

# 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  $\pi$  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

### Gibbs sampler

- ▶ Initialize  $\mathbf{x} = (x_1, \dots, x_D)$
- ▶  $\mathbf{x}^{(0)} \leftarrow \mathbf{x}$
- ▶ 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  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(S)}$

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 satisfies 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.