CS188: Announcements

Self Grade Drop: You have 1 for the semester.
Self Grade Drop: You have 1 for the semester.

See Piazza
CS188: Announcements

Self Grade Drop: You have 1 for the semester.

See Piazza

Homework due tonight.
CS188: Announcements

Self Grade Drop: You have 1 for the semester.
See Piazza
Homework due tonight.
Project 3 on Friday.
Self Grade Drop: You have 1 for the semester.

See Piazza

Homework due tonight.

Project 3 on Friday.

Discussing new attendance policy.

Bayes’ Nets

Bayes’ Nets
Conditional probability:

\[ P(x \mid y) = \frac{P(x, y)}{P(y)} \]

Product rule:

\[ P(x, y) = P(x \mid y) P(y) \]

Bayes Rule:

\[ P(y \mid x) = \frac{P(x \mid y) P(y)}{P(x)} \]

Chain rule:

\[ P(x_1, x_2, \ldots, x_n) = P(x_1) P(x_2 \mid x_1) P(x_3 \mid x_1, x_2) \ldots = \prod_{i=1}^{n} P(x_i \mid x_1, \ldots, x_{i-1}) \]

\( X, Y \) in independent if and only if:

\[ \forall x, y: P(x, y) = P(x) P(y) \]

\( X \) and \( Y \) are conditionally independent given \( Z \) if and only if:

\[ \forall x, y, z: P(x, y \mid z) = P(x \mid z) P(y \mid z) \]
Conditional probability:

\[
P(x | y) = \frac{P(x, y)}{P(y)}.
\]

Product rule:

\[
P(x, y) = P(x | y) P(y).
\]

Bayes Rule:

\[
P(y | x) = \frac{P(x | y) P(y)}{P(x)}.
\]

Chain rule:

\[
P(x_1, x_2, \ldots, x_n) = P(x_1) P(x_2 | x_1) P(x_3 | x_1, x_2) \ldots = \prod_{i=1}^{n} P(x_i | x_1, \ldots, x_{i-1}).
\]
Probability Recap


Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).
Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule:
Conditional probability: \( P(x \mid y) = \frac{P(x, y)}{P(y)} \).

Product rule: \( P(x, y) = P(x \mid y)P(y) \).
Conditional probability: $P(x|y) = \frac{P(x,y)}{P(y)}$.

Product rule: $P(x,y) = P(x|y)P(y)$.

Bayes Rule:

Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:

Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:
\[
P(x_1, x_2, \ldots, x_n) = P(x_1)
\]
Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:
\[
P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)
\]
Probability Recap


Conditional probability: $P(x|y) = \frac{P(x,y)}{P(y)}$.

Product rule: $P(x,y) = P(x|y)P(y)$.

Bayes Rule: $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$.

Chain rule:

$$P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots$$

Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:
\[
P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots \\
= \prod_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}).
\]

Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:
\[
P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots \\
= \prod_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}).
\]

\( X, Y \) independent if and only if:
Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:
\[
P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots
= \prod_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}).
\]

\( X, Y \) independent if and only if: \( \forall x, y : P(x, y) = P(x)P(y) \).
Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:
\[
P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots
\]
\[
= \prod_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}).
\]

\( X, Y \) independent if and only if: \( \forall x, y : P(x,y) = P(x)P(y) \).

\( X \) and \( Y \) are conditionally independent given \( Z \) if and only if:
Probability Recap


Conditional probability: $P(x|y) = \frac{P(x,y)}{P(y)}$.

Product rule: $P(x,y) = P(x|y)P(y)$.

Bayes Rule: $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$.

Chain rule:

$$P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots$$
$$= \prod_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}).$$

$X, Y$ independent if and only if: $\forall x, y : P(x,y) = P(x)P(y)$.

$X$ and $Y$ are conditionally independent given $Z$ if and only if:

$\forall x, y, z : P(x,y|z) = P(x|z)P(y|z)$. 
Ghostbusters Chain Rule

[Demo: Ghostbuster – with probability (L12D2) ]
Ghostbusters Chain Rule

[Demo: Ghostbuster – with probability (L12D2) ]

Each sensor depends only on where the ghost is.

Givens:
- \( P( +g ) = 0.5 \)
- \( P( -g ) = 0.5 \)
- \( P( +t | +g ) = 0.8 \)
- \( P( +t | -g ) = 0.4 \)
- \( P( +b | +g ) = 0.4 \)
- \( P( +b | -g ) = 0.8 \)

\[
P(T, B, G) = P(G)P(T | G)P(B | G)
\]
Each sensor depends only on where the ghost is.
That means, the two sensors are conditionally independent, given the ghost position.
Ghostbusters Chain Rule

[Demo: Ghostbuster – with probability (L12D2) ]

Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.
Ghostbusters Chain Rule

[Demo: Ghostbuster – with probability (L12D2) ]

Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:
Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:

\[ P(+g) = 0.5 \]
Ghostbusters Chain Rule


[Demo: Ghostbuster – with probability (L12D2) ]

Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:

\[
P( +g ) = 0.5 \\
P( -g ) = 0.5
\]
Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:

\[ P( +g ) = 0.5 \]
\[ P( -g ) = 0.5 \]
\[ P( +t | +g ) = 0.8 \]
\[ P( +t | -g ) = 0.4 \]
\[ P( +b | +g ) = 0.4 \]
\[ P( +b | -g ) = 0.8 \]
Ghostbusters Chain Rule

[Demo: Ghostbuster – with probability (L12D2) ]

Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:

\[
P( +g ) = 0.5 \\
P( -g ) = 0.5 \\
P( +t \mid +g ) = 0.8 \\
P( +t \mid -g ) = 0.4 \\
P( +b \mid +g ) = 0.4 \\
P( +b \mid -g ) = 0.8
\]

\[
P(T, B, G) = P(G)P(T \mid G)P(B \mid G)
\]
Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:

\[ \begin{align*}
P( +g ) &= 0.5 \\
P( -g ) &= 0.5 \\
P( +t | +g ) &= 0.8 \\
P( +t | -g ) &= 0.4 \\
P( +b | +g ) &= 0.4 \\
P( +b | -g ) &= 0.8
\end{align*} \]

\[ P(T, B, G) = P(G)P(T|G)P(B|G) \]
Bayes’Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint is WAY too big to represent explicitly.
- Hard to learn (estimate) anything empirically about more than a few variables at a time.

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities).

- More properly called graphical models.
- We describe how variables locally interact.
- Local interactions chain together to give global, indirect interactions.
- For now, we’ll be vague about how these interactions are specified.
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint distribution is WAY too big to represent explicitly.
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time
Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint is WAY too big to represent explicitly.
- Hard to learn (estimate) anything empirically about more than a few variables at a time.
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)

- More properly called graphical models
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
- More properly called graphical models
- We describe how variables locally interact
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)

- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions
- For now, we’ll be vague about how these interactions are specified
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)

- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions
- For now, we’ll be vague about how these interactions are specified
Example Bayes’ Net: Car
Ghostbusters Bayes Net
Ghostbusters Bayes Net

$R_1 \quad \ldots \quad R_n$
Graphical Model Notation

Nodes: variables (with domains)
- Variables can be assigned (observed) or unassigned (unobserved).

Arcs: interactions
- Similar to CSP constraints.
- Indicate "direct influence" between variables.
- Formally: encode conditional independence (more later).

For now: imagine that arrows mean direct causation (in general, they don't!)

Diagram:
- Weather
- Cavity
- Toothache
- Catch

Diagram shows a causal chain:
- Weather influences toothache.
- Toothache can lead to cavity.
- Cavity can lead to catch.
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate "direct influence" between variables
- Formally: encode conditional independence (more later)

For now: imagine that arrows mean direct causation (in general, they don’t!)
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

[Diagrams with nodes and arrows illustrating the concepts of nodes and arcs in graphical models]
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate “direct influence” between variables
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate “direct influence” between variables
- Formally: encode conditional independence (more later)
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate “direct influence” between variables
- Formally: encode conditional independence (more later)
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate “direct influence” between variables
- Formally: encode conditional independence (more later)

For now: imagine that arrows mean direct causation (in general, they don’t!)
Example: Coin Flips

N independent coin flips
No interactions between variables: absolute independence

\[X_1, X_2, \ldots, X_n\]
Example: Coin Flips

N independent coin flips
Example: Coin Flips

N independent coin flips
No interactions between variables: absolute independence

$X_1 \quad X_2 \quad \cdots \quad X_n$
Example: Traffic

Variables:
- R: It rains
Example: Traffic

Variables:
- R: It rains
- T: There is traffic
Example: Traffic

Variables:
- R: It rains
- T: There is traffic
Example: Traffic

Variables:
- R: It rains
- T: There is traffic

Model 1: independence

Model 2: rain causes traffic

Why is an agent using model 2 better?
Example: Traffic

Variables:
- R: It rains
- T: There is traffic

Model 1: independence

Model 2: rain causes traffic

Why is an agent using model 2 better?
Example: Traffic

Variables:
- R: It rains
- T: There is traffic

Model 1: independence

Model 2: rain causes traffic

Why is an agent using model 2 better?
Example: Traffic

Variables:
- R: It rains
- T: There is traffic

Model 1: independence

Model 2: rain causes traffic

Why is an agent using model 2 better?
Example: Traffic II

Let’s build a causal graphical model!
Example: Traffic II

Let’s build a causal graphical model!
Example: Traffic II

Let's build a causal graphical model!

Variables
- T: Traffic
Example: Traffic II

Let’s build a causal graphical model!

Variables
- T: Traffic
- R: It rains
Example: Traffic II

Let’s build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
Example: Traffic II

Let’s build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
Example: Traffic II

Let's build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
Example: Traffic II

Let’s build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity
Example: Traffic II

Let's build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity
Example: Traffic II

Let’s build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity
Example: Traffic II

Let’s build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity
Example: Traffic II

Let's build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity

Diagram:

- Low Pres → Rain → Traffic
- Low Pres → Ballgame
- Rain → Traffic
- Ballgame → Drips
- Drips → Cavity
Example: Traffic II

Let’s build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity
Example: Alarm Network

- Variables
  - $t_B$: Burglary
  - $t_A$: Alarm goes off
  - $t_M$: Mary calls
  - $t_J$: John calls
  - $t_E$: Earthquake!
Example: Alarm Network

Variables
- B: Burglary
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!

