Approximate Inference Queries

Sampling

- A lot like repeated simulation
- Basic Idea
  1. Draw N samples from a sampling distribution S
  2. Compute an approximate posterior probability
  3. Show this converges to the true probability P

Sampling from a Given Distribution

1. Get sample u from uniform distribution over $[0, 1)$ (use PRNG or something)
2. Convert this sample u into an outcome for the given distribution by having each target outcome associated with a sub-interval of $[0, 1)$ with sub-interval size equal to probability of the outcome

- eg:

C	P(C)
red	0.6
green	0.1
blue	0.3

$$0 \leq u < 0.6, \rightarrow C = red$$
$$0.6 \leq u < 0.7, \rightarrow C = green$$
$$0.7 \leq u < 1, \rightarrow C = blue$$

Prior Sampling

- Sample topologically in order of bayes net
- All sub-samples taken given the parents
- Return $(x_1, x_2, ..., x_n) \mid x_i := \text{sample}(x_i, \mathbb{P}(X_i \mid \text{parents}(X_i)))$
- Generates samples with probability $S_{PS}(x_1, ..., x_n) = \prod_{i=1}^{n} \mathbb{P}(x_i \mid \text{parents}(X_i))$
  - Same as $\mathbb{P}(x_1, ..., x_n)$
  - The probability of a sample is equal to the probability of the sample in the joint distribution
- $$\lim_{n \to \infty} \hat{P}(x_1, ..., x_n) = \frac{\lim_{n \to \infty} N_{PS}(x_1, ..., x_n)}{N} = S_{PS}(x_1, ..., x_n) = \mathbb{P}(x_1, ..., x_n)$$
  - N_{PS} := number of samples of an event
  - Basically as your samples increase, you get more accurate
- Consistent := the samples match the frequencies in the joint distribution
- To get the probability from the samples, you normalize the counts for the variable you're looking for
- Good because you can estimate any probability in the BN given the samples
- Good because you can choose the amount of samples you want to take
- If you don't have any samples for something, you don't know what it is and need more samples (no answer)
  - Bad if evidence is rare

Rejection Sampling

- Like prior sampling
- Just keep a count of the amount of times things show up
  - Means you don't have to maintain a long list of samples

- If there is a given (evidence provided) and the sample does not fulfill the evidence, then ignore the sample and do not add it to the counts
- Bad if evidence is rare
- Consistent

Likelihood Weighting

- Fix evidence variables and sample the rest
- Problem: sample distribution is not consistent
- Solution: weight by probability of evidence given parents
  - Each sample also has a weight
  - Weight is initially 1 and then multiplied by the weights of the fixed variables in the sample
- Sample Values:

$$S_{WS}(z,e) := \prod_{i=1}^{l} \mathbb{P}(z_i \,|\, \text{parents}(Z_i)) \Bigg| \begin{aligned} z &:= \text{ sampled values} \\ e &:= \text{ evidence values fixed} \\ l &:= \text{ number of samples retained} \end{aligned}$$

- Sample weights:

$$w(z,e) := \prod_{i=1}^{m} \mathbb{P}(e_i \,|\, \text{parents}(E_i)) \Bigg|_{m := \text{number of samples 'rigged'}}$$

- Weighted sampling distribution is consistent:

$$S_{WS}(z,e) * w(z,e) = \prod_{i=1}^{l} \mathbb{P}(z_i \,|\, \text{parents}(z_i)) \prod_{i=1}^{m} \mathbb{P}(e_{i \,|\, \text{parents}(e_i)})$$

- Good when evidence is at the top of the BN
  - Generally lower weight the further 'downstream' something is in a BN
  - Evidence influences the choice of downstream variables, but not upstream ones

Gibbs Sampling

- Mostly not required to know
- Keep track of a full instantiation of $x_1, x_2, ..., x_n$. Start with an arbitrary instantiation consistent with the evidence. Sample one variable at a time, conditioned on all the rest, but keep evidence fixed. Keep repeating this for a long time.
- In the limit of repeating this infinitely many times, the resulting samples come from the correct distribution
- Both upstream and downstream variables condition on evidence
  - Have to compute the cost of all non-fixed elements conditioned on every other element
  -
    - Note: Only need to consider CPTs with S remain

- Sample from P(S | +c, +r, -w)

$$
\begin{aligned}
P(S|+c,+r,-w) &= \frac{P(S,+c,+r,-w)}{P(+c,+r,-w)} && \text{def. of conditional probability} \\
&= \frac{P(S,+c,+r,-w)}{\sum_s P(s,+c,+r,-w)} && \text{introduce summation} \\
&= \frac{P(+c)P(S|+c)P(+r|+c)P(-w|S,+r)}{\sum_s P(+c)P(s|+c)P(+r|+c)P(-w|s,+r)} && \text{def. of Bayes' Nets} \\
&= \frac{P(+c)P(S|+c)P(+r|+c)P(-w|S,+r)}{P(+c)P(+r|+c)\sum_s P(s|+c)P(-w|s,+r)} && \text{move summation term out} \\
&= \frac{P(S|+c)P(-w|S,+r)}{\sum_s P(s|+c)P(-w|s,+r)} && \text{cancel out terms}
\end{aligned}
$$

Figure 1: 525