CS 188: Artificial Intelligence
Spring 2009

Lecture 19: Sampling
3/31/2009

John DeNero – UC Berkeley
Slides adapted from Dan Klein

Announcements

- Course contest was posted over break
  - Yes, it's fun
  - Lots of baseline infrastructure is given to you
  - Extra credit for participation, in addition to glory

- Written 3 released this week

- Project 4 released this weekend
  - Due date subject to change
Midterm

- It was long and challenging. Good work!

- Course-wide statistics are available on glookup if you're interested. Mean: 68.5 & Std. Dev: 17.3

- If you want your test back, find me after class or in office hours

- Grading questions should be directed to me

Today: Inference Methods

- Agent beliefs are encoded as conditional probabilities and related via a Bayes’ net
- Inference is reasoning from a core set of beliefs to related probabilities of interest
- Techniques for Bayes’ net inference:
  - **Enumeration**: Build the joint distribution over all variables, then compute what you want
  - **Variable Elimination**: Successively eliminate hidden variables, then normalize what’s left
  - **Sampling**: Simulate the world and count
Example Domain

- New Finance Industry
  - B: Bank nears failure
  - A: America rescues
  - C: Congress complains

\[ P(B) \]
\[
\begin{array}{c|c}
B & P \\
\hline
b & 0.1 \\
\neg b & 0.9 \\
\end{array}
\]

\[ P(A|B) \]
\[
\begin{array}{c|c|c}
B & A & P \\
\hline
b & a & 0.8 \\
b & \neg a & 0.2 \\
\neg b & a & 0.1 \\
\neg b & \neg a & 0.9 \\
\end{array}
\]

\[ P(C|A) \]
\[
\begin{array}{c|c|c}
A & C & P \\
\hline
a & c & 0.7 \\
a & \neg c & 0.3 \\
\neg a & c & 0.5 \\
\neg a & \neg c & 0.5 \\
\end{array}
\]

Example: Applying Bayes Rule

Query: \( P(B|a) \)

Do not underestimate the power of Bayes Rule!

\[
P(b|a) = \frac{P(a|b)P(b)}{P(a)} = \frac{P(a|b)P(b)}{P(b, a) + P(\neg b, a)}
\]

\[ P(B|a) \]
\[
\begin{array}{c|c|c}
A & B & P \\
\hline
a & b & 8/17 \\
\neg b & a & 9/17 \\
\end{array}
\]
Variable Elimination Bayes Rule

Start / Select

\[
P(B)
\]

<table>
<thead>
<tr>
<th>B</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>0.1</td>
</tr>
<tr>
<td>~b</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Join on B

\[
\begin{align*}
P(a, B) & \\
A & B & P \\
a & b & 0.08 \\
a & ~b & 0.09 \\
\end{align*}
\]

\[
P(B|a)
\]

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>8/17</td>
</tr>
<tr>
<td>a</td>
<td>~b</td>
<td>9/17</td>
</tr>
</tbody>
</table>

Normalize

\[
P(A|B) \rightarrow P(a|B)
\]

<table>
<thead>
<tr>
<th>B</th>
<th>A</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>a</td>
<td>0.8</td>
</tr>
<tr>
<td>b</td>
<td>~a</td>
<td>0.2</td>
</tr>
<tr>
<td>~b</td>
<td>a</td>
<td>0.1</td>
</tr>
<tr>
<td>~b</td>
<td>~a</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Example: Multiple Joins

\[
P(B)
\]

<table>
<thead>
<tr>
<th>B</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>0.1</td>
</tr>
<tr>
<td>~b</td>
<td>0.9</td>
</tr>
</tbody>
</table>

\[
P(A|B)
\]

<table>
<thead>
<tr>
<th>B</th>
<th>A</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>a</td>
<td>0.8</td>
</tr>
<tr>
<td>b</td>
<td>~a</td>
<td>0.2</td>
</tr>
<tr>
<td>~b</td>
<td>a</td>
<td>0.1</td>
</tr>
<tr>
<td>~b</td>
<td>~a</td>
<td>0.9</td>
</tr>
</tbody>
</table>

\[
P(B, A)
\]

<table>
<thead>
<tr>
<th>B</th>
<th>A</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>a</td>
<td>0.08</td>
</tr>
<tr>
<td>b</td>
<td>~a</td>
<td>0.02</td>
</tr>
<tr>
<td>~b</td>
<td>a</td>
<td>0.09</td>
</tr>
<tr>
<td>~b</td>
<td>~a</td>
<td>0.81</td>
</tr>
</tbody>
</table>

\[
P(C|A)
\]

<table>
<thead>
<tr>
<th>A</th>
<th>C</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>c</td>
<td>0.7</td>
</tr>
<tr>
<td>a</td>
<td>~c</td>
<td>0.3</td>
</tr>
<tr>
<td>~a</td>
<td>c</td>
<td>0.5</td>
</tr>
<tr>
<td>~a</td>
<td>~c</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[
P(B, A)
\]

<table>
<thead>
<tr>
<th>A</th>
<th>C</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>c</td>
<td>0.7</td>
</tr>
<tr>
<td>a</td>
<td>~c</td>
<td>0.3</td>
</tr>
<tr>
<td>~a</td>
<td>c</td>
<td>0.5</td>
</tr>
<tr>
<td>~a</td>
<td>~c</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Example: Multiple Joins

\[ P(B, A) \]
\[
\begin{array}{ccc}
B & A & P \\
b & a & 0.08 \\
b & \neg a & 0.02 \\
\neg b & a & 0.09 \\
\neg b & \neg a & 0.81 \\
\end{array}
\]

\[ P(C|A) \]
\[
\begin{array}{ccc}
A & C & P \\
a & c & 0.7 \\
a & \neg c & 0.3 \\
\neg a & c & 0.5 \\
\neg a & \neg c & 0.5 \\
\end{array}
\]

Join on A

\[ P(B, A, C) \]
\[
\begin{array}{ccc}
B & A & C & P \\
b & a & c & 0.056 \\
b & a & \neg c & 0.024 \\
b & \neg a & c & 0.010 \\
b & \neg a & \neg c & 0.010 \\
\neg b & a & c & 0.063 \\
\neg b & a & \neg c & 0.027 \\
\neg b & \neg a & c & 0.405 \\
\neg b & \neg a & \neg c & 0.405 \\
\end{array}
\]

Inference by Enumeration

Sum out B

\[ P(A, C) \]
\[
\begin{array}{ccc}
A & C & P \\
a & c & 0.119 \\
a & \neg c & 0.051 \\
\neg a & c & 0.415 \\
\neg a & \neg c & 0.415 \\
\end{array}
\]

Sum out A

\[ P(C) \]
\[
\begin{array}{cc}
C & P \\
c & 0.534 \\
\neg c & 0.466 \\
\end{array}
\]
P(C) : Marginalizing Early

\[
\begin{array}{c|c|c}
B & A & P \\
\hline
b & a & 0.08 \\
\hline
b & \neg a & 0.02 \\
\hline
\neg b & a & 0.09 \\
\hline
\neg b & \neg a & 0.81 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
A & C & P \\
\hline
a & c & 0.7 \\
\hline
a & \neg c & 0.3 \\
\hline
\neg a & c & 0.5 \\
\hline
\neg a & \neg c & 0.5 \\
\end{array}
\]

Sum out B

\[
\begin{array}{c|c}
A & P \\
\hline
a & 0.17 \\
\hline
\neg a & 0.83 \\
\end{array}
\]

Marginalizing Early (aka VE*)

\[
\begin{array}{c|c|c}
A & P \\
\hline
a & 0.17 \\
\hline
\neg a & 0.83 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
A & C & P \\
\hline
a & c & 0.7 \\
\hline
a & \neg c & 0.3 \\
\hline
\neg a & c & 0.5 \\
\hline
\neg a & \neg c & 0.5 \\
\end{array}
\]

Join on A

\[
\begin{array}{c|c|c}
A & C & P \\
\hline
a & c & 0.119 \\
\hline
a & \neg c & 0.051 \\
\hline
\neg a & c & 0.415 \\
\hline
\neg a & \neg c & 0.415 \\
\end{array}
\]

Sum out A

\[
\begin{array}{c|c}
C & P \\
\hline
c & 0.534 \\
\hline
\neg c & 0.466 \\
\end{array}
\]

* VE is variable elimination
Approximate Inference

- Simulation has a name: sampling
- Sampling is a hot topic in machine learning, and it’s really simple
- Basic idea:
  - Draw N samples from a sampling distribution S
  - Compute an approximate posterior probability
  - Show this converges to the true probability P

Prior Sampling

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

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

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

- This process generates samples with probability
  \[
  S_{PS}(x_1 \ldots x_n) = \prod_{i=1}^{n} P(x_i|\text{Parents}(X_i)) = P(x_1 \ldots x_n)
  \]
  ...i.e. the BN’s joint probability

- Let the number of samples of an event be \( N_{PS}(x_1 \ldots x_n) \)

- Then
  \[
  \lim_{N \to \infty} \hat{P}(x_1, \ldots, x_n) = \lim_{N \to \infty} \frac{N_{PS}(x_1, \ldots, x_n)}{N} = S_{PS}(x_1, \ldots, x_n) = P(x_1 \ldots x_n)
  \]
  i.e., the sampling procedure is consistent

Example

- We’ll get a bunch of samples from the BN:
  - \( c, \neg s, r, w \)
  - \( c, s, r, w \)
  - \( \neg c, s, r, \neg w \)
  - \( c, \neg s, r, w \)
  - \( \neg c, s, \neg r, w \)

- If we want to know \( P(W) \)
  - We have counts \( <w:4, \neg w:1> \)
  - Normalize to get \( P(W) = <w:0.8, \neg w:0.2> \)
  - This will get closer to the true distribution with more samples
  - Can estimate anything else, too
  - What about \( P(C|\neg r) \)? \( P(C|\neg r, \neg w) \)?
Rejection Sampling

- Let's say we want $P(C)$
  - No point keeping all samples around
  - Just tally counts of $C$ outcomes
- Let's say we want $P(C|s)$
  - Same thing: tally $C$ outcomes, but ignore (reject) samples which don’t have $S=s$
  - This is rejection sampling
  - It is also consistent for conditional probabilities (i.e., correct in the limit)

Likelihood Weighting

- Problem with rejection sampling:
  - If evidence is unlikely, you reject a lot of samples
  - You don’t exploit your evidence as you sample
  - Consider $P(B|a)$
- Idea: fix evidence variables and sample the rest
- Problem: sample distribution not consistent!
- Solution: weight by probability of evidence given parents
**Likelihood Sampling**

\[ w = 1.0 \times 0.1 \times 0.99 \]

**Likelihood Weighting**

- Sampling distribution if \( z \) sampled and \( e \) fixed evidence
  \[
  S_{WS}(z, e) = \prod_{i=1}^{l} P(z_i | \text{Parents}(Z_i))
  \]
- Now, samples have weights
  \[
  w(z, e) = \prod_{i=1}^{m} P(e_i | \text{Parents}(E_i))
  \]
- Together, weighted sampling distribution is consistent
  \[
  S_{WS}(z, e) \cdot w(z, e) = \prod_{i=1}^{l} P(z_i | \text{Parents}(z_i)) \prod_{i=1}^{m} P(e_i | \text{Parents}(e_i))
  = P(z, e)
  \]
**Sampling Example**

- There are 2 cups.
  - The first contains 1 penny and 1 quarter
  - The second contains 2 quarters

- Say I pick a cup uniformly at random, then pick a coin randomly from that cup. It's a quarter (yes!). What is the probability that the other coin in that cup is also a quarter?

---

**Likelihood Weighting**

- Likelihood weighting is good
  - we have taken evidence into account as we generate the sample
  - More of our samples will reflect the state of the world suggested by the evidence

- Likelihood weighting doesn’t solve all our problems
  - Evidence is taken into account for downstream variables, but not upstream ones

- We would like to consider evidence when we sample every variable
Markov Chain Monte Carlo*

- **Idea**: instead of sampling from scratch, create samples that are each like the last one.

- **Procedure**: resample one variable at a time, conditioned on all the rest, but keep evidence fixed. E.g., for $P(b|c)$:

  - $b \rightarrow a \rightarrow c$
  - $\neg b \rightarrow a \rightarrow c$
  - $\neg b \rightarrow \neg a \rightarrow c$

- **Properties**: Now samples are not independent, but sample averages are still consistent estimators!

- **What’s the point**: both upstream and downstream variables condition on evidence.