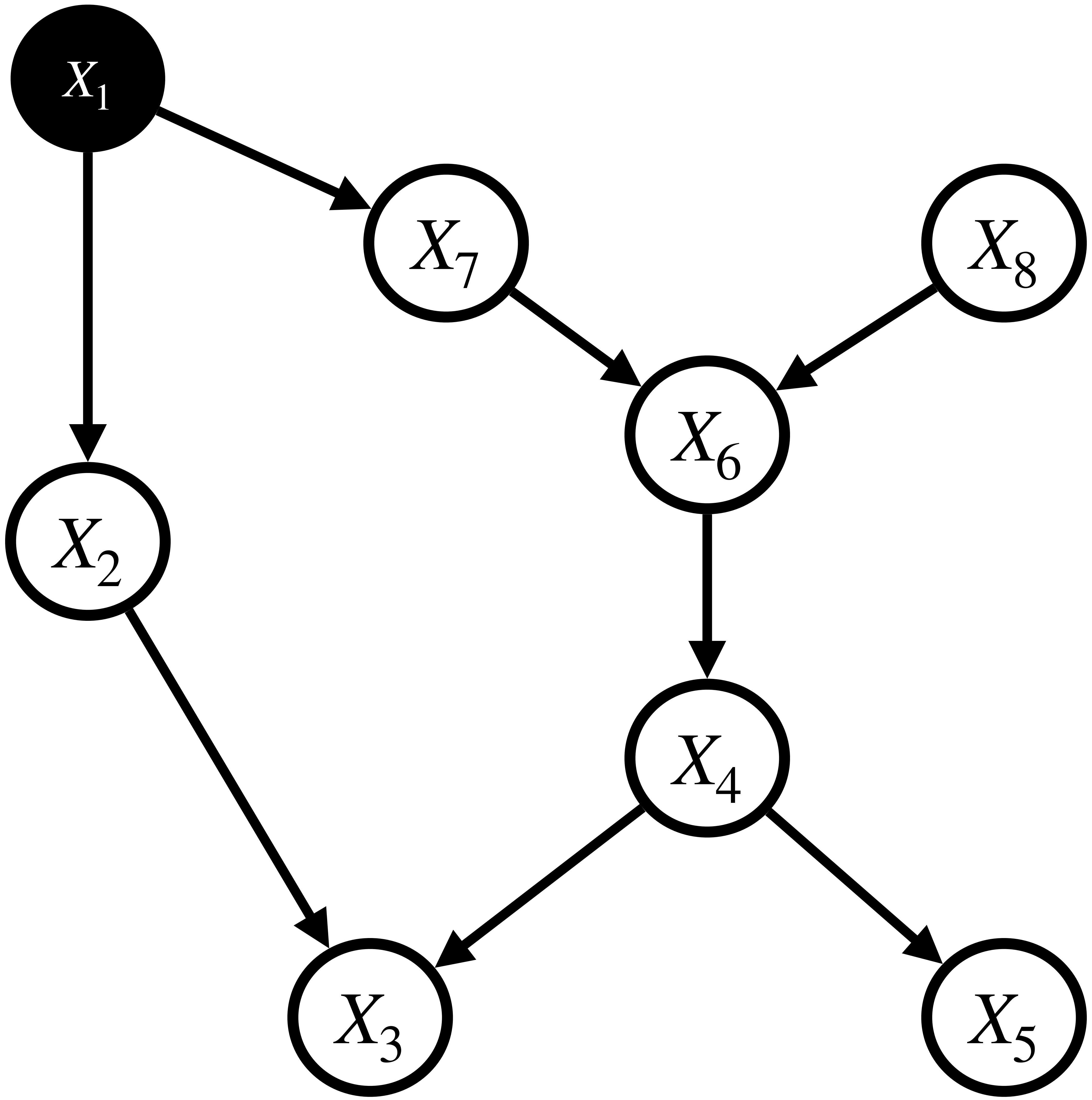
D-Separation (Example 1)



$X_2 \perp\!\!\!\perp X_7 \,|\, X_1$?

Credit: Petr Pošík

D-Separation (Example 1)

