CSE 473: Artificial Intelligence
Spring 2012

Bayesian Networks

Dan Weld

Many slides adapted from Dan Klein, Stuart Russell, Andrew Moore & Luke Zettlemoyer

Outline

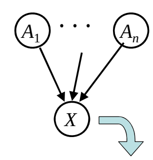
- Probabilistic models (and inference)
  - Bayesian Networks (BNs)
  - Independence in BNs
  - Efficient Inference in BNs
  - Learning

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' Net Semantics

Formally:



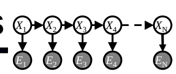
$P(X|A_1 \ldots A_n)$

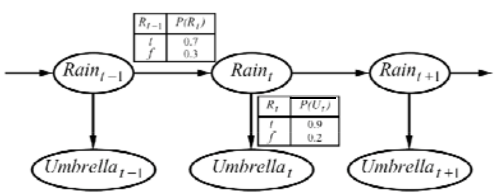
- A set of nodes, one per variable X
- A directed, acyclic graph
- A CPT for each node
  - CPT = "Conditional Probability Table"
  - Collection of distributions over X, one for each combination of parents' values

$P(X|a_1 \ldots a_n)$

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

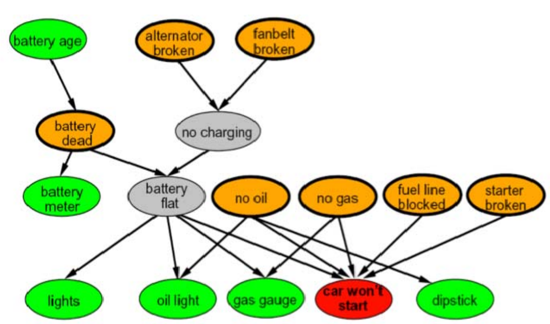
Hidden Markov Models





- An HMM is defined by:
  - Initial distribution: $P(X_1)$
  - Transitions: $P(X_t|X_{t-1})$
  - Emissions: $P(E|X)$

Example Bayes' Net: Car



Probabilities in BNs

- 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:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i))$$

- This lets us reconstruct any entry of the full joint
- Not every BN can represent every joint distribution
  - The topology enforces certain *independence* assumptions
  - Compare to the exact decomposition according to the chain rule!

Example: Independence

- N fair, independent coin flips:

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

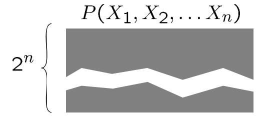
$\cdots$

$P(X_n)$

h	0.5
t	0.5

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

2^n

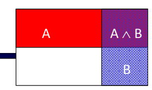


Example: Coin Flips

- N independent coin flips

- No interactions between variables: absolute independence

Independence

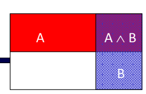


- Two variables are *independent* if:

$$\forall x, y : P(x, y) = \quad ?$$

  - This says that their joint distribution *factors* into a product of two simpler distributions
  - We write: $X \perp\!\!\!\perp Y$

Independence



- Two variables are *independent* if:

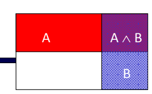
$$\forall x, y : P(x, y) = P(x)P(y)$$

  - This says that their joint distribution *factors* into a product of two simpler distributions
  - Another form:

$$\forall x, y : P(x|y) = \quad ?$$

  - We write: $X \perp\!\!\!\perp Y$

Independence



- Two variables are *independent* if:

$$\forall x, y : P(x, y) = P(x)P(y)$$

  - This says that their joint distribution *factors* into a product two simpler distributions
  - Another form:

$$\forall x, y : P(x|y) = P(x)$$

  - We write: $X \perp\!\!\!\perp Y$
- Independence is a simplifying *modeling assumption*
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity, Toothache}?

Example: Independence?

