Bayesian Networks

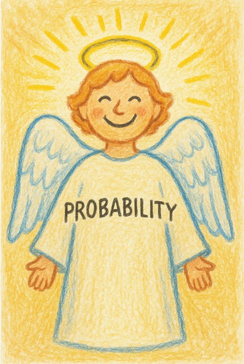
Probabilistic Decision Making — Lecture 6

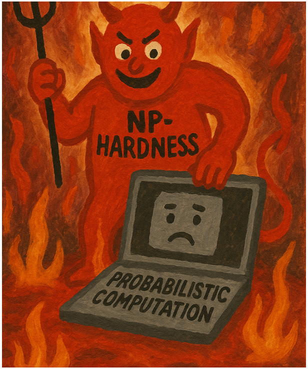
5th November 2025

Robert Peharz
Institute of Machine Learning and Neural Computation
Graz University of Technology

Everything is wonderful. . .

Probability is a "faithful guardian of common sense" and "common sense reduced to computation", but . . .





. . . it is also a **computational nightmare.**

Single Bernoulli Variable

Example

 A coin can be modeled using a Bernoulli RV X with success probability θ. The PMF is

x	$p(x)$
0	$1-\theta$
1	θ

1 parameter

representation, **learning**, and **inference** are trivial

Let S and C be RVs indicating whether a person is *smoker* (S) and develops *lung cancer* (C), respectively. The PMF is

s	c	$p(s, c)$
0	0	θ_1
0	1	θ_2
1	0	θ_3
1	1	θ_4

3 free parameters (as $\theta_4 = 1 - \sum_{k=1}^{3} \theta_k$)
still trivial

10 Bernoulli Variables

Example

Consider a system of 10 physical particles with 'spin up' or 'spin down,' modeled with binary RVs $X_1, X_2, \ldots, X_{10}$.

$\boldsymbol{x}$	$p(\boldsymbol{x})$
$(0, 0, \ldots, 0, 0)$	θ_1
$(0, 0, \ldots, 0, 1)$	θ_2
$(0, 0, \ldots, 1, 0)$	θ_3
$\ldots$	$\ldots$
$(1, 1, \ldots, 1, 1)$	$\theta_{\mathbf{1024}}$

1023 free parameters

easy

20 Bernoulli Variables

Example

Consider a system of 20 physical particles with 'spin up' or 'spin down,' modeled with binary RVs $X_1, X_2, \ldots, X_{20}$.

$\boldsymbol{x}$	$p(\boldsymbol{x})$
$(0, 0, \ldots, 0, 0)$	π_1
$(0, 0, \ldots, 0, 1)$	π_2
$(0, 0, \ldots, 1, 0)$	π_3
$\vdots$	$\vdots$
$(1, 1, \ldots, 1, 1)$	$\pi_{\mathbf{1048576}}$

1 million parameters
not really a problem

30 Bernoulli Variables

Example

Consider a system of 30 physical particles with 'spin up' or 'spin down,' modeled with binary RVs $X_1, X_2, \ldots, X_{30}$.

$\boldsymbol{x}$	$p(\boldsymbol{x})$
$(0, 0, \ldots, 0, 0)$	π_1
$(0, 0, \ldots, 0, 1)$	π_2
$(0, 0, \ldots, 1, 0)$	π_3
$\ldots$	$\ldots$
$(1, 1, \ldots, 1, 1)$	$\pi_{\mathbf{1073741824}}$

1 billion parameters
getting delicate

40 Bernoulli Variables

Example

Consider a system of 40 physical particles with 'spin up' or 'spin down,' modeled with binary RVs $X_1, X_2, \ldots, X_{40}$.

$\boldsymbol{x}$	$p(\boldsymbol{x})$
$(0, 0, \ldots, 0, 0)$	π_1
$(0, 0, \ldots, 0, 1)$	π_2
$(0, 0, \ldots, 1, 0)$	π_3
$\ldots$	$\ldots$
$(1, 1, \ldots, 1, 1)$	$\pi_{\mathbf{1099511627776}}$

1 trillion parameters

game over

(except for Google and OpenAI)

Curse of Dimensionality

High dimensional distributions "are hard."

- hard to **represent**
- hard to **learn**
- hard to perform **inference** in
- holds true for continuous and discrete RVs

Graphical models address these challenges. We will focus on **directed graphical models**, that utilize two ingredients:

- **chain rule of probability**
- **conditional independence**

Naturally represented with graphs over random variables, hence the name "graphical models".

Let $p(x_1, x_2, \ldots, x_D)$ be an **arbitrary** joint distribution. Due to the product rule we can write:

$$p(x_1, x_2, x_3 \ldots, x_D) = p(x_D \mid x_1, \ldots, x_{D-1})\, p(x_1, \ldots, x_{D-1})$$

Applying the chain rule again to the right-most factor

$$\begin{aligned} p(x_1, x_2, x_3 \ldots, x_D) = {} & p(x_D \mid x_1, \ldots, x_{D-1}) \\ & p(x_{D-1} \mid x_1, \ldots, x_{D-2}) \\ & p(x_1, \ldots, x_{D-2}) \end{aligned}$$

Thus, by induction

$$p(x_1, x_2, \ldots, x_D) = \prod_{k=1}^{D} p(x_k \mid x_1, \ldots, x_{k-1}).$$

For $k = 1$, we define $p(x_1 \mid \ldots) \coloneqq p(x_1)$.

The chain rule works for any joint.

Moreover, we can apply it to any variable ordering.

Chain Rule of Probability

Let $p(x_1, x_2, \ldots, x_D)$ be an **arbitrary** joint distribution, and let $i_1, i_2, \ldots, i_D$ be an **arbitrary** permutation of integers $1, 2, \ldots, D$. The joint can be written as

$$p(x_1, x_2, \ldots, x_D) = \prod_{k=1}^{D} p(x_{i_k} \mid x_{i_1}, \ldots, x_{i_{k-1}}).$$

For $k = 1$, $p(x_{i_k} \mid x_{i_1}, \ldots, x_{i_{k-1}}) := p(x_{i_1})$.

Chain Rule of Probability

Example

Let $p(x_1, x_2, x_3, x_4)$ be **any** joint distribution. We can write it as

$$\begin{aligned} p(x_1, x_2, x_3, x_4) &= p(x_4 \mid x_1, x_2, x_3)\, p(x_1, x_2, x_3) \\ &= p(x_4 \mid x_1, x_2, x_3)\, p(x_3 \mid x_1, x_2)\, p(x_1, x_2) \\ &= p(x_4 \mid x_1, x_2, x_3)\, p(x_3 \mid x_1, x_2)\, p(x_2 \mid x_1)\, p(x_1) \end{aligned}$$

Or, we might also use ordering X_4, X_2, X_3, X_1 and write

$$\begin{aligned} p(x_1, x_2, x_3, x_4) &= p(x_1 \mid x_2, x_3, x_4)\, p(x_2, x_3, x_4) \\ &= p(x_1 \mid x_2, x_3, x_4)\, p(x_3 \mid x_2, x_4)\, p(x_2, x_4) \\ &= p(x_1 \mid x_2, x_3, x_4)\, p(x_3 \mid x_2, x_4)\, p(x_2 \mid x_4)\, p(x_4) \end{aligned}$$

Note that all conditional distributions are over **single RVs**.

Is the Chain Rule Enough?

Course of dimensionality: joint distributions grow exponentially in the number of RVs. **Does the chain rule already help? No...**

For example, assume 4 binary RVs and consider the chain rule:

$$\overbrace{p(x_1, x_2, x_3, x_4)}^{\text{15 params}} = \overbrace{p(x_4 \mid x_1, x_2, x_3)}^{\text{8 params}} \overbrace{p(x_3 \mid x_1, x_2)}^{\text{4 params}} \overbrace{p(x_2 \mid x_1)}^{\text{2 params}} \overbrace{p(x_1)}^{\text{1 param}}$$

The exhaustive joint distribution (left) and the factorized distribution (right) require the same number of parameters.

