Inference in Bayes Nets via Sampling

• If the Bayes net is too complex to do Variable Elimination on, or we want to answer queries that VE can’t easily model.

• The Bayes net defines a joint distribution $P(X_1, X_2, ..., X_n)$ over its features.

• Idea is to draw atomic events from this distribution in such a way that the probability we obtain the atomic event $<X_1=d_1, X_2=d_2, ..., X_n=d_n>$ is exactly $P(X_1=d_1, X_2=d_2, ..., X_n=d_n)$
Inference in Bayes Nets via Sampling

• Then if we collect a set of m samples
  
  <X1=d11, X2=d12, ..., Xn=d1n>
  <X1=d21, X2=d32, ..., Xn=d2n>
  <X1=d31, X2=d32, ..., Xn=d3n>
  <X1=d41, X2=d42, ..., Xn=d4n>
  <X1=d51, X2=d52, ..., Xn=d5n>
  <X1=d61, X2=d62, ..., Xn=d6n>
  ...
  <X1=dm1, X2=dm2, ..., Xn=dmn>
Inference in Bayes Nets via Sampling

• We can estimate many different probabilities by looking that the frequency among the m samples.

• e.g. to estimate $P(X_1=a)$ we count how many samples have $X_1=a$ and divide by m

• to estimate $P(X_1=a \lor X_2 = b)$ we count how many samples have $X_1=a$ or $X_2=b$ and divide by m.

• Notice that the 2$^{nd}$ query can’t easily be answered with VE, even if we had the computational resources to do so.
Prior Sampling

• Start at the roots of the Bayes nets (i.e., those variables that have no parents)
• Randomly select a value for each of these variables using their CPTs.
• Work your way down the net, when you have a value for all of a variable X’s parents randomly select a value for X using the CPT \( P(X|\text{par}(X)) \) --- you know the value of \( \text{par}(X) \).

• When all variables have a value---you get one sample.
### Conditional Probability Table

**Cloudy**

<table>
<thead>
<tr>
<th>C</th>
<th>P(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>.50</td>
</tr>
</tbody>
</table>

**Sprinkler**

| C | P(S|C) |
|---|------|
| T | .10  |
| F | .50  |

**Rain**

| C | P(R|C) |
|---|------|
| T | .80  |
| F | .20  |

**Wet Grass**

| S | R | P(W|S,R) |
|---|---|--------|
| T | T | .99    |
| T | F | .90    |
| F | T | .90    |
| F | F | .01    |
Rejection Sampling

• Suppose we want to estimate $P(X_k = d_k \mid X_1=d_1, X_2=d_2)$
• For this query all of the samples where $X_1$ is not equal to $d_1$ or $X_2 = d_2$ are useless
• $P(X_k = d_k \mid X_1=d_1, X_2=d_2) = \frac{P(X_k = d_k, X_1=d_1, X_2=d_2)}{P(X_1=d_1, X_2=d_2)}$

• So we have to count how many samples have $X_k = d_k$ and $X_1=d_1$ and $X_2=d_2$ and then divide by the number of samples that have $X_1=d_1$ and $X_2=d_2$

• This is called rejection sampling---the samples were $X_1$ is not equal to $d_1$ or $X_2$ is not equal to $d_2$ are useless—they are rejected.
Rejection Sampling

• Problem is if $X_1=d_1$, $X_2=d_2$ has low probability we will reject almost all samples!
Likelihood Weighting

• If we want to estimate a conditional probability like
  \[ P(X_k = d_k \mid X_1 = d_1, X_2 = d_2) \]
  we force all samples to satisfy the condition.

• Problem—in doing so we bias the samples, so that we are no longer sampling from the Bayes net distribution.

• “Solution” we reweigh the samples so that we undo this introduced bias.
Likelihood Weighting

• Say the conditioning is \( X_1 = d_1, X_2 = d_2 \)
• Set wt of the sample = 1
• start at the roots as before
  • if the variable is fixed by the condition, we set it to the required value and multiply wt by the probability it takes on that value.
  • Else we set it at random according to its probability
• Moving down, we set the value of a variable \( X \) after we have set the values of all of its parents.
  • Again if \( X \) is fixed by the condition we set it to the required value and multiply wt by the probability it takes on that value
  • Else we set it at random according to its probability.
• At the end we have a single sample and a weight.
Likelihood Weighting

• To estimate the probabilities we want we now take the sum of the weights of the good samples divided by the total weight of the samples.

• \[ \Pr(X_k = d_k | X_1 = d_1, X_2 = d_2) = \frac{\text{sum of weights of samples where } X_k = d_k}{\text{total sum of weights of samples}} \]