Independence

- Propositions *a* and *b* are independent if and only if $P(a \wedge b) = P(a)\,P(b)$
- Equivalently (by product rule): $P(a|b) = P(a)$
- Equivalently: $P(b|a) = P(b)$

Illustration of Independence

- We know (product rule) that

$P(toothache, catch, cavity, Weather = cloudy)$ $=$ $P(Weather = cloudy|toothache, catch, cavity)$ $\times$ $P(toothache, catch, cavity)$. By independence: $P(Weather = cloudy|toothache, catch, cavity)$ $=$ $P(Weather = cloudy)$. Therefore we have that $P(toothache, catch, cavity, Weather = cloudy)$ $=$ $P(Weather = cloudy) \times P(toothache, catch, cavity)$.

Illustration continued

- Allows us to represent a 32-element table for full joint on *Weather*, *Toothache*, *Catch*, *Cavity* by an 8-element table for the joint of *Toothache*, *Catch*, *Cavity*, and a 4-element table for *Weather*.
- If we add a Boolean variable X to the 8-element table, we get 16 elements. A new 2-element table suffices with independence.

Bayes' Rule

Recall product rule:

$$P(a \wedge b) = P(a|b)\,P(b)$$

$$P(a \wedge b) = P(b|a)\,P(a)$$

Equating right-hand sides and dividing by $P(a)$:

$$P(b|a) = \frac{P(a|b)\,P(b)}{P(a)}$$

For multi-valued variables X and Y:

$$\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\,\mathbf{P}(Y)}{\mathbf{P}(X)}$$

Bayes' Rule with Background Evidence

Often we'll want to use Bayes' Rule conditionalized on some background evidence **e**:

$$\mathbf{P}(Y|X,\mathbf{e}) = \frac{\mathbf{P}(X|Y,\mathbf{e})\,\mathbf{P}(Y|\mathbf{e})}{\mathbf{P}(X|\mathbf{e})}$$

Example of Bayes' Rule

- P(*stiff neck*|*meningitis*) = 0.5
- P(*meningitis*) = 1/50,000
- P(*stiff neck*) = 1/20
- Then P(*meningitis*|*stiff neck*) =

$$\frac{P(\textit{stiff neck}|\textit{meningitis})\ P(\textit{meningitis})}{P(\textit{stiff neck})} =$$

$$\frac{(0.5)(1/50{,}000)}{1/20} = 0.0002$$

Normalization with Bayes' Rule

- P(*stiff neck*|*meningitis*) and P(*meningitis*) are relatively easy to estimate from medical records.
- Prior probability of *stiff neck* is harder to estimate accurately.
- Bayes' rule with normalization:

$$\mathbf{P}(Y|X) = \alpha \mathbf{P}(X|Y)\,\mathbf{P}(Y)$$

Normalization with Bayes' Rule (continued)

Might be easier to compute

P(*stiff neck*|*meningitis*) P(*meningitis*) and

P(*stiff neck*|¬*meningitis*) P(¬*meningitis*)

than to directly estimate

P(*stiff neck*).

Why Use Bayes' Rule

- Causal knowledge such as P(*stiff neck*|*meningitis)* often is more reliable than diagnostic knowledge such as P*(meningitis*|*stiff neck*).
- Bayes' Rule lets us use causal knowledge to make diagnostic inferences (derive diagnostic knowledge).

Difficulty with Bayes' Rule with More than Two Variables

The definition of Bayes' Rule extends naturally to multiple variables:

$$\mathbf{P}(X_1,\ldots,X_m|Y_1,\ldots,Y_n) = \alpha\mathbf{P}(Y_1,\ldots,Y_n|X_1,\ldots,X_m)\ \mathbf{P}(X_1,\ldots,X_m).$$

But notice that to apply it we must know conditional probabilities like

$$P(y_1,\ldots,y_n|x_1,\ldots,x_m)$$

for all 2^n settings of the *Y*s and all 2^m settings of the *X*s (assuming Booleans). Might as well use full joint.

Conditional Independence

- X and Y are *conditionally independent given* Z if and only if $\mathbf{P}(X,Y|Z) = \mathbf{P}(X|Z)\ \mathbf{P}(Y|Z)$.
- $Y_1,\ldots,Y_n$ are *conditionally independent given* $X_1,\ldots,X_m$ if and only if $\mathbf{P}(Y_1,\ldots,Y_n|X_1,\ldots,X_m)= \mathbf{P}(Y_1|X_1,\ldots,X_m)\ \mathbf{P}(Y_2|X_1,\ldots,X_m) \ldots \mathbf{P}(Y_m|X_1,\ldots,X_m)$.
- We've reduced 2^n2^m to $2n2^m$. Additional conditional independencies may reduce 2^m.