Diagram:
- Burglary
- Earthquake
- Alarm
- John calls
- Mary calls
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!

Burglary ➔ Alarm ➔ Earthqk
- John calls
- Mary calls
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!

Diagram:
- Burglary
- Earthquake
- Alarm
- John calls
- Mary calls
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!

Diagram:
- Burglary
- Earthqk
- Alarm
  - John calls
  - Mary calls
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!

Diagram:

- Burglary → Alarm
- Earthquake → Alarm
- Alarm → John calls
- Alarm → Mary calls
Bayes’ Net Semantics
Bayes’ Net Semantics

A set of nodes, one per variable $X$
A directed, acyclic graph
A conditional distribution for each node
A collection of distributions over $X$, one for each combination of parents’ values
CPT: conditional probability table
Description of a noisy “causal” process

A Bayes net = Topology (graph) + Local Conditional Probabilities
Bayes’ Net Semantics

A set of nodes, one per variable $X$

$A_1 \quad \ldots \quad A_n$
Bayes’ Net Semantics

A set of nodes, one per variable \( X \)
A directed, acyclic graph

\[ A_1 \quad \ldots \quad A_n \]

\[ X \]
Bayes’ Net Semantics

A set of nodes, one per variable $X$
A directed, acyclic graph
A conditional distribution for each node

$A_1 \ldots A_n$
Bayes’ Net Semantics

A set of nodes, one per variable $X$
A directed, acyclic graph
A conditional distribution for each node

- A collection of distributions over $X$, one for each combination of parents’ values
Bayes’ Net Semantics

- A set of nodes, one per variable $X$
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over $X$, one for each combination of parents’ values
  - CPT: conditional probability table
Bayes’ Net Semantics

- A set of nodes, one per variable $X$
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over $X$, one for each combination of parents’ values
  - CPT: conditional probability table
  - Description of a noisy “causal” process
Bayes’ Net Semantics

A set of nodes, one per variable $X$
A directed, acyclic graph
A conditional distribution for each node
- A collection of distributions over $X$, one for each combination of parents’ values
- CPT: conditional probability table
- Description of a noisy “causal” process
Bayes’ Net Semantics

A set of nodes, one per variable $X$
A directed, acyclic graph
A conditional distribution for each node
- A collection of distributions over $X$, one for each combination of parents’ values
- CPT: conditional probability table
- Description of a noisy “causal” process

A Bayes net = Topology (graph) + Local Conditional Probabilities
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions
- As a product of local conditional distributions
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions

- As a product of local conditional distributions

\[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i | \text{Parents}(X_i)) \]
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions

- As a product of local conditional distributions

\[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i | \text{Parents}(X_i)) \]

- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions
- As a product of local conditional distributions
  \[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i|\text{Parents}(X_i)) \]
- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions

- As a product of local conditional distributions

\[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i | Parents(X_i)) \]

- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

\[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i | Parents(X_i)) \]
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions
- As a product of local conditional distributions
  \[ P(x_1, \ldots, x_n) = \prod_i \Pr(x_i | \text{Parents}(X_i)) \]
- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
  \[ P(x_1, \ldots, x_n) = \prod_i \Pr(x_i | \text{Parents}(X_i)) \]

Example:
\[
\Pr(G = b | T = r, B = r) = \Pr(G = r | T = r)P(S = s | B = r)
\]
Bayes’ nets implicitly encode joint distributions

- As a product of local conditional distributions
  \[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i|Parents(X_i)) \]

- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
  \[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i|Parents(X_i)) \]

Example:
\[
Pr(G = b|T = r, B = r) = Pr(G = r|T = r)P(S = s|B = r)
\]
\[
Pr(G = b|T = r, B = r) = Pr(G = r|T = r, B = r)P(S = s|B = r)
\]
Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_{i-1}) \]

Assume conditional independences:

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i | \text{parents}(X_i)) \]

Consequence:

Not every BN can represent every joint distribution.

The topology enforces certain conditional independencies!
Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?
Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

\[
P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_{i-1})
\]

Assume conditional independences:

\[
P(x_1, \ldots, x_n) = \prod_i P(x_i | \text{parents}(X_i))
\]

Consequence:

Not every BN can represent every joint distribution.