A distribution over D binary RVs requires $2^D - 1$ free params. Conditioning on C binary RVs means that we really have 2^C distribution over the "stuff on the left of the conditioning bar".

Independence

$\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y}$ if the **equivalent** conditions hold for all $\boldsymbol{x}, \boldsymbol{y}$:

- $p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x})\, p(\boldsymbol{y})$ (joint factorizes)
- $p(\boldsymbol{x} \mid \boldsymbol{y}) = p(\boldsymbol{x})$ (marginal equals conditional)
- $p(\boldsymbol{y} \mid \boldsymbol{x}) = p(\boldsymbol{y})$ (marginal equals conditional)

Conditional Independence

$\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{Z}$ if for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$:

- $p(\boldsymbol{x}, \boldsymbol{y} \mid \boldsymbol{z}) = p(\boldsymbol{x} \mid \boldsymbol{z})\, p(\boldsymbol{y} \mid \boldsymbol{z})$
- $p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{z}) = p(\boldsymbol{x} \mid \boldsymbol{z})$
- $p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{z}) = p(\boldsymbol{y} \mid \boldsymbol{z})$

Factorize the joint of 4 binary RVs:

$$p(x_1, x_2, x_3, x_4) = p(x_4 \mid x_1, x_2, x_3)\, p(x_3 \mid x_1, x_2)\, p(x_2 \mid x_1)\, p(x_1)$$

Let's make some **conditional independence (CI) assumptions:**

- $X_4 \perp\!\!\!\perp X_2, X_3 \mid X_1$
- $X_3 \perp\!\!\!\perp X_1 \mid X_2$

This allows us to **delete** some conditioned RVs:

$$\overbrace{p(x_1, x_2, x_3, x_4)}^{15 \text{ params}} = \overbrace{p(x_4 \mid x_1, \cancel{x_2, x_3})}^{2 \text{ params}}\, \overbrace{p(x_3 \mid \cancel{x_1}, x_2)}^{2 \text{ params}}\, \overbrace{p(x_2 \mid x_1)}^{2 \text{ params}}\, \overbrace{p(x_1)}^{1 \text{ param}}$$

7 **parameters instead of** 15**!**

Chain Rule + Conditional Independence

- CI is very common in real data
- savings in representation size can be dramatic in high dimensions
- CI also helps in **learning** and **inference**
- this "trick" works with **any** variable ordering

Let's assume an ordering X_3, X_1, X_4, X_2. The chain rule following this ordering yields:

$$p(x_1, x_2, x_3, x_4) = p(x_2 \mid x_1, x_3, x_4)\, p(x_4 \mid x_1, x_3)\, p(x_1 \mid x_3)\, p(x_3)$$

Conditional independence (CI) assumptions:

- $X_1 \perp\!\!\!\perp X_3$
- $X_2 \perp\!\!\!\perp X_4 \mid X_1, X_3$

$$\overbrace{p(x_1, x_2, x_3, x_4)}^{\text{15 params}} = \overbrace{p(x_2 \mid x_1, x_3, \cancel{x_4})}^{\text{4 params}}\, \overbrace{p(x_4 \mid x_1, x_3)}^{\text{4 params}}\, \overbrace{p(x_1 \cancel{\mid x_3})}^{\text{1 param}}\, \overbrace{p(x_3)}^{\text{1 param}}$$

10 **parameters instead of** 15**.**

Graphical Language

In both examples, we

- assumed an **ordering** of the RVs
- wrote the joint using the chain rule
- introduced CI assumptions which "deleted" conditioning RVs

This can be nicely represented with **directed graphs**. Specifically, we construct a directed graph

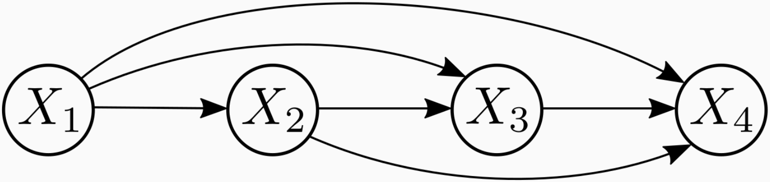
- whose nodes are the RVs
- introduce a directed edge from X_i to X_j, if X_i is a "conditioner" for X_j

Order: X_1, X_2, X_3, X_4

Order: X_1, X_2, X_3, X_4

Chain Rule:

$$p(x_1, x_2, x_3, x_4) = p(x_4 \mid x_1, x_2, x_3)\, p(x_3 \mid x_1, x_2)\, p(x_2 \mid x_1)\, p(x_1)$$



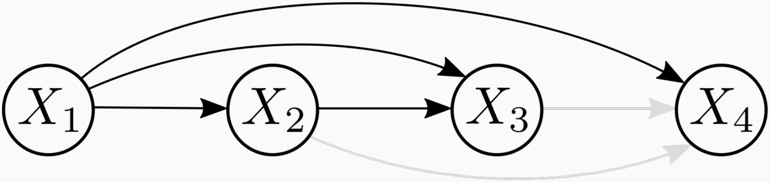
Order: X_1, X_2, X_3, X_4

Chain Rule:

$$p(x_1, x_2, x_3, x_4) = p(x_4 \mid x_1, \cancel{x_2, x_3})\, p(x_3 \mid x_1, x_2)\, p(x_2 \mid x_1)\, p(x_1)$$

Conditional independence (CI) assumptions:

- $X_4 \perp\!\!\!\perp X_2, X_3 \mid X_1$



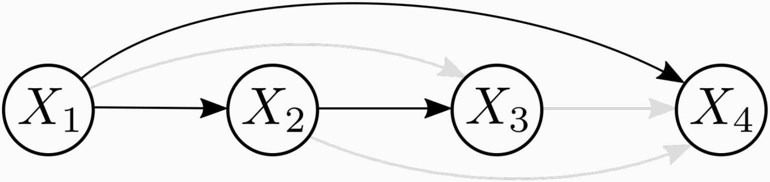
Order: X_1, X_2, X_3, X_4

Chain Rule:

$$p(x_1, x_2, x_3, x_4) = p(x_4 \mid x_1, \cancel{x_2, x_3})\, p(x_3 \mid \cancel{x_1}, x_2)\, p(x_2 \mid x_1)\, p(x_1)$$

Conditional independence (CI) assumptions:

- $X_4 \perp\!\!\!\perp X_2, X_3 \mid X_1$
- $X_3 \perp\!\!\!\perp X_1 \mid X_2$



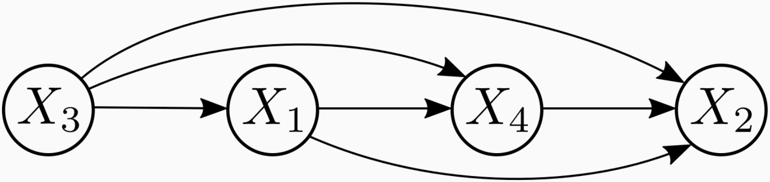
Order: X_3, X_1, X_4, X_2

Using Conditional Independence II – revisited Example

Order: X_3, X_1, X_4, X_2

Chain Rule:

$$p(x_1, x_2, x_3, x_4) = p(x_2 \mid x_1, x_3, x_4)\, p(x_4 \mid x_1, x_3)\, p(x_1 \mid x_3)\, p(x_3)$$



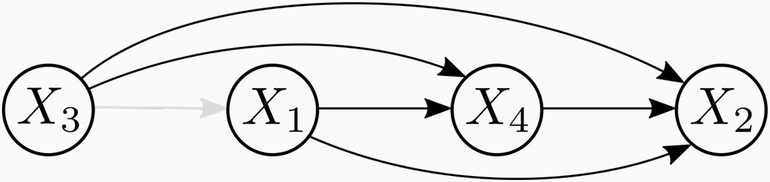
Order: X_3, X_1, X_4, X_2

Chain Rule:

$$p(x_1, x_2, x_3, x_4) = p(x_2 \mid x_1, x_3, x_4)\, p(x_4 \mid x_1, x_3)\, p(x_1 \mid \cancel{x_3})\, p(x_3)$$

