COMPSCI 688: Probabilistic Graphical Models

Lecture 15: Gibbs Sampler Correctness

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)

Gibbs Sampler Correctness

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 π if $\forall x, x'$,

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

  Deatiled balance implies π is a stationary distribution of T.
- MCMC idea: given π, design a regular Markov chain that satisfies detailed balance with repect to π. Then samples from the Markov chain converge to π. (Specifiy the transitions $T(x'|x)$ "algorithmically", since the state space is huge.)

Gibbs Sampler Algorithm

Gibbs sampler

- Initialize $\mathbf{x} = (x_1, \ldots, x_D)$
- $\mathbf{x}^{(0)} \leftarrow \mathbf{x}$
- For $t = 1$ to S
  - For $i = 1$ to D
    - Sample r from $p(X_i \,|\, \mathbf{X}_{-i} = \mathbf{x}_{-i})$
    - $x_i \leftarrow r$
  - $\mathbf{x}^{(t)} \leftarrow \mathbf{x}$
- Return $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(S)}$

We need to show

1. Regularity
2. Detailed Balance

Gibbs Sampling Picture

Regularity for Gibbs Sampling

We need to show it is possible to transition from $\mathbf{x}$ to $\mathbf{x}'$ in exactly t time steps for some t and arbitrary $\mathbf{x}, \mathbf{x}'$.

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

Answer: It is true for $t = 1$. Recall that we sweep through all variables in a single time step, sweeps. For each i there is positive probability of moving from x_i to x_i'

Detailed Balance for Gibbs Sampling

- The Gibbs sampler re-samples the value of every variable X_i in sequence from the full conditional $p(X_i|\mathbf{X}_{-i} = \mathbf{x}_{-i})$
- We can view this as simulating a Markov chain with a *sequence* of transition operators, one for every variable:

$$T_i(\mathbf{x}'|\mathbf{x}) = p(x_i'|\mathbf{x}_{-i})\mathbb{I}[\mathbf{x}_{-i} = \mathbf{x}'_{-i}]$$

- 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 satsifies detailed balance.

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

Proof:

$$\begin{aligned} p(\mathbf{x}')T_i(\mathbf{x}|\mathbf{x}') &= p(\mathbf{x}')p(x_i|\mathbf{x}'_{-i})\mathbb{I}[\mathbf{x}_{-i} = \mathbf{x}'_{-i}] \\ &= p(\mathbf{x}'_{-i})p(x_i'|\mathbf{x}'_{-i})p(x_i|\mathbf{x}'_{-i})\mathbb{I}[\mathbf{x}_{-i} = \mathbf{x}'_{-i}] \\ &= p(\mathbf{x}_{-i})p(x_i'|\mathbf{x}_{-i})p(x_i|\mathbf{x}_{-i})\mathbb{I}[\mathbf{x}_{-i} = \mathbf{x}'_{-i}] \\ &= p(\mathbf{x})T_i(\mathbf{x}'|\mathbf{x}). \end{aligned}$$

Gibbs Sampling Picture 2

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?
- 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.

Metropolis-Hastings

The Metropolis-Hastings Sampler

- The Metropolis Hastings sampler is an extremely general sampler based on the idea of "proposing" a new state with a *proposal distribution* $q(\mathbf{x}'|\mathbf{x})$, and then "accepting" or "rejecting"
- Like the Gibbs sampler, it can be used with continuous or discrete distributions and avoids computation of the partition function.
- Unlike the Gibbs sampler, it doesn't require the ability to sample from the conditional distributions.

Proposal and Acceptance MCMC Illustration

Proposal and Acceptance MCMC

Initialize x
for $t = 1, 2, 3, \dots, S$:
  Sample $x' \sim q(x'|x)$
  Look at x and x', and calculate a
   probability $\alpha(x, x')$ of keeping x'.
  Choose $r \in [0, 1]$ uniformly
  If $r < \alpha(x, x')$**then**
   $x \leftarrow x'$
  $x^{(t)} \leftarrow x$
return $x^{(1)}, x^{(2)}, \dots, x^{(S)}$

How to Choose Acceptance Probability?

