Approximate Inference in Bayes Nets
Sampling based methods

Mausam

(Based on slides by Jack Breese and Daphne Koller)
Bayes Nets is a generative model

- We can easily generate samples from the distribution represented by the Bayes net
  - Generate one variable at a time in topological order

Use the samples to compute marginal probabilities, say $P(c)$
Stochastic simulation $P(B|C)$

- $P(b) = 0.03$
- $P(e) = 0.001$
- $P(a|b) = 0.98$
- $P(a|\bar{b}) = 0.7$
- $P(a|\bar{e}) = 0.4$
- $P(a|\bar{b},\bar{e}) = 0.01$
- $P(c|a) = 0.8$
- $P(c|\bar{a}) = 0.05$
- $P(n|e) = 0.3$
- $P(n|\bar{e}) = 0.001$

The diagram shows the conditional probabilities and the relationships between events $B$, $C$, $a$, $e$, $n$, and the alarm and newscast. The event $Call = c$ is the outcome of interest.
Stochastic simulation $P(B|C)$

- $P(b) = 0.03$
- $P(e) = 0.001$
- $P(a|b) = 0.98$
- $P(a|\neg b) = 0.7$
- $P(a|\neg b, \neg e) = 0.4$
- $P(a|\neg b, e) = 0.01$
- $P(c|a) = 0.8$
- $P(c|\neg a) = 0.05$

**Samples:**

- $B, E, A, C, N$

**Call** = $c$

**Alarm**

**Burglary**

**Earthquake**
Stochastic simulation \( P(B|C) \)

- **Burglary**
  - \( P(b) = 0.03 \)
  - \( P(a|b) \):
    - \( b \rightarrow e \): 0.98
    - \( b \rightarrow \bar{e} \): 0.7
    - \( \bar{b} \rightarrow e \): 0.4
    - \( \bar{b} \rightarrow \bar{e} \): 0.01

- **Earthquake**
  - \( P(e) = 0.001 \)
  - \( P(a|e) \):
    - \( e \rightarrow a \): 0.3
    - \( e \rightarrow \bar{a} \): 0.001

- **Alarm**
  - \( P(c) \):
    - \( a \rightarrow c \): 0.8
    - \( \bar{a} \rightarrow c \): 0.05

- **Newscast**
  - \( P(n) \):
    - \( e \rightarrow n \): 0.3
    - \( \bar{e} \rightarrow n \): 0.001

**Samples:**

- \( B \) \( E \) \( A \) \( C \) \( N \)
Stochastic simulation $P(B|C)$

- $P(B) = 0.03$
- $P(e) = 0.001$
- $P(B|C) = \frac{P(B \cap C)}{P(C)}$
- $P(B \cap C)$ can be calculated as $P(B) \times P(C|B)$
- $P(C|B)$ can be calculated as $\frac{P(B \cap C)}{P(B)}$ (assuming $P(B)$ and $P(C)$ are independent)

**Bayesian Network**

- **Burglary** ($B$)
  - $P(b) = 0.03$
  - $P(b|\neg b) = 0.8$
  - $P(b|b) = 0.2$
- **Earthquake** ($E$)
  - $P(e) = 0.001$
  - $P(e|\neg e) = 0.99$
  - $P(e|e) = 0.1$
- **Alarm** ($A$)
  - $P(a|b) = 0.98$
  - $P(a|\neg b) = 0.02$
- **Newscast** ($N$)
  - $P(n|e) = 0.3$
  - $P(n|\neg e) = 0.001$
- **Call** ($C$)
  - $P(c|a) = 0.8$
  - $P(c|\neg a) = 0.05$
  - $P(c|\neg a) = 0.05$

**Samples:**

- Sample: $b$

**Probability Calculations:**

- $P(B \cap C) = P(B) \times P(C|B)$
- $P(C) = P(B) \times P(C|B)$
- $P(B|C) = \frac{P(B \cap C)}{P(C)}$

**Independence Assumption:**

- If $B$ and $C$ are independent, then $P(B \cap C) = P(B) \times P(C)$.

**Conditional Independence:**

- If $B$ and $C$ are independent, then $P(B|C) = P(B)$.

**Bayesian Inference:**

- $P(B|C)$ can be updated using Bayes' theorem:
  
  $$P(B|C) = \frac{P(C|B) \times P(B)}{P(C)}$$

- $P(C) = P(C|B) \times P(B) + P(C|\neg B) \times P(\neg B)$
Stochastic simulation $P(B|C)$

- $P(b) = 0.03$
- $P(e) = 0.001$

**Burglary**
- $P(a|b) = 0.98$
- $P(a|\neg b) = 0.7$
- $P(\neg a|b) = 0.4$
- $P(\neg a|\neg b) = 0.01$

**Earthquake**
- $P(e|a) = 0.8$
- $P(e|\neg a) = 0.05$

**Alarm**
- $P(\neg c|b) = 0.97$
- $P(c|\neg b) = 0.3$

**Newscast**
- $P(\neg e|a) = 0.3$
- $P(e|\neg a) = 0.001$

**Samples:**
- $\bar{b}$
Stochastic simulation \( P(B|C) \)

\[
\begin{align*}
P(b) & = 0.03 \\
P(a) & = 0.98 \quad 0.7 \quad 0.4 \quad 0.01 \\
P(c) & = 0.8 \quad 0.05 \\
P(n) & = 0.3 \quad 0.001
\end{align*}
\]

Samples:

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>A</th>
<th>C</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\overline{b} )</td>
<td>e</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Stochastic simulation $P(B|C)$

$P(b) = 0.03$

<table>
<thead>
<tr>
<th>$b$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.98</td>
<td>0.7</td>
</tr>
</tbody>
</table>

$P(a|b) = 0.4$

$P(a|\bar{b}) = 0.01$

$P(c|a) = 0.8$

$P(c|\bar{a}) = 0.05$

$P(e|a) = 0.3$

$P(e|\bar{a}) = 0.001$

Samples:

$\bar{b} \quad e$

© Jack Breese (Microsoft) & Daphne Koller (Stanford)
Stochastic simulation $P(B|C)$

$P(b) = 0.03$

$P(e) = 0.001$

$P(a|b,e) = 0.98$
$P(a|b,\not{e}) = 0.7$
$P(a|\not{b},e) = 0.4$
$P(a|\not{b},\not{e}) = 0.01$

$P(c|a) = 0.8$
$P(c|\not{a}) = 0.05$

$P(B) = P(b) \cdot P(a|b) + P(\not{b}) \cdot P(a|\not{b})$

$P(E) = P(e) \cdot P(a|e) + P(\not{e}) \cdot P(a|\not{e})$

Samples:

$\bar{b} e a$

$Call = c$
Stochastic simulation $P(B|C)$

$P(b) = 0.03$

**Burglary**

<table>
<thead>
<tr>
<th>$b$</th>
<th>$e$</th>
<th>$\bar{b}$</th>
<th>$\bar{e}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(a)$</td>
<td>0.98</td>
<td>0.7</td>
<td>0.4</td>
</tr>
</tbody>
</table>

**Earthquake**

$P(e) = 0.001$

**Alarm**

**Call**

$P(c) = 0.8$

**Newscast**

<table>
<thead>
<tr>
<th>$e$</th>
<th>$\bar{e}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(n)$</td>
<td>0.3</td>
</tr>
</tbody>
</table>

**Samples:**

<table>
<thead>
<tr>
<th>$B$</th>
<th>$E$</th>
<th>$A$</th>
<th>$C$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{b}$</td>
<td>$e$</td>
<td>$a$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Stochastic simulation $P(B|C)$

- **Burglary**
  - $P(b) = 0.03$
  - $P(a) = 0.98$
  - $P(c) = 0.8$

- **Earthquake**
  - $P(e) = 0.001$
  - $P(n) = 0.3$

- **Alarm**
  - Call = $c$

- **Newscast**
  - $P(c|e) = 0.001$
  - $P(c|\bar{e}) = 0.05$

Samples:

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>A</th>
<th>C</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b)</td>
<td>(e)</td>
<td>(a)</td>
<td>(c)</td>
<td></td>
</tr>
</tbody>
</table>
Stochastic simulation $P(B|C)$

- $P(b) = 0.03$
- $P(e) = 0.001$
- $P(a|b) = 0.98$
- $P(a|\overline{b}) = 0.7$
- $P(\overline{a}|b) = 0.4$
- $P(\overline{a}|\overline{b}) = 0.01$
- $P(c|a) = 0.8$
- $P(c|\overline{a}) = 0.05$
- $P(n|e) = 0.3$
- $P(n|\overline{e}) = 0.001$

Samples:

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>A</th>
<th>C</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>\overline{b}</td>
<td>e</td>
<td>a</td>
<td>c</td>
<td></td>
</tr>
</tbody>
</table>
Stochastic simulation $P(B|C)$

- $P(b) = 0.03$
- $P(e) = 0.001$
- $P(a) = 0.98$
- $P(c) = 0.8$

- Burglary
- Earthquake
- Alarm
- Newscast

Samples:

- $B, E, A, C, N$
- $\overline{b}, e, a, c, \overline{n}$
Stochastic simulation $P(B|C)$

- $P(b) = 0.03$
- $P(e) = 0.001$

### Evidence
- $b$: Burglary
- $e$: Earthquake

### Probabilities
- $P(a) = 0.98$
- $P(c) = 0.8$

### Call and Newscast
- $P(n) = 0.3$

### Samples:
- $B E A C N$
- $\bar{b} e a c n$
- $b \bar{e} a c n$

© Jack Breese (Microsoft) & Daphne Koller (Stanford)
Stochastic simulation $P(B | C)$

- $P(b) = 0.03$
- $P(e) = 0.001$
- $P(a | b) = 0.98$, $P(a | \bar{b}) = 0.7$
- $P(n | a) = 0.8$, $P(n | \bar{a}) = 0.05$
- $P(\bar{n} | a) = 0.3$, $P(\bar{n} | \bar{a}) = 0.001$

Samples:

- $\bar{b} \ e \ a \ c \ n$
- $b \ e \ a \ c \ n$

$P(B | C)$ = Call

$P(E | C)$ = Newscast

$P(A | B)$ = Alarm

$P(C | B)$ = Burglary

$P(C | \bar{B})$ = Earthquake

© Jack Breese (Microsoft) & Daphne Koller (Stanford)
Stochastic simulation $P(B|C)$

**Samples:**

- $b\overline{e}a\overline{c}\overline{n}$
- $b\overline{e}a\overline{c}n$
- $\ldots$

**P(b)**: 0.03

**P(a)**: 0.98
- $b\overline{e}$: 0.7
- $\overline{b}\overline{e}$: 0.4
- $\overline{b}e$: 0.01

**P(c)**: 0.8
- $a$: 0.8
- $\overline{a}$: 0.05

**P(e)**: 0.001

**Burglary**

**Earthquake**

**Alarm**

**Call** = c

**Newscast**

© Jack Breese (Microsoft) & Daphne Koller (Stanford)
Stochastic simulation $P(B|C)$

$P(b) = 0.03$

$P(a) = 0.98$

$P(c) = 0.8$

$P(e) = 0.001$

$P(n) = 0.3$

Samples:

$P(b|c) \sim \frac{\text{# of live samples with } B=b}{\text{total # of live samples}}$

© Jack Breese (Microsoft) & Daphne Koller (Stanford)
Rejection Sampling

• Sample from the prior
  – reject if do not match the evidence

• Returns consistent posterior estimates

• Hopelessly expensive if $P(e)$ is small
  – $P(e)$ drops off exponentially with no. of evidence vars
Likelihood Weighting

• Idea:
  – fix evidence variables
  – sample only non-evidence variables
  – weight each sample by the likelihood of evidence
Likelihood weighting $P(B|C)$

- **Burglary**
  - $a$: 0.8
  - $\bar{a}$: 0.05

- **Earthquake**

- **Alarm**
  - Call = $c$

- **Newscast**

Samples:

```
B E A C N
```
Likelihood weighting $P(B|C)$
Likelihood weighting $P(B|C)$

<table>
<thead>
<tr>
<th></th>
<th>$a$</th>
<th>$\neg a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(c)$</td>
<td>0.8</td>
<td>0.05</td>
</tr>
<tr>
<td>$P(\neg c)$</td>
<td>0.2</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Samples:

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>A</th>
<th>C</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\neg b$</td>
<td>$e$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$B$ for Burglary, $E$ for Earthquake, $A$ for Alarm, $C$ for Call, $N$ for Newscast.
Likelihood weighting $P(B|C)$

<table>
<thead>
<tr>
<th></th>
<th>$a$</th>
<th>$\bar{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(c)$</td>
<td>0.8</td>
<td>0.05</td>
</tr>
<tr>
<td>$P(\bar{c})$</td>
<td>0.2</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Samples:

<table>
<thead>
<tr>
<th>$B$</th>
<th>$E$</th>
<th>$A$</th>
<th>$C$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{b}$</td>
<td>$e$</td>
<td>$a$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Likelihood weighting \( P(B|C) \)

\[
\begin{array}{c|cc}
 & a & \overline{a} \\
P(c) & 0.8 & 0.05 \\
P(\overline{c}) & 0.2 & 0.95 \\
\end{array}
\]

Samples:

\[
\begin{array}{cccccc}
B & E & A & C & N \\
\overline{b} & e & a & c & \text{ } \\
\end{array}
\]
Likelihood weighting $P(B | C)$

Samples:

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>A</th>
<th>C</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{b}$</td>
<td>$e$</td>
<td>$a$</td>
<td>$c$</td>
<td>$\overline{n}$</td>
</tr>
</tbody>
</table>

$P(c) = 0.8 \quad P(\overline{c}) = 0.2$

$P(a) = 0.05 \quad P(\overline{a}) = 0.95$
Likelihood weighting $P(B|C)$

Samples:

<table>
<thead>
<tr>
<th>$B$</th>
<th>$E$</th>
<th>$A$</th>
<th>$C$</th>
<th>$N$</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>$e$</td>
<td>$a$</td>
<td>$c$</td>
<td>$n$</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Likelihood weighting $P(B | C)$

$P(B | C)$

<table>
<thead>
<tr>
<th></th>
<th>$a$</th>
<th>$\bar{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(B)$</td>
<td>0.8</td>
<td>0.05</td>
</tr>
<tr>
<td>$P(\overline{B})$</td>
<td>0.2</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Samples:

<table>
<thead>
<tr>
<th>$B$</th>
<th>$E$</th>
<th>$A$</th>
<th>$C$</th>
<th>$N$</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{b}$</td>
<td>$e$</td>
<td>$a$</td>
<td>$c$</td>
<td>$\overline{n}$</td>
<td>0.8</td>
</tr>
<tr>
<td>$b$</td>
<td>$\overline{e}$</td>
<td>$\overline{a}$</td>
<td>$c$</td>
<td>$n$</td>
<td>0.05</td>
</tr>
</tbody>
</table>

$\text{Call} = c$
Likelihood weighting $P(B|C)$

Samples:

<table>
<thead>
<tr>
<th>$B$</th>
<th>$E$</th>
<th>$A$</th>
<th>$C$</th>
<th>$N$</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{b}$</td>
<td>$e$</td>
<td>$a$</td>
<td>$c$</td>
<td>$\overline{n}$</td>
<td>0.8</td>
</tr>
<tr>
<td>$b$</td>
<td>$e$</td>
<td>$a$</td>
<td>$c$</td>
<td>$n$</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>
Likelihood weighting $P(B|C)$

<table>
<thead>
<tr>
<th></th>
<th>$a$</th>
<th>$\bar{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(c)$</td>
<td>0.8</td>
<td>0.05</td>
</tr>
<tr>
<td>$P(\bar{c})$</td>
<td>0.2</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Samples:

<table>
<thead>
<tr>
<th>$B$</th>
<th>$E$</th>
<th>$A$</th>
<th>$C$</th>
<th>$N$</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{b}$</td>
<td>$\bar{e}$</td>
<td>$a$</td>
<td>$c$</td>
<td>$\bar{n}$</td>
<td>0.8</td>
</tr>
<tr>
<td>$b$</td>
<td>$\bar{e}$</td>
<td>$\bar{a}$</td>
<td>$c$</td>
<td>$n$</td>
<td>0.05</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

$P(b|c) = \frac{\text{weight of samples with } B=b}{\text{total weight of samples}}$
Likelihood Weighting

- **Sampling probability**: \( S(z,e) = \prod_i P(z_i | \text{Parents}(Z_i)) \)
  - Neither prior nor posterior
- **Wt for a sample \(<z,e>\)**: \( w(z,e) = \prod_i P(e_i | \text{Parents}(E_i)) \)
- **Weighted Sampling probability** \( S(z,e)w(z,e) \)
  \[
  = \prod_i P(z_i | \text{Parents}(Z_i)) \prod_i P(e_i | \text{Parents}(E_i))
  = P(z,e)
  \]
  - returns consistent estimates
- performance degrades w/ many evidence vars
  - but a few samples have nearly all the total weight
  - late occurring evidence vars do not guide sample generation
MCMC with Gibbs Sampling

• Fix the values of observed variables
• Set the values of all non-observed variables randomly
• Perform a random walk through the space of complete variable assignments. On each move:
  1. Pick a variable $X$
  2. Calculate $\Pr(X=\text{true} \mid \text{all other variables})$
  3. Set $X$ to true with that probability
• Repeat many times. Frequency with which any variable $X$ is true is it’s posterior probability.
• Converges to true posterior when frequencies stop changing significantly
  – stable distribution, mixing
Markov Blanket Sampling

• How to calculate $\Pr(X=\text{true} \mid \text{all other variables})$?

• Recall: a variable is independent of all others given it’s Markov Blanket
  – parents
  – children
  – other parents of children

• So problem becomes calculating $\Pr(X=\text{true} \mid \text{MB}(X))$
  – We solve this sub-problem exactly
  – Fortunately, it is easy to solve

$$P(X) = \alpha P(X \mid \text{Parents}(X)) \prod_{Y \in \text{Children}(X)} P(Y \mid \text{Parents}(Y))$$
Example

\[ P(X) = \alpha P(X \mid Parents(X)) \prod_{Y \in \text{Children}(X)} P(Y \mid Parents(Y)) \]

\[
P(X \mid A, B, C) = \frac{P(X, A, B, C)}{P(A, B, C)}
\]

\[
= \frac{P(A)P(X \mid A)P(C)P(B \mid X, C)}{P(A, B, C)}
\]

\[
= \left[ \frac{P(A)P(C)}{P(A, B, C)} \right] P(X \mid A)P(B \mid X, C)
\]

\[
= \alpha P(X \mid A)P(B \mid X, C)
\]
Example

**Smoking**

<table>
<thead>
<tr>
<th>P(s)</th>
<th>0.2</th>
</tr>
</thead>
</table>

**Heart disease**

<table>
<thead>
<tr>
<th>P(h)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>0.6</td>
</tr>
<tr>
<td>~s</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Lung disease**

<table>
<thead>
<tr>
<th>P(g)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>0.8</td>
</tr>
<tr>
<td>~s</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Shortness of breath**

<table>
<thead>
<tr>
<th>H</th>
<th>G</th>
<th>P(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>g</td>
<td>0.9</td>
</tr>
<tr>
<td>h</td>
<td>~g</td>
<td>0.8</td>
</tr>
<tr>
<td>~h</td>
<td>g</td>
<td>0.7</td>
</tr>
<tr>
<td>~h</td>
<td>~g</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Example