The topology enforces certain conditional independencies!
Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_i) \]
Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_i)$$

Assume conditional independences:
Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_i)$$

Assume conditional independences:

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | \text{parents}(X_i))$$
Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_i) \]

Assume conditional independences:

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i | \text{parents}(X_i)) \]

**Consequence:** Not every BN can represent every joint distribution.

- The topology enforces certain conditional independencies!
Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_i) \]

Assume conditional independences:

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i | \text{parents}(X_i)) \]

**Consequence:** Not every BN can represent every joint distribution.

- The topology enforces certain conditional independencies!
Example: Coin Flips

<table>
<thead>
<tr>
<th>X_1</th>
<th>h</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Only distributions whose variables are absolutely independent can be represented by a Bayes’ net with no arcs.
Only distributions whose variables are absolutely independent can be represented by a Bayes’ net with no arcs.
Example: Traffic

\[ P(\pm r, \pm t) = P(\pm r) \times P(\mp t | \pm r) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \]
Example: Traffic

\[ P(+r, -t) = P(+r) \times P(-t | +r) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \]

<table>
<thead>
<tr>
<th>R</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>1/4</td>
</tr>
<tr>
<td>-r</td>
<td>3/4</td>
</tr>
</tbody>
</table>
Example: Traffic

\[
P(+r,-t) = \frac{1}{16}
\]

\[
\begin{array}{|c|c|}
\hline
R & T \\
\hline
+r & 1/4 \\
-r & 3/4 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
R & T & P(T|R) \\
\hline
+r & +t & 3/4 \\
+r & -t & 1/4 \\
-r & +t & 1/2 \\
-r & -t & 1/2 \\
\hline
\end{array}
\]
Example: Traffic

\[ P(+r, -t) = P(+r) \times P(-t \mid +r) \]

<table>
<thead>
<tr>
<th>R</th>
<th>T</th>
<th>( P(T \mid R) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>+t</td>
<td>3/4</td>
</tr>
<tr>
<td>+r</td>
<td>-t</td>
<td>1/4</td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>1/2</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
<td>1/2</td>
</tr>
</tbody>
</table>
Example: Traffic

\[ P(+r,-t) = P(+r) \times P(-t \mid +r) \]

\[ \begin{array}{c|c|c|c}
 R & T & P(T \mid R) \\
\hline
 +r & +t & 3/4 \\
 +r & -t & 1/4 \\
 -r & +t & 1/2 \\
 -r & -t & 1/2 \\
\end{array} \]

\[ = (1/4) \times (1/4) = 1/16 \]
Example: Traffic

\[ P(+r,-t) = P(+r) \times P(-t \mid +r) = (1/4) \times (1/4) \]
Example: Traffic

\[
P(+r,-t) = P(+r) \times P(-t \mid +r) \\
= (1/4) \times (1/4) = 1/16
\]
Example: Traffic

\[ \begin{align*} 
R & \quad T \\
+r & \quad 1/4 \\
-r & \quad 3/4 \\
\end{align*} \]

\[ \begin{align*} 
R & \quad T & \quad \text{P}(T|R) \\
+r & \quad +t & \quad 3/4 \\
+r & \quad -t & \quad 1/4 \\
-r & \quad +t & \quad 1/2 \\
-r & \quad -t & \quad 1/2 \\
\end{align*} \]

\[ P(+r,-t) = P(+r) \times P(-t | +r) \\
= (1/4) \times (1/4) = 1/16 \]
Example: Alarm Network

- Burglary
- Earthqk
- Alarm
- John calls
- Mary calls
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

- **Burglary**
- **Earthquake**
- **Alarm**
- **John calls**
- **Mary calls**
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>
Example: Alarm Network

- **Burglary**
  - \( P(B) \)
  - \( +b \): 0.001
  - \( -b \): 0.999

- **Earthquake**

- **Alarm**

- **John calls**

- **Mary calls**

- **Table of Probabilities**:
  | B  | E  | A  | \( P(A|B,E) \) |
  |----|----|----|----------------|
  | +b | +e | +a | 0.95           |
  | +b | +e | -a | 0.05           |
  | +b | -e | +a | 0.94           |
  | +b | -e | -a | 0.06           |
  | -b | +e | +a | 0.29           |
  | -b | +e | -a | 0.71           |
  | -b | -e | +a | 0.001          |
  | -b | -e | -a | 0.999          |
Example: Alarm Network

- **B** (Burglary) and **E** (Earthquake) are the root causes that can trigger an alarm.
- **A** (Alarm) is the alarm itself, which can be triggered by **B** or **E**.
- **John calls** and **Mary calls** are the outcomes of the alarm.

**Probabilities**:

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| B | E | A   | P(A|B,E) |
|---|---|-----|--------|
| +b| +e| +a  | 0.95   |
| +b| +e| -a  | 0.05   |
| +b| -e| +a  | 0.94   |
| +b| -e| -a  | 0.06   |
| -b| +e| +a  | 0.29   |
| -b| +e| -a  | 0.71   |
| -b| -e| +a  | 0.001  |
| -b| -e| -a  | 0.999  |

- **A** and **J** (John calls) are the outcomes of **B** (Burglary) and **E** (Earthquake).

| A | J | P(J|A) |
|---|---|------|
| +a| +j| 0.9  |
| +a| -j| 0.1  |
| -a| +j| 0.05 |
| -a| -j| 0.95 |
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

| A  | J  | P(J|A) |
|----|----|------|
| +a | +j | 0.9  |
| +a | -j | 0.1  |
| -a | +j | 0.05 |
| -a | -j | 0.95 |

| A  | M  | P(M|A) |
|----|----|------|
| +a | +m | 0.7  |
| +a | -m | 0.3  |
| -a | +m | 0.01 |
| -a | -m | 0.99 |