Conditional independence (CI) assumptions:

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



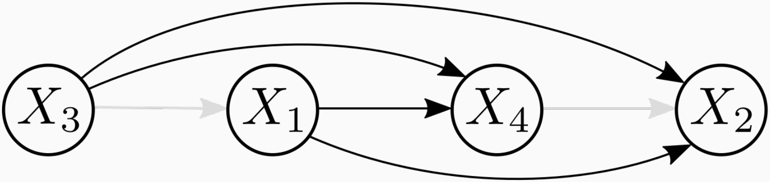
Order: X_3, X_1, X_4, X_2

Chain Rule:

$$p(x_1, x_2, x_3, x_4) = p(x_2 \mid x_1, x_3, \cancel{x_4})\, p(x_4 \mid x_1, x_3)\, p(x_1 \mid \cancel{x_3})\, p(x_3)$$

Conditional independence (CI) assumptions:

- $X_1 \perp\!\!\!\perp X_3$
- $X_2 \perp\!\!\!\perp X_4 \mid X_1, X_3$



- note that the directed graphs in the examples were **acyclic**, i.e. there was no directed path from any X_i to itself—such graphs are called **directed acyclic graphs (DAGs)**
- this is no coincidence, as the graphs were following the **order** assumed in the chain rule: edges can only be drawn from lower-order RVs to higher-order RVs
- in the context of directed graphs, this is called a **topological order** of the nodes (RVs).

Equivalence Topological Ordering and DAG

A directed graph is **acyclic** (i.e. a DAG) if and only if it has at least one topological ordering.

DAGs Represent Chain Rule + CI

DAGs over RVs can be used to **analyze**, **describe** and **represent** joint distributions, as they represent

- an RV ordering of the chain rule
- CIs corresponding to missing edges
- factorize the joint according to

$$p(\boldsymbol{x}) = \prod_{i=1}^{D} p(x_i \mid \mathbf{pa}_i)$$

  where $\mathbf{pa}_i$ are the states of X_i's **parents** in the graph, i.e. all nodes that have an edge directed towards X_i
- this can be used as a language to compactly represent high-dimensional joint distribution—**Bayesian networks**, also called **directed graphical models**

A **Bayesian network** (BN) over RVs $\boldsymbol{X}$ is a pair $(\mathcal{G}, \mathcal{P})$, where

- $\mathcal{G}$ is a DAG which has RVs $\boldsymbol{X}$ as nodes
- $\mathcal{P}$ is a collection of distributions $p(x \mid \mathbf{pa}_X)$, one for each $X \in \boldsymbol{X}$, and conditional on X's parents in $\mathcal{G}$; they are called **CPDs** (conditional probability distributions)

The BN represents the joint distribution

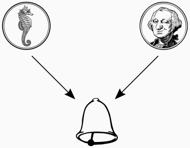
$$p(\boldsymbol{x}) = \prod_{i=1}^{D} p(x_i \mid \mathbf{pa}_i)$$

- completely specified by a graph $\mathcal{G}$ and a set of CPDs $\mathcal{P}$
- nicely represented on a computer, if the conditionals are small, i.e. if the graph is sparse enough
- many real-world problems **do** translate into sparse graphs.
- often, the graph can be derived from **causal knowledge**
- we can **learn** the joint from data, by optimizing over $\mathcal{G}$ and $\mathcal{P}$

Coins and Bells cont'd

We already saw an example of a BN earlier:

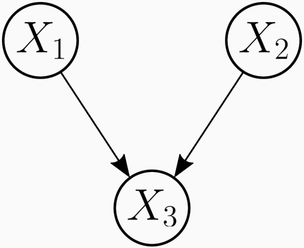
- Two fair coins are tossed, modeled with Bernoulli X_1 and X_2.
- If both show heads, a Bell (Bernoulli X_3) rings.
- If exactly one shows heads, the Bell rings with 50% probability.
- If both show tails, the Bell rings with 1% probability.



(x_1, x_2, x_3)	$p(x_1, x_2, x_3)$
(0, 0, 0)	0.2475
(0, 0, 1)	0.0025
(0, 1, 0)	0.125
(0, 1, 1)	0.125
(1, 0, 0)	0.125
(1, 0, 1)	0.125
(1, 1, 0)	0
(1, 1, 1)	0.25

Two Coins and a Bell

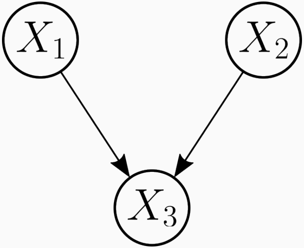
Example



x_1	$p(x_1)$
0	0.5
1	0.5

x_2	$p(x_2)$
0	0.5
1	0.5

x_3	$p(x_3 \mid x_1, x_2)$	x_1	x_2
0	0.99	0	0
1	0.01		
0	0.5	0	1
1	0.5		
0	0.5	1	0
1	0.5		
0	0	1	1
1	1		



x_1	$p(x_1)$
0	0.5
1	0.5

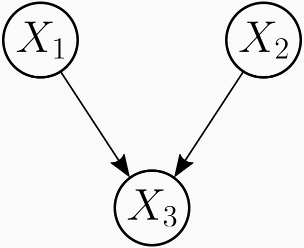
x_2	$p(x_2)$
0	0.5
1	0.5

x_3	$p(x_3 \mid x_1, x_2)$	x_1	x_2
0	0.99	0	0
1	0.01		
0	0.5	0	1
1	0.5		
0	0.5	1	0
1	0.5		
0	0	1	1
1	1		

x_1	x_2	b	$p(x_1, x_2, x_3)$
0	0	0	
0	0	1	
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	

Two Coins and a Bell

Example



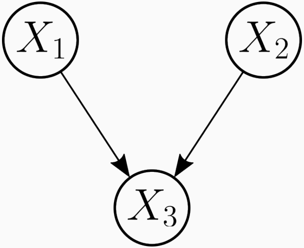
x_1	$p(x_1)$
0	0.5
1	0.5

x_2	$p(x_2)$
0	0.5
1	0.5

x_3	$p(x_3 \mid x_1, x_2)$	x_1	x_2
0	0.99	0	0
1	0.01		
0	0.5	0	1
1	0.5		
0	0.5	1	0
1	0.5		
0	0	1	1
1	1		

x_1	x_2	b	$p(x_1, x_2, x_3)$
0	0	0	$p(x_1 = 0) \times p(x_2 = 0) \times p(x_3 = 0 \mid x_1 = 0, x_2 = 0)$
0	0	1	
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	

Two Coins and a Bell



x_1	$p(x_1)$
0	0.5
1	0.5

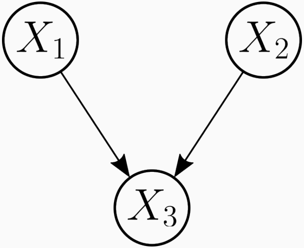
x_2	$p(x_2)$
0	0.5
1	0.5

x_3	$p(x_3 \mid x_1, x_2)$	x_1	x_2
0	0.99	0	0
1	0.01		
0	0.5	0	1
1	0.5		
0	0.5	1	0
1	0.5		
0	0	1	1
1	1		

x_1	x_2	b	$p(x_1, x_2, x_3)$
0	0	0	$0.5 \times 0.5 \times 0.99$
0	0	1	
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	

