Poisson-Minibatching for Gibbs Sampling with Convergence Rate Guarantees

Ruqi Zhang and Christopher De Sa

Overview

Gibbs sampling
+ Convergence guarantees
– Not scale to large datasets or models

Subsampling methods
+ Scale MCMC
– No theoretical guarantees

Poisson-Gibbs – a novel minibatch Gibbs sampler

- ▹ Low computational cost
- ▹ ***Unbiased*** even without the M-H step
- ▹ Support both discrete and ***continuous*** distributions

We provide convergence rate guarantees

- ▹ Hyper-parameter selection guided by theoretical bounds

Gibbs Sampling on Graphical Models

Consider factor graphs

$$\pi(x) = \frac{1}{Z} \cdot \prod_{\phi \in \Phi} \exp\left(\phi(x)\right)$$

Sample from π by Gibbs sampling

Loop
Select a variable x_i to sample at random
Compute the conditional distribution of x_i based on ***all factors*** ϕ that depend on x_i
Resample variable x_i from the conditional distribution
End Loop

Very expensive when the factor set is large!

Future Work

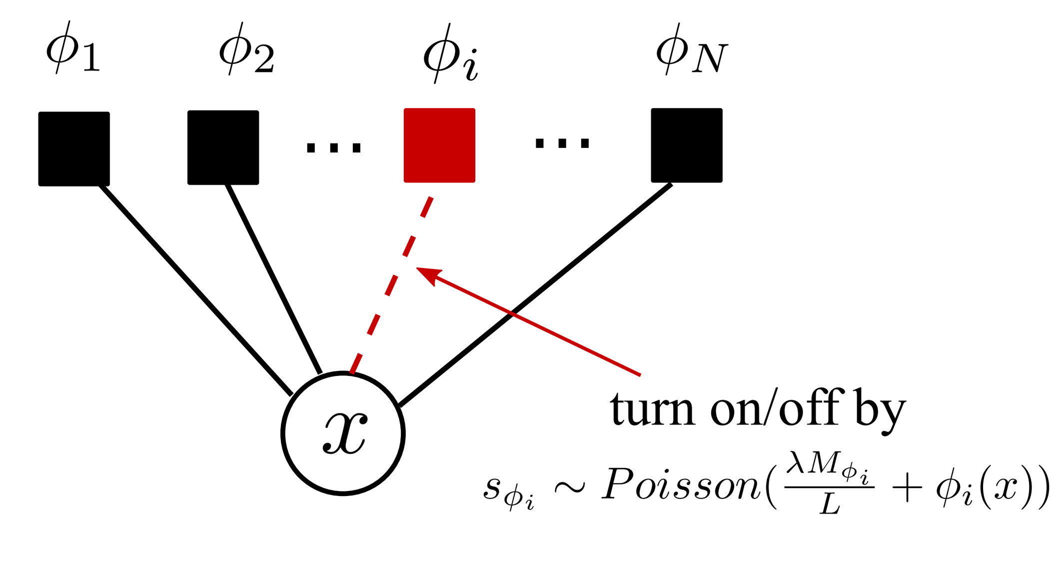
Apply Poisson-minibatching to other MCMC methods

- ▹ We applied Poisson-minibatching to Metropolis-Hasting sampling—***Poisson-MH***
- ▹ Poisson-MH is ***unbiased***
- ▹ Poisson-MH has a ***guaranteed convergence rate***
- ▹ An experimental demonstration is in the paper

Interesting to combine Poisson-minibatching with more MCMC methods

Poisson-Gibbs

A new minibatch method based on auxiliary variables



The joint distribution

- ▹ Keep the marginal distribution of x ***unchanged***
- ▹ Allow a factor ϕ to contribute to the energy only when $s_\phi > 0$

$$\pi(x, s) \propto \exp\left(\sum_{\phi\in\Phi}\left(s_\phi \log\left(1 + \frac{L}{\lambda M_\phi}\phi(x)\right) + s_\phi \log\left(\frac{\lambda M_\phi}{L}\right) - \log\left(s_\phi !\right)\right)\right)$$

An *upper bound* on the expected number of factors being used

$$\mathbf{E}\left[\left|\{\phi \in A[i] \mid s_\phi > 0\}\right|\right] \leq \lambda + L$$