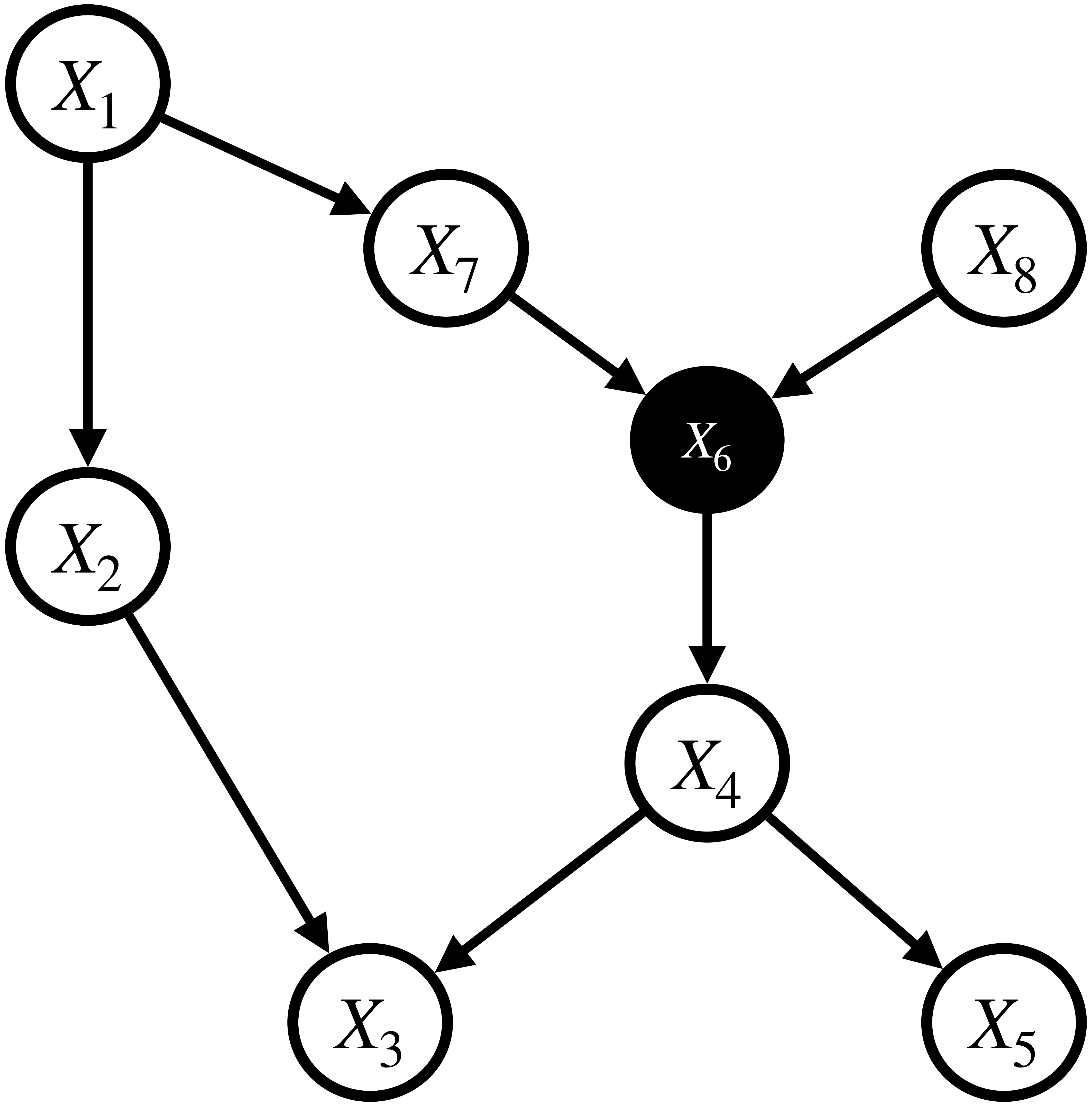


$X_2 \perp\!\!\!\perp X_7 \mid X_1$*?*

Yes, (X_2, X_1, X_7) *is blocked and* (X_2, X_3, X_4) *is also blocked.*

Credit: Petr Pošík

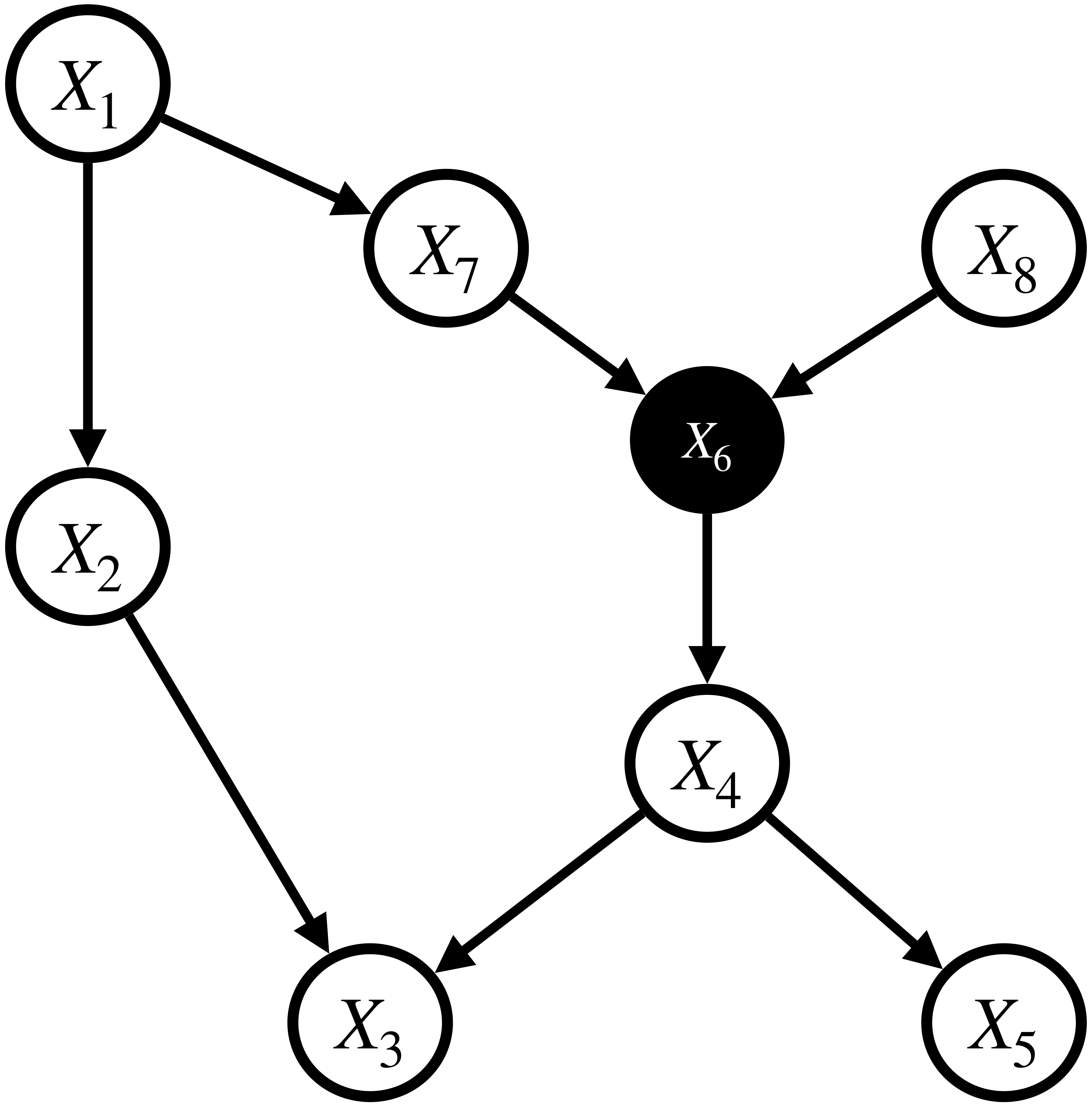
D-Separation (Example 2)



