COMPSCI 688: Probabilistic Graphical Models  
Lecture 15: Gibbs and Metropolis-Hastings Samplers

Dan Sheldon  
Manning College of Information and Computer Sciences  
University of Massachusetts Amherst

Partially based on materials by Benjamin M. Marlin (marlin@cs.umass.edu) and Justin Domke (domke@cs.umass.edu)

Review

A Markov chain is regular if there is a $t$ such that $(T^t)_{ij} > 0$ for all $i, j$. It is possible to get from any state $i$ to any state $j$ in exactly $t$ steps. A regular Markov chain has a unique stationary distribution and is guaranteed to converge to it.

A Markov chain $T$ satisfies detailed balance with respect to $\pi$ if $\forall x, x'$,

$$\pi(x)T(x'|x) = \pi(x')T(x|x').$$

Detailed balance implies $\pi$ is a stationary distribution of $T$.

MCMC idea: given $\pi$, design a regular Markov chain that satisfies detailed balance with respect to $\pi$. Then samples from the Markov chain converge to $\pi$. (Specify the transitions $T(x'|x)$ “algorithmically”, since the state space is huge.)

Gibbs Sampler Algorithm

\begin{itemize}
    \item Initialize $x = (x_1, \ldots, x_D)$
    \item $x^{(0)} \leftarrow x$
    \item For $t = 1$ to $S$
        \begin{itemize}
            \item For $i = 1$ to $D$
                \begin{itemize}
                    \item Sample $r$ from $p(X_i | X_{-i} = x_{-i})$
                    \item $x^{(t)}_i \leftarrow r$
                \end{itemize}
            \end{itemize}
    \end{itemize}
\end{itemize}

We need to show

1. Regularity
2. Detailed Balance
Gibbs Sampler Correctness

Regularity for Gibbs Sampling

We need to show it is possible to transition from \( x \) to \( x' \) in exactly \( t \) time steps for some \( t \) and arbitrary \( x, x' \).

**Question:** Assume the full conditionals satisfy \( p(x_i|x_{-i}) > 0 \) always, e.g. because \( p(x) > 0 \). Is this condition true for Gibbs sampling? For what \( t \) is it true?

**Answer:** True for \( t = 1 \). Update \( x_i = x'_i \) in one loop through all variables.

Gibbs Sampler Correctness

Detailed Balance for Gibbs Sampling

- The Gibbs sampler re-samples the value of every variable \( X_i \) in sequence from the full conditional \( p(X_i|X_{-i} = x_{-i}) \).
- We can view this as simulating a Markov chain with a sequence of transition operators, one for every variable:
  \[
  T_i(x'|x) = p(x'_i|x_{-i}) I[x_{-i} = x'_{-i}]
  \]
- We’ll show that each of these operators satisfies detailed balance with respect to the full distribution \( p \). The full result then follows from the fact that the composition of operators satisfying detailed balance also satisfies detailed balance.

Claim: For all \( i \), the operator \( T_i \) satisfies detailed balance with respect to \( p \).

Proof:

\[
\text{RHS} = p(x) T_i(x'|x) = \left[ p(x) T_i(x'|x) \right] I[x_i = x'_i]
\]

\[
= p(x') T_i(x'|x') I[x_i = x'_i]
\]

\[
= p(x') T_i(x'|x') I[x_i = x'_i]
\]

\[
= p(x^0) T_i(x'|x^0)
\]

\[
= p(x^0) T_i(x'|x^0)
\]

\[
\text{Want} : p(x) T_i(x'|x) = p(x) T_i(x|x')
\]
Applications and Limitations of The Gibbs Sampler

- The Gibbs sampler is great for graphical models because the single variable conditionals only depend on factors involving that variable.
- The Gibbs sampler can work with unnormalized densities, including Markov networks, without needing to compute the partition function. Why? $p(x_i|x_j) \propto p(x_j|x_i)$. 
- The Gibbs sampler can always be used with discrete distributions, because the conditionals are always available in exact form.
- For continuous distributions, it may be harder or impossible to sample from the conditional distributions.
- The Gibbs sampler can be “slow mixing” (take a long time to converge) if correlations between variables are high.