- Smoking
  - Heart disease
  - Lung disease
- Shortness of breath

<table>
<thead>
<tr>
<th>H</th>
<th>G</th>
<th>P(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>g</td>
<td>0.9</td>
</tr>
<tr>
<td>h</td>
<td>~g</td>
<td>0.8</td>
</tr>
<tr>
<td>~h</td>
<td>g</td>
<td>0.7</td>
</tr>
<tr>
<td>~h</td>
<td>~g</td>
<td>0.1</td>
</tr>
</tbody>
</table>

- Evidence: s, b
Example

- Evidence: s, b
- Randomly set: h, b

<table>
<thead>
<tr>
<th>H</th>
<th>G</th>
<th>P(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>g</td>
<td>0.9</td>
</tr>
<tr>
<td>h</td>
<td>~g</td>
<td>0.8</td>
</tr>
<tr>
<td>~h</td>
<td>g</td>
<td>0.7</td>
</tr>
<tr>
<td>~h</td>
<td>~g</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
</tr>
<tr>
<td>~s</td>
</tr>
<tr>
<td>0.6</td>
</tr>
<tr>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
</tr>
<tr>
<td>~s</td>
</tr>
<tr>
<td>0.8</td>
</tr>
<tr>
<td>0.1</td>
</tr>
</tbody>
</table>
Example

- Evidence: s, b
- Randomly set: h, g
- Sample H using $P(H|s,g,b)$

<table>
<thead>
<tr>
<th>s</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>~s</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(h)</th>
<th>s</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>~s</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

$P(g)$

<table>
<thead>
<tr>
<th>s</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>~s</td>
<td>0.1</td>
</tr>
</tbody>
</table>

$P(b)$

<table>
<thead>
<tr>
<th>H</th>
<th>G</th>
<th>P(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>g</td>
<td>0.9</td>
</tr>
<tr>
<td>h</td>
<td>~g</td>
<td>0.8</td>
</tr>
<tr>
<td>~h</td>
<td>g</td>
<td>0.7</td>
</tr>
<tr>
<td>~h</td>
<td>~g</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Example

- Evidence: s, b
- Randomly set: ~h, g
- Sample H using P(H|s,g,b)
- Suppose result is ~h

<table>
<thead>
<tr>
<th>P(s)</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(g)</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>0.8</td>
</tr>
<tr>
<td>~s</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(h)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>0.6</td>
</tr>
<tr>
<td>~s</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(b)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>h g</td>
<td>0.9</td>
</tr>
<tr>
<td>h ~g</td>
<td>0.8</td>
</tr>
<tr>
<td>~h g</td>
<td>0.7</td>
</tr>
<tr>
<td>~h ~g</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Example

- Evidence: s, b
- Randomly set: ~h, g
- Sample H using $P(H|s,g,b)$
  - Suppose result is ~h
  - Sample G using $P(G|s,\sim h,b)$
Evidence: s, b
Randomly set: ~h, g
Sample H using $P(H|s,g,b)$
Suppose result is ~h
Sample G using $P(G|s,\sim h,b)$
⇒ Suppose result is g
Example

- Evidence: s, b
- Randomly set: ~h, g
- Sample H using $P(H|s,g,b)$
  - Suppose result is ~h
  - Sample G using $P(G|s,~h,b)$
  - Suppose result is g
  - Sample G using $P(G|s,~h,b)$
Example

- Evidence: s, b
- Randomly set: ~h, g

Sample H using $P(H|s,g,b)$
Suppose result is ~h
Sample G using $P(G|s,\sim h,b)$
⇒ Suppose result is g
Sample G using $P(G|s,\sim h,b)$
⇒ Suppose result is ~g
Gibbs MCMC Summary

\[ P(X|E) = \frac{\text{number of samples with } X=x}{\text{total number of samples}} \]

• Advantages:
  – No samples are discarded
  – No problem with samples of low weight
  – Can be implemented very efficiently
    • 10K samples @ second

• Disadvantages:
  – Can get stuck if relationship between two variables is deterministic
  – Many variations have been devised to make MCMC more robust
Other inference methods

• Exact inference
  – Junction tree

• Approximate inference
  – Belief Propagation
  – Variational Methods