CS/CNS/EE 155: Probabilistic Graphical Models
Problem Set 3

Handed out: 9 Nov 2009
Due: 23 Nov 2009

Gibbs sampling

Preamble. Suppose that we are given a full joint distribution $p(Y_1, \ldots, Y_n)$ over random variables, but are only interested in drawing samples $Y_i \sim p(Y_i)$. The natural solution to this problem would be to actually calculate the marginal by integrating over $\{Y_1, \ldots Y_n\} - Y_i$:

$$p(Y_i) = \sum_{\{Y_1, \ldots Y_n\} - Y_i} p(Y_1, \ldots, Y_n),$$

from which samples Y_i might be drawn.

Unfortunately, in many cases (e.g., where the joint distribution has high treewidth), the required marginalization is intractable. In Homework 1, you showed that if $p(Y_1, \ldots, Y_n)$ is given as a graphical model, and $Y_1, \ldots, Y_n$ are all discrete, and you have observations of all variables *except* Y_k, computing the conditional distribution

$$p(Y_k \mid Y_1 = y_1, \ldots, Y_{k-1} = y_{k-1}, Y_{k+1} = y_{k+1}, \ldots, Y_n = y_n)$$

is very efficient. In this case, sampling from this conditional distribution only requires sampling from a Bernoulli random variable. Gibbs sampling is an approximate inference method peculiarly well suited to such situation, where one can sample from the conditional distributions

$$p(Y_k | \{Y_1, \ldots Y_n\} - Y_k) \quad \forall k.$$

As outlined in both [KF09, Bis06], the algorithm for Gibbs sampling is quite straightforward:

1. Generate an initial guess ($t = 0$) about the state $(Y_1 = y_1^{t=0}, \ldots, Y_n = y_n^{t=0})$.
2. Repeat, over t, until convergence (a very loaded statement—see mixing times, stationary distributions, etc):

   Draw $y_1^{t+1} \sim p(Y_1 | Y_2 = y_2^t, \ldots Y_n = y_n^t)$
   Draw $y_2^{t+1} \sim p(Y_2 | Y_1 = y_1^{t+1}, Y_3 = y_3^t \ldots Y_n = y_n^t)$
   $\vdots$
   Draw $y_n^{t+1} \sim p(Y_n | Y_1 = y_1^{t+1}, Y_2 = y_2^{t+1} \ldots Y_{n-1} = y_{n-1}^{t+1})$

As it turns out, these samples $\{y_i^1, y_i^2, \ldots, y_i^T\}$ not only converge (in the sense that given a long enough burn in time, the samples are all coming from the same distribution), but converge to the "correct" distribution, which in this case is the marginal $p(Y_i)$.

Problem.

1. Write a Gibbs sampler to estimate the marginal distribution generated from the conditional distributions

$$p(x|y) \propto ye^{-yx}, \quad 0 < x < B < \infty$$
$$p(y|x) \propto xe^{-xy}, \quad 0 < y < B < \infty$$

   where B is a known positive constant (as we will see later, the restriction to the interval $x, y \in (0, B)$ ensures that the marginal $p(x)$ exists).

   For $B = 5$, and for sample sizes $T = 500, 5000, 50000$, plot the histogram of values for x.

   Note that although the simulation of $p(x)$ is straightforward using a Gibbs sampler, the marginal is not obvious from inspection of the above marginals.

2. Importantly, simulation based approximations provide straightforward methods of calculating the moments of the marginal distributions. Provide an estimate of the expectation of X, $\mathbb{E}_{p(X)}[X]$ by using the 500, 5000 and 50000 samples from 1).

3. [*Extra Credit:*] Consider the bivariate case, where we are given $p_{X|Y}(X|Y)$ and $p_{Y|X}(Y|X)$. As it turns out, we can derive a *fixed point integral equation* to determine the marginal density $p_X(X)$, and thus the joint density, using repeated applications of the chain rule and marginalization: first, notice that

$$p_X(X) = \int p_{XY}(X,Y)dY = \int p_{X|Y}(X|Y)p_Y(Y)\,dY.$$

   However, we can similarly substitute for $p_Y(Y)$:

$$\begin{aligned} p_X(X) &= \int p_{X|Y}(X|Y) \int p_{Y|X}(Y|Z)p_X(Z)\,dZ\,dY \quad (Z \text{ is a dummy variable}) \\ &= \int \left[\int p_{X|Y}(X|Y)p_{Y|X}(Y|Z)\,dY \right] p_X(Z)\,dZ \\ &= \int h(X,Z)p_X(Z)\,dZ. \end{aligned}$$

   Using the normalized conditionals from problem 1

$$p(x|y) = \frac{ye^{-yx}}{1 - e^{-By}}, \quad 0 < x < B < \infty$$
$$p(y|x) = \frac{xe^{-xy}}{1 - e^{-Bx}}, \quad 0 < y < B < \infty$$

   calculate $p_x(x)$. *Hint:* $p_x(x) \propto (1 - e^{-Bx})/x$ *– find the normalizer and verify that* p_x *solves the fixed point equations.*

4. [*Extra Credit:*] Now that we have an analytic form for the marginal $p_x(x)$ from problem 3, we can compare with our earlier estimates.

   (a) Plot $p_x(x)$ and the histogram of collected samples of $p_x(x)$.

   (b) Compute $\mathbb{E}_{p_x(x)}[x]$ using the analytic form found in problem 3.

References

[Bis06] C. Bishop. *Pattern Recognition and Machine Learning.* Springer Science+Business Media, LLC, New York, NY, 2006.

[KF09] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques.* The MIT Press, Cambridge, MA, 2009.