Directed Acyclic Graphical Models, and Causal Models
9.S918: Statistical Inference in Brain and Cognitive Sciences
Spring 2024

Roger Levy

Massachusetts Institute of Technology

9 April 2024

Today's content

- Conditional Independence
- Bayes Nets (a.k.a. directed acyclic graphical models, DAGs)

(Conditional) Independence

Events A and B are said to be Conditionally Independent given information C if

$$P(A, B|C) = P(A|C)P(B|C)$$

Conditional independence of A and B given C is often expressed as

$$A \perp B|C$$

Directed graphical models

- A lot of the interesting joint probability distributions that arise in science and practical applications alike involve *conditional independencies* among the variables

Directed graphical models

- A lot of the interesting joint probability distributions that arise in science and practical applications alike involve *conditional independencies* among the variables
- So next is an introduction to a general framework for specifying conditional independencies among collections of random variables

Directed graphical models

- A lot of the interesting joint probability distributions that arise in science and practical applications alike involve *conditional independencies* among the variables
- So next is an introduction to a general framework for specifying conditional independencies among collections of random variables
- It won't allow us to express *all possible* independencies that may hold, but it goes a long way

Directed graphical models

- A lot of the interesting joint probability distributions that arise in science and practical applications alike involve *conditional independencies* among the variables
- So next is an introduction to a general framework for specifying conditional independencies among collections of random variables
- It won't allow us to express *all possible* independencies that may hold, but it goes a long way
- And I hope that you'll agree that the framework is intuitive too!

Directed graphical models

- A lot of the interesting joint probability distributions that arise in science and practical applications alike involve *conditional independencies* among the variables
- So next is an introduction to a general framework for specifying conditional independencies among collections of random variables
- It won't allow us to express *all possible* independencies that may hold, but it goes a long way
- And I hope that you'll agree that the framework is intuitive too!
- The intuitiveness is because a causal interpretation of the framework is natural—and, indeed, this is formalized in the causal treatment of Bayes nets

The coin factory

- Imagine a factory that produces three types of coins in equal volumes:

The coin factory

- Imagine a factory that produces three types of coins in equal volumes:
    - Fair coins;

The coin factory

- Imagine a factory that produces three types of coins in equal volumes:
  - Fair coins;
  - 2-headed coins;

The coin factory

- Imagine a factory that produces three types of coins in equal volumes:
  - Fair coins;
  - 2-headed coins;
  - 2-tailed coins.

The coin factory

- Imagine a factory that produces three types of coins in equal volumes:
  - Fair coins;
  - 2-headed coins;
  - 2-tailed coins.
- Generative process:

The coin factory

- Imagine a factory that produces three types of coins in equal volumes:
  - Fair coins;
  - 2-headed coins;
  - 2-tailed coins.
- Generative process:
  - The factory produces a coin of type X and sends it to you;

The coin factory

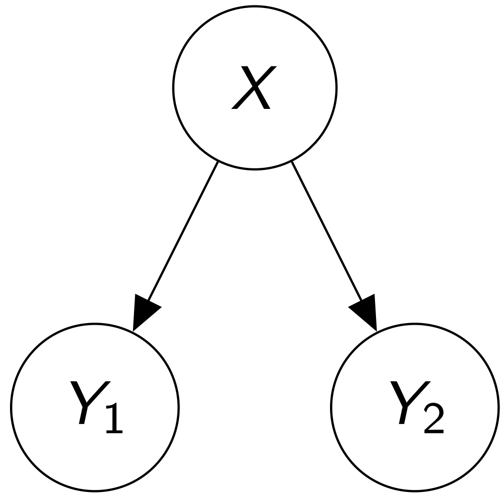
- Imagine a factory that produces three types of coins in equal volumes:
  - Fair coins;
  - 2-headed coins;
  - 2-tailed coins.
- Generative process:
  - The factory produces a coin of type X and sends it to you;
  - You receive the coin and flip it twice, with H(eads)/T(ails) outcomes Y_1 and Y_2

The coin factory

- Imagine a factory that produces three types of coins in equal volumes:
  - Fair coins;
  - 2-headed coins;
  - 2-tailed coins.
- Generative process:
  - The factory produces a coin of type X and sends it to you;
  - You receive the coin and flip it twice, with H(eads)/T(ails) outcomes Y_1 and Y_2
- Receiving a coin from the factory and flipping it twice is **sampling** (or **taking a sample**) from the joint distribution $P(X, Y_1, Y_2)$