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| B  | E  | A  | P(A|B,E) |
|----|----|----|---------|
| +b | +e | +a | 0.95    |
| +b | +e | -a | 0.05    |
| +b | -e | +a | 0.94    |
| +b | -e | -a | 0.06    |
| -b | +e | +a | 0.29    |
| -b | +e | -a | 0.71    |
| -b | -e | +a | 0.001   |
| -b | -e | -a | 0.999   |
Example: Traffic

Causal direction

\[ P(T|R) \]
Example: Traffic

Causal direction

<table>
<thead>
<tr>
<th>R</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>1/4</td>
</tr>
<tr>
<td>-r</td>
<td>3/4</td>
</tr>
</tbody>
</table>
Example: Traffic

Causal direction

\[
P(T|R) = \begin{array}{c|c}
R & T \\
+\alpha & 1/4 \\
-\alpha & 3/4 \\
\end{array}
\]
Example: Traffic

Causal direction

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>1/4</td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>3/4</td>
<td></td>
</tr>
</tbody>
</table>

|   |   | P(T|R) |
|---|---|--------|
| +r| t | 3/4    |
| +r| t | 1/4    |
| -r| t | 1/2    |
| -r| t | 1/2    |
Example: Traffic

Causal direction

\[
\begin{array}{c|c}
R & T \\
+\epsilon & \frac{1}{4} \\
-\epsilon & \frac{3}{4} \\
\end{array}
\]

\[
P(T|R)
\]

\[
\begin{array}{c|c|c}
+\epsilon & +t & \frac{3}{4} \\
+\epsilon & -t & \frac{1}{4} \\
-\epsilon & +t & \frac{1}{2} \\
-\epsilon & -t & \frac{1}{2} \\
\end{array}
\]
Example: Traffic

Causal direction

\[
\begin{array}{c|c}
 R & T \\
+ r & 1/4 \\
- r & 3/4 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
 & +t & 3/4 \\
+ r & +t & 3/4 \\
+ r & -t & 1/4 \\
- r & +t & 1/2 \\
- r & -t & 1/2 \\
\end{array}
\]

\[
\begin{array}{c|c}
 P(T|R) & P(R,T) \\
+ r & -r \\
+ t & + t \\
- t & - t \\
\end{array}
\]
Example: Traffic

Causal direction

\[
\begin{array}{|c|c|}
\hline
R & T \\
\hline
+r & 1/4 \\
-r & 3/4 \\
\hline
\end{array}
\]

\[
P(T|R)
\]

\[
\begin{array}{|c|c|c|}
\hline
+r & +t & 3/4 \\
+r & -t & 1/4 \\
-r & +t & 1/2 \\
-r & -t & 1/2 \\
\hline
\end{array}
\]

\[
P(R,T)
\]

\[
\begin{array}{|c|c|c|}
\hline
+r & +t & 3/16 \\
+r & -t & 1/16 \\
-r & +t & 6/16 \\
-r & -t & 6/16 \\
\hline
\end{array}
\]
Example: Reverse Traffic

Reverse causality?

\[
\begin{align*}
P(T|+t) &= \frac{9}{16} + t + \frac{1}{3} + t + \frac{1}{7} - t - \frac{6}{7} - t \\
P(T|-t) &= \frac{7}{16} - t - \frac{2}{3} - t - \frac{6}{16} - t \\
P(R,T) &= \frac{3}{16} + r + t + \frac{1}{16} + r - t + \frac{6}{16} - r + t - \frac{6}{16} - r - t
\end{align*}
\]
Example: Reverse Traffic

Reverse causality?

\[ P(T) \]

\[ R \]

\[ T \]
Example: Reverse Traffic

Reverse causality?

\[ P(T) \]

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+t</td>
<td>9/16</td>
</tr>
<tr>
<td>-t</td>
<td>7/16</td>
</tr>
</tbody>
</table>
Example: Reverse Traffic

Reverse causality?

\[
P(T) \\
\begin{array}{|c|c|}
  \hline
  +t & 9/16 \\
  -t & 7/16 \\
  \hline
\end{array}
\]

\[P(R|T)\]
Example: Reverse Traffic

Reverse causality?

\[ P(T) \]

\begin{array}{c|c}
+ t & 9/16 \\
- t & 7/16 \\
\end{array}

\[ P(R | T) \]

\begin{array}{c|c|c}
+ t & + r & 1/3 \\
+ t & - r & 2/3 \\
- t & + r & 1/7 \\
- t & - r & 6/7 \\
\end{array}
Example: Reverse Traffic

Reverse causality?

\[
P(T)\]

<table>
<thead>
<tr>
<th></th>
<th>+t</th>
<th>-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>+t</td>
<td>9/16</td>
<td></td>
</tr>
<tr>
<td>-t</td>
<td>7/16</td>
<td></td>
</tr>
</tbody>
</table>

\[
P(R|T)\]

