

Veural Information Processing Systems Foundation

1. Gibbs Sampling

- Machine learning systems use probabilistic inference to cope with uncertainty
- Exact inference is often intractable
- -Approximate Markov chain Monte Carlo techniques are used instead
- -Gibbs sampling is one of the most popular MCMC techniques

Algorithm 1 Gibbs sampler

input Variables x_i for $i \in [n]$, and distribution π Initialize x_1, \ldots, x_n arbitrarily loop Select variable index s from $\{1, \ldots, n\}$ Sample x_s from conditional distribution $\mathbf{P}_{\pi}\left(X_{s} \mid X_{\{1,\ldots,n\}\setminus\{s\}}\right)$ end loop

2. Scan Order

-What order do you sample the variables in?

-Two common scan orders:

Random scan: sample uniformly and independently Systematic scan:

sample in a fixed permutation

-Systematic scan has better hardware efficiency due to spatial locality

- Most theoretical results only for random

-Which scan has better statistical efficiency? (smaller mixing time)

SCAN ORDER IN GIBBS SAMPLING: Models in Which it Matters and Bounds on How Much

Bryan He, Christopher De Sa, Ioannis Mitliagkas, Christopher Ré {bryanhe,cdesa,imit,chrismre}@stanford.edu

3. Folklore

- Scan order does not really matter, but systematic is slightly better.
- -Random can only be constant factors faster than systematic
- -Systematic can only be log factors faster than random

4. Our Contributions

5. Models

We introduce two models to show where the folklore breaks down.

Sequence of Dependencies



Variables: x_1, \ldots, x_n

 $-x_i$ is never true unless x_i is true

-Variables are independently likely to be true (M is large)

$$P(x) \propto \begin{cases} 0 & \text{if } x_i \text{ and not } x_{i-1} \\ M^{\sum_{i=1}^n x_i} & \text{otherwise} \end{cases}$$

- Random mixes in $O(n^2)$
- -Systematic x_1, x_2, \ldots, x_n mixes in O(n)
- -Systematic $x_n, x_{n-1}, \ldots, x_1$ mixes in $O(n^2)$

-Two models showing that

- Systematic can mix much faster than random - Random can mix much faster than systematic - Permutation used by systematic scan matters

- Analysis techniques for comparing mixing times

-Bounds on relative mixing times of different scans



Variables: $x_1, \ldots, x_n, y_1, \ldots, y_n$

-x variables and y variables contradict (never true at the same time)

$$P(x, y) \propto \begin{cases} 0 & \text{if } \exists x_i \text{ true and } \exists y_j \text{ true} \\ 1 & \text{otherwise} \end{cases}$$

-Systematic $x_1, \ldots, x_n, y_1, \ldots, y_n$ takes O(n) times as long as random to mix

-Systematic $x_1, y_1, x_2, y_2, \ldots, x_n, y_n$ mixes a constant factor faster than random

-We introduce techniques for comparing relative mixing times with conductance

(1/2 -

-Often imply that the relative mixing times differ by only polynomial factors

Our experiments analyze how different scans behave on our models.

(thousands)

Island ON Mass



6. Mixing Time Bounds

$$(1/2 - \epsilon)^2 t_{\min}(R, \epsilon) \le 2t_{\min}^2(S, \epsilon) \log\left(\frac{1}{\epsilon \pi_{\min}}\right)$$
$$+ \epsilon)^2 t_{\min}(S, \epsilon) \le \frac{8n^2}{\left(\min_{x,i} P_i(x, x)\right)^2} t_{\min}^2(R, \epsilon) \log\left(\frac{1}{\epsilon \pi_{\min}}\right)$$

7. Experiments

