Statistical Techniques in Robotics (16-831, F10) Lecture#08 (Thursday September 16)

Gibbs Fields: Inference and Relation to Bayes Networks

Lecturer: Drew Bagnell *Scribe:Debadeepta Dey* [1]

1 Inference on Gibbs Fields

Once we have 'learnt' a Gibbs Field or a Markov Random Field (both of them are equivalent according to the Hammersly-Clifford theorem) the main question is how to do inference on the learnt graphical model. There are two major types of queries that are useful in general.

1. Marginals
2. Most probable assignment given some observed nodes

1.1 Marginals

$$P(X_i|observedX's) \tag{1}$$

The naive way of doing marginals is to sum over all possible combination of assignments to the unobserved variables.

Assuming that we have a graph whose clique potentials are Gaussians we can write the marginal in the conventional way as:

$$p(\vec{X}|observedX's) = (1/Z) * exp(-\sum_i \sum_{j \in N_i} f(X_i, X_j)) \tag{2}$$

Here N_i are the neighbors of the i^{th} node. If the random variables in the node are binary and there are n such nodes and k of these nodes are observed then the number of elements is of the order of 2^{n-k}. This is generally intractable as the number of nodes in a graph can easily be of the order of a few thousands or more.

We will explore 2 methods of doing this:

1. Importance Sampling
2. Gibbs Sampling

[1] Some content adapted from previous scribes: Byron Boots

1.1.1 Importance Sampling

Generate samples from

$$P(X) = \prod_i p(X_i) \tag{3}$$

while clamping nodes observed to the observed values. This means that we are not sampling the original distribution and hence we must weight each of the samples by the probability of the particular joint assignment.

The marginal is then obtained from all the weighted samples by the ratio of the sum of the weights of the samples which satisfy the query divided to the total weight of all samples.

$$p(\vec{X}|observedX's) = \frac{N_c}{N} \tag{4}$$

where N_c = sum of weights of samples which satisfy the query and N is the total weight of all samples.

1.1.2 Gibbs Sampling

The best way to explain Gibbs Sampling is to go over an example of how to do it and then try to get the intuition behind the procedure.

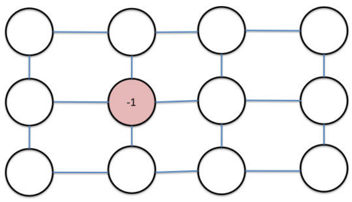

Figure 1: Example lattice graph with a single node observed as -1. This kind of single layer graph is also called an 'Ising' model.

Refer to Fig.1.1.2. It is a simple lattice structure also commonly called as the 'Ising' model. Here the node in red has been observed to be the value -1. Given this observation we want to know what the values of the rest of the nodes in the graph.

The pseudo-code for the Gibbs sampling algorithm is given in Algorithm.1. In line 1 you can randomly assign any value to all the unobserved nodes while fixing the observed nodes to the values observed . In lines $3-5$ you move over the graph fixing the value of every i^{th} node under consideration by sampling the conditional distribution for that particular node which is easy to do as the node under consideration is independent of the rest of the graph given its Markov blanket which in the case of Gibbs Fields are just its neighbors. This takes the form $P(x_i|N(x_i)) =$

$\frac{1}{Z}exp(\sum_{i,j \in N} f(x_i, x_j))$ where N is the set of neighbors of x_i. Please note that the updated value of x_i is used as soon as it is computed while computing the value of its neighbor. This fact is noted in Algorithm.1 in line 4 where some of the nodes have $t+1$ in the index while others have t. Once you have gone over the whole graph once this completes the inner loop and you start over with the current updated state of the graph. All through the procedure the observed nodes remain fixed.

Algorithm 1 Pseudo-code for the Gibbs sampling method on Gibbs Fields

1: $X^{(0)} :=< x_1^0, ..., x_k^0 >$
2: **for** $t = 1$ to T **do**
3:     **for** $i = 1$ to K **do**
4:         $x_i^{(t+1)} = P(x_i | x_1^{(t+1)}, ..., x_{i-1}^{t+1}, x_{i+1}^t, ..., x_k^t)$
5:     **end for**
6: **end for**

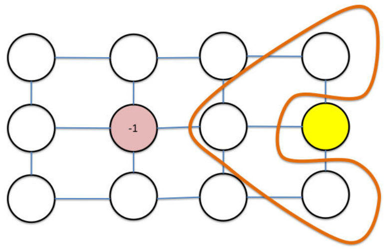

Figure 2: For the node under consideration (yellow), its value is determined by the conditional distribution over all its neighboring nodes shown in the figure. Then this process is repeated over all the nodes in turn while keeping the observed nodes fixed to observed values.