$X_1 \perp\!\!\!\perp X_4 \mid X_6$?

Credit: Petr Pošík

D-Separation (Example 2)

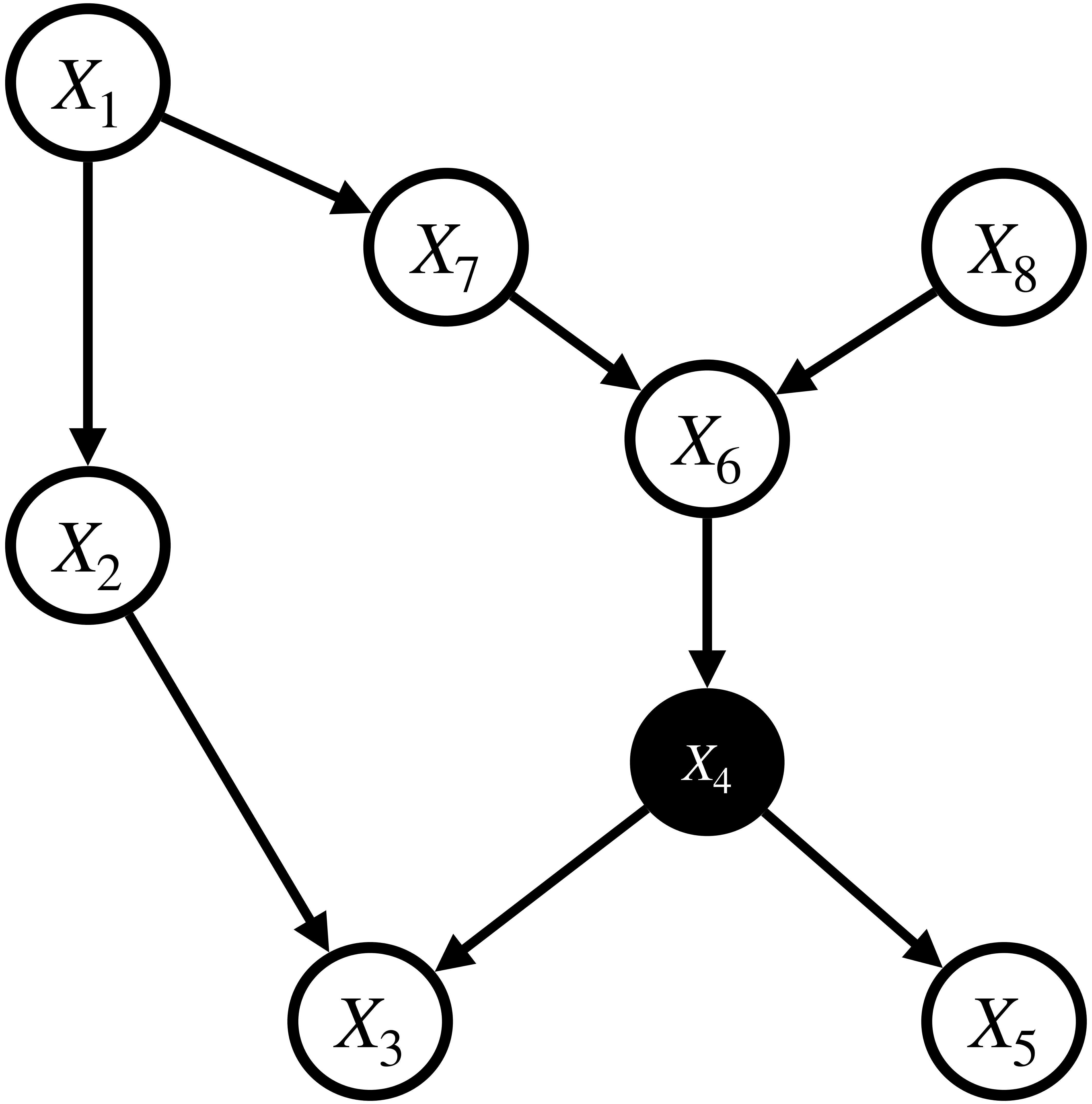


$X_1 \perp\!\!\!\perp X_4 \,|\, X_6$?

