BIN 504
Probabilistic and Statistical Modeling for Bioinformatics
Fall 2010-2011

Week 13
Bayesian Networks
(slides adapted from Matthew Turk’s Statistical Models and Methods in Computer Science course)
Preliminaries

• Bayes rule
• Marginalization
• Chain rule
• Independence
• Conditional independence
Joint Probability Distribution

- Random variables encode attributes
- Not all possible combinations of attributes are equally likely
  - Joint probability distributions quantify this
- $P(X = x, Y = y) = P(x, y)$
  - How probable is it to observe these two attributes together?
- Generalizes to N-RVs
- How can we manipulate joint probability distributions?
The Chain Rule

\[
P(X_1, X_2, \ldots, X_N) = P(X_1 | X_2, \ldots, X_N) \ P(X_2, \ldots, X_N)
\]

\[
P(X_2, \ldots, X_N) = P(X_2 | X_3, \ldots, X_N) \ P(X_3, \ldots, X_N)
\]

\[
P(X_3, \ldots, X_N) = P(X_3 | X_4, \ldots, X_N) \ P(X_4, \ldots, X_N)
\]

\[ \vdots \]

\[
P(X_{N-1}, X_N) = P(X_{N-1} | X_N) \ P(X_N)
\]

\[
P(X_1, X_2, \ldots, X_N) = P(X_1 | X_2, \ldots, X_N) \ P(X_2 | X_3, \ldots, X_N) \ P(X_3 | X_4, \ldots, X_N) \ldots P(X_{N-1} | X_N) \ P(X_N)
\]
Chain Rule

- Always true
  - \( P(x, y, z) = p(x) \ p(y | x) \ p(z | x, y) \)
    
    \[
    = p(z) \ p(y | z) \ p(x | y, z)
    \]
    
    \( =... \)
Conditional Probability

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

But we will always write it this way:

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

- We know $p(X,Y)$, what is $P(X=x)$?
- We can use the law of total probability.

\[
p(x) = \sum_y P(x, y) = \sum_y P(y)P(x \mid y)
\]
Marginalization Cont.

• Another example

\[
p(x) = \sum_{y, z} P(x, y, z) = \sum_{y, z} P(y, z)P(x \mid y, z)
\]
Bayes Rule

• We know that $P(\text{smart}) = .7$

• If we also know that the students grade is A+, then how this affects our belief about his intelligence?

$$P(x \mid y) = \frac{P(x)P(y \mid x)}{P(y)}$$
Bayes Rule cont.

- You can condition on more variables

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

• X is independent of Y means that knowing Y does not change our belief about X.
  • \( P(X \mid Y=y) = P(X) \)
  • \( P(X=x, Y=y) = P(X=x) \cdot P(Y=y) \)
    • Why this is true?
• The above should hold for all \( x, y \)
• It is symmetric and written as \( X \perp Y \)
CI: Conditional Independence

• RV are rarely independent but we can still leverage local structural properties like CI.

• $X \perp Y \mid Z$ if once $Z$ is observed, knowing the value of $Y$ does not change our belief about $X$
  • The following should hold for all $x,y,z$
    • $P(X=x \mid Z=z, Y=y) = P(X=x \mid Z=z)$
    • $P(Y=y \mid Z=z, X=x) = P(Y=y \mid Z=z)$
    • $P(X=x, Y=y \mid Z=z) = P(X=x \mid Z=z) \cdot P(Y=y \mid Z=z)$

We call these factors: very useful concept!!
Monty Hall Problem

- You're given the choice of three doors: Behind one door is a car; behind the others, goats.
- You pick a door, say No. 1
- The host, who knows what's behind the doors, opens another door, say No. 3, which has a goat.
- Do you want to pick door No. 2 instead?
Host must reveal Goat A or
Host reveals Goat B

Host must reveal Goat B

Host must reveal Goat A
Monty Hall Problem: Bayes Rule

- $C_i$ : the car is behind door $i$, $i = 1, 2, 3$
- $P(C_i) = 1/3$
- $H_{ij}$ : the host opens door $j$ after you pick door $i$

$$P(H_{ij} | C_k) = \begin{cases} 
0 & i = j \\
0 & j = k \\
1/2 & i = k \\
1 & i \neq k, j \neq k
\end{cases}$$
Monty Hall Problem: Bayes Rule cont.