This generative process is a Bayes Net

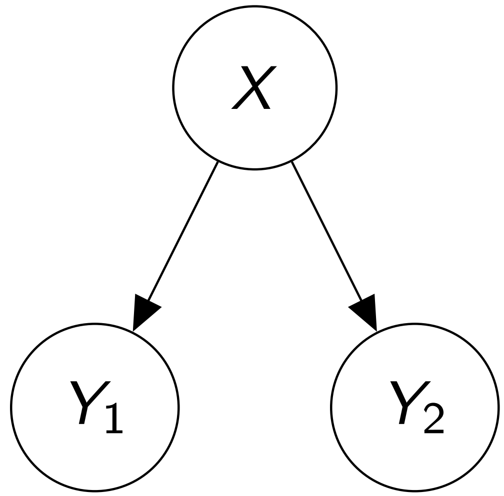
The directed acyclic graphical model (DAG), or Bayes net:



- Semantics of a Bayes net: the joint distribution can be expressed as the product of the conditional distributions of each variable **given only its parents**

This generative process is a Bayes Net

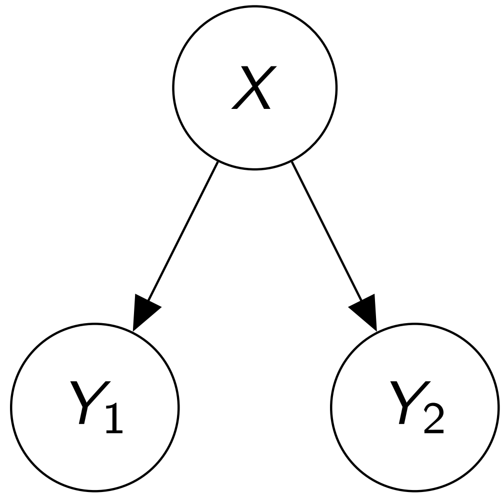
The directed acyclic graphical model (DAG), or Bayes net:



- Semantics of a Bayes net: the joint distribution can be expressed as the product of the conditional distributions of each variable **given only its parents**
- In this DAG, $P(X, Y_1, Y_2) = P(X)P(Y_1|X)P(Y_2|X)$

This generative process is a Bayes Net

The directed acyclic graphical model (DAG), or Bayes net:

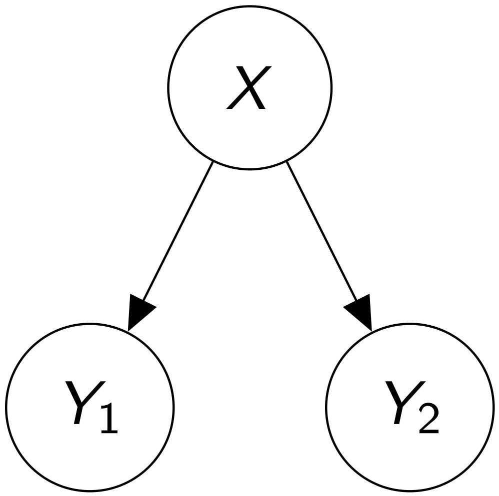


- Semantics of a Bayes net: the joint distribution can be expressed as the product of the conditional distributions of each variable **given only its parents**
- In this DAG, $P(X, Y_1, Y_2) = P(X)P(Y_1|X)P(Y_2|X)$

X	$P(X)$
Fair	$\frac{1}{3}$
2-H	$\frac{1}{3}$
2-T	$\frac{1}{3}$

This generative process is a Bayes Net

The directed acyclic graphical model (DAG), or Bayes net:



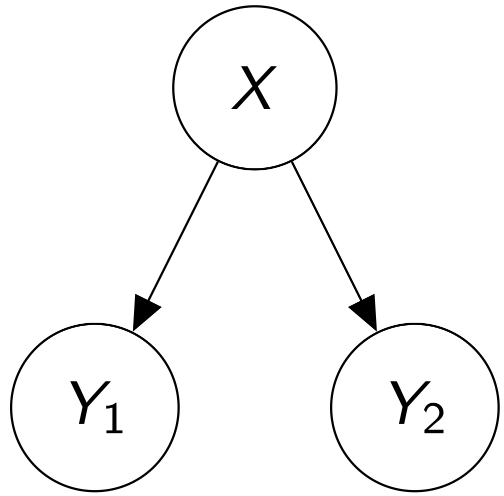
- Semantics of a Bayes net: the joint distribution can be expressed as the product of the conditional distributions of each variable **given only its parents**
- In this DAG, $P(X, Y_1, Y_2) = P(X)P(Y_1|X)P(Y_2|X)$