Benefits of Conditional Independence

- Allows probabilistic systems to scale up (tabular representations of full joint distributions quickly become too large.)
- Conditional independence is much more commonly available than is absolute independence.

Decomposing a Full Joint by Conditional Independence

- Might assume *Toothache* and *Catch* are conditionally independent given *Cavity*: $\mathbf{P}(Toothache, Catch | Cavity) = \mathbf{P}(Toothache | Cavity)\ \mathbf{P}(Catch | Cavity)$.
- Then $\mathbf{P}(Toothache, Catch, Cavity) =_{\text{[product rule]}} \mathbf{P}(Toothache, Catch | Cavity)\ \mathbf{P}(Cavity) =_{\text{[conditional independence]}} \mathbf{P}(Toothache | Cavity)\ \mathbf{P}(Catch | Cavity)\ \mathbf{P}(Cavity)$.

Bayesian Networks: Motivation

- Capture independence and conditional independence where they exist.
- Among variables where dependencies exist, encode the relevant portion of the full joint.
- Use a graphical representation for which we can more easily investigate the complexity of inference and can search for efficient inference algorithms.

A Bayesian Network is a ...

- Directed Acyclic Graph (DAG) in which …
- … the nodes denote random variables
- … each node *X* has a conditional probability distribution P(*X*|*Parents(X)*).
- The intuitive meaning of an arc from *X* to *Y* is that *X directly influences Y*.

Additional Terminology

- If *X* and its parents are discrete, we can represent the distribution **P(***X*|*Parents(X)*) by a *conditional probability table (CPT)* specifying the probability of each value of *X* given each possible combination of settings for the variables in *Parents(X)*.
- A *conditioning case* is a row in this CPT (a setting of values for the parent nodes).

Bayesian Network Semantics

- A Bayesian Network completely specifies a full joint distribution over its random variables, as below -- this is its meaning.
- $P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i))$
- In the above, $P(x_1,\ldots,x_n)$ is shorthand notation for $P(X_1=x_1,\ldots,X_n=x_n)$.

Inference Example

- What is probability alarm sounds, but neither a burglary nor an earthquake has occurred, and both John and Mary call?
- Using *j* for *John Calls*, *a* for *Alarm*, etc.:

$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) =$$
$$P(j|a)\,P(m|a)\,P(a|\neg b \wedge \neg e)\,P(\neg b)\,P(\neg e) =$$
$$(0.9)(0.7)(0.001)(0.999)(0.998) = 0.00062$$

Chain Rule

- Generalization of the product rule, easily proven by repeated application of the product rule.
- Chain Rule: $P(x_1,\ldots x_n) =$

$$P(x_n|x_{n-1},\ldots,x_1)P(x_{n-1}|x_{n-2},\ldots,x_1)\ldots P(x_2|x_1)P(x_1)$$

$$= \prod_{i=1}^{n} P(x_i|x_{i-1},\ldots,x_i)$$

Chain Rule and BN Semantics

BN semantics: $P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i|\mathit{Parents}(X_i))$

Key Property: $\mathbf{P}(X_i|X_{i-1},\ldots,X_1) = \mathbf{P}(X_i|\mathit{Parents}(X_i))$ provided $\mathit{Parents}(X_i) \subseteq X_1,\ldots,X_{i-1}$. Says a node is conditionally independent of its predecessors in the node ordering given its parents, and suggests incremental procedure for network construction.

Procedure for BN Construction

- **Choose** relevant random variables.
- While there are variables left:
  1. **Choose** a next variable X_i and add a node for it.
  2. Set *Parents*(X_i) to **some** minimal set of nodes such that the Key Property (previous slide) is satisfied.
  3. Define the conditional distribution **P**(X_i|*Parents*(X_i)).

Principles to Guide Choices

- Goal: build a *locally structured (sparse)* network -- each component interacts with a bounded number of other components.
- Add *root causes* first, then the variables that they influence.

Conditional Independence Again

- Recall that *X* is conditionally independent of its predecessors given *Parents(X)*.
- *Markov Blanket* of *X*: set consisting of the parents of *X*, the children of *X*, and the other parents of the children of *X*.
- *X* is conditionally independent of all nodes in the network given its Markov Blanket.