STAT 534
Lecture of 05/21

mmp@stat.washington.edu
Scribes: Ashwin T.A.N

1 Gibbs Sampling

The goal of Gibbs sampling is to obtain samples from a complex multivariate probability distribution of the form

$$\pi(x_1, x_2, \ldots, x_n) = \frac{\varphi(x_1, \ldots, x_n)}{Z}, \qquad x_i \in \Omega_i,$$

where φ is some non-negative function on $\Omega = \prod_{i=1}^{n} \Omega_i$ and Z is the normalization constant. The sets Ω_i can be discrete or continuous, but we will deal only with the discrete case for simplicity. The idea behind Gibbs sampling is that it might be much easier to sample from the univariate conditional distributions $P[X_j | x_1, \ldots x_{j-1}, x_{j+1}, \ldots, x_n]$ than it is to draw i.i.d samples from the multivariate distribution π. For instance, we will consider below the example of the Ising model, where directly computing the normalization constant Z turns out to be intractable.

The Gibbs sampling algorithm generates a Markov chain $\mathbf{X}^i = (X_1^i, \ldots, X_n^i)$, $i = 0, 1, 2, \ldots, T$ such that the empirical distribution of the samples $\{\mathbf{X}^1, \ldots, \mathbf{X}^T\}$ approximates the joint probability distribution π. Henceforth, we let $\mathbf{x}_{-i}$ denote $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

Algorithm 1: Gibbs Sampling

Initialize $\mathbf{X}^0 = (X_1^0, \ldots, X_n^0)$
for $t = 1, 2, \ldots, T$ **do**
    Select i randomly from $\{1, \ldots, n\}$ with uniform probability
    Sample $X_i^t \sim P[X_i | \mathbf{X}_{-i}^{t-1}]$
    Set $\mathbf{X}_{-i}^t = \mathbf{X}_{-i}^{t-1}$
end
Output: $\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^T$.

Under suitable assumptions on the transition probability matrix $(T_{\mathbf{x},\mathbf{x}'})$ (to be discussed later), $\mathbf{X}^0, \mathbf{X}^1, \mathbf{X}^2, \ldots$ will be a Markov chain with stationary distribution π. Note however that $\mathbf{X}^1, \ldots, \mathbf{X}^T$ need not be i.i.d.

2 The Ising model

The Ising model is a graphical probabilistic model with origins in statistical physics. Let G be a graph with vertex set $V = \{1, 2, \ldots, n\}$ and edge set $\mathcal{E}$. Each $i \in V$ is associated with a random variable X_i that takes values in $\{+1, -1\}$. The Ising model assumes that $(X_1, \ldots, X_n)$ has a joint probability distribution of the form

$$\pi(x) = \frac{1}{Z}\exp\left(-\sum_{i\in V} h_i x_i - \sum_{\{i,j\}\in\mathcal{E}} h_{ij} x_i x_j\right), \qquad x \in \Omega = \{\pm 1\}^n,$$

where h_i, h_{ij} are parameters and Z is the appropriate normalization constant. Choosing $h_{ij} < 0$ for all $\{i,j\} \in \mathcal{E}$ amounts to assuming that states where a vertex has the same sign as its neighbors have a higher probability. We sometimes use the notation $i \sim j$ to indicate that $\{i,j\} \in \mathcal{E}$.

Note that an explicit computation of Z would require $\Omega(2^{\max(n,|\mathcal{E}|)})$ calculations and hence is not tractable for large n. However, the conditional probability distributions have a relatively simple form and are easy to sample from:

$$\begin{aligned}
P[X_i = 1 | \mathbf{X}_{-i} = \mathbf{x}_{-i}] &= \frac{P[X_i = 1, \mathbf{X}_{-i} = \mathbf{x}_{-i}]}{P[\mathbf{X}_{-i} = \mathbf{x}_{-i}]} \\
&= \frac{P[X_i = 1, \mathbf{X}_{-i} = \mathbf{x}_{-i}]}{P[X_i = 1, \mathbf{X}_{-i} = \mathbf{x}_{-i}] + P[X_i = -1, \mathbf{X}_{-i} = \mathbf{x}_{-i}]} \\
&= \frac{\exp\left(-h_i - \sum_{j\sim i} h_{ij} x_j\right)}{\exp\left(-h_i - \sum_{j\sim i} h_{ij} x_j\right) + \exp\left(h_i + \sum_{j\sim i} h_{ij} x_j\right)} \\
&= \sigma\left(-2\left(h_i + \sum_{j\sim i} h_{ij} x_j\right)\right),
\end{aligned}$$

where σ is the logistic function $\sigma(z) = \frac{1}{1+e^{-z}}$. Therefore, it is natural to use the Gibbs sampling algorithm to sample from π. We should also that this model satisfies the so-called local Markov property: for every $i \in V$,

$$X_i \perp \text{ all other } X_k \,|\, \{X_j | j \sim i\}.$$