X	$P(X)$
Fair	$\frac{1}{3}$
2-H	$\frac{1}{3}$
2-T	$\frac{1}{3}$

X	$P(Y_1 = \text{H}\|X)$	$P(Y_1 = \text{T}\|X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

This generative process is a Bayes Net

The directed acyclic graphical model (DAG), or Bayes net:



- Semantics of a Bayes net: the joint distribution can be expressed as the product of the conditional distributions of each variable **given only its parents**
- In this DAG, $P(X, Y_1, Y_2) = P(X)P(Y_1|X)P(Y_2|X)$

X	$P(X)$
Fair	$\frac{1}{3}$
2-H	$\frac{1}{3}$
2-T	$\frac{1}{3}$

X	$P(Y_1 = \mathrm{H}|X)$	$P(Y_1 = \mathrm{T}|X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

X	$P(Y_2 = \mathrm{H}|X)$	$P(Y_2 = \mathrm{T}|X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

Conditional independence in Bayes nets

X	$P(X)$
Fair	$\frac{1}{3}$
2-H	$\frac{1}{3}$
2-T	$\frac{1}{3}$

X	$P(Y_1 = \mathrm{H} \mid X)$	$P(Y_1 = \mathrm{T} \mid X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

X	$P(Y_2 = \mathrm{H} \mid X)$	$P(Y_2 = \mathrm{T} \mid X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

Question:

- *Conditioned on not having any further information, are the two coin flips Y_1 and Y_2 in this generative process independent?*

Conditional independence in Bayes nets

X	$P(X)$
Fair	$\frac{1}{3}$
2-H	$\frac{1}{3}$
2-T	$\frac{1}{3}$

X	$P(Y_1 = \text{H}\|X)$	$P(Y_1 = \text{T}\|X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

X	$P(Y_2 = \text{H}\|X)$	$P(Y_2 = \text{T}\|X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

Question:

- *Conditioned on not having any further information, are the two coin flips Y_1 and Y_2 in this generative process independent?*
- "Independent" needs further interpretation! It might mean: is it the case that $Y_1 \perp Y_2 | \{\}$?

Conditional independence in Bayes nets

X	$P(X)$
Fair	$\frac{1}{3}$
2-H	$\frac{1}{3}$
2-T	$\frac{1}{3}$

X	$P(Y_1 = \mathrm{H} \vert X)$	$P(Y_1 = \mathrm{T} \vert X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

X	$P(Y_2 = \mathrm{H} \vert X)$	$P(Y_2 = \mathrm{T} \vert X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

Question:

- *Conditioned on not having any further information, are the two coin flips Y_1 and Y_2 in this generative process independent?*
- "Independent" needs further interpretation! It might mean: is it the case that $Y_1 \perp Y_2 | \{\}$?
- The answer to this question is **No!**

Conditional independence in Bayes nets

X	$P(X)$
Fair	$\frac{1}{3}$
2-H	$\frac{1}{3}$
2-T	$\frac{1}{3}$

X	$P(Y_1 = \mathrm{H}\|X)$	$P(Y_1 = \mathrm{T}\|X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

X	$P(Y_2 = \mathrm{H}\|X)$	$P(Y_2 = \mathrm{T}\|X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

Question:

- *Conditioned on not having any further information, are the two coin flips Y_1 and Y_2 in this generative process independent?*
- "Independent" needs further interpretation! It might mean: is it the case that $Y_1 \perp Y_2 | \{\}$?
- The answer to this question is **No!**
  - $P(Y_2 = H) = \frac{1}{2}$ (you can see this by symmetry)

Conditional independence in Bayes nets

X	$P(X)$
Fair	$\frac{1}{3}$
2-H	$\frac{1}{3}$
2-T	$\frac{1}{3}$

X	$P(Y_1 = \mathrm{H}\mid X)$	$P(Y_1 = \mathrm{T}\mid X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

X	$P(Y_2 = \mathrm{H}\mid X)$	$P(Y_2 = \mathrm{T}\mid X)$
Fair	$\frac{1}{2}$	$\frac{1}{2}$
2-H	1	0
2-T	0	1

Question:

- *Conditioned on not having any further information, are the two coin flips Y_1 and Y_2 in this generative process independent?*
- "Independent" needs further interpretation! It might mean: is it the case that $Y_1 \perp Y_2 | \{\}$?
- The answer to this question is **No!**
  - $P(Y_2 = H) = \frac{1}{2}$ (you can see this by symmetry)
  - But $P(Y_2 = H | Y_1 = H) = \overbrace{\frac{1}{3} \times \frac{1}{2}}^{\text{Coin was fair}} + \overbrace{\frac{2}{3} \times 1}^{\text{Coin was 2-H}} = \frac{5}{6}$

Formally assessing conditional independence in Bayes Nets

- The comprehensive criterion for assessing conditional independence is known as D-separation.

