CSci 5512:
Gibbs Sampling for Approximate Inference in Bayesian Networks

Let $p(X_1, \ldots, X_n | e_1, \ldots, e_m)$ denote the joint distribution of a set of random variables $(X_1, \ldots, X_n)$ conditioned on a set of evidence variables $(e_1, \ldots, e_m)$. Gibbs sampling is an algorithm to generate a sequence of samples from such a joint probability distribution. The purpose of such a sequence is to approximate the joint distribution (as with a histogram), or to compute an integral (such as an expected value).

Gibbs sampling is applicable when the joint distribution is not known explicitly, but the conditional distribution of each variable is known. The Gibbs sampling algorithm is used to generate an instance from the distribution of each variable in turn, conditional on the current values of the other variables. It can be shown that the sequence of samples comprises a Markov chain, and the stationary distribution of that Markov chain is just the sought-after joint distribution. Gibbs sampling is particularly well-adapted to sampling the posterior distribution of a Bayesian network, since Bayesian networks are typically specified as a collection of conditional distributions.

1 The Gibbs Sampler

A Gibbs sampler runs a Markov chain on $(X_1, \ldots, X_n)$. For convenience of notation, we denote the set $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ as $X_{(-i)}$, and $\mathbf{e} = (e_1, \ldots, e_m)$. Then, the following method gives one possible way of creating a Gibbs sampler:

1. Initialize:
   (a) Instantiate X_i to one of its possible values $x_i, 1 \leq i \leq n$.
   (b) Let $\mathbf{x}^{(0)} = (x_1, \ldots, x_n)$
2. For $t = 1, 2, \ldots$
   (a) Pick an index $i, 1 \leq i \leq n$ uniformly at random.[1]
   (b) Sample x_i from $P(X_i | \mathbf{x}^{(t-1)}_{(-i)}, \mathbf{e})$.
   (c) Let $\mathbf{x}^{(t)} = (\mathbf{x}_{(-i)}, x_i)$

The sampler generates a sequence of samples $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(t)}, \ldots$ from the Markov chain over all possible states. The stationary distribution of the Markov chain is the joint distribution $P(X_1, \ldots, X_n | \mathbf{e})$. Thus, drawing samples from the Markov chain at *long enough* intervals, i.e., allowing enough time for the chain to reach the stationary distribution, gives independent samples from the distribution $P(X_1, \ldots, X_n | \mathbf{e})$.

2 Gibbs sampling in the Rain Network

The rain network is shown in Figure 2. Consider the inference problem of estimating $P(Rain | Sprinkler = true, WetGrass = true)$. Since *Sprinkler* and *WetGrass* is set to true, the Gibbs sampler draws

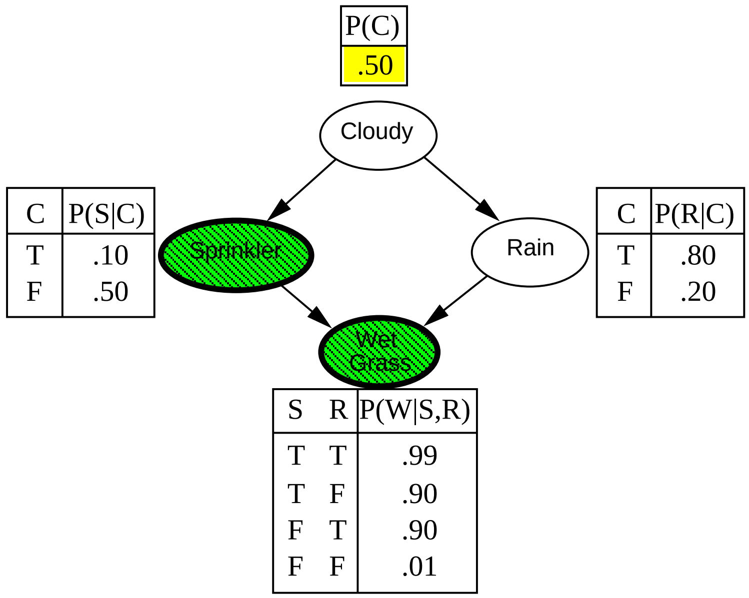

Figure 2.1: The Rain network

samples from $P(Rain, Cloudy|Sprinkler = true, WetGrass = true)$. The Gibbs sampler for the rain network works as follows:

1. Initialize:
   (a) Instantiate $Rain = true$, $Cloudy = true$
   (b) Let $\mathbf{x}^{(0)} = (Rain = true, Cloudy = true)$
2. For $t = 1, 2, \dots$
   (a) Pick variable to update from $\{Rain, Cloudy\}$ uniformly at random.
   (b) If $Rain$ was picked
      i. Sample $Rain$ from $P(Rain|Cloudy = [value]_{t-1}, Sprinkler = true, WetGrass = true)$
      ii. Let $\mathbf{x}^{(t)} = (Rain = [value]_t, Cloudy = [value]_{t-1})$
   (c) Else
      i. Sample $Cloudy$ from $P(Cloudy|Rain = [value]_{t-1}, Sprinkler = true, WetGrass = true)$
      ii. Let $\mathbf{x}^{(t)} = (Cloudy = [value]_t, Rain = [value]_{t-1})$

Thus, we need conditional probabilities of $Rain$ and $Cloudy$ given values for all other variables. Such conditional probabilities can be obtained using the CPT tables and the Markov blanket of the variable of interest. For convenience, we denote $Sprinkler = true$ by s, $WetGrass = true$ by

[1]Several other schedules are possible here. One of the simplest schedules is to go cyclically over all the indices.

w, $Rain = true$ by r, $Rain = false$ by $\neg r$, $Cloudy = true$ by c and $Cloudy = false$ by $\neg c$. With this notation, from Figure 2 it follows that

$$\begin{aligned}
P(c|r,s,w) &= P(c|s,r) = \frac{P(c)P(s,r|c)}{P(s,r)} \\
&= \frac{P(c)P(s|c)P(r|c)}{P(c)P(s|c)P(r|c) + P(\neg c)P(s|\neg c)P(r|\neg c)} \\
&= \frac{0.5 \times 0.1 \times 0.8}{0.5 \times 0.1 \times 0.8 + 0.5 \times 0.5 \times 0.2} \\
&= 0.4444
\end{aligned}$$

Clearly, $P(\neg c|r,s,w) = 1 - P(c|r,s,w) = \cancel{0.6923}$ 0.5556. The probabilities $P(c|\neg r,s,w)$ and $P(\neg c|\neg r,s,w)$ can be similarly computed. Also,

$$\begin{aligned}
P(r|c,s,w) &= \frac{P(r|c,s)P(w|r,c,s)}{P(w|c,s)} = \frac{P(r|c)P(w|r,s)}{P(w|c,s)} \\
&= \frac{P(r|c)P(w|r,s)}{P(r|c)P(w|r,s) + P(\neg r|c)P(w|\neg r,s)} \\
&= \frac{0.8 \times 0.99}{0.8 \times 0.99 + 0.2 \times 0.9} \\
&= 0.8148\ .
\end{aligned}$$

Clearly, $P(\neg r|c,s,w) = 1 - P(r|c,s,w) = 0.1852$. The probabilities $P(r|\neg c,s,w)$ and $P(\neg r|\neg c,s,w)$ can be similarly computed.

The above analysis gives the exact conditional probabilities needed by the Gibbs sampler described earlier in the section. Samples are drawn by running the Markov chain implemented by the Gibbs sampler to convergence (to the stationary distribution).