$P_1(T,W)$

T	W	P
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P(T)$

T	P
warm	0.25
cold	0.75

$P(W)$

W	P
sun	0.6
rain	0.4

$P_2(T,W)$

T	W	
warm	sun	0.15
warm	rain	0.10
cold	sun	0.45
cold	rain	0.30

Conditional Independence

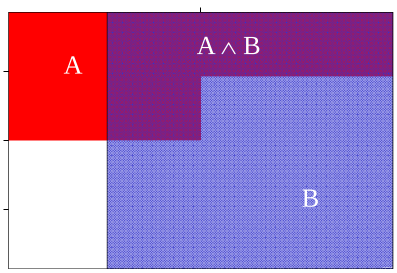
- Unconditional (absolute) independence very rare (why?)
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments:

$$\forall x,y,z : P(x,y|z) = P(x|z)P(y|z)$$
$$\forall x,y,z : P(x|z,y) = P(x|z)$$
$$X \perp\!\!\!\perp Y | Z$$

- What about fire, smoke, alarm?

Conditional Independence

Are A & B independent? P(A|B) ≤ P(A)



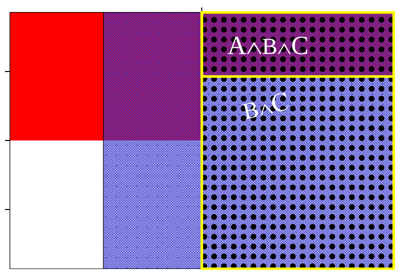
P(A)=(.25+.5)/2
= .375

P(B)= .75

P(A|B)=(.25+.25+.5)/3
=.3333

A, B Conditionally Independent Given C

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



P(A|C) =.25
P(A|B,C)=.25

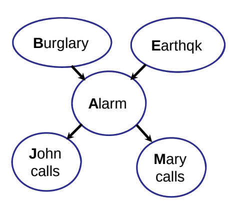
Example: Alarm Network

- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!
- How big is joint distribution?
  - 2^n-1 = 31 parameters

Example: Alarm Network

Only 10 params

B	P(B)
+b	0.001
←b	0.999



E	P(E)
+e	0.002
←e	0.998

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	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

Example: Traffic II

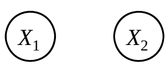
- Let's build a graphical model
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

Changing Bayes' Net Structure

- The same joint distribution can be encoded in many different Bayes' nets
- Analysis question: given some edges, what other edges do you need to add?
  - One answer: fully connect the graph
  - Better answer: don't make any false conditional independence assumptions

Example: Independence

- For this graph, you can fiddle with (the CPTs) all you want, but you won't be able to represent any distribution in which the flips are dependent!

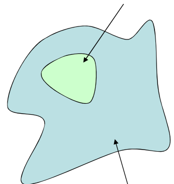


$P(X_1)$

h	0.5
t	0.5

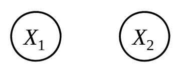
$P(X_2)$

h	0.5
t	0.5



Example: Coins

- Extra arcs don't prevent representing independence, just allow non-independence

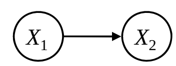


$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5



$P(X_1)$

h	0.5
t	0.5

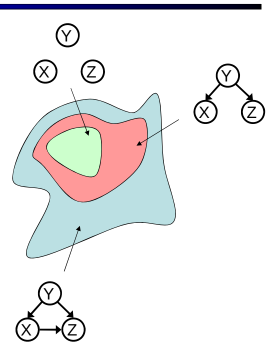
$P(X_2|X_1)$

h | h	0.5
t | h	0.5
h | t	0.5
t | t	0.5

- Adding unneeded arcs isn't wrong, it's just inefficient

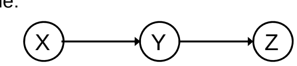
Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- The graph structure guarantees certain (conditional) independences
- (There might be more independence)
- Adding arcs increases the set of distributions, but has several costs
- Full conditioning can encode any distribution



Independence in a BN

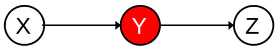
- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:



- Question: are X and Z independent?
  - Answer: no.
    - Example: low pressure causes rain, which causes traffic.
  - Knowledge about X may change belief in Z,
  - Knowledge about Z may change belief in X (via Y)
  - Addendum: they *could* be independent: how?