Two Coins and a Bell

Example



x_1	$p(x_1)$
0	0.5
1	0.5

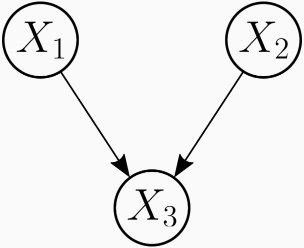
x_2	$p(x_2)$
0	0.5
1	0.5

x_3	$p(x_3 \mid x_1, x_2)$	x_1	x_2
0	0.99	0	0
1	0.01		
0	0.5	0	1
1	0.5		
0	0.5	1	0
1	0.5		
0	0	1	1
1	1		

x_1	x_2	b	$p(x_1, x_2, x_3)$
0	0	0	0.2475
0	0	1	
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	

Two Coins and a Bell

Example



x_1	$p(x_1)$
0	0.5
1	0.5

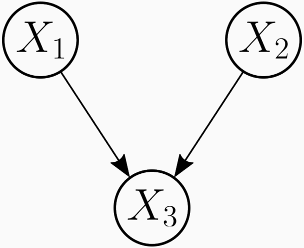
x_2	$p(x_2)$
0	0.5
1	0.5

x_3	$p(x_3 \mid x_1, x_2)$	x_1	x_2
0	0.99	0	0
1	0.01		
0	0.5	0	1
1	0.5		
0	0.5	1	0
1	0.5		
0	0	1	1
1	1		

x_1	x_2	b	$p(x_1, x_2, x_3)$
0	0	0	0.2475
0	0	1	$p(x_1 = 0) \times p(x_2 = 0) \times p(x_3 = 1 \mid x_1 = 0, x_2 = 0)$
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	

Two Coins and a Bell

Example



x_1	$p(x_1)$
0	0.5
1	0.5

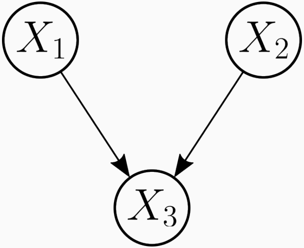
x_2	$p(x_2)$
0	0.5
1	0.5

x_3	$p(x_3 \mid x_1, x_2)$	x_1	x_2
0	0.99	0	0
1	0.01		
0	0.5	0	1
1	0.5		
0	0.5	1	0
1	0.5		
0	0	1	1
1	1		

x_1	x_2	b	$p(x_1, x_2, x_3)$
0	0	0	0.2475
0	0	1	$0.5 \times 0.5 \times 0.01$
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	

Two Coins and a Bell

Example



x_1	$p(x_1)$
0	0.5
1	0.5

x_2	$p(x_2)$
0	0.5
1	0.5

x_3	$p(x_3 \mid x_1, x_2)$	x_1	x_2
0	0.99	0	0
1	0.01		
0	0.5	0	1
1	0.5		
0	0.5	1	0
1	0.5		
0	0	1	1
1	1		

x_1	x_2	b	$p(x_1, x_2, x_3)$
0	0	0	0.2475
0	0	1	0.0025
0	1	0	0.125
0	1	1	0.125
1	0	0	0.125
1	0	1	0.125
1	1	0	0
1	1	1	0.25

You have a smoke detector in your apartment, where

- each day, there is a 0.01 probability of fire (F)
- if there is fire, there is with probability 0.995 smoke (S); Even if there is no fire, you might have smoke in the apartment with probability 0.02
- the smoke detector detects smoke with probability 0.98 (D); it has a false alarm probability of 0.015

Fire Detector cont'd

f	$p(f)$
0	0.99
1	0.01

s	$p(s \mid f)$	f
0	0.98	0
1	0.02	0
0	0.005	1
1	0.995	1

d	$p(d \mid s)$	s
0	0.985	0
1	0.015	0
0	0.02	1
1	0.98	1

f	s	d	$p(f, s, d)$
0	0	0	
0	0	1	
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	

Fire Detector cont'd

Example

f	$p(f)$
0	0.99
1	0.01

s	$p(s \mid f)$	f
0	0.98	0
1	0.02	0
0	0.005	1
1	0.995	1

d	$p(d \mid s)$	s
0	0.985	0
1	0.015	0
0	0.02	1
1	0.98	1

f	s	d	$p(f, s, d)$
0	0	0	$p(f=0) \times p(s=0 \mid f=0) \times p(d=0 \mid s=0)$
0	0	1	$p(f=0) \times p(s=0 \mid f=0) \times p(d=1 \mid s=0)$
0	1	0	$p(f=0) \times p(s=1 \mid f=0) \times p(d=0 \mid s=1)$
0	1	1	$p(f=0) \times p(s=1 \mid f=0) \times p(d=1 \mid s=1)$
1	0	0	$p(f=1) \times p(s=0 \mid f=1) \times p(d=0 \mid s=0)$
1	0	1	$p(f=1) \times p(s=0 \mid f=1) \times p(d=1 \mid s=0)$
1	1	0	$p(f=1) \times p(s=1 \mid f=1) \times p(d=0 \mid s=1)$
1	1	1	$p(f=1) \times p(s=1 \mid f=1) \times p(d=1 \mid s=1)$

f	$p(f)$
0	0.99
1	0.01

s	$p(s \mid f)$	f
0	0.98	0
1	0.02	
0	0.005	1
1	0.995	

d	$p(d \mid s)$	s
0	0.985	0
1	0.015	
0	0.02	1
1	0.98	

f	s	d	$p(f, s, d)$
0	0	0	$0.99 \times 0.98 \times 0.985$
0	0	1	$0.99 \times 0.98 \times 0.015$
0	1	0	$0.99 \times 0.02 \times 0.02$
0	1	1	$0.99 \times 0.02 \times 0.98$
1	0	0	$0.01 \times 0.005 \times 0.985$
1	0	1	$0.01 \times 0.005 \times 0.015$
1	1	0	$0.01 \times 0.995 \times 0.02$
1	1	1	$0.01 \times 0.995 \times 0.98$

Fire Detector cont'd

Example

f	$p(f)$
0	0.99
1	0.01

s	$p(s \mid f)$	f
0	0.98	0
1	0.02	
0	0.005	1
1	0.995	

d	$p(d \mid s)$	s
0	0.985	0
1	0.015	
0	0.02	1
1	0.98	

f	s	d	$p(f, s, d)$
0	0	0	0.955647
0	0	1	0.0145530
0	1	0	0.000396
0	1	1	0.019404
1	0	0	0.00004925
1	0	1	0.00000075
1	1	0	0.000199
1	1	1	0.009751

f	$p(f)$
0	0.99
1	0.01

s	$p(s \mid f)$	f
0	0.98	0
1	0.02	
0	0.005	1
1	0.995	

d	$p(d \mid s)$	s
0	0.985	0
1	0.015	
0	0.02	1
1	0.98	

Joint can be used for inference

f	s	d	$p(f, s, d)$
0	0	0	0.955647
0	0	1	0.0145530
0	1	0	0.000396
0	1	1	0.019404
1	0	0	0.00004925
1	0	1	0.00000075
1	1	0	0.000199
1	1	1	0.009751

Fire Detector cont'd

Example

f	$p(f)$
0	0.99
1	0.01

s	$p(s \mid f)$	f
0	0.98	0
1	0.02	
0	0.005	1
1	0.995	

d	$p(d \mid s)$	s
0	0.985	0
1	0.015	
0	0.02	1
1	0.98	

Joint can be used for inference

f	s	d	$p(f,s,d)$
0	0	0	0.955647
0	0	1	0.0145530
0	1	0	0.000396
0	1	1	0.019404
1	0	0	0.00004925
1	0	1	0.00000075
1	1	0	0.000199
1	1	1	0.009751