- WLOG, $i=1, j=3$

- $P(C_1|H_{13}) = \frac{P(H_{13}|C_1)P(C_1)}{P(H_{13})}$

- $P(H_{13}|C_1)P(C_1) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
Monty Hall Problem: Bayes Rule cont.

- \( P(H_{13}) = P(H_{13}, C_1) + P(H_{13}, C_2) + P(H_{13}, C_3) \)
  
  \[ = P(H_{13} | C_1) P(C_1) + P(H_{13} | C_2) P(C_2) \]
  
  \[ = \frac{1}{6} + 1 \cdot \frac{1}{3} \]
  
  \[ = \frac{1}{2} \]

- \( P(C_1 | H_{13}) = \frac{1/6}{1/2} = \frac{1}{3} \)
Belief Networks (a.k.a. Bayesian Networks)

a.k.a. Probabilistic networks, Belief nets, Bayes nets, etc.

- **Belief network**
  - A data structure (depicted as a graph) that represents the dependence among variables and allows us to concisely specify the joint probability distribution
  - The graph itself is known as an “influence diagram”

- **A belief network is a directed acyclic graph where:**
  - The nodes represent the set of random variables (one node per random variable)
  - Arcs between nodes represent influence, or causality
    - A link from node X to node Y means that X “directly influences” Y
  - Each node has a *conditional probability table* (CPT) that defines $P(\text{node} \mid \text{parents})$
Example

- Random variables $X$ and $Y$
  - $X$ – It is raining
  - $Y$ – The grass is wet
- $X$ has a *causal effect* on $Y$
- Or, $Y$ is a *symptom* of $X$
- Draw two nodes and link them
- Define the CPT for each node
  - $P(X)$ and $P(Y|X)$

- Typical use: we observe $Y$ and we want to query $P(X|Y)$
  - $Y$ is an *evidence variable*
  - $X$ is a *query variable*
Try it…

- What is $P(X \mid Y)$?
  - Given that we know the CPTs of each node in the graph

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}$$

$$= \frac{P(Y \mid X)P(X)}{\sum_Y P(X, Y)}$$

$$= \frac{P(Y \mid X)P(X)}{\sum_Y P(Y \mid X)P(X)}$$
Conditional Independence

- Can we determine conditional independence of variables directly from the graph?
- A set of nodes $X$ is independent of another set of nodes $Y$, given a set of (evidence) nodes $E$, if every path from $X$ to $Y$ is d-separated, or blocked, by $E$

\[
P(X \mid Y, E) = P(X \mid E) \quad P(Y \mid X, E) = P(Y \mid E)
\]

The set of nodes $E$ d-separates sets $X$ and $Y$

3 ways to block paths from $X$ to $Y$, given $E$
X and Y are independent, given E

- The variable Z is in E
- This variable is not in E! (Nor are its descendants)
Examples

\[ P(W \mid R, G) = P(W \mid G) \]

\[ P(T \mid C, F) = P(T \mid F) \]

\[ P(W \mid I, M) \neq P(W \mid M) \]
\[ P(W \mid I) = P(W) \]
Examples

```
X  Z  Y
X ind. of Y?  X ind. of Y given Z?
Yes  Yes

X → Z → Y
X ind. of Y?  X ind. of Y given Z?
No  Yes

X ← Z → Y
X ind. of Y?  X ind. of Y given Z?
No  Yes

X → Z ← Y
X ind. of Y?  X ind. of Y given Z?
Yes  No

X → Z → Y
X ind. of Y?  X ind. of Y given Z?
No  No
```
Examples (cont.)

X – wet grass
Y – rainbow
Z – rain

\[ P(X, Y) \neq P(X) P(Y) \]
\[ P(X | Y, Z) = P(X, Z) \]

Are X and Y ind.? Are X and Y cond. ind. given…?

