Representation Size

- At first glance it would appear that the space to represent a Bayes Net is necessarily quadratic (possible number of arcs) in the number of random variables.
- But recall we also must represent the CPT of each node, which in general will have size exponential in the number of parents of the node.

Canonical Distributions

- Often CPTs fall into one of several common categories or canonical distributions.
- These canonical forms are based on regularities that permit much more compact representations.

Deterministic Nodes

- A deterministic node has its value specified exactly by the values of its parents, with no uncertainty.
- Different types of deterministic nodes, including logical and numerical.

Deterministic Nodes: Logical

- Value of a child node is determined by a logical expression over the parent nodes.
- Might be negation, disjunction, conjunction, or a more complex expression.
- Example: The relationship between Boolean variables *Canadian*, *US*, and *Mexican* as parents and the child node *North American* is the child is a disjunction of the parents.

Deterministic Nodes: Numerical

- The value of the child node is a numerical function of the parent values.
- If the parent nodes are prices of the same model car at different dealers, and the child is the price paid by a careful shopper, the child is the minimum of the parents.
- Might also have differences, sums, or more complex numerical functions.

Noisy-OR

- Analogous to logical-OR, except we are uncertain about the degree to which each parent can cause the child to be true.
- Captures the intuition that the causal relationship between parent and child can be *inhibited*.
- The degree of uncertainty or inhibition can vary from one parent to the next.

Noisy-OR Assumptions

- Parents and child are Boolean variables.
- Inhibition of one parent is independent of the inhibitions of any other parents; e.g., whatever prevents *Malaria* from causing *Fever* is independent of whatever inhibits *Flu* from causing *Fever*.
- All possible causes are listed. In practice this constraint is not an issue because...

…Can Add a Leak Node

- A leak node is an additional parent of a Noisy-OR node.
- Covers “miscellaneous cases.”
- Easier to see after a definition of Noisy-OR.

Definition of Noisy-OR

- A child node is *false* only if its *true* parents are inhibited.
- The probability of such inhibition is the product of the inhibition probabilities for each parent.
- So the probability the child node is *true* is 1 minus the product of the inhibition probabilities for the *true* parents.

Value of Noisy-OR

- To represent a Noisy-OR CPT, we need only specify the inhibition probability for each parent node.
- If a node has n parents, then this representation requires only n probabilities in constrast with the 2^n ordinarily required.

The Basic Inference Task in Bayesian Networks

- Given some observed **event** (some assignment of values to a set of **evidence variables**), compute the posterior probability distribution over a set of **query variables**.
- Variables that are neither evidence variables nor query variables are **hidden variables**.
- Most common query: $\mathbf{P}(X|\mathbf{e})$.

Generality of Approach

- A Bayes Net is flexible enough that any variable can be the query variables, and any variables can be evidence variables.
- Algorithms we consider are easily extended to handle queries with multiple query variables.

Example Query

- $\mathbf{P}$(*Burglary* | *JohnCalls=true, MaryCalls=true)*
  = <0.284,0.716>
- How can we compute such answers?
- One approach is to compute entire full joint distribution represented by the network, and use our earlier algorithm. But this would defeat the entire purpose of Bayes Nets.

Inference By Enumeration

- $\mathbf{P}(X|\mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$
- Recall the ENUMERATEJOINTASK procedure based on the preceding equation, which computed the answers to queries given the full joint represented as a table.
- We will modify ENUMERATEJOINTASK to use the Bayes Net instead of the table. Need only get desired table entries using Bayes Net.

Getting a Full Joint Table Entry from a Bayes Net

- Recall: $P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i|Parents(X_i))$
- A table entry for $X_1 = x_1,\ldots,X_n = x_n$ is simply $P(x_1,\ldots,x_n)$ which can be calculated based on the Bayes Net semantics above.
- Recall example:

$$P(a,j,m,\neg b,\neg e) = P(j|a)\,P(m|a)\,P(a|\neg b,\neg e)P(\neg b)P(\neg e)$$

Example of Full Procedure

- Query: **P**(*Burglary* | *JohnCalls=true, MaryCalls=true)*
- $\mathbf{P}(B|j,m) = \alpha \sum_{\mathbf{e}} \sum_{\mathbf{a}} \mathbf{P}(B,e,a,j,m)$
- Solve without the normalization constant for both *B* = *true* and *B* = *false*, and then compute the normalization constant and the final probabilities.

Example (Continued)

$P(b|j,m) = \alpha \sum_e \sum_a P(b)P(e)P(a|b,e)P(j|a)P(m|a)$

$= P(b) \sum_e P(e) \sum_a P(a|b,e)P(j|a)P(m|a)$. Go to CPTs in Bayes Net to pull out numbers (Fig. 16.8). Repeat with $\neg b$ instead of b throughout. Resulting numbers are 0.000592 and 0.001494 respectively.

Example (Continued)

Normalizing: $0.000592\alpha + 0.001494\alpha = 1.0$.

$\alpha = 1.0 / 0.002086 \approx 479$.

$\mathbf{P}(Burglary) = \langle 0.284, 0.716 \rangle$.

Some Observations

- The number of terms in the sum is exponential in the number of *hidden variables*.
- Many sub-expressions are repeated on multiple branches. Each could be computed once and saved … leads to the idea of *variable elimination*.

Variable Elimination by Example

$$P(B|j,m) = \alpha\, P(B) \sum_e P(e) \sum_a P(a|B,e)\, P(j|a)\, P(m|a)$$

$f_M(A) \times f_J(A) = f_M(A) \times f_J(A)$

$f_M(A)$		$f_J(A)$		$f_M(A) \times f_J(A)$	
.70	$m|a$	.90	$j|a$	.63	$m, j|a$
.01	$m|\neg a$	.05	$j|\neg a$	.0005	$m, j|\neg a$

$f_A(A, B, E)$

	b		$\neg b$	
	e	$\neg e$	e	$\neg e$
a	.95	.94	.29	.001
$\neg a$	.05	.06	.71	.999

V.E. by Example (Continued)

$$f_A(A,B,E) \times (f_J(A) \times f_M(A))$$

	b, e	b, $\neg e$	$\neg b$, e	$\neg b$, $\neg e$
$a(m, j \mid a)$	(.63)(.95)	(.63)(.94)	(.63)(.29)	(.63)(.001)
$\neg a(m, j \mid \neg a)$	(.0005)(.05)	(.0005)(.06)	(.0005)(.71)	(.0005)(.999)

Because A is a hidden variable, we sum out A (for each column, sum the cases based on a and $\neg a$) to get:

	$b \wedge e$	$b \wedge \neg e$	$\neg b \wedge e$	$\neg b \wedge \neg e$
$f_{\bar{A}JM}(B,E)$	.59853	.59223	.18310	.00113

Variable Elimination by Example (Continued): Now Eliminate (Sum Out) E

$f_E(e)$

e	.002
$\neg e$	.998

$\times$ $f_{\bar{A}JM}(B,E)$

	b	$\neg b$
e	.59853	.18310
$\neg e$	.59223	.00113

$=$ $f_E(e) \times f_{\bar{A}JM}(B,E)$

	b	$\neg b$
e	.001197	.000366
$\neg e$	.598	.00113

After summing out E:

$f_{\bar{E}\bar{A}JM}(B)$

b	.5992
$\neg b$	.0015

Multiply by B and Normalize

	$f_{\overline{E}AJM}(B)$		$f_B(b)$		
b	.5992	$\times$	.001	$=$	.000599
$\neg b$	.0015		.999		.001499

Complete the solution by normalizing the final table.