1 handout: slides

asst 5 was due

EOLQs

Bayesian Networks

Probabilistic Knowledge Representation

Representation: variables, connectives
Inference: approximate, exact

The Alarm Domain

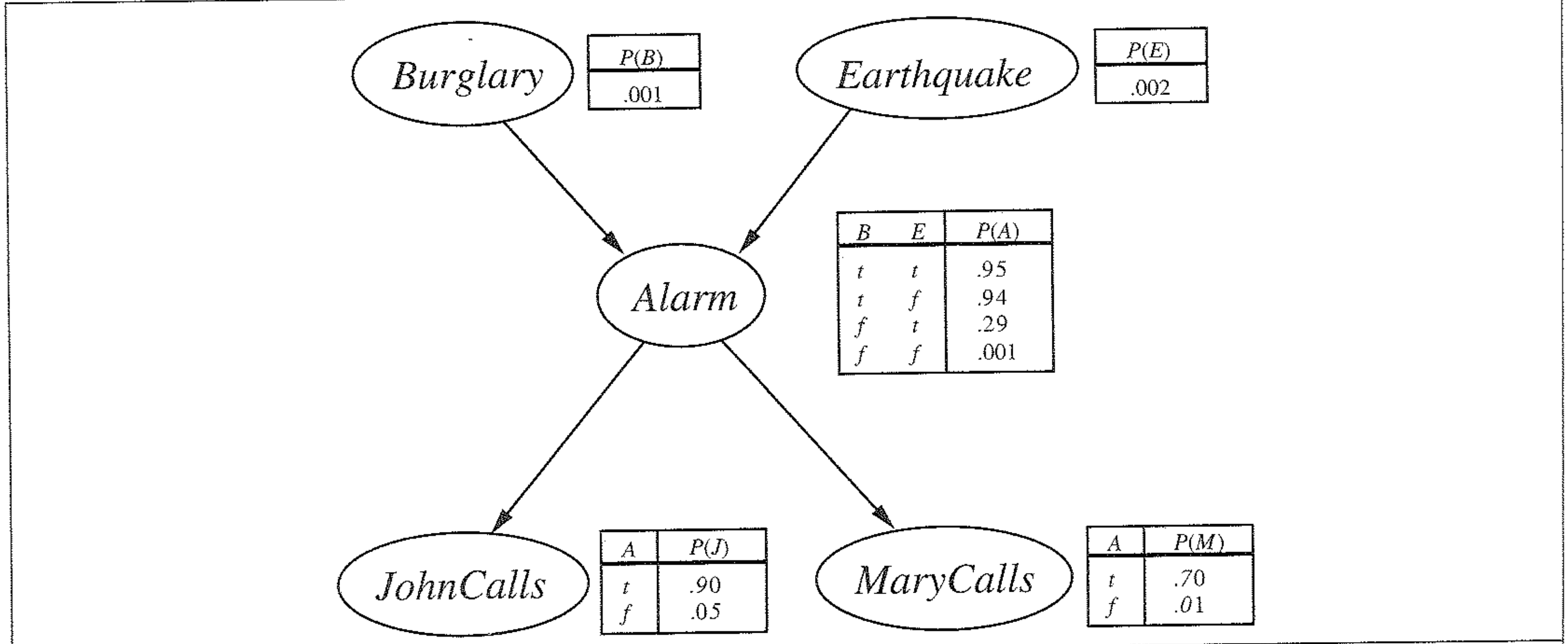

Figure 14.2 A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters B, E, A, J, and M stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

Probabilistic Models

MDPs:
Naive Bayes:
k-**Means:**

Representation: variables, connectives
Inference: approximate, exact

The Full Joint Distribution

ultimate power: knowing the probability of every possible atomic event (combination of values)

The Full Joint Distribution

ultimate power: knowing the probability of every possible atomic event (combination of values)

simple inference via enumeration over the joint:
what is distribution of X given evidence e and unobserved Y

$$P(X|e) = \frac{P(e|X)P(X)}{P(e)} = \alpha P(X, e) = \alpha \sum_y P(X, e, y)$$

Bayes Net = joint probability distribution

The Magic of Independence

In general:

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

The Magic of Independence

In general:

$$\begin{aligned} P(x_1, \ldots, x_n) &= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1) \\ &= \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1) \end{aligned}$$

The Magic of Independence

In general:

$$\begin{aligned} P(x_1, \ldots, x_n) &= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1) \\ &= \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1) \end{aligned}$$

A Bayesian net specifies independence:

$$P(X_i | X_{i-1}, \ldots, X_1) = P(X_i | parents(X_i))$$

The Magic of Independence

In general:

$$
\begin{aligned}
P(x_1, \ldots, x_n) &= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1) \\
&= \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1)
\end{aligned}
$$

A Bayesian net specifies independence:

$$P(X_i | X_{i-1}, \ldots, X_1) = P(X_i | parents(X_i))$$

So we get:

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

The Magic of Independence

In general:

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

A Bayesian net specifies independence:

$$P(X_i | X_{i-1}, \ldots, X_1) = P(X_i | parents(X_i))$$

So we get:

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

For n b-ary variables with p parents, that's nb^p instead of b^n!

Example

$$
\begin{aligned}
P(S=1|W=1) &= \frac{P(S=1,W=1)}{P(W=1)} \\
&= \frac{\sum_{c,r} P(C=c,S=1,R=r,W=1)}{P(W=1)} \\
&= 0.2781/0.6471 = 0.430
\end{aligned}
$$

$$
P(W=1) = \sum_{c,r,s} P(C=c,S=s,R=r,W=1) = 0.6471
$$

$$
\begin{aligned}
P(R=1|W=1) &= \frac{P(R=1,W=1)}{P(W=1)} \\
&= \frac{\sum_{c,s} P(C=c,S=s,R=1,W=1)}{P(W=1)} \\
&= 0.4581/0.6471 = 0.708
\end{aligned}
$$

Break

- exam 2
- exam 3
- projects: presentations, paper, paper drafts

Approximate Inference

Rejection Sampling

What is distribution of X given evidence e and unobserved Y?

Draw worlds from the joint, rejecting those that do not match e. Look at distribution of X.

each sample is linear time, but overall slow if e is unlikely

Likelihood Weighting

What is distribution of X given evidence e and unobserved Y?

ChooseSample (e)
$w \leftarrow 1$
for each variable V_i in topological order:
  if $(V_i = v_i) \in e$ then
    $w \leftarrow w \cdot P(v_i|parents(v_i))$
  else
    $v_i \leftarrow$ sample from $P(V_i|parents(V_i))$

(afterwards, normalize samples so all w's sum to 1)

uses all samples, but needs lots of samples if e are late in ordering

EOLQs

- What question didn't you get to ask today?
- What's still confusing?
- What would you like to hear more about?

Please write down your most pressing question about AI and put it in the box on your way out.
Thanks!