$\Rightarrow$

f	d	$p(f,d)$
0	0	0.956043
0	1	0.033957
1	0	0.00024825
1	1	0.00975175

f	$p(f)$
0	0.99
1	0.01

s	$p(s \mid f)$	f
0	0.98	0
1	0.02	
0	0.005	1
1	0.995	

d	$p(d \mid s)$	s
0	0.985	0
1	0.015	
0	0.02	1
1	0.98	

Joint can be used for inference

f	s	d	$p(f, s, d)$
0	0	0	0.955647
0	0	1	0.0145530
0	1	0	0.000396
0	1	1	0.019404
1	0	0	0.00004925
1	0	1	0.00000075
1	1	0	0.000199
1	1	1	0.009751

$\Rightarrow$

f	d	$p(f, d)$
0	0	0.956043
0	1	0.033957
1	0	0.00024825
1	1	0.00975175

$\Rightarrow$

d	$p(d)$
0	0.95629125
1	0.04370875

f	$p(f)$
0	0.99
1	0.01

s	$p(s \mid f)$	f
0	0.98	0
1	0.02	
0	0.005	1
1	0.995	

d	$p(d \mid s)$	s
0	0.985	0
1	0.015	
0	0.02	1
1	0.98	

Joint can be used for inference

f	s	d	$p(f, s, d)$
0	0	0	0.955647
0	0	1	0.0145530
0	1	0	0.000396
0	1	1	0.019404
1	0	0	0.00004925
1	0	1	0.00000075
1	1	0	0.000199
1	1	1	0.009751

$\Rightarrow$

f	d	$p(f, d)$
0	0	0.956043
0	1	0.033957
1	0	0.00024825
1	1	0.00975175

$\Rightarrow$

d	$p(d)$
0	0.95629125
1	0.04370875

$\Downarrow$

f	$p(f \mid d)$	d
0	0.99974	0
1	0.00026	
0	0.776892	1
1	0.223108	

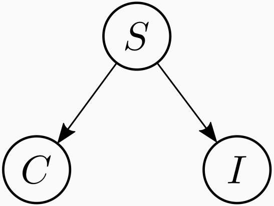
How are **sun shine, crime, ice cream** related?

- each day, there is a 0.9 probability of sun-shine (S)[1]
- on sunny days, the probability of high ice-cream sales (I) is 0.8; on cloudy days, the probability of high ice-cream sales is 0.03
- on sunny days the probability of an increased crime rate (C) is 0.3 (due to hot temper); on cloudy days the probability of an increased crime rate is 0.08

[1] This example is from Judea Pearl, who lives in LA.

Sun, Crime, Ice Cream cont'd

Example



s	$p(s)$
0	0.1
1	0.9

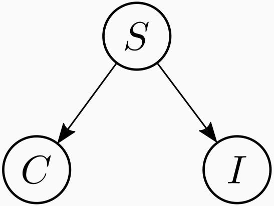
c	$p(c \mid s)$	s
0	0.92	0
1	0.08	
0	0.7	1
1	0.3	

i	$p(i \mid s)$	s
0	0.97	0
1	0.03	
0	0.2	1
1	0.8	

s	c	i	$p(s, c, i)$
0	0	0	
0	0	1	
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	

Sun, Crime, Ice Cream cont'd

Example



s	$p(s)$
0	0.1
1	0.9

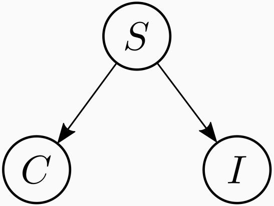
c	$p(c \mid s)$	s
0	0.92	0
1	0.08	
0	0.7	1
1	0.3	

i	$p(i \mid s)$	s
0	0.97	0
1	0.03	
0	0.2	1
1	0.8	

s	c	i	$p(s, c, i)$
0	0	0	$p(s=0) \times p(c=0 \mid s=0) \times p(i=0 \mid s=0)$
0	0	1	$p(s=0) \times p(c=0 \mid s=0) \times p(i=1 \mid s=0)$
0	1	0	$p(s=0) \times p(c=1 \mid s=0) \times p(i=0 \mid s=0)$
0	1	1	$p(s=0) \times p(c=1 \mid s=0) \times p(i=1 \mid s=0)$
1	0	0	$p(s=1) \times p(c=0 \mid s=1) \times p(i=0 \mid s=1)$
1	0	1	$p(s=1) \times p(c=0 \mid s=1) \times p(i=1 \mid s=1)$
1	1	0	$p(s=1) \times p(c=1 \mid s=1) \times p(i=0 \mid s=1)$
1	1	1	$p(s=1) \times p(c=1 \mid s=1) \times p(i=1 \mid s=1)$

Sun, Crime, Ice Cream cont'd

Example



s	$p(s)$
0	0.1
1	0.9

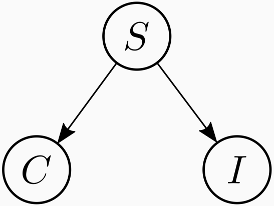
c	$p(c \mid s)$	s
0	0.92	0
1	0.08	
0	0.7	1
1	0.3	

i	$p(i \mid s)$	s
0	0.97	0
1	0.03	
0	0.2	1
1	0.8	

s	c	i	$p(s, c, i)$
0	0	0	$0.1 \times 0.92 \times 0.97$
0	0	1	$0.1 \times 0.92 \times 0.03$
0	1	0	$0.1 \times 0.08 \times 0.97$
0	1	1	$0.1 \times 0.08 \times 0.03$
1	0	0	$0.9 \times 0.7 \times 0.2$
1	0	1	$0.9 \times 0.7 \times 0.8$
1	1	0	$0.9 \times 0.3 \times 0.2$
1	1	1	$0.9 \times 0.3 \times 0.8$

Sun, Crime, Ice Cream cont'd

Example



s	$p(s)$
0	0.1
1	0.9

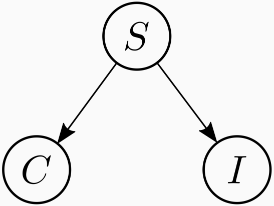
c	$p(c \mid s)$	s
0	0.92	0
1	0.08	0
0	0.7	1
1	0.3	1

i	$p(i \mid s)$	s
0	0.97	0
1	0.03	0
0	0.2	1
1	0.8	1

s	c	i	$p(s, c, i)$
0	0	0	0.08924
0	0	1	0.00276
0	1	0	0.00776
0	1	1	0.00024
1	0	0	0.126
1	0	1	0.504
1	1	0	0.054
1	1	1	0.216

Sun, Crime, Ice Cream cont'd

Example



s	$p(s)$
0	0.1
1	0.9

c	$p(c \mid s)$	s
0	0.92	0
1	0.08	0
0	0.7	1
1	0.3	1

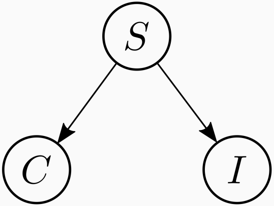
i	$p(i \mid s)$	s
0	0.97	0
1	0.03	0
0	0.2	1
1	0.8	1

Joint can be used for inference

s	c	i	$p(s, c, i)$
0	0	0	0.08924
0	0	1	0.00276
0	1	0	0.00776
0	1	1	0.00024
1	0	0	0.126
1	0	1	0.504
1	1	0	0.054
1	1	1	0.216

Sun, Crime, Ice Cream cont'd

Example



s	$p(s)$
0	0.1
1	0.9

c	$p(c \mid s)$	s
0	0.92	0
1	0.08	
0	0.7	1
1	0.3	

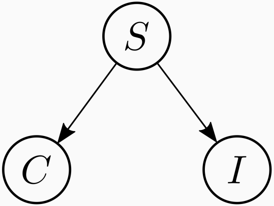
i	$p(i \mid s)$	s
0	0.97	0
1	0.03	
0	0.2	1
1	0.8	

Joint can be used for inference

s	c	i	$p(s, c, i)$
0	0	0	0.08924
0	0	1	0.00276
0	1	0	0.00776
0	1	1	0.00024
1	0	0	0.126
1	0	1	0.504
1	1	0	0.054
1	1	1	0.216

$\Rightarrow$

c	i	$p(c, i)$
0	0	0.21524
0	1	0.50676
1	0	0.06176
1	1	0.21624



s	$p(s)$
0	0.1
1	0.9

c	$p(c \mid s)$	s
0	0.92	0
1	0.08	0
0	0.7	1
1	0.3	1

i	$p(i \mid s)$	s
0	0.97	0
1	0.03	0
0	0.2	1
1	0.8	1

Joint can be used for inference

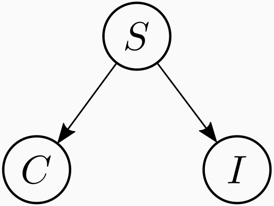
s	c	i	$p(s, c, i)$
0	0	0	0.08924
0	0	1	0.00276
0	1	0	0.00776
0	1	1	0.00024
1	0	0	0.126
1	0	1	0.504
1	1	0	0.054
1	1	1	0.216

$\Rightarrow$

c	i	$p(c, i)$
0	0	0.21524
0	1	0.50676
1	0	0.06176
1	1	0.21624

$\Rightarrow$

i	$p(i)$
0	0.277
1	0.723



s	$p(s)$
0	0.1
1	0.9

c	$p(c \mid s)$	s
0	0.92	0
1	0.08	
0	0.7	1
1	0.3	

i	$p(i \mid s)$	s
0	0.97	0
1	0.03	
0	0.2	1
1	0.8	

Joint can be used for inference

s	c	i	$p(s, c, i)$
0	0	0	0.08924
0	0	1	0.00276
0	1	0	0.00776
0	1	1	0.00024
1	0	0	0.126
1	0	1	0.504
1	1	0	0.054
1	1	1	0.216

$\Rightarrow$

c	i	$p(c, i)$
0	0	0.21524
0	1	0.50676
1	0	0.06176
1	1	0.21624

$\Rightarrow$

i	$p(i)$
0	0.277
1	0.723

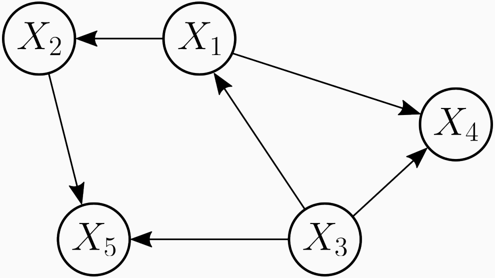
$\Downarrow$ $\Searrow$

c	$p(c \mid i)$	i
0	0.777	0
1	0.223	
0	0.701	1
1	0.299	

D-Separation

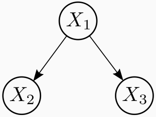
Conditional Independencies in BNs

- what are the CI assumptions in a given BN?
- generally, they depend on both the graph $\mathcal{G}$ and the CPDs $\mathcal{P}$
- however, there are CIs which always follow from the structure $\mathcal{G}$, regardless of $\mathcal{P}$ $\Rightarrow$ **structural CIs**
- how can we read them off the graph?

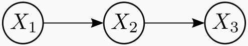


Canonical Examples

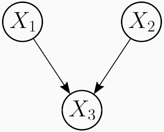
Let's look at the three examples of BNs we have discussed so far. They will turn out to be **canonical examples** for CI in BNs:



tail-to-tail node, **fork**



head-to-tail node, **chain**

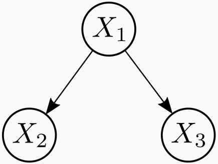


head-to-head node, **collider**

Theorem

Neighboring nodes are in general **not** independent, regardless of the conditioning set. That is, if X and Y are connected by an edge then there exist CPDs such that $X \not\perp\!\!\!\perp Y \mid \boldsymbol{Z}$ for any $\boldsymbol{Z} \subseteq \boldsymbol{X} \setminus \{X, Y\}$.

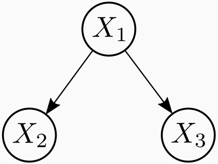
Fork, Unconditional Independence



$$p(x_1, x_2, x_3) = p(x_2 \mid x_1)\, p(x_3 \mid x_1)\, p(x_1)$$

- neighbors are in general dependent
- but does $X_2 \perp\!\!\!\perp X_3$ hold?

Fork, Unconditional Independence



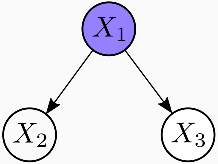
$$p(x_1, x_2, x_3) = p(x_2 \mid x_1)\, p(x_3 \mid x_1)\, p(x_1)$$

- neighbors are in general dependent
- but does $X_2 \perp\!\!\!\perp X_3$ hold?
- no, recall the Sun-Crime-Ice Cream example:

c	$p(c \mid i)$	i
0	0.777	0
1	0.223	
0	0.701	1
1	0.299	

- information "flows" over X_1

Fork, Conditional Independence

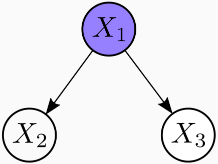


$$p(x_1, x_2, x_3) = p(x_2 \mid x_1)\, p(x_3 \mid x_1)\, p(x_1)$$

blue nodes in BNs are observed/conditioned on

- neighbors are in general dependent
- $X_2 \perp\!\!\!\perp X_3 \mid X_1$?

Fork, Conditional Independence



$$p(x_1, x_2, x_3) = p(x_2 \mid x_1)\, p(x_3 \mid x_1)\, p(x_1)$$

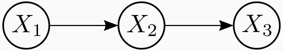
blue nodes in BNs are observed/conditioned on

- neighbors are in general dependent
- $X_2 \perp\!\!\!\perp X_3 \mid X_1$?

$$p(x_2 \mid x_1, x_3) = \frac{p(x_1, x_2, x_3)}{p(x_1, x_3)} = \frac{p(x_2 \mid x_1)\, p(x_3 \mid x_1)\, p(x_1)}{p(x_1, x_3)}$$
$$= p(x_2 \mid x_1)\, \frac{p(x_1, x_3)}{p(x_1, x_3)} = p(x_2 \mid x_1)$$

- thus, $X_2 \perp\!\!\!\perp X_3 \mid X_1 \Rightarrow$ information gets "blocked" by observing X_1

Chain, Unconditional Independence



$$p(x_1, x_2, x_3) = p(x_3 \mid x_2)\, p(x_2 \mid x_1)\, p(x_1)$$

- $X_1 \perp\!\!\!\perp X_3$?

Chain, Unconditional Independence

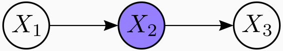
$$X_1 \longrightarrow X_2 \longrightarrow X_3 \qquad p(x_1, x_2, x_3) = p(x_3 \mid x_2)\, p(x_2 \mid x_1)\, p(x_1)$$

- $X_1 \perp\!\!\!\perp X_3$?
- no, recall the Fire Detector example:

f	$p(f \mid d)$	d
0	0.99974	0
1	0.00026	
0	0.776892	1
1	0.223108	

- information "flows" over X_2

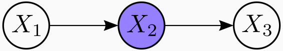
Chain, Conditional Independence



$$p(x_1, x_2, x_3) = p(x_3 \mid x_2)\, p(x_2 \mid x_1)\, p(x_1)$$

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

Chain, Conditional Independence



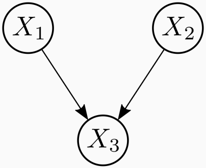
$$p(x_1, x_2, x_3) = p(x_3 \mid x_2)\, p(x_2 \mid x_1)\, p(x_1)$$

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

$$p(x_3 \mid x_1, x_2) = \frac{p(x_1, x_2, x_3)}{p(x_1, x_2)} = \frac{p(x_3 \mid x_2)\, p(x_2 \mid x_1)\, p(x_1)}{p(x_1, x_2)}$$
$$= p(x_3 \mid x_2) \frac{p(x_1, x_2)}{p(x_1, x_2)} = p(x_3 \mid x_2)$$

- thus, $X_1 \perp\!\!\!\perp X_3 \mid X_2 \Rightarrow$ information gets "blocked" by observing X_2

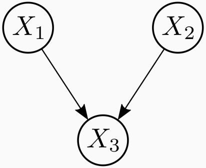
Collider, Unconditional Independence



$$p(x_1, x_2, x_3) = p(x_3 \mid x_1, x_2)\, p(x_2)\, p(x_1)$$

- $X_1 \perp\!\!\!\perp X_2$?