<table>
<thead>
<tr>
<th></th>
<th>+t</th>
<th>+r</th>
<th>-t</th>
<th>-r</th>
<th>1/3</th>
</tr>
</thead>
<tbody>
<tr>
<td>+t</td>
<td></td>
<td>+r</td>
<td></td>
<td>-r</td>
<td>2/3</td>
</tr>
<tr>
<td>-t</td>
<td>+r</td>
<td></td>
<td>-r</td>
<td></td>
<td>1/7</td>
</tr>
<tr>
<td>-t</td>
<td>-r</td>
<td></td>
<td></td>
<td></td>
<td>6/7</td>
</tr>
</tbody>
</table>
Example: Reverse Traffic

Reverse causality?

\[ \begin{align*}
  P(T) & \\
  +t & 9/16 \\
  -t & 7/16 \\
\end{align*} \]

\[ \begin{align*}
  P(R|T) & \\
  +t & +r & 1/3 \\
  +t & -r & 2/3 \\
  -t & +r & 1/7 \\
  -t & -r & 6/7 \\
\end{align*} \]
Example: Reverse Traffic

Reverse causality?

\[
\begin{array}{c|c|c}
  R & +t & 9/16 \\
  \ \ & -t & 7/16 \\
\end{array}
\]

\[
\begin{array}{c|c|c|c|c}
  P(R|T) & +r & +t & 1/3 \\
  +t & -r & 2/3 \\
  -t & +r & 1/7 \\
  -t & -r & 6/7 \\
\end{array}
\]

\[
\begin{array}{c|c|c|c|c}
  P(R,T) & +r & +t & 3/16 \\
  +r & -t & 1/16 \\
  -r & +t & 6/16 \\
  -r & -t & 6/16 \\
\end{array}
\]
Causality?

When Bayes' nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal:
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation

What do the arrows really mean?
- Topology may happen to encode causal structure
- Topology really encodes conditional independence
Causality?

When Bayes’ nets reflect the true causal patterns:

- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

Bayes’ nets need not actually be causal:
- Sometimes no causal net exists over the domain (especially if variables are missing)

E.g. consider the variables Traffic and Drips:
- End up with arrows that reflect correlation, not causation

What do the arrows really mean?
- Topology may happen to encode causal structure
- Topology really encodes conditional independence
When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
When Bayes’ nets reflect the true causal patterns:

- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts
When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation

\[ P(x_i|x_1, \ldots, x_{i-1}) = P(x_i|\text{parents}(X_i)) \]
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation

\[ P(x_i|x_1, \ldots, x_{i-1}) = P(x_i|\text{parents}(X_i)) \]

What do the arrows really mean?
- Topology may happen to encode causal structure
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation

$$P(x_i|x_1, \ldots, x_{i-1}) = P(x_i|\text{parents}(X_i))$$

What do the arrows really mean?
- Topology may happen to encode causal structure
- Topology really encodes conditional independence
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation

\[ P(x_i|x_1,\ldots,x_{i-1}) = P(x_i|\text{parents}(X_i)) \]

What do the arrows really mean?
- Topology may happen to encode causal structure
- **Topology really encodes conditional independence**
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

First assembled BNs using an intuitive notion of conditional independence as causality.

Then saw that key property is conditional independence.

Main goal: answer queries about conditional independence and influence.

After that: how to answer numerical queries (inference).
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
  - First assembled BNs using an intuitive notion of conditional independence as causality.
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
- First assembled BNs using an intuitive notion of conditional independence as causality.
- Then saw that key property is conditional independence
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
- First assembled BNs using an intuitive notion of conditional independence as causality.
- Then saw that key property is conditional independence
- Main goal: answer queries about conditional independence and influence
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
  - First assembled BNs using an intuitive notion of conditional independence as causality.
  - Then saw that key property is conditional independence
  - Main goal: answer queries about conditional independence and influence
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
- First assembled BNs using an intuitive notion of conditional independence as causality.
- Then saw that key property is conditional independence
- Main goal: answer queries about conditional independence and influence

After that: how to answer numerical queries (inference)
Bayes’ Nets

A Bayes’ net is an efficient encoding of a probabilistic model of a domain.

Questions we can ask:
- Inference: given a fixed BN, what is $P(X|e)$?
- Representation: given a BN graph, what kinds of distributions can it encode?
- Modeling: what BN is most appropriate for a given domain?
Bayes’ Nets

A Bayes’ net is an efficient encoding of a probabilistic model of a domain.
Bayes’ Nets

A Bayes’ net is an efficient encoding of a probabilistic model of a domain

Questions we can ask:

- Inference: given a fixed BN, what is \( P(X | e) \)?
A Bayes’ net is an efficient encoding of a probabilistic model of a domain.

Questions we can ask:
- Inference: given a fixed BN, what is $P(X \mid e)$?
- Representation: given a BN graph, what kinds of distributions can it encode?
Bayes’ Nets

A Bayes’ net is an efficient encoding of a probabilistic model of a domain

Questions we can ask:
- Inference: given a fixed BN, what is $P(X \mid e)$?
- Representation: given a BN graph, what kinds of distributions can it encode?
- Modeling: what BN is most appropriate for a given domain?
A Bayes’ net is an efficient encoding of a probabilistic model of a domain.