Causal Chains

- This configuration is a "causal chain"



X: Low pressure
Y: Rain
Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Is X independent of Z given Y?

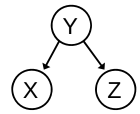
$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)}$$

$$= P(z|y)$$ *Yes!*

Evidence along the chain "blocks" the influence

Common Parent

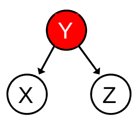
- Another basic configuration: two effects of the same parent
  - Are X and Z independent?



Y: Project due
X: Forum busy
Z: Lab full

Common Parent

- Another basic configuration: two effects of the same parent
  - Are X and Z independent?
  - Are X and Z independent given Y?



$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

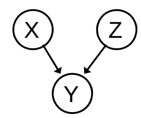
$$= P(z|y)$$ *Yes!*

Y: Project due
X: Forum busy
Z: Lab full

- Observing the cause blocks influence between effects.

Common Effect

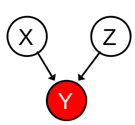
- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - Yes: the ballgame and the rain cause traffic, but they are not correlated
    - Still need to prove they must be (try it!)



X: Raining
Z: Ballgame
Y: Traffic

Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - Yes: the ballgame and the rain cause traffic, but they are not correlated
    - Still need to prove they must be (try it!)
  - Are X and Z independent given Y?
    - ***No:*** seeing traffic puts the rain and the ballgame in competition as explanation!
  - This is backwards from the other cases
    - Observing an effect ***activates*** influence between possible causes.



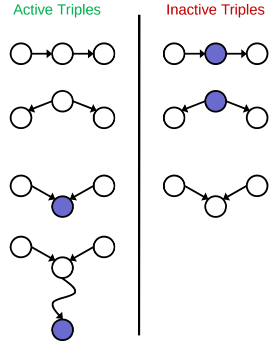
X: Raining
Z: Ballgame
Y: Traffic

The General Case

- Any complex example can be analyzed using these three canonical cases
- General question: in a given BN, are two variables independent (given evidence)?
- Solution: analyze the graph

Reachability (D-Separation)

- Question: Are X and Y conditionally independent given evidence vars {Z}?
  - Yes, if X and Y "separated" by Z
  - Look for active paths from X to Y
  - No active paths = independence!
- A path is active if each triple is active:
  - Causal chain A → B → C where B is **unobserved** (either direction)
  - Common cause A ← B → C where B is **unobserved**
  - Common effect (aka v-structure) A → B ← C where B *or one of its descendents* is **observed**
- All it takes to block a path is a single inactive segment

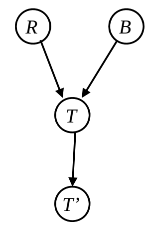


Example: Independent?

$R \perp\!\!\!\perp B$ *Yes*

$R \perp\!\!\!\perp B | T$

$R \perp\!\!\!\perp B | T'$



Example: Independent?

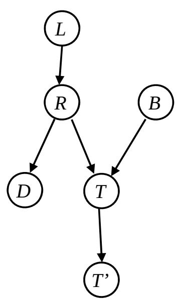
$L \perp\!\!\!\perp T' | T$ *Yes*

$L \perp\!\!\!\perp B$ *Yes*

$L \perp\!\!\!\perp B | T$

$L \perp\!\!\!\perp B | T'$

$L \perp\!\!\!\perp B | T, R$ *Yes*



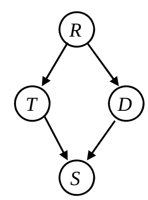
Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I'm sad
- Questions:

  $T \perp\!\!\!\perp D$

  $T \perp\!\!\!\perp D | R$ *Yes*

  $T \perp\!\!\!\perp D | R, S$



Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence guarantees from graph alone
- A Bayes' net's joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution