# CSC520 - Artificial Intelligence
## Lecture 18

Dr. Scott N. Gerard

North Carolina State University

Mar 20, 2025

# Class Exercise

Given:

| $D$ | $W$ | $P(D \mid W)$ |
|---|---|---|
| wet | sun | 0.1 |
| dry | sun | 0.9 |
| wet | rain | 0.7 |
| dry | rain | 0.3 |

| $W$ | $P(W)$ |
|---|---|
| sun | 0.8 |
| rain | 0.2 |

- What is $P(W \mid dry)$ ?

| $D$ | $W$ | $P(W \mid D) = \alpha P(D \mid W) P(W)$ |
|---|---|---|
| wet | sun | 0.08 |
| dry | sun | 0.72 |
| wet | rain | 0.14 |
| dry | rain | 0.06 |

| $W$ | $\alpha P(W \mid dry)$ |
|---|---|
| sun | 0.72 $\alpha$ |
| rain | 0.06 $\alpha$ |

| $W$ | $P(W \mid dry)$ |
|---|---|
| sun | 0.92 |
| rain | 0.08 |

# Agenda

- Bayesian network
- Bayesian network construction
- Inference by enumeration
- Inference by variable elimination
- Inference by sampling

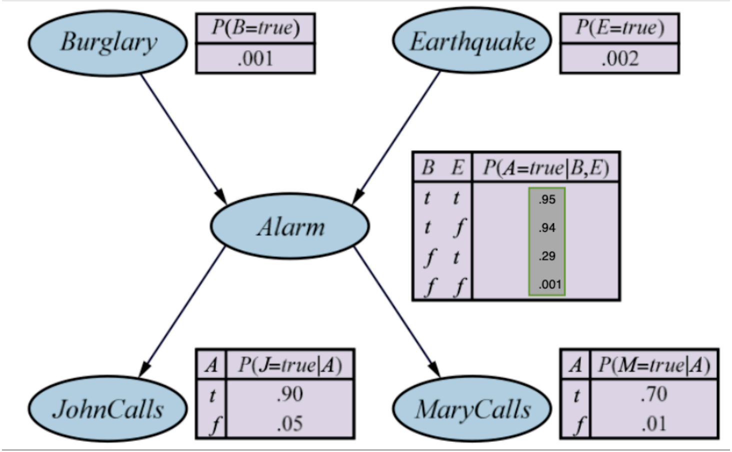
# Bayesian Network

- Problem: Full joint distribution representation and inferencing requires exponential space
- Bayesian networks addresses this problem by employing conditional independence among the variables
- Bayesian network is a probabilistic graphical model
  - A directed acyclic graph (DAG)
  - A node represents a random variable
  - A directed link from $X$ (parent) to $Y$ (child) means $X$ has <u>direct</u> influence on $Y$
  - Each node $X_i$ has a conditional probability table (CPT): $P(X_i|Parents(X_i))$

# Bayesian Network Example

- Your home has a burglar alarm which goes off in case of a burglary or when there is an earthquake
- John and Mary are your neighbors who have promised to call in case of an alarm
- John may fail to call if he confuses alarm with telephone ringing
- Mary may fail to call if she is listening to loud music

# Bayesian Network Example




Only $1 + 1 + 4 + 2 + 2 = 10$ entries compared to $2^5 - 1 = 31$ entries in the full joint distribution

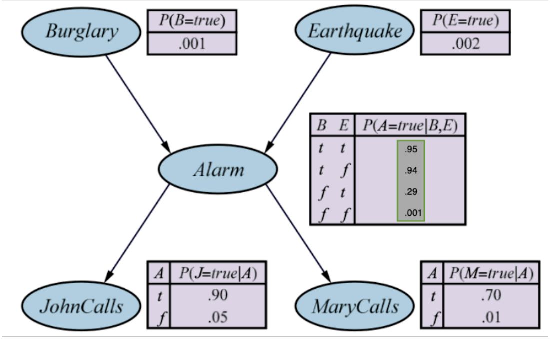
# Semantics of Bayesian Network

- By chain rule, we get:

$$
\begin{aligned}
&P(J, M, A, B, E) \\
&= P(J|M, A, B, E)P(M|A, B, E)P(A|B, E)P(B|E)P(E) \\
&= P(J|A)P(M|A)P(A|B, E)P(B)P(E)
\end{aligned}
$$

- In general, Bayes network defines the joint distribution as:
$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|parents(X_i))$

# Bayesian Network Example




- E.g. Probability that: alarms goes off, John calls, Mary calls, no burglary, no earthquake

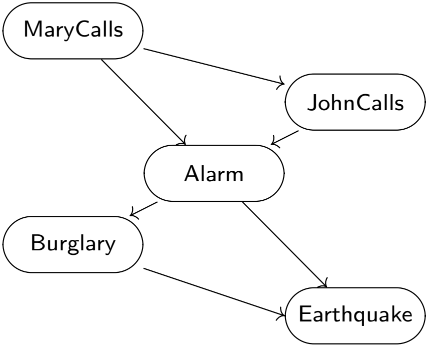
$$
\begin{aligned}
P(j, m, a, \neg b, \neg e) &= P(j|a)P(m|a)P(a|\neg b \wedge \neg e)P(\neg b)P(\neg e) \\
&= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.000628
\end{aligned}
$$

# Bayes Net Construction

- Nodes
  - Identify the variables required to model the domain
  - Order the variables $\{X_1, X_2, \ldots, X_n\}$; any ordering will work but network will be more compact if causes precede effects
- Links
  - For each node $X_i$ select a minimal set of parents from $\{X_1, \ldots, X_{i-1}\}$ such that: $P(X_i|X_{i-1}, \ldots, X_1) = P(X_i|Parents(X_i))$
  - Add a link from each *parent* $\in Parents(X_i)$ node to $X_i$
  - Write CPT for each $X_i$, $P(X_i|Parents(X_i))$

# Bayes Net Construction

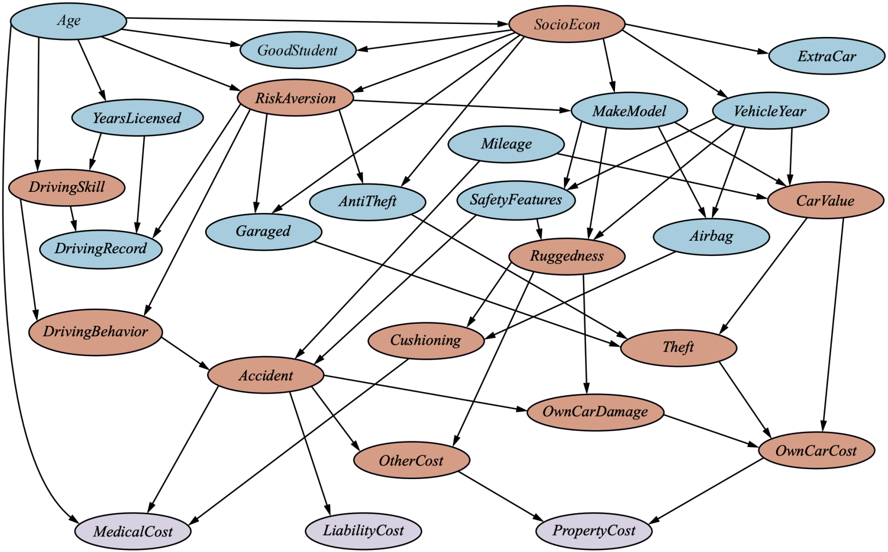
$$P(J|M) = P(J)? \quad \text{No}$$
$$P(A|J,M) = P(A)? \quad \text{No}$$
$$P(A|J,M) = P(A|J)? \quad \text{No}$$
$$P(B|A,J,M) = P(B)? \quad \text{No}$$
$$P(B|A,J,M) = P(B|A)? \quad \text{Yes}$$
$$P(E|B,A,J,M) = P(E|A)? \quad \text{No}$$
$$P(E|B,A,J,M) = P(E|A,B)? \quad \text{Yes}$$




- Deciding conditional indepedence is hard in noncausal direction
- Network is less compact: $1 + 2 + 4 + 2 + 4 = 13$ numbers are needed

# Bayes Network

## Insurance Bayesian Network



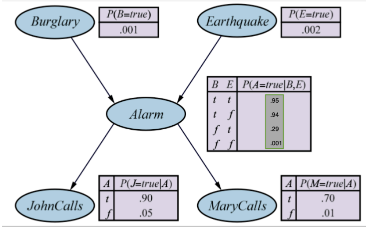
# Inference in Bayes Network

- Queries are of the form: $P(Q|e_1, \ldots, e_k)$
  - Evidence variables: $E_1, \ldots, E_k = e_1, \ldots, e_k$
  - Query variables: $Q$
  - Hidden variables: $H_1, \ldots, H_r$
- Inference by enumeration
- Inference by variable elimination
- Inference by sampling

# Inference by Enumeration

- Using joint distribution for inferencing without actually constructing the joint distribution fully
- E.g. Query: $P(Burglary|JohnCalls = true, MaryCalls = true)$

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



# Inference by Enumeration

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

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

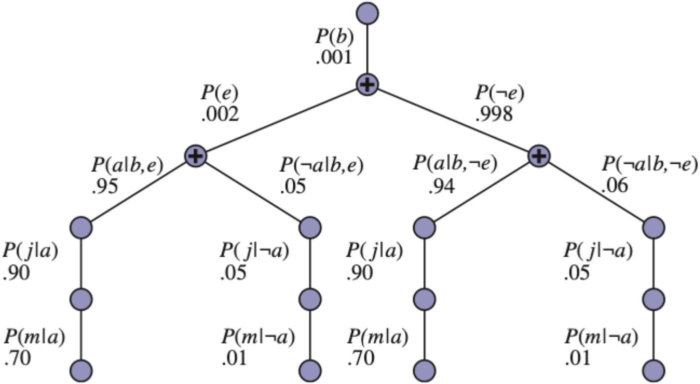
$$= \alpha P(b)\{P(e)[P(a|b,e)P(j|a)P(m|a) + P(\neg a|b,e)P(j|\neg a)P(m|\neg a)] +$$
$$P(\neg e)[P(a|b,\neg e)P(j|a)P(m|a) + P(\neg a|b,\neg e)P(j|\neg a)P(m|\neg a)]\}$$
$$= \alpha 0.001\{0.002 * [(0.95 * 0.9 * 0.7) + (0.05 * 0.05 * 0.01)] +$$
$$0.998 * [(0.94 * 0.9 * 0.7) + (0.06 * 0.05 * 0.01)]\}$$
$$= 0.00059224\alpha$$

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

$$P(\neg b|j,m) = 0.0014919\alpha$$
$$\alpha = 1/(0.00059224 + 0.0014919) = 479.814216$$
$$P(b|j,m) = 479.814216 * 0.00059224 = 0.284$$
$$P(\neg b|j,m) = 479.814216 * 0.0014919 = 0.716$$

$$P(B|j,m) = \langle 0.284, 0.716 \rangle$$

# Inference by Enumeration



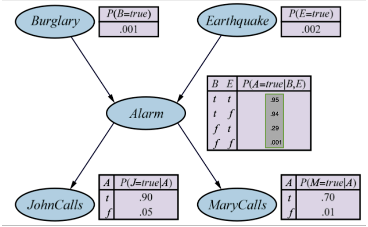
# Inference by Variable Elimination

- Queries are of the form: $P(Q|e_1, \ldots, e_k)$
  - Evidence variables: $E_1, \ldots, E_k = e_1, \ldots, e_k$
  - Query variables: $Q$
  - Hidden variables: $H_1, \ldots, H_r$
- A factor is a function that maps a set of variables to a specific value
- Step 1: Start with initial factors
  - CPTs are factors, e.g. $P(A|E, B)$
- Step 2: While there are hidden variables:
  - Pick a hidden variable $H$
  - Join ALL factors mentioning $H$
  - Eliminate (sum out) $H$
- Step 3: Join all remaining factors and normalize

# Inference by Variable Elimination

- Query: $P(Burglary | JohnCalls = true, MaryCalls = true)$

$$
\begin{aligned}
P(B|j,m) &\propto P(B,j,m) \\
&= \sum_{e,a} P(B,j,m,e,a) \\
&= \sum_{e,a} P(B)P(e)P(a|B,e)P(j|a)P(m|a) \\
&= \sum_{e} P(B)P(e) \sum_{a} P(a|B,e)P(j|a)P(m|a) \\
&= \sum_{e} P(B)P(e) f_1(B,e) \\
&= P(B) \sum_{e} P(e) f_1(B,e) \\
&= P(B) f_2(B)
\end{aligned}
$$



# Inference by Sampling

- Inference by enumeration or variable elimination can be time consuming on large Bayesian network
- Approximate answers to queries can be computed using randomized sampling (Monte Carlo method)
- Idea: Generate several random samples from the Bayes net
  - Use the samples to compute answers to the queries
- Simplest sampling method is direct or prior sampling
  - Random sampling from Bayes net with no associated evidence

# Sampling from Distribution

- Suppose we want to generate a sample from the below distribution

| $W$ | $P(W)$ |
|---|---|
| rain | 0.6 |
| sun | 0.3 |
| fog | 0.1 |

- Generate a sample $x$ from uniform distribution over [0, 1)
  - For e.g., Python function: random.random()
- Assign the outcome based on the value of $x$

$$0 \leq x < 0.6 \rightarrow W = rain$$
$$0.6 \leq x < 0.9 \rightarrow W = sun$$
$$0.9 \leq x < 1 \rightarrow W = fog$$

- E.g. sampling 4 times gives:

$$\mathrm{x} = 0.151,\ \mathrm{W} = \mathrm{rain}$$
$$\mathrm{x} = 0.059,\ \mathrm{W} = \mathrm{rain}$$
$$\mathrm{x} = 0.217,\ \mathrm{W} = \mathrm{rain}$$
$$\mathrm{x} = 0.810,\ \mathrm{W} = \mathrm{sun}$$

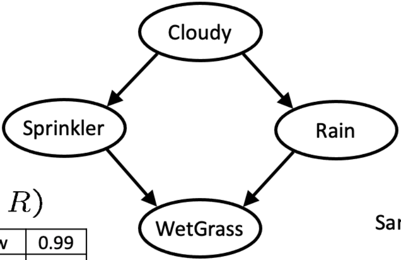
# Inference by Sampling

$P(C)$

| | |
|---|---|
| +c | 0.5 |
| -c | 0.5 |

$P(S|C)$

| | | |
|---|---|---|
| +c | +s | 0.1 |
| | -s | 0.9 |
| -c | +s | 0.5 |
| | -s | 0.5 |




$P(R|C)$

| | | |
|---|---|---|
| +c | +r | 0.8 |
| | -r | 0.2 |
| -c | +r | 0.2 |
| | -r | 0.8 |

$P(W|S,R)$

| | | | |
|---|---|---|---|
| +s | +r | +w | 0.99 |
| | | -w | 0.01 |
| | -r | +w | 0.90 |
| | | -w | 0.10 |
| -s | +r | +w | 0.90 |
| | | -w | 0.10 |
| | -r | +w | 0.01 |
| | | -w | 0.99 |

Samples:

# Inference by Sampling

- Suppose we get the following samples

$$
\begin{aligned}
&+c, -s, +r, +w \\
&+c, +s, +r, +w \\
&-c, +s, +r, -w \\
&+c, -s, +r, +w \\
&-c, -s, -r, +w
\end{aligned}
$$

- What is the estimate of $P(W)$?
- What is the estimate of $P(C|+w)$?
- What is the estimate of $P(C|+r,+w)$
- What is the estimate of $P(C|-r,-w)$
- With infinite samples, the estimated distributions will equal the true distributions

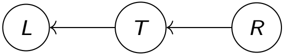
# Inference by Sampling

- Suppose we get the following samples

$$
\begin{aligned}
&+c, -s, +r, +w \\
&+c, +s, +r, +w \\
&-c, +s, +r, -w \\
&+c, -s, +r, +w \\
&-c, -s, -r, +w
\end{aligned}
$$

- What is the estimate of $P(W)$?
  - $P(+w) = 0.8, P(-w) = 0.2$
- What is the estimate of $P(C|+w)$?
  - $P(+c) = 0.75, P(-c) = 0.25$
- What is the estimate of $P(C|+r, +w)$
  - $P(+c) = 1, P(-c) = 0$
- What is the estimate of $P(C|-r, -w)$
  - Cannot be computed
- With infinite samples, the estimated distributions will equal the true distributions

- Random variables
  - $R$: Raining
  - $T$: Traffic
  - $L$: Late for class




Compute $P(L) = \alpha \sum_l \sum_r P(L, T, R) = \alpha \sum_l P(L|T) \sum_r P(T|R)P(R)$.

| $T$ | $P(T)$ |
|---|---|
| $+t$ | |
| $-t$ | |

$\sum_r$

| $T$ | $R$ | $P(T,R)$ |
|---|---|---|
| $+t$ | $+r$ | |
| $-t$ | $+r$ | |
| $+t$ | $-r$ | |
| $-t$ | $-r$ | |

| $T$ | $R$ | $P(T\|R)$ |
|---|---|---|
| $+t$ | $+r$ | 0.8 |
| $-t$ | $+r$ | 0.2 |
| $+t$ | $-r$ | 0.1 |
| $-t$ | $-r$ | 0.9 |

| $R$ | $P(R)$ |
|---|---|
| $+r$ | 0.1 |
| $-r$ | 0.9 |

| $L$ | $P(L)$ |
|---|---|
| $+l$ | |
| $-l$ | |

$\sum_t$

| $L$ | $T$ | $P(L,T)$ |
|---|---|---|
| $+l$ | $+t$ | |
| $-l$ | $+t$ | |
| $+l$ | $-t$ | |
| $-l$ | $-t$ | |

| $L$ | $T$ | $P(L\|T)$ |
|---|---|---|
| $+l$ | $+t$ | 0.7 |
| $-l$ | $+t$ | 0.3 |
| $+l$ | $-t$ | 0.1 |
| $-l$ | $-t$ | 0.9 |

| $T$ | $P(T)$ |
|---|---|
| $+t$ | |
| $-t$ | |