Yes, (X_2, X_3, X_3) *is blocked and so is* (X_7, X_6, X_4).

Credit: Petr Pošík

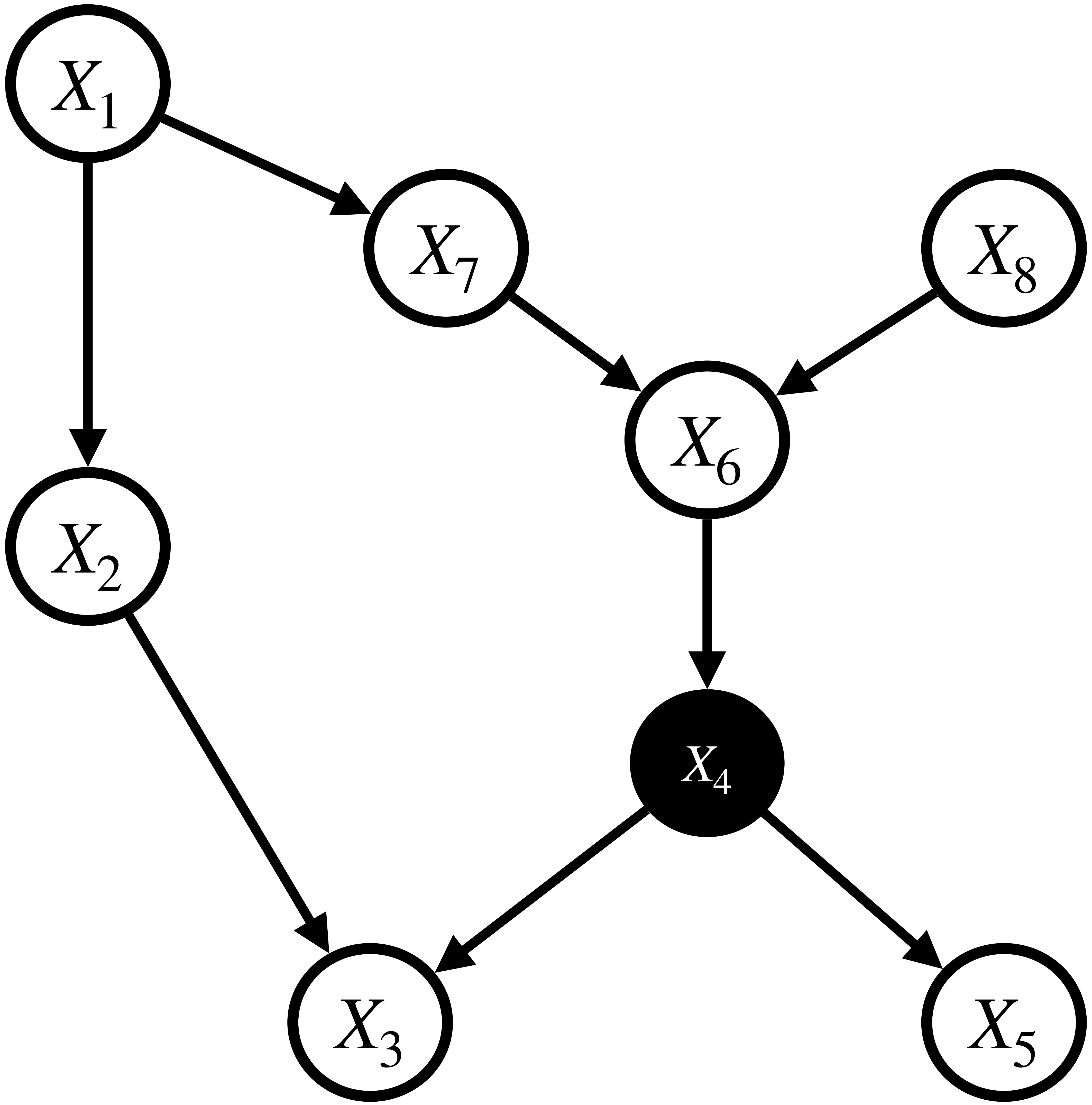
D-Separation (Example 3)



$X_7 \perp\!\!\!\perp X_8 \mid X_4?$

Credit: Petr Pošík

D-Separation (Example 3)