X – rain
Y – sprinkler
Z – wet grass
W – worms

\[ P(X, Y) = P(X) P(Y) \]
\[ P(X | Y, Z) \neq P(X | Z) \]
\[ P(X | Y, W) \neq P(X | W) \]
Example: Flu and measles

To create the belief net:
- Choose variables (evidence and query)
- Choose an ordering and create links (direct influences)
- Fill in probabilities
Example: Flu and measles

P(Flu) = 0.01
P(Measles) = 0.001
P(Flu) = 0.01
P(Spots | Measles) = [0, 0.9]

P(Fever | Flu, Measles) = [0.01, 0.8, 0.9, 1.0]

Compute P(Flu | Fever) and P(Flu | Fever, Spots).
Are they equivalent?
BN example #1

- A patient comes in with a stiff neck. What is the probability that he has meningitis? Whiplash?
- We know that:
  - Meningitis causes a stiff neck 50% of the time
  - Whiplash causes a stiff neck 80% of the time
  - The prior probability of a patient having meningitis is 1/50,000
  - The prior probability of a patient having a stiff neck is 1/20
  - The prior probability of a patient having whiplash is 1/1000
BN example #2

- A test for a new, deadly strain of anthrax (that has no symptoms) is known to be 99.9% accurate. Should you get tested?
- We know that:
  - One out of a million people in the population has this anthrax

- What if you work in the post office?
- What if you get tested a second time? A third time?
Belief nets represent the joint probability

- The joint probability function can be calculated directly from the network
  - It’s the product of the CPTs of all the nodes
  - \( P(\text{var}_1, \ldots, \text{var}_N) = \prod_i P(\text{var}_i|\text{Parents}(\text{var}_i)) \)

\[
\begin{align*}
P(X, Y) &= P(X) \cdot P(Y|X) \\
P(X, Y, Z) &= P(X) \cdot P(Y) \cdot P(Z|X, Y)
\end{align*}
\]
Example

I’m at work and my neighbor John calls to say my home alarm is ringing, but my neighbor Mary doesn’t call. The alarm is sometimes triggered by minor earthquakes. Was there a burglar at my house?

• Random (boolean) variables:
  – JohnCalls, MaryCalls, Earthquake, Burglar, Alarm

• The belief net shows the causal links

• This defines the joint probability
  – \( P(\text{JohnCalls, MaryCalls, Earthquake, Burglar, Alarm}) \)

• What do we want to know?

\[
P(B \mid J, \neg M)
\]
Example

Burglary

Earthquake

Alarm

JohnCalls

MaryCalls

Links and CPTs?
Example

Joint probability? \( P(J, \neg M, A, B, \neg E) \)?
Calculate $P(J, \neg M, A, B, \neg E)$

Read the joint pf from the graph:

$P(J, M, A, B, E) = P(B) \cdot P(E) \cdot P(A|B,E) \cdot P(J|A) \cdot P(M|A)$

Plug in the desired values:

$P(J, \neg M, A, B, \neg E) = P(B) \cdot P(\neg E) \cdot P(A|B,\neg E) \cdot P(J|A) \cdot P(\neg M|A)$

$= 0.001 \cdot 0.998 \cdot 0.94 \cdot 0.9 \cdot 0.3$

$= 0.0002532924$

How about $P(B \mid J, \neg M)$?

Remember, this means $P(B=\text{true} \mid J=\text{true}, M=\text{false})$
Calculate $P(B \mid J, \neg M)$

$$P(B \mid J, \neg M) = \frac{P(B, J, \neg M)}{P(J, \neg M)}$$

By marginalization:

$$\sum_i \sum_j P(J, \neg M, A_i, B, E_j)$$

$$= \sum_i \sum_j \sum_k P(J, \neg M, A_i, B_j, E_k)$$

$$= \sum_i \sum_j P(B)P(E_j)P(A_i \mid B, E_j)P(J \mid A_i)P(\neg M \mid A_i)$$

$$= \sum_i \sum_j \sum_k P(B_j)P(E_k)P(A_i \mid B_j, E_k)P(J \mid A_i)P(\neg M \mid A_i)$$
Example

- Conditional independence is seen here
  - \( P(\text{JohnCalls} | \text{MaryCalls}, \text{Alarm}, \text{Earthquake}, \text{Burglary}) = P(\text{JohnCalls} | \text{Alarm}) \)
  - So JohnCalls is independent of MaryCalls, Earthquake, and Burglary, given Alarm

- Does this mean that an earthquake or a burglary do not influence whether or not John calls?
  - No, but the influence is already accounted for in the Alarm variable
  - JohnCalls is \textit{conditionally} independent of Earthquake, but not \textit{absolutely} independent of it
Drawing belief nets