Questions we can ask:

- Inference: given a fixed BN, what is $P(X \mid e)$?
- Representation: given a BN graph, what kinds of distributions can it encode?
- Modeling: what BN is most appropriate for a given domain?
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable
A conditional probability table (CPT) for each node
A collection of distributions over $X$, one for each combination of parents’ values

$P(x_1, \ldots, x_n) = \prod_i P(x_i | \text{parents}(X_i))$.
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable

A conditional probability table (CPT) for each node

- A collection of distributions over X, one for each combination of parents’ values
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable

A conditional probability table (CPT) for each node

- A collection of distributions over X, one for each combination of parents’ values
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable

A conditional probability table (CPT) for each node

- A collection of distributions over X, one for each combination of parents’ values

\[ P(x_1 | a_1, \ldots, a_n). \]
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable

A conditional probability table (CPT) for each node

- A collection of distributions over $X$, one for each combination of parents’ values

\[ P(x_1 | a_1, \ldots, a_n). \]
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable

A conditional probability table (CPT) for each node

- A collection of distributions over $X$, one for each combination of parents’ values

$$P(x_1 | a_1, \ldots, a_n).$$

Bayes’ nets implicitly encode joint distributions
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable

A conditional probability table (CPT) for each node

- A collection of distributions over $X$, one for each combination of parents’ values

$$P(x_1|a_1,\ldots,a_n).$$

Bayes’ nets implicitly encode joint distributions

- As a product of local conditional distributions
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable

A conditional probability table (CPT) for each node
- A collection of distributions over $X$, one for each combination of parents’ values

$$P(x_1|a_1, \ldots, a_n).$$

Bayes’ nets implicitly encode joint distributions
- As a product of local conditional distributions
- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable

A conditional probability table (CPT) for each node

- A collection of distributions over $X$, one for each combination of parents’ values

$$P(x_1 | a_1, \ldots, a_n).$$

Bayes’ nets implicitly encode joint distributions

- As a product of local conditional distributions
- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable
A conditional probability table (CPT) for each node
- A collection of distributions over $X$, one for each combination of parents’ values

$$P(x_1|a_1,\ldots,a_n).$$

Bayes’ nets implicitly encode joint distributions
- As a product of local conditional distributions
- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1,\ldots,x_n) = \prod_i P(x_i|\text{parents}(X_i)).$$
Bayes’ Net Semantics

A directed, acyclic graph, one node per random variable

A conditional probability table (CPT) for each node
- A collection of distributions over X, one for each combination of parents’ values

\[ P(x_1|a_1,\ldots,a_n). \]

Bayes’ nets implicitly encode joint distributions
- As a product of local conditional distributions
- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

\[ P(x_1,\ldots,x_n) = \prod_i P(x_i|\text{parents}(X_i)), \]
Example: Alarm Network

\[
P(+b, -e, +a, -j, +m) = P(+b) P(-e) P(+a | +b, -e) P(-j | +a) P(+m | +a)
\]

\[
= 0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7
\]
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>( P(B) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

\[
P(+b, -e, +a, -j, +m) = P(+b) P(-e) P(+a | +b, -e) P(-j | +a) P(+m | +a)
\]

\[
= 0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7
\]
Example: Alarm Network

<table>
<thead>
<tr>
<th></th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>+b: 0.001, -b: 0.999</td>
</tr>
</tbody>
</table>
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

\[
P(+b, -e, +a, -j, +m) = P(+b) \times P(-e) \times P(+a | +b, -e) \times P(-j | +a) \times P(+m | +a) = -0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7
\]
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| B   | E   | A   | P(A|B,E) |
|-----|-----|-----|---------|
| +b  | +e  | +a  | 0.95    |
| +b  | +e  | -a  | 0.05    |
| +b  | -e  | +a  | 0.94    |
| +b  | -e  | -a  | 0.06    |
| -b  | +e  | +a  | 0.29    |
| -b  | +e  | -a  | 0.71    |
| -b  | -e  | +a  | 0.001   |
| -b  | -e  | -a  | 0.999   |
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

| A  | J  | P(J|A) |
|----|----|-------|
| +a | +j | 0.9   |
| +a | -j | 0.1   |
| -a | +j | 0.05  |
| -a | -j | 0.95  |

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| B  | E  | A  | P(A|B,E) |
|----|----|----|---------|
| +b | +e | +a | 0.95    |
| +b | +e | -a | 0.05    |
| +b | -e | +a | 0.94    |
| +b | -e | -a | 0.06    |
| -b | +e | +a | 0.29    |
| -b | +e | -a | 0.71    |
| -b | -e | +a | 0.001   |
| -b | -e | -a | 0.999   |
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

| A | J | P(J|A) |
|---|---|------|
| +a | +j | 0.9  |
| +a | -j | 0.1  |
| -a | +j | 0.05 |
| -a | -j | 0.95 |

| A | M | P(M|A) |
|---|---|------|
| +a | +m | 0.7  |
| +a | -m | 0.3  |
| -a | +m | 0.01 |
| -a | -m | 0.99 |