Convergence Rate Guarantees

Issue: possibly intractable continuous conditional distributions

Solution: Double Chebyshev Approximation

- ▹ Get polynomial approximation of the PDF by using ***Chebyshev polynomial approximation*** twice
- ▹ Generate a sample by ***inverse transform sampling***

Convergence rate guarantees: higher spectral gap, higher convergence rate

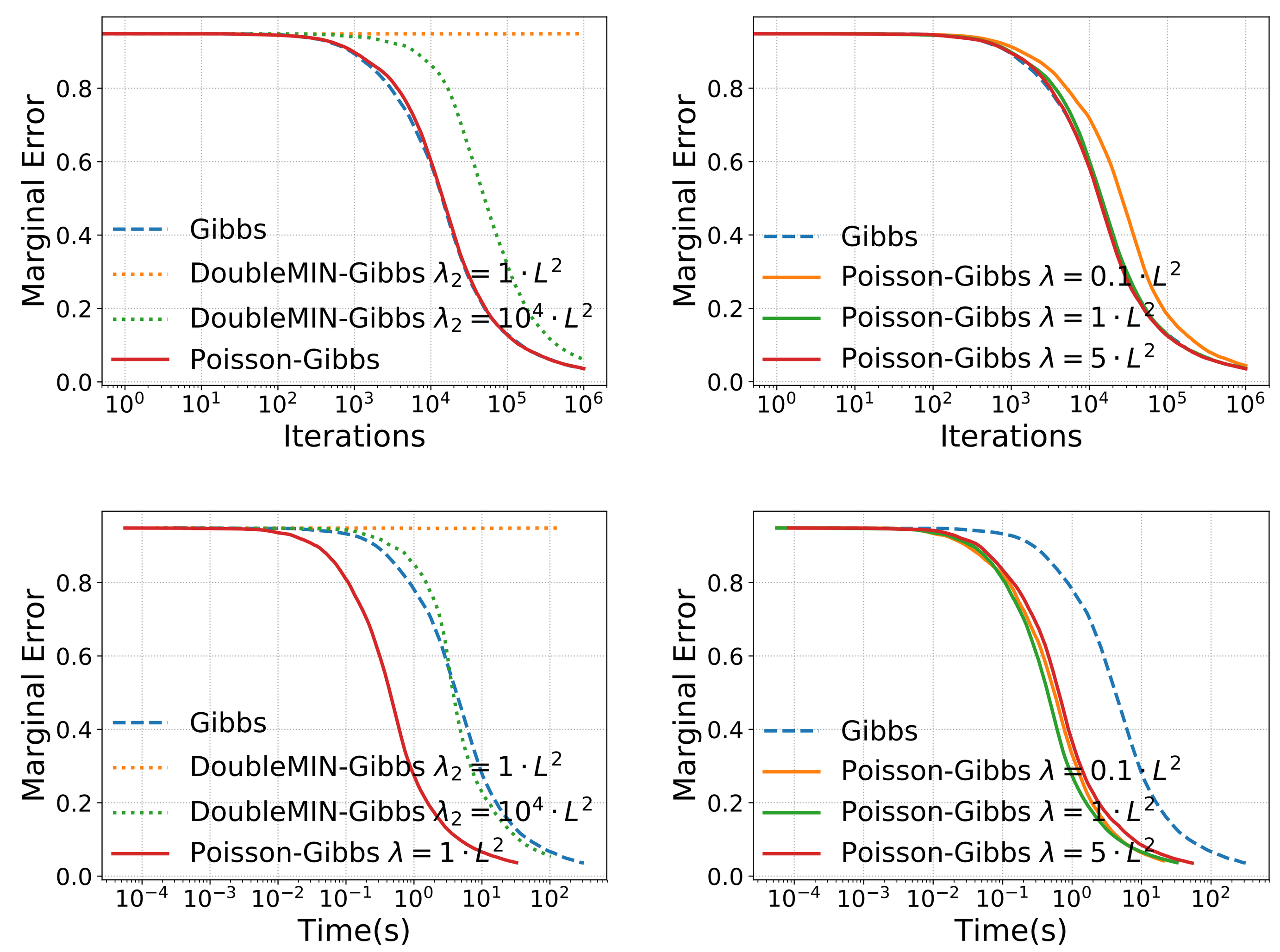
$$\underbrace{\bar{\gamma}}_{\text{spectral gap of Poisson-Gibbs}} \geq \alpha \cdot \exp\left(-\frac{4L^2}{\lambda}\right) \cdot \underbrace{\gamma}_{\text{spectral gap of Gibbs}}$$