- What would a belief net look like if all the variables were fully dependent?

\[ P(X_1, X_2, X_3, X_4, X_5) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3)P(X_5|X_1, X_2, X_3, X_4) \]

- But this isn’t the only way to draw the belief net when all the variables are fully dependent.
Fully connected belief net

- In fact, there are $N!$ ways of connecting up a fully-connected belief net
  - That is, there are $N!$ ways of ordering the nodes

For $N=2$

$$X_1 \rightarrow X_2$$
$$X_1 \leftarrow X_2 \quad P(X_1,X_2) = ?$$

For $N=5$

$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_5$$
$$P(X_1,X_2,X_3,X_4,X_5) = ?$$

and 119 others…
Drawing belief nets (cont.)

Fully-connected net displays the joint distribution

\[ P(X_1, X_2, X_3, X_4, X_5) = P(X_1) P(X_2|X_1) P(X_3|X_1,X_2) P(X_4|X_1,X_2,X_3) P(X_5|X_1, X_2, X_3, X_4) \]

But what if there are conditionally independent variables?

\[ P(X_1, X_2, X_3, X_4, X_5) = P(X_1) P(X_2|X_1) P(X_3|X_1,X_2) P(X_4|X_2,X_3) P(X_5|X_3, X_4) \]
Drawing belief nets (cont.)

What if the variables are all independent?
\[ P(X_1, X_2, X_3, X_4, X_5) = P(X_1) P(X_2) P(X_3) P(X_4) P(X_5) \]

What if the links are drawn like this:

Not allowed – not a DAG
Drawing belief nets (cont.)

What if the links are drawn like this:

\[ P(X_1, X_2, X_3, X_4, X_5) = P(X_1) P(X_2 | X_3) P(X_3 | X_1) P(X_4 | X_2) P(X_5 | X_4) \]

It can be redrawn like this:

All arrows going left-to-right
How to construct a belief net

- Choose the random variables that describe the domain
  - These will be the nodes of the graph

- Choose a left-to-right ordering of the variables that indicates a general order of influence
  - “Root causes” to the left, symptoms to the right

\[ X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5 \]

Causes \quad \rightarrow \quad \text{Symptoms}
How to construct a belief net (cont.)

- Draw arcs from left to right to indicate “direct influence” (causality) among variables
  - May have to reorder some nodes

![Belief Net Diagram]

- Define the conditional probability table (CPT) for each node
  - $P(\text{node} \mid \text{parents})$
    
    $P(X_1)$  $P(X_4 \mid X_2, X_3)$
    $P(X_2)$  $P(X_5 \mid X_4)$
    $P(X_3 \mid X_1, X_2)$
How to construct a belief net (cont.)

- To calculate *any* probability, use (1) definition of conditional probability and (2) marginalization
  - \( P(\text{red vars} \mid \text{green vars}) = ? \)

\[
P(\{r\} \mid \{g\}) = \frac{P(\{r\}, \{g\})}{P(\{g\})}
\]

\[
= \frac{P(\{r\}, \{g\})}{\sum_{i} P(\{r\}, \{g\})}
\]

where \( P(\{r\}, \{g\}) = \prod P(n_i \mid \text{parents}(n_i)) \)
Belief nets

• General assumptions
  – A DAG is a reasonable representation of the influences among the variables
    ♦ Leaves of the DAG have no direct influence on other variables
  – Conditional independences cause the graph to be much less than fully connected (the system is *locally structured*, or *sparse*)
  – The CPTs are relatively easy to state
    ♦ Many can be estimated as a *canonical distribution* (a standard pattern – just specify the parameters) or as a deterministic node (direct function – logical or numerical combination – of parents)
What are belief nets for?

- Given the structure, we can now pose queries

- Typically: \( P(\text{Query} \mid \text{Evidence}) \) or \( P(\text{Cause} \mid \text{Symptoms}) \)
  - \( P(X_1 \mid X_4, X_5) \)
  - \( P(\text{Earthquake} \mid \text{JohnCalls}) \)
  - \( P(\text{Burglary} \mid \text{JohnCalls}, \neg \text{MaryCalls}) \)