Intuition The Gibbs sampling procedure is an instance of the Metropolis-Hastings algorithm which is a type of Markov Chain Monte Carlo method. It is a Markov Chain method since it is generating current samples depending only the last state of the entire graph. It is a Monte Carlo method since the sampling is a probabilistic process.

The idea is that by repeatedly sampling in the manner outlined the algorithm will converge to the actual distribution of the graph. The actual distribution of the graph can be thought of as the stationary distribution underlying the markov chain we are generating in our sampling procedure.

The reader is directed to the read the first part of the excellent tutorial [1] for more intuition and the second part for all the related mathematical details.

Gibbs Sampling Problems

Critical Phenomena This is also sometimes referred to as 'Critical Slowing Down'. This happens when one node starts determining the value of another node really far away. This can start happening when the sampling has gone through many iterations but there are regions of high probablity and low probability now in the graph. These regions may be very highly correlated and hence will cause the sampling to produce series of values which are similar.

In general the stronger the connections between nodes the slower Gibbs sampling will run to converge in distribution.

Another good reference which goes into great detail on the topic is [2].

1.2 Most probable assignment given some observed nodes

$$\arg\max_{\vec{X}} P(\vec{X}|observedX's) \tag{5}$$

We will deal with this later.

2 Gibbs Fields and Bayesian Networks

You can go from Bayesian Networks to Gibbs Fields but not vice versa. Even when you convert a Bayesian Network to a Gibbs field you are going to lose information as will be illustrated in the example below.

You can not necessarily convert a Bayes' net into a Gibbs field. For example, consider the Bayes' net in Figure 3. If you remove the arrows (Figure 4), then the graph is *not* equivalent. In particular observing node B causes nodes A and C to become independent. This is the opposite of what the original graph represented. Instead, we need to *moralize* the graph. Whenever there are two parents that are not connected (married), we connect them. Thus, Figure 5 shows the correct representation of the original Bayes net. Note that during this conversion we actually lost information, namely that A and C are marginally independent.

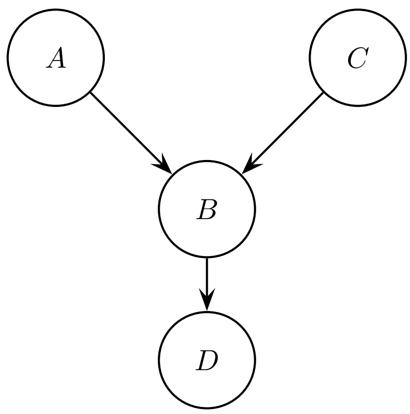

Figure 3: Bayes' Net

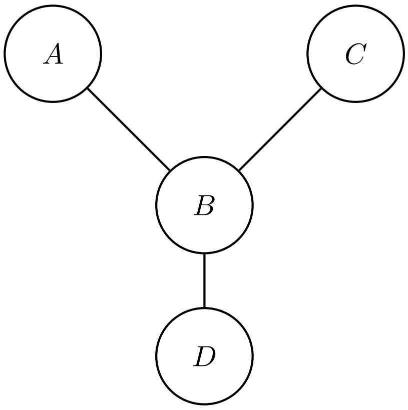

Figure 4: INCORRECT

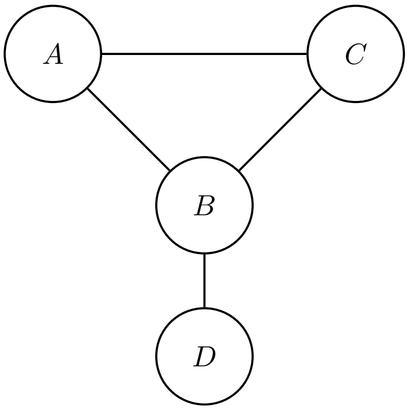

Figure 5: CORRECT: Moralizing

2.1 Gibbs Sampling of Bayesian Networks

References

[1] Philip Resnik and Eric Hardisty, *Gibbs Sampling for the Uninitiated*, UMIACS-TR-2010-04, (2010).

[2] Gilks, W.R. and Richardson, S. and Spiegelhalter, D.J., *Markov chain Monte Carlo in practice*, Chapman & Hall/CRC, (1996).