Formally assessing conditional independence in Bayes Nets

- The comprehensive criterion for assessing conditional independence is known as D-separation.
- A path between two disjoint node sets A and B is a sequence of edges connecting some node in A with some node in B

Formally assessing conditional independence in Bayes Nets

- The comprehensive criterion for assessing conditional independence is known as D-separation.
- A path between two disjoint node sets A and B is a sequence of edges connecting some node in A with some node in B
- Any node on a given path has converging arrows if two edges on the path connect to it and point to it.

Formally assessing conditional independence in Bayes Nets

- The comprehensive criterion for assessing conditional independence is known as D-separation.
- A path between two disjoint node sets A and B is a sequence of edges connecting some node in A with some node in B
- Any node on a given path has converging arrows if two edges on the path connect to it and point to it.
- A node on the path has non-converging arrows if two edges on the path connect to it, but at least one does not point to it.

Formally assessing conditional independence in Bayes Nets

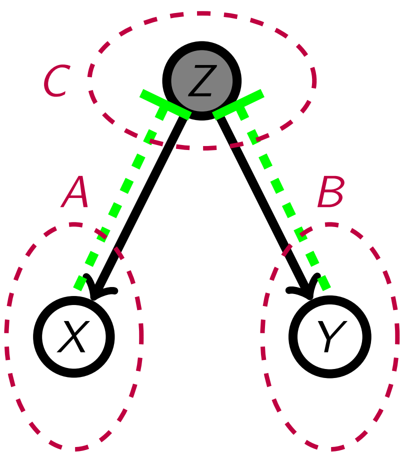
- The comprehensive criterion for assessing conditional independence is known as D-separation.
- A path between two disjoint node sets A and B is a sequence of edges connecting some node in A with some node in B
- Any node on a given path has converging arrows if two edges on the path connect to it and point to it.
- A node on the path has non-converging arrows if two edges on the path connect to it, but at least one does not point to it.
- A third disjoint node set C d-separates A and B if for every path between A and B, either:
  1. there is some node N on the path whose arrows do not converge and which *is* in C; or
  2. there is some node N on the path with converging arrows, and neither N nor any of its descendants is in C.

Major types of d-separation

A node set C d-separates A and B if for every path between A and B, either:

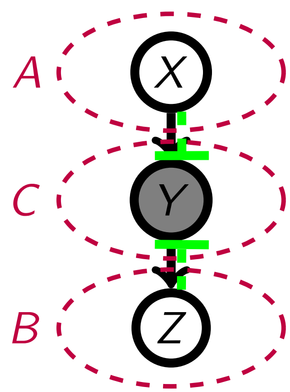
1. there is some node N on the path whose arrows do not converge and which *is* in C; or
2. there is some node N on the path with converging arrows, and neither N nor any of its descendants is in C.

Common-cause d-separation (from knowing Z)

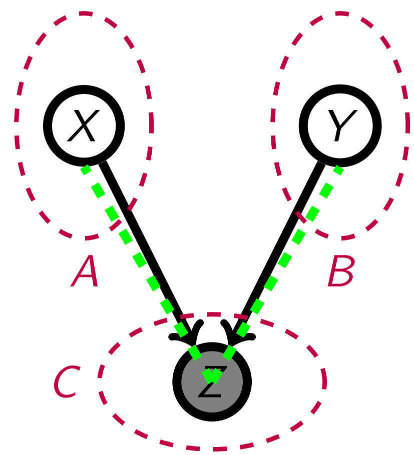


(Shaded node=in C)

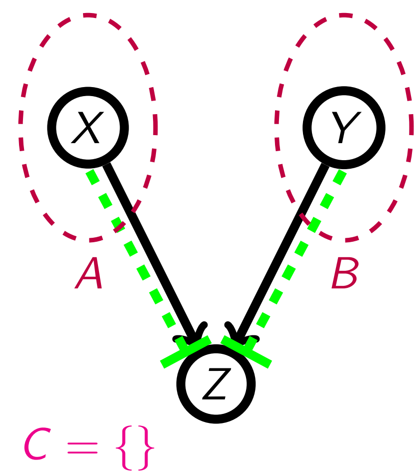
Intervening d-separation (from knowing Y)



Explaining away: knowing Z prevents d-separation



D-separation in the absence of knowledge of Z



D-separation and conditional independence

A node set C d-separates A and B if for every path between A and B, either:

1. there is some node N on the path whose arrows do not converge and which *is* in C; or
2. there is some node N on the path with converging arrows, and neither N nor any of its descendants is in C.