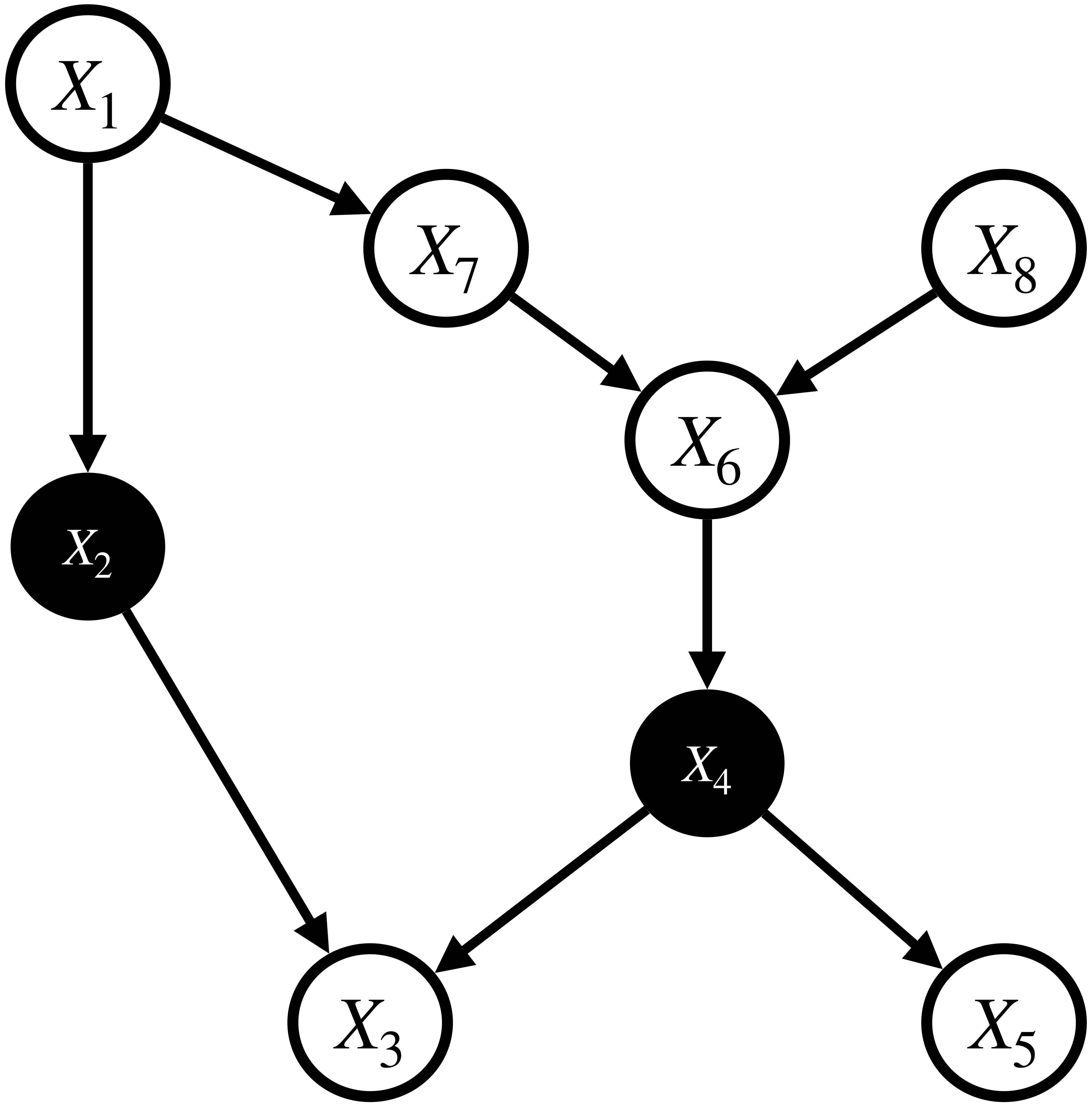


$X_7 \perp\!\!\!\perp X_8 \mid X_4$?

No! There is an active path: X_7, X_6, X_8 (observed descendant)