where

$$\alpha = \begin{cases} 1, & \text{exact sampling from conditionals (e.g. discrete)} \\ \underbrace{(1 - 4\sqrt{F})}_{\text{Chebyshev approx. error}}, & \text{continuous} \end{cases}$$

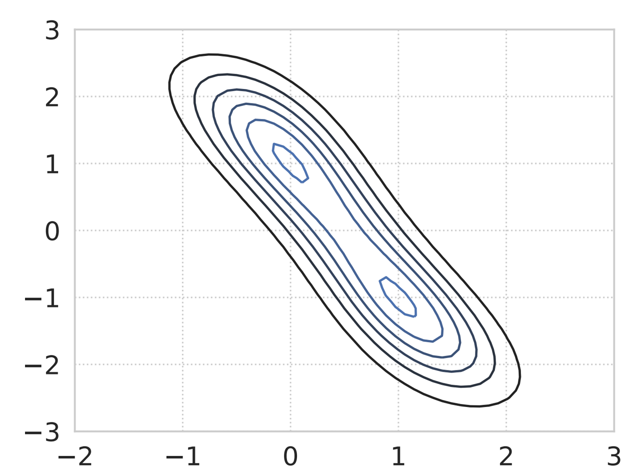
- ▹ Recipe for minibatch size: $\lambda = \Theta(L^2)$, then the constant becomes $O(1)$

Experiments

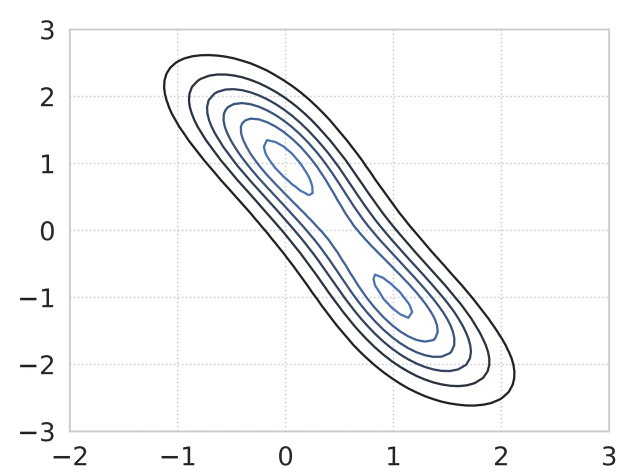


Potts model

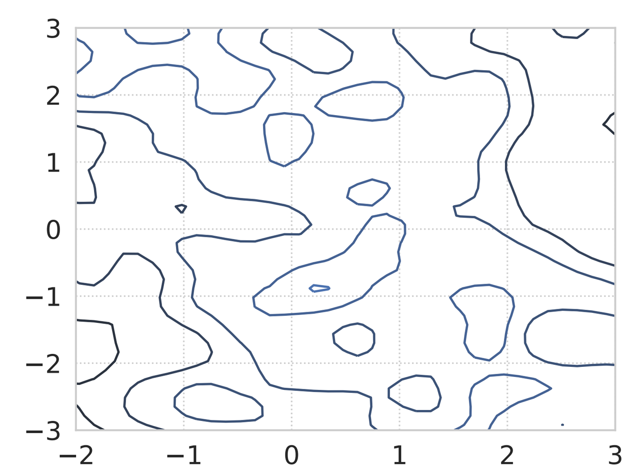
- ▹ Poisson-Gibbs is faster; validate our theory



(a) Ground Truth



(b) PGDA



(c) Gibbs-rejection

Gaussian Mixture (10^6 factors)

- ▹ PGDA only uses 1800 factors on average
- ▹ PGDA estimates the density accurately while rejection sampling fails