- If C d-separates A and B, then

$$A \perp B | C$$

D-separation and conditional independence

A node set C d-separates A and B if for every path between A and B, either:

1. there is some node N on the path whose arrows do not converge and which *is* in C; or
2. there is some node N on the path with converging arrows, and neither N nor any of its descendants is in C.

- If C d-separates A and B, then

$$A \perp B | C$$

- **Caution:** the converse is *not* the case: $A \perp B | C$ does not necessarily imply that the joint distribution on all the random variables in $A \cup B \cup C$ can be represented with a Bayes Net in which C d-separates A and B.

D-separation and conditional independence

A node set C d-separates A and B if for every path between A and B, either:

1. there is some node N on the path whose arrows do not converge and which *is* in C; or
2. there is some node N on the path with converging arrows, and neither N nor any of its descendants is in C.

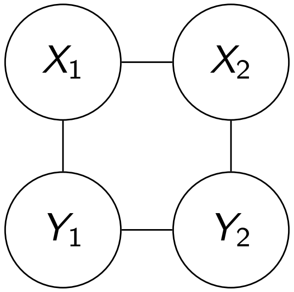
- If C d-separates A and B, then

$$A \perp B | C$$

- **Caution:** the converse is *not* the case: $A \perp B | C$ does not necessarily imply that the joint distribution on all the random variables in $A \cup B \cup C$ can be represented with a Bayes Net in which C d-separates A and B.
  - **Example:** let X_1, X_2, Y_1, Y_2 each be $0/1$ random variable, and let the joint distribution reflect the constraint that $Y_1 = (X_1 == X_2)$ and $Y_2 = \text{xor}(X_1, X_2)$. This gives us $Y_1 \perp Y_2 | \{X_1, X_2\}$, but you won't be able to write a Bayes net involving these four variables such that $\{X_1, X_2\}$ d-separates Y_1 and Y_2.

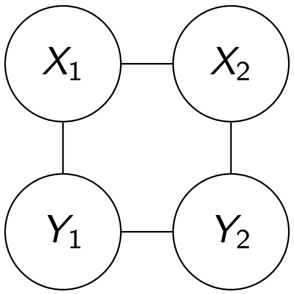
Conditional independencies not expressible in a Bayes net

- **Example:** let X_1, X_2, Y_1, Y_2 each be binary 0/1 random variables, in the following arrangement on an **undirected** graph:



Conditional independencies not expressible in a Bayes net

- **Example:** let X_1, X_2, Y_1, Y_2 each be binary 0/1 random variables, in the following arrangement on an **undirected** graph:



$$
\begin{aligned}
f_1(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_1 \neq X_2) \\
f_2(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_1 \neq Y_1) \\
f_3(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_2 \neq Y_2) \\
f_4(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(Y_1 \neq Y_2)
\end{aligned}
$$

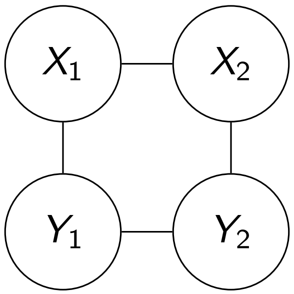
- Suppose the joint distribution is determined entirely by adjacent nodes "liking" to have the same value. Formally, for example:

$$P(X_1, X_2, Y_1, Y_2) \propto \prod_{i=1}^{4} \left(\frac{1}{2}\right)^{f_i(X_1, X_2, Y_1, Y_2)}$$

(Most probable outcomes, each with prob. 0.195: either all 0s, or all 1s)

Conditional independencies not expressible in a Bayes net

- **Example:** let X_1, X_2, Y_1, Y_2 each be binary 0/1 random variables, in the following arrangement on an **undirected** graph:



$$\begin{aligned}
f_1(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_1 \neq X_2) \\
f_2(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_1 \neq Y_1) \\
f_3(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_2 \neq Y_2) \\
f_4(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(Y_1 \neq Y_2)
\end{aligned}$$