Credit: Petr Pošík

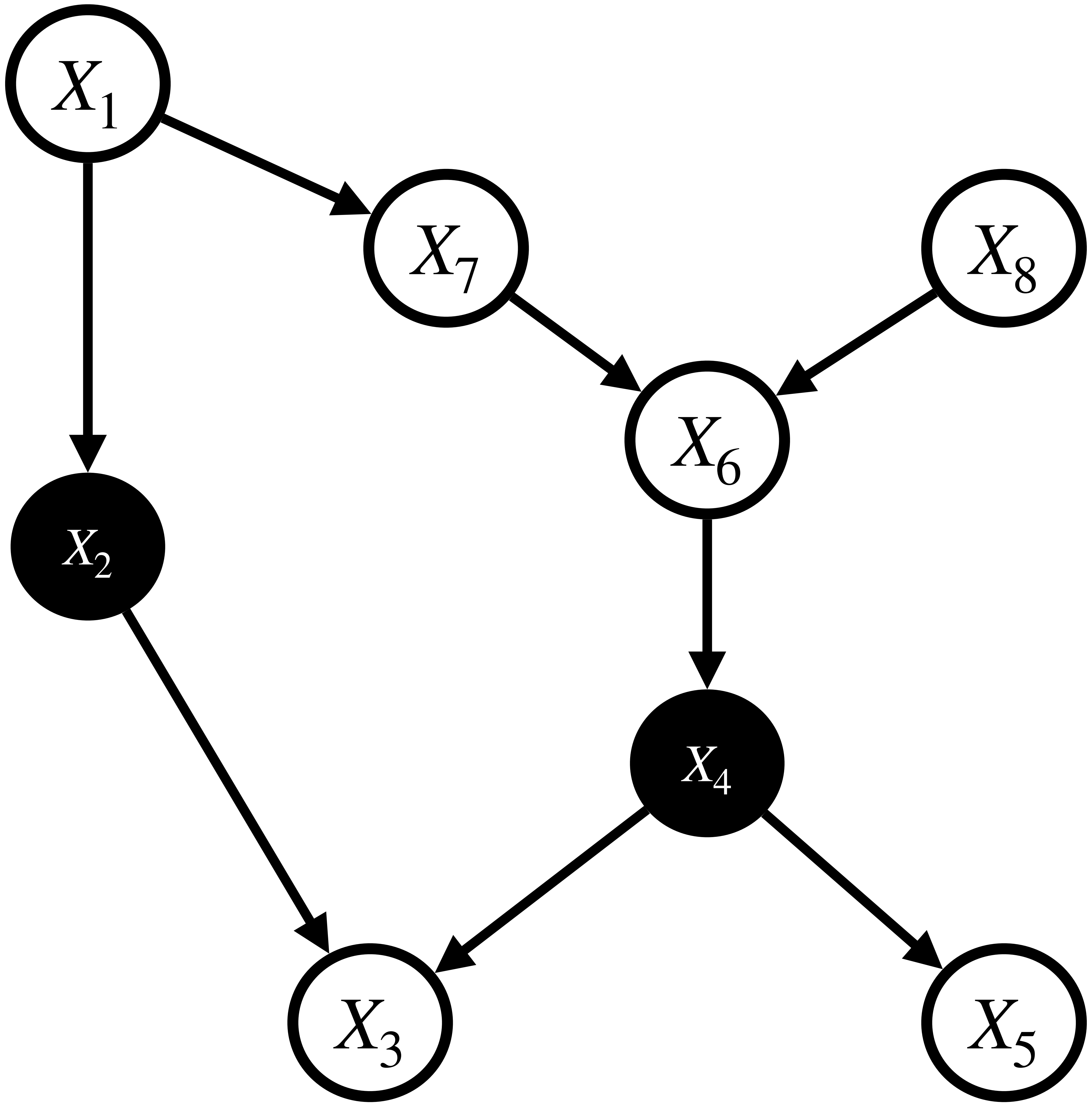
D-Separation (Example 4)



$$X_1 \perp\!\!\!\perp X_3 \mid \{X_2, X_4\}?$$

Credit: Petr Pošík

D-Separation (Example 4)



$X_1 \perp\!\!\!\perp X_3 \mid \{X_2, X_4\}$?

