Review

• Inference in BNs
  Evidence nodes E
  Query nodes Q
  How to compute \( P(Q|E) \)?

• Polytrees
  Singly connected networks
  Polynomial time inference.

• Loopy BNs
  Exact inference: node clustering
  Approximate inference: stochastic simulation
  **Today**

• Friday: QUIZ
Sample a discrete random variable

- Given: distribution $P(X = x_i)$
  - Uniform random number generator $r \in [0, 1]$
  - How to sample $X$ from $P(X = x_i)$?

- Intuitively:
  - $P(X = x_i)$ defines a partition of unity: \[ \sum_{i} P(X = x_i) = 1 \]
  - $\leftarrow P(X = x_1) \rightarrow \leftarrow P(X = x_2) \rightarrow \ldots \leftarrow P(X = x_n) \rightarrow 0$
  - This partition maps $r \in [0, 1]$ into a discrete value of $X$.

- Formally:
  - Define cumulative distribution:
    \[ c_i = \begin{cases} 
      \sum_{j=1}^{i} P(X = x_j) & \text{for } i = 1 \ldots n \\
      \emptyset & \text{for } i = 0
    \end{cases} \]
  - Generate $r \in [0, 1]$
  - Output $x_i$ if $c_{i-1} < r < c_i$
Sampling in discrete BNs

* Given: $BN = DAG + CPTs$
  How to estimate $P(Q=q | E=e)$?

* Markov blanket $B_x = $ parents, children, spouses of $X$
  Conditional independence:
  $P(X|B_x) = P(X|B_x, Y)$ for $Y \notin \{X, B_x\}$

* Markov chain Monte Carlo (MCMC)
  - sampling scheme for approximate inference in loopy BNs
  - Based on repeated sampling of $P(X|B_x)$

* To estimate $P(Q=q | E=e)$:
  - Fix evidence nodes to observed values $E=e$.
  - Initialize non-evidence nodes at random.
  - Repeat $N$ times
    - Pick non-evidence node at random, $X$
    - Use Bayes rule to compute $P(X=x | B_x)$
    - Resample $X$ based on $P(X=x | B_x)$
  - Count "snapshots" of $BN$
    $N(q)$ in which $Q=q$
    In general: $N(q) \geq 0$, $N(q) \leq N$

Estimate $P(Q=q | E=e) \approx \frac{N(q)}{N}$
Converges in limit $N \to \infty$ to correct answer.
• BN = DAG + CPT not always available from experts. How to learn from examples?

• Issues
  • Structure (DAG) — known or unknown
  • Evidence: complete vs. incomplete/partial instantiation of the nodes in BN
  • Optimization: combinatorial vs. continuous (e.g., learning DAG) vs. continuous (e.g., learning CPTs)
  • Algorithms: iterative vs. non-iterative
  • Solution: local vs. global optimum
- Maximum likelihood (ML) estimation
  - simplest form of learning in BNs
  - choose ("estimate") the model (DAG+CPTs) to maximize $P(\text{observed data} \mid \text{model})$

**Example:** biased coin

$X \in \{\text{heads, tails}\}$

$P(X=\text{heads}) = p$

$P(X=\text{tails}) = 1-p$

Trivial BN: $\times P(x)$

- How to estimate $p$ from observed samples (results of $T$ coin tosses)?
- I.I.D. assumption
  - samples are independently, identically, distributed according to $P(x)$.
  - $\{x^{(1)}, x^{(2)}, \ldots, x^{(T)}\}$

- Probability of I.I.D. data set:
  
  $P(\text{DATA}) = P(x=x^{(1)})P(x=x^{(2)})\cdots P(x=x^{(T)}) = \prod_{t=1}^{T} P(x=x^{(t)})$

- Log-likelihood $L$

  $$L = \log P(\text{DATA}) = \log \prod_{t=1}^{T} P(x=x^{(t)}) = \sum_{t=1}^{T} \log P(x=x^{(t)})$$

Let $N_H = \text{count}(X=\text{heads})$. Let $N_T = \text{count}(X=\text{tails})$.

In terms of counts:

$$L(p) = N_H \log p + N_T \log (1-p)$$

- Maximum likelihood estimation

  $$\frac{\delta L}{\delta p} = \frac{N_H}{p} + \frac{N_T}{1-p} (-1) = 0$$

  $$N_H (1-p) - N_T p = 0$$

  $$(N_H + N_T) \rho = N_H$$

  $$\rho = \frac{N_H}{N_H + N_T} = \frac{N_H}{T}$$

Intuitively: max likelihood estimate of $P(x=\text{heads})$ is relative frequency of heads in observed coin tosses.
Fully observed BNs

- Structure is known; DAG is fixed a priori over discrete nodes \( \{X_1, X_2, \ldots, X_n\} \)

- CPTs enumerate \( P(X_i = x_i \mid \text{pa}(X_i) = \text{PT}) \) as lookup tables

- Data is \( T \) complete instantiations of nodes in BN \( \{X_1^{(t)}, X_2^{(t)}, \ldots, X_n^{(t)}\}_{t=1}^T \)

Example:

```
  X_1  
 /    
X_2  X_3
```

\( X_i \in \{0, 1\} \)
\( n = 3 \)

<table>
<thead>
<tr>
<th>Sample</th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>