The key missing step is how to set the acceptance probability $\alpha(x, x')$. It can depend on p and q. The transition probability density is

$$T(x'|x) = \begin{cases} q(x'|x)\alpha(x, x') & \text{if } x \neq x' \\ ? & \text{if } x = x' \end{cases}$$

Our goal is to satisfy detailed balance, i.e., for all x, x':

$$p(x)T(x'|x) = p(x')T(x|x')$$

$$\iff p(x)q(x'|x)\alpha(x, x') = p(x')q(x|x')\alpha(x', x)$$

We don't care about $T(x'|x)$ when $x = x'$, because the detailed balance condition is always satsified for $x = x'$.

There are different acceptance rules $\alpha(x, x')$ that ensure detailed balance. Metropolis-Hastings is based on the adjusting the larger "flow" to be equal to the smaller one.

$$\underbrace{p(x)q(x'|x)}_{x \to x' \text{ flow}} \underbrace{\alpha(x, x')}_{\text{adjustment}} = \underbrace{p(x')q(x|x')}_{x' \to x \text{ flow}} \underbrace{\alpha(x', x)}_{\text{adjustment}}$$

The rule is

- If $x \to x'$ flow $>$ $x' \to x$ flow, set $\alpha(x, x')$ equal to their ratio, and set $\alpha(x', x) = 1$
- If $p(x)q(x'|x) > p(x')q(x|x')$, set $\alpha(x, x') = \dfrac{p(x')q(x|x')}{p(x)q(x'|x)}$ and set $\alpha(x', x) = 1$.

By symmetry, the general Metropolis-Hastings acceptance rule is:

$$\boxed{\alpha(x, x') = \min\left\{1, \frac{p(x')q(x|x')}{p(x)q(x'|x)}\right\}}$$

Proof of Detailed Balance

Claim: detailed balance holds with $\alpha(x, x') = \min\left\{1, \frac{p(x')q(x|x')}{p(x)q(x'|x)}\right\}$

Proof: First, consider when $p(x)q(x'|x) > p(x')q(x|x')$. Then $\alpha(x, x') = \frac{p(x')q(x|x')}{p(x)q(x'|x)}$ and $\alpha(x', x) = 1$, and we have

$$\begin{aligned} p(x)T(x'|x) &= p(x)q(x'|x)\alpha(x, x') \\ &= p(x)q(x'|x)\frac{p(x')q(x|x')}{p(x)q(x'|x)} \\ &= p(x')q(x|x') \\ &= p(x')q(x|x')\alpha(x', x) \\ &= p(x')T(x|x'). \end{aligned}$$

For the second case, we have $p(x')q(x|x') > p(x)q(x'|x)$. The proof is the same as the first case, with x and x' swapped.

Metropolis-Hastings Algorithm

Initialize x
for $t = 1, 2, 3, \ldots, S$:
  Sample $x' \sim q(x'|x)$
  Choose $r \in [0, 1]$ uniformly
  if $r < \frac{p(x')Q(x|x')}{p(x)Q(x'|x)}$ **then**
    $x \leftarrow x'$
  $x^{(t)} \leftarrow x$
return $x^{(1)}, x^{(2)}, \ldots, x^{(S)}$

Gaussian Random Walk Sampler

A simple proposal uses a Gaussian random walk as the proposal distribution:

$$\mathbf{x}' \sim \mathcal{N}(\mathbf{x}'|\mathbf{x}, \sigma^2 I)$$

By symmetry, the acceptance probability simplifies

$$\alpha(\mathbf{x}', \mathbf{x}) = \frac{p(\mathbf{x}')\mathcal{N}(\mathbf{x}|\mathbf{x}', \sigma^2 I)}{p(\mathbf{x})\mathcal{N}(\mathbf{x}'|\mathbf{x}, \sigma^2 I)} = \frac{p(\mathbf{x}')}{p(\mathbf{x})}$$

Demo: Gaussian Random Walk Sampler