Yes! (X_1, X_2, X_3) *and* (X_6, X_4, X_3) *are both blocked.*

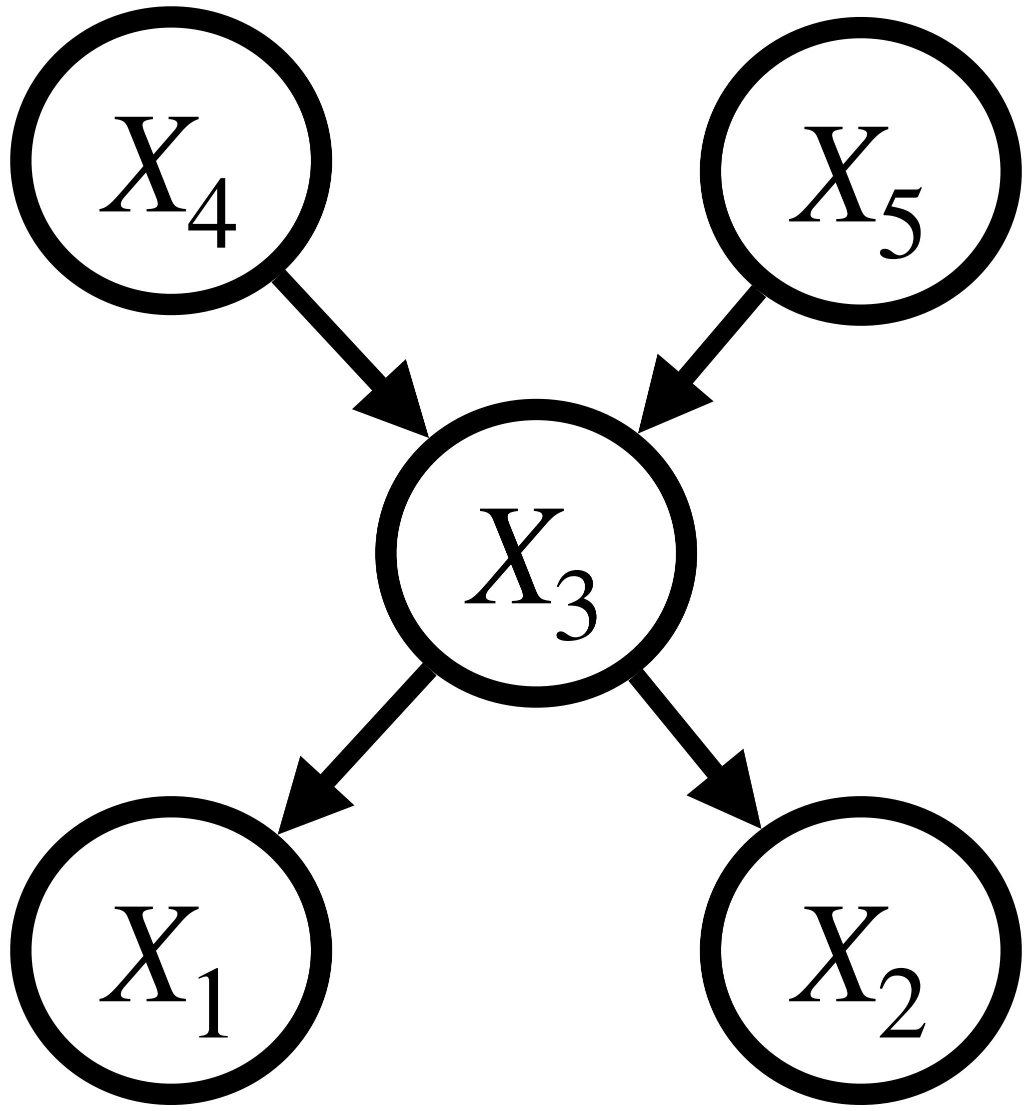
Credit: Petr Pošík

Part 5: Variable Elimination
(First Look Into Inference)

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:



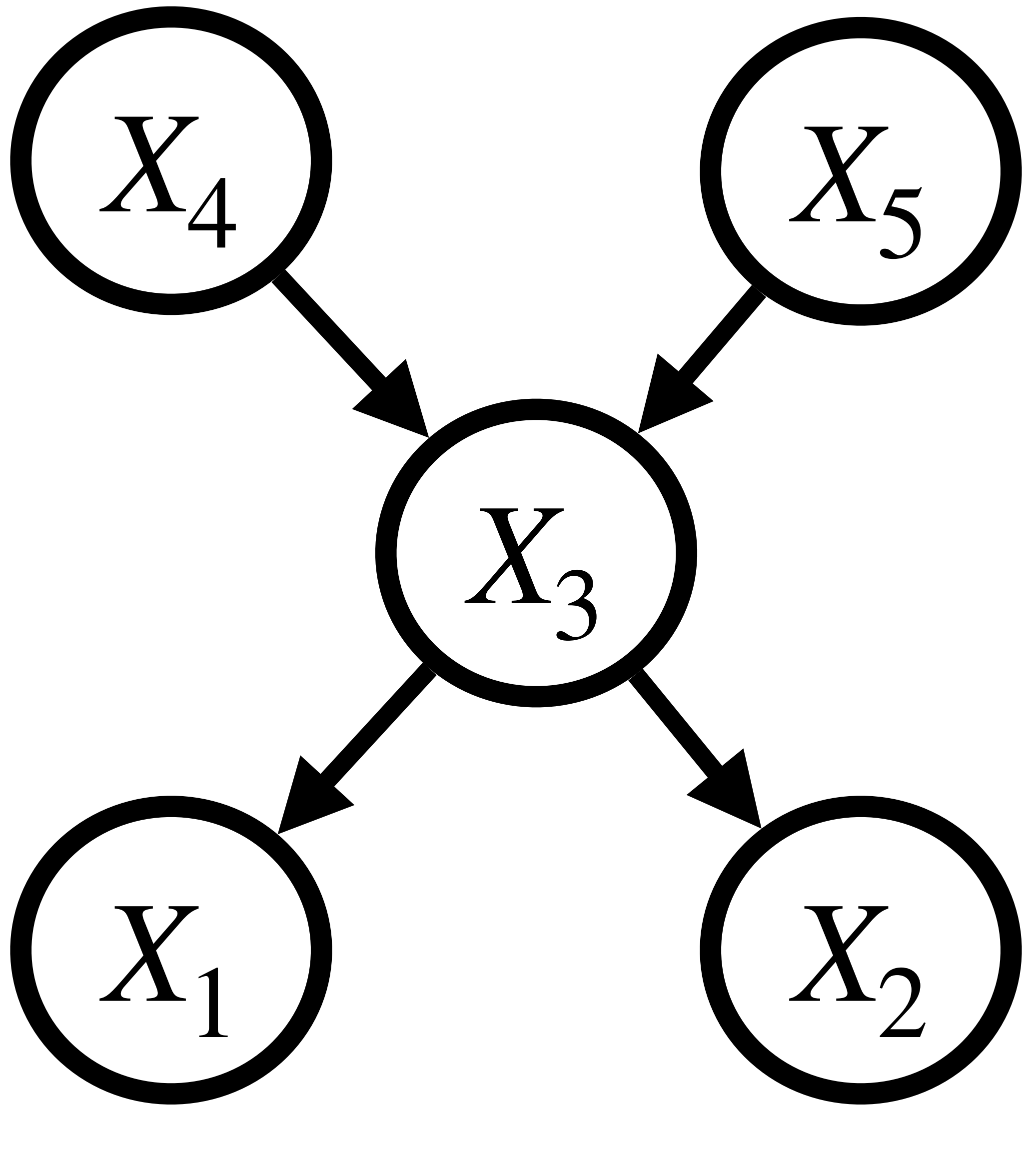
Naive Approach

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

Compute the following sum explicitly:

$$P_{X_1}(x_1) = \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} P_{X_1,\ldots,X_5}(x_1, x_2, x_3, x_4, x_5).$$

