Advanced topics in MCMC

KALANI H

GUEST LECTURE – STAT951

Overview

Auxiliary variable methods

- Simulated Tempering
- Slice Sampler

Perfect sampling

- Coupling from the past (CFTP)

Auxiliary variable methods

01

Standard MCMC can take too long to converge.

02

Augment the state space of the variable of interest

03

Leads to chains that converge faster and require less tuning

04

Mostly used in Bayesian spatial lattice models

General idea

$\boldsymbol{X} \sim f(\boldsymbol{x})$

Let $\boldsymbol{U}$ be a vector

Then we augment $\boldsymbol{X}$ with $\boldsymbol{U}$, $(\boldsymbol{X}, \boldsymbol{U}) \sim f(\boldsymbol{x}, \boldsymbol{u})$

Construct the Markov chain over the joint distribution of $(\boldsymbol{X}, \boldsymbol{U})$

Simulate $(\boldsymbol{X}^{(t)}, \boldsymbol{U}^{(t)})$ and then discard $\boldsymbol{U}^{(t)}$ and infer about the marginal distribution of $\boldsymbol{X}$

Simulated Tempering

High number of dimensions, multimodality or slow MCMC with extremely long chains

Potential solution

General Idea

Based on a sequence of densities, f_i for $i = 1 \ldots m$

Cold $(i = 1)$ and hot $(i = m)$

For $(\boldsymbol{X}, I)$, where $I \sim \mathrm{p(i)}$ – random prior

Starting values $(\boldsymbol{x}^{(0)}, i^{(0)})$

Use Metropolis-Hastings or Gibbs sampler to draw $\boldsymbol{X}^{(t+1)} | i^{(t)}$ from a stationary distribution $f_{i