Akanksha Bansal (ab45563)
bansal@cs.utexas.edu

Assignment 4
April 11, 2015

Rejection Sampling is based of the idea that when you have a probability it is possible to sample from it. So given a Bayesian network and its conditional probabilities, it is possible to sample an atomic event. (An atomic event is an assignment of values to the random variables). The basic operation of direct sampling is that we sample each variable in topological order according to the conditional probability over its parents. Then we will reject those samples that do not match the evidence. Let $P(X|e)$ be the estimated distribution that the algorithm returns, then:

$P(X|e) = \frac{N(X,e)}{N(e)} \approx P(X|e)$

For example, consider that we want to compute the probability $P(Alarm|Earthquake = true)$ using 100 samples. Assume also that in 86 of these samples, the variable Earthquake appears with value false and in the remaining 14 it appears with value true. Then we reject the 86 samples where Earthquake appears with value false and considers only the rest. Now lets say that in 8 of these the value of Alarm has the equal to true. In this case the value that rejection sampling is going to return is:

$P(Alarm|Earthquake = true) =< \frac{8}{27}$

This sampling process is able to only approximate the correct value. The error of the approach has a standard deviation of $\frac{1}{n}$, where n is the number of samples used.

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, or to compute an integral. 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.

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

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

The sampler generates a sequence of samples $x_{(0)}, x_{(1)}, ..., x_{(t)}$, . . . from the Markov chain over all possible states. The stationary distribution of the Markov chain is the joint distribution

$P(X_1, ..., X_n|e)$. 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 $P(X_1, ..., X_n|e)$.

Results After running various tests these are my findings :

- The estimates provided by the Gibbs sampling are more accurate than Rejection Sampling. Across all my tests the accuracy of Gibbs Sampling varied from $95\% - 99\%$. However the accuracy of Rejection Sampling varied from $50\% - 90\%$
- The increase in accuracy with increase of sample size if more in Rejection sample than with Gibbs sampling for one Case.

Sample Size	Accuracy with Rejection Sampling	Accuracy with Gibbs Sampling
100	54.8%	95%
1000	78.3%	97.3%
10000	88.3%	98.6%

- The problem with rejection sampling is that it rejects too many samples. There was a lot of time invested in the generation of these samples that is not utilized by the method.
- When the results are very small, Rejection Sampling might not be able to provide an accurate result. In the babes Net I have taken as an example in my code base, with nodes such that $Recession- > Burglary$, the probability of recession being the case of Burglary is very low < 0.00001. Rejection Sample is able to provide a correct and only when the sample size is huge. When the Sample size is just 1000 Rejection sampling is not able to provide any results.