Collider, Unconditional Independence



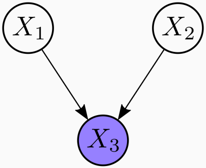
$$p(x_1, x_2, x_3) = p(x_3 \mid x_1, x_2)\, p(x_2)\, p(x_1)$$

- $X_1 \perp\!\!\!\perp X_2$?

$$\begin{aligned} p(x_1, x_2) &= \sum_{x_3} p(x_3 \mid x_1, x_2)\, p(x_2)\, p(x_1) \\ &= p(x_2)\, p(x_1) \underbrace{\sum_{x_3} p(x_3 \mid x_1, x_2)}_{=1} \\ &= p(x_2)\, p(x_1) \end{aligned}$$

- thus, $X_1 \perp\!\!\!\perp X_2$!

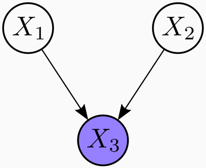
Head-to-Head Node, Collider



$$p(X_1, X_2, X_3) = p(X_3 \mid X_1, X_2)\, p(X_2)\, p(X_1)$$

Conditional independence: Is $X_1 \perp\!\!\!\perp X_2 \mid X_3$?

Head-to-Head Node, Collider



$$p(X_1, X_2, X_3) = p(X_3 \mid X_1, X_2)\, p(X_2)\, p(X_1)$$

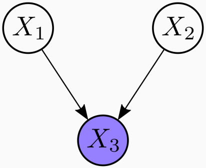
Conditional independence: Is $X_1 \perp\!\!\!\perp X_2 \mid X_3$?

No, consider the "coins and bell" example:

c_1	c_2	b	$p(c_1, c_2, b)$
0	0	0	0.2475
0	0	1	0.0025
0	1	0	0.125
0	1	1	0.125
1	0	0	0.125
1	0	1	0.125
1	1	0	0
1	1	1	0.25

b	$p(b)$
0	0.4975
1	0.5025

Head-to-Head Node, Collider



$$p(X_1, X_2, X_3) = p(X_3 \mid X_1, X_2)\, p(X_2)\, p(X_1)$$

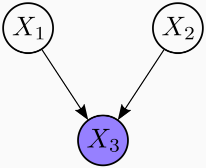
Conditional independence: Is $X_1 \perp\!\!\!\perp X_2 \mid X_3$?

No, consider the "coins and bell" example:

c_1	c_2	b	$p(c_1, c_2, b)$
0	0	0	0.2475
0	0	1	0.0025
0	1	0	0.125
0	1	1	0.125
1	0	0	0.125
1	0	1	0.125
1	1	0	0
1	1	1	0.25

b	$p(b)$
0	0.4975
1	0.5025

c_1	c_2	$p(c_1, c_2 \mid b)$	b
0	0	0.498	0
0	1	0.251	
1	0	0.251	
1	1	0	
0	0	0.005	1
0	1	0.2488	
1	0	0.2488	
1	1	0.4974	



$$p(X_1, X_2, X_3) = p(X_3 \mid X_1, X_2)\, p(X_2)\, p(X_1)$$

Conditional independence: Is $X_1 \perp\!\!\!\perp X_2 \mid X_3$?

No, consider the “coins and bell” example:

c_1	c_2	b	$p(c_1, c_2, b)$
0	0	0	0.2475
0	0	1	0.0025
0	1	0	0.125
0	1	1	0.125
1	0	0	0.125
1	0	1	0.125
1	1	0	0
1	1	1	0.25

b	$p(b)$
0	0.4975
1	0.5025

<table>
<tr><th>c_1</th><th>c_2</th><th>$p(c_1, c_2 \mid b)$</th><th>b</th></tr>
<tr><td>0</td><td>0</td><td>0.498</td><td rowspan="4">0</td></tr>
<tr><td>0</td><td>1</td><td>0.251</td></tr>
<tr><td>1</td><td>0</td><td>0.251</td></tr>
<tr><td>1</td><td>1</td><td>0</td></tr>
<tr><td>0</td><td>0</td><td>0.005</td><td rowspan="4">1</td></tr>
<tr><td>0</td><td>1</td><td>0.2488</td></tr>
<tr><td>1</td><td>0</td><td>0.2488</td></tr>
<tr><td>1</td><td>1</td><td>0.4974</td></tr>
</table>

$p(C_1, C_2 \mid B)$ does not factorize. E.g., $p(C_1 = 1 \mid B = 0) = 0.251$ and $p(C_2 = 1 \mid B = 0) = 0.251$, but $p(C_1 = 1, C_2 = 1 \mid B = 0) = 0$.

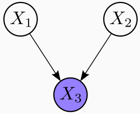
Explaining Away

We see that the **collider** introduces an interesting independence pattern, that is opposite to **forks** and **chains**.

This corresponds to a common sense human reasoning pattern, called **explaining away**.

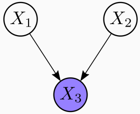
Explaining Away I

Example



- you have a car, where the battery might be dead ($X_1 = 0$, low probability) or alive ($X_1 = 1$, high probability)
- the gas tank might be empty ($X_2 = 0$, low probability) or full ($X_2 = 1$, high probability)
- you observe that the car does not start ($X_3 = 0$)
- "gosh, is the battery dead?" (it's expensive to replace)
- you check the gas tank and see it is empty $X_2 = 0$
- "oh, fortunately, the battery is probably fine!"
- **gas tank X_2 became predictive for X_1 when X_3 is observed, although X_1 and X_2 are independent a-priori**
- $X_2 = 0$ **explained away** the fact that $X_3 = 0$

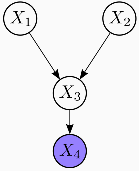
Explaining Away II



- you might have a burglary in your home ($X_1 = 1$, low probability) or not ($X_1 = 0$, high probability)
- there might be an earthquake ($X_2 = 1$, low probability) or not ($X_2 = 0$, high probability)
- both X_1 or X_2 have a high chance to trigger your alarm X_3
- you get a call in your office from your neighbor that ($X_3 = 1$)
- "oh dear, a burglary in my home?"
- your drive home and hear in the radio that there was an earthquake in your neighborhood ($X_2 = 1$)
- "oh fortunately, it was only an earthquake, not a burglary!"
- $X_2 = 1$ **explained away the fact that** $X_3 = 1$

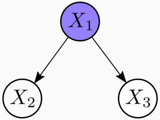
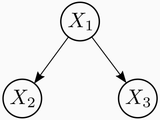
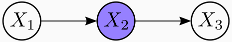
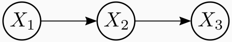
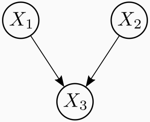
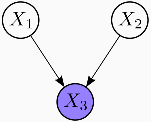
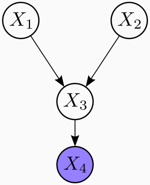
Explaining Away II cont'd

Example



- say your neighbor X_4 is not completely reliable
- he might miss an alarm with probability 0.05
- also, he has a funny sense of humor and reports a false alert with probability 0.01
- thus, you are not observing X_3 but a "noisy version" X_4
- yet, X_2 explains away X_4, decreasing the probability of X_1
- **explaining away extends to descendants of colliders!**

Basic Independencies

blocked	unblocked
fork	
	
chain	
	
collider	
	 

These patterns generalize to arbitrary BNs!