- Suppose the joint distribution is determined entirely by adjacent nodes "liking" to have the same value. Formally, for example:

$$P(X_1, X_2, Y_1, Y_2) \propto \prod_{i=1}^{4} \left(\frac{1}{2}\right)^{f_i(X_1, X_2, Y_1, Y_2)}$$

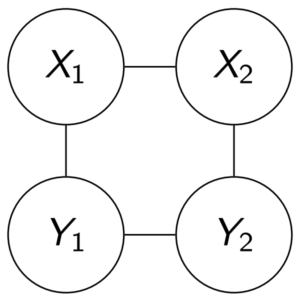
(Most probable outcomes, each with prob. 0.195: either all 0s, or all 1s)

- In this model, both the following conditional independencies hold:

$$X_1 \perp Y_2 | \{X_2, Y_1\} \qquad\qquad X_2 \perp Y_1 | \{X_1, Y_2\}$$

Conditional independencies not expressible in a Bayes net

- **Example:** let X_1, X_2, Y_1, Y_2 each be binary 0/1 random variables, in the following arrangement on an **undirected** graph:



$$
\begin{aligned}
f_1(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_1 \neq X_2) \\
f_2(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_1 \neq Y_1) \\
f_3(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_2 \neq Y_2) \\
f_4(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(Y_1 \neq Y_2)
\end{aligned}
$$

- Suppose the joint distribution is determined entirely by adjacent nodes "liking" to have the same value. Formally, for example:

$$P(X_1, X_2, Y_1, Y_2) \propto \prod_{i=1}^{4} \left(\frac{1}{2}\right)^{f_i(X_1, X_2, Y_1, Y_2)}$$

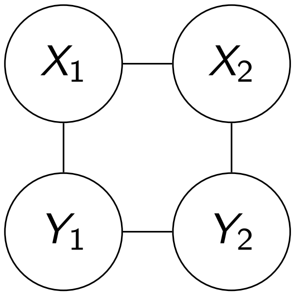
(Most probable outcomes, each with prob. 0.195: either all 0s, or all 1s)

- In this model, both the following conditional independencies hold:

$$X_1 \perp Y_2 | \{X_2, Y_1\} \qquad X_2 \perp Y_1 | \{X_1, Y_2\}$$

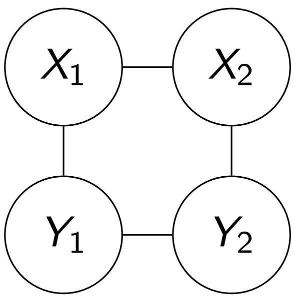
- But this set of conditional independencies cannot be expressed in a Bayes Net.

Conditional independencies not expressable in a Bayes net



$$\begin{aligned}
f_1(X_1, X_2, Y_1, Y_2) &= \mathbf{I}\,(X_1 \neq X_2) \\
f_2(X_1, X_2, Y_1, Y_2) &= \mathbf{I}\,(X_1 \neq Y_1) \\
f_3(X_1, X_2, Y_1, Y_2) &= \mathbf{I}\,(X_2 \neq Y_2) \\
f_4(X_1, X_2, Y_1, Y_2) &= \mathbf{I}\,(Y_1 \neq Y_2)
\end{aligned}$$

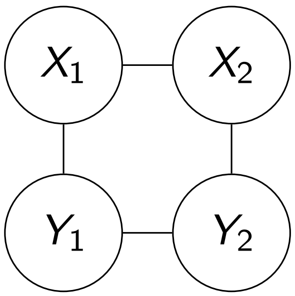
Conditional independencies not expressable in a Bayes net



$$
\begin{aligned}
f_1(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_1 \neq X_2)\\
f_2(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_1 \neq Y_1)\\
f_3(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_2 \neq Y_2)\\
f_4(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(Y_1 \neq Y_2)
\end{aligned}
$$

- This example is an instance of an Ising model, the prototypical case of a Markov random field, a model class that can be represented as undirected graphs

Conditional independencies not expressable in a Bayes net



$$
\begin{aligned}
f_1(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_1 \neq X_2) \\
f_2(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_1 \neq Y_1) \\
f_3(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(X_2 \neq Y_2) \\
f_4(X_1, X_2, Y_1, Y_2) &= \mathbf{I}(Y_1 \neq Y_2)
\end{aligned}
$$

- This example is an instance of an Ising model, the prototypical case of a Markov random field, a model class that can be represented as undirected graphs
- We won't look at these further, but you can read about them in books and papers about graphical models (e.g., (Bishop, 2006, Section 8.3)