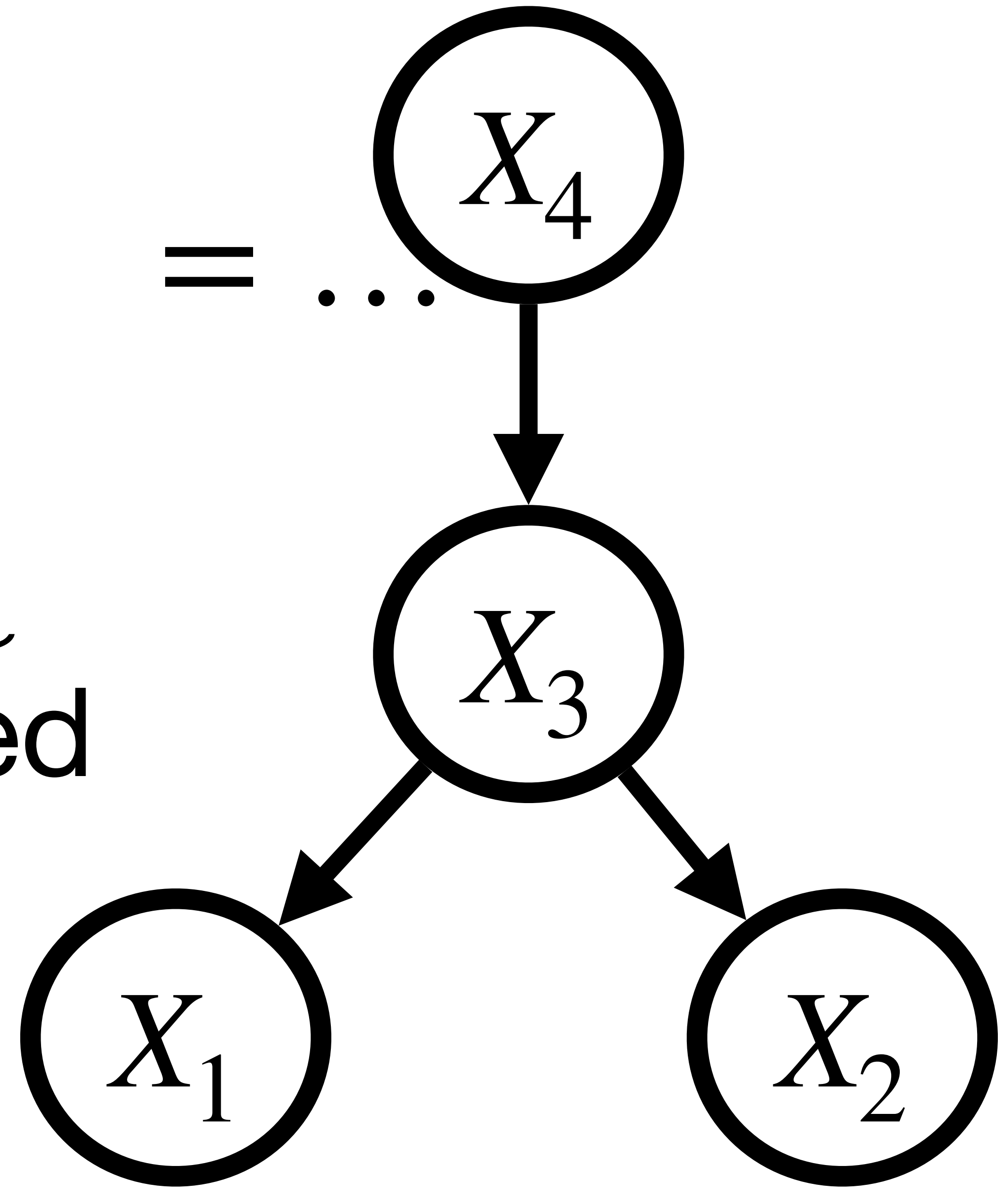
This will have exponential complexity in the number of random variables.



Variable Elimination: Basic Idea

$$P_{X_1}(x_1) = \sum_{x_2} \sum_{x_3} \sum_{x_4} P_{X_4}(x_4) P_{X_3|X_4}(x_3 \mid x_4) P_{X_1|X_3}(x_1 \mid x_3)$$

$$= \sum_{x_3} P_{X_1|X_3}(x_1 \mid x_3) \sum_{x_2} P_{X_2|X_3}(x_2 \mid x_3) \underbrace{\left(\sum_{x_4} P_{X_4}(x_4) P_{X_3|X_4}(x_3 \mid x_4) \right)}_{\text{function of } x_3\text{, it can be cached}} = \ldots$$



Next Lecture

We will finish variable elimination… And we will talk about inference in general.