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| B | E | A | P(A|B,E) |
|---|---|---|---------|
| +b | +e | +a | 0.95    |
| +b | +e | -a | 0.05    |
| +b | -e | +a | 0.94    |
| +b | -e | -a | 0.06    |
| -b | +e | +a | 0.29    |
| -b | +e | -a | 0.71    |
| -b | -e | +a | 0.001   |
| -b | -e | -a | 0.999   |

\[ P(+b,-e,+a,-j,+m) = \]
Example: Alarm Network

\[
P(+b, -e, +a, -j, +m) = P(+b) \cdot P(-e) \cdot P(+a | +b, -e) \cdot P(-j | +a) \cdot P(+m | +a)
\]

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

\[
\begin{array}{c|c}
  A & J & P(J|A) \\
  +a & +j & 0.9 \\
  +a & -j & 0.1 \\
  -a & +j & 0.05 \\
  -a & -j & 0.95 \\
\end{array}
\]

\[
\begin{array}{c|c|c|c}
  A & M & P(M|A) \\
  +a & +m & 0.7 \\
  +a & -m & 0.3 \\
  -a & +m & 0.01 \\
  -a & -m & 0.99 \\
\end{array}
\]

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| B   | E   | A    | P(A|B,E) |
|-----|-----|------|----------|
| +b  | +e  | +a   | 0.95     |
| +b  | +e  | -a   | 0.05     |
| +b  | -e  | +a   | 0.94     |
| +b  | -e  | -a   | 0.06     |
| -b  | +e  | +a   | 0.29     |
| -b  | +e  | -a   | 0.71     |
| -b  | -e  | +a   | 0.001    |
| -b  | -e  | -a   | 0.999    |
Example: Alarm Network

\[
P(+b, -e, +a, -j, +m) = P(+b) \cdot P(-e) \cdot P(+a | +b, -e) \cdot P(-j | +a) \cdot P(+m | +a)
\]
\[
= 0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7
\]
Size of a Bayes’ Net

How big is a joint distribution over $N$ Boolean variables?

$2^N$

How big is an $N$-node net if nodes have up to $k$ parents?

$O(N \times 2^k + 1)$

Both give you the power to calculate $P(X_1, X_2, \cdots, X_N)$.

BNs: Huge space savings!

Also easier to elicit local CPTs

Also faster to answer queries

(coming)
Size of a Bayes’ Net

How big is a joint distribution over N Boolean variables?

- $2^N$
Size of a Bayes’ Net

How big is a joint distribution over $N$ Boolean variables?

- $2^N$
Size of a Bayes’ Net

How big is a joint distribution over N Boolean variables?
- $2^N$

How big is an N-node net if nodes have up to k parents?
- $O(N \times 2^{k+1})$
Size of a Bayes’ Net

How big is a joint distribution over N Boolean variables?
- $2^N$

How big is an N-node net if nodes have up to k parents?
- $O(N \times 2^{k+1})$
Size of a Bayes’ Net

How big is a joint distribution over N Boolean variables?
- $2^N$

How big is an N-node net if nodes have up to k parents?
- $O(N \times 2^{k+1})$

Both give you the power to calculate
Size of a Bayes’ Net

How big is a joint distribution over N Boolean variables?
- $2^N$

How big is an N-node net if nodes have up to k parents?
- $O(N \times 2^{k+1})$

Both give you the power to calculate

$$P(X_1, X_2, \ldots, X_N)$$
Size of a Bayes’ Net

How big is a joint distribution over N Boolean variables?
- $2^N$

How big is an N-node net if nodes have up to k parents?
- $O(N \times 2^{k+1})$

Both give you the power to calculate

$$P(X_1, X_2, \ldots, X_N)$$

BNs: Huge space savings!
Size of a Bayes’ Net

How big is a joint distribution over N Boolean variables?

- $2^N$

How big is an N-node net if nodes have up to k parents?

- $O(N \times 2^{k+1})$

Both give you the power to calculate

$$P(X_1, X_2, \ldots, X_N)$$

BNs: Huge space savings!
Also easier to elicit local CPTs
Size of a Bayes’ Net

How big is a joint distribution over $N$ Boolean variables?
- $2^N$

How big is an $N$-node net if nodes have up to $k$ parents?
- $O(N \times 2^{k+1})$

Both give you the power to calculate

$$P(X_1, X_2, \ldots, X_N)$$

BNs: Huge space savings!
Also easier to elicit local CPTs
Also faster to answer queries (coming)
Bayes’ Nets

Representation. ✓
Bayes’ Nets

Representation. ✓

Conditional Independences (Next.)
Bayes’ Nets

Representation. ✓

Conditional Independences (Next.)

Probabilistic Inference
Bayes’ Nets

Representation.

Conditional Independences (Next.)

Probabilistic Inference

Learning Bayes’ Nets from Data