- consider two arbitrary RVs $X, Y \in \boldsymbol{X}$
- let $\boldsymbol{Z} \subseteq \boldsymbol{X} \setminus \{X, Y\}$ (might also be empty)
- let $\Pi = (X, \ldots, Y)$ be an arbitrary path from X to Y in $\mathcal{G}$

We say that path Π is **blocked by $\boldsymbol{Z}$** if for any $V_k \notin \{X, Y\}$ in Π

- (V_{k-1}, V_k, V_{k+1}) is a **fork** or a **chain** and V_k is in $\boldsymbol{Z}$.
- (V_{k-1}, V_k, V_{k+1}) is a **collider** and neither V_k nor any of its descendant is in $\boldsymbol{Z}$.

d-separation

Let $\mathcal{G}$ be a DAG over RVs $\boldsymbol{X}$. Consider arbitrary two RVs $X, Y \in \boldsymbol{X}$ and let $\boldsymbol{Z} \subseteq \boldsymbol{X} \setminus \{X, Y\}$. We say that X and Y are **d-separated by $\boldsymbol{Z}$** if **all** paths from X to Y are blocked by $\boldsymbol{Z}$.

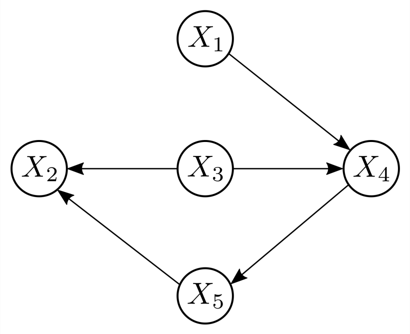
Theorem (Geiger, Verma, and Pearl, 1990)

Let $(\mathcal{G}, \mathcal{P})$ be a Bayesian network. If X and Y are d-separated by $\boldsymbol{Z}$ in $\mathcal{G}$, then $X \perp\!\!\!\perp Y \mid \boldsymbol{Z}$ in the BN distribution.

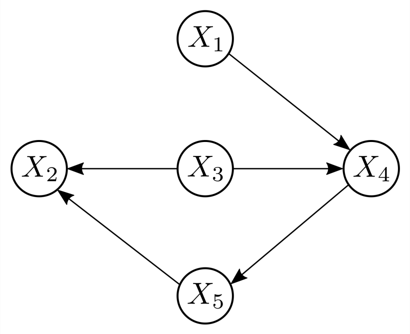
Note: this is a **graphical condition** and holds for any collection of CPDs $\mathcal{P}$.

Note: it is a sufficient, not necessary condition—even a fully connected DAG can have arbitrary CIs encoded in $\mathcal{P}$.

Is X_1 d-separated from X_2 by $\emptyset$?

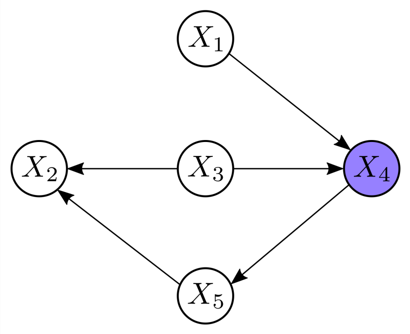


Is X_1 d-separated from X_2 by $\emptyset$?

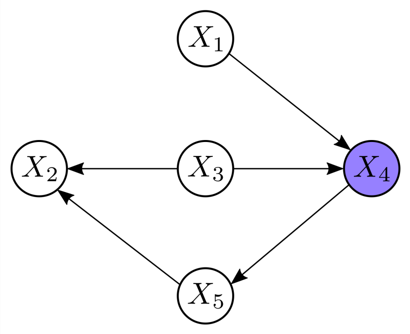


No, since path (X_1, X_4, X_5, X_2) is not blocked.

Is X_1 d-separated from X_2 by $\{X_4\}$?

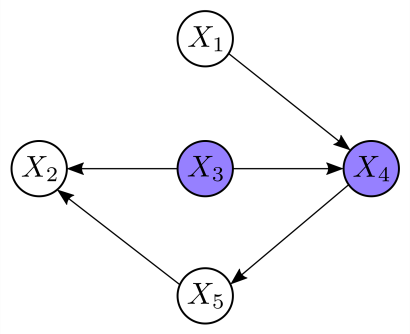


Is X_1 d-separated from X_2 by $\{X_4\}$?

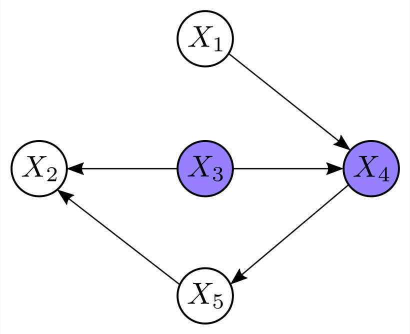


No, since path (X_1, X_4, X_3, X_2) is not blocked.

Is X_1 d-separated from X_2 by $\{X_3, X_4\}$?

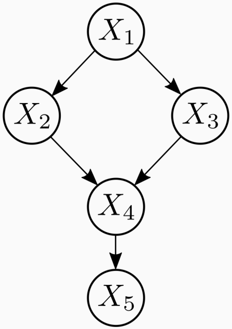


Is X_1 d-separated from X_2 by $\{X_3, X_4\}$?

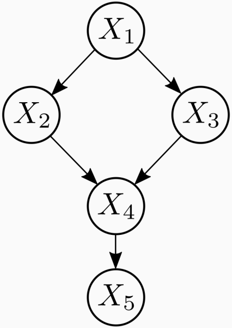


Yes! Thus any BN with this graph will have have $X_1 \perp\!\!\!\perp X_2 \mid X_3, X_4$.

Is X_1 d-separated from X_5?

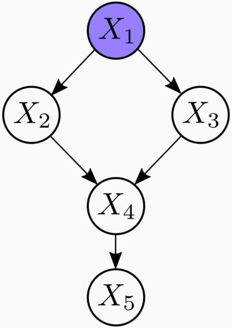


Is X_1 d-separated from X_5?

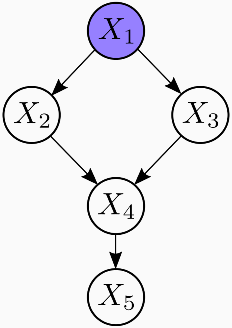


No, the paths (X_1, X_2, X_4, X_5) and (X_1, X_3, X_4, X_5) are not blocked.

Is X_2 d-separated from X_3 by $\{X_1\}$?

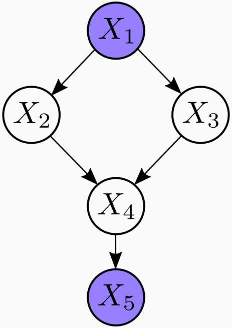


Is X_2 d-separated from X_3 by $\{X_1\}$?

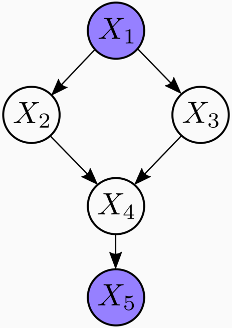


Yes! Thus any BN with this graph will have have $X_2 \perp\!\!\!\perp X_3 \mid X_1$.

Is X_2 d-separated from X_3 by $\{X_1, X_5\}$?



Is X_2 d-separated from X_3 by $\{X_1, X_5\}$?



No, the path X_2, X_4, X_3 is not blocked, as X_5 is a descendant of collider X_4.

Graphical Models—A Lingua Franca of Probabilistic Modeling

So, why does this matter?

- **graphical models** such as **Bayesian networks** are usually understood as a "type of model"
- however, the real benefit is to understand **dependency structure** and information flow
- basically, all probabilistic models can be interpreted through a graphical models lens
- authors and researchers can easily **communicate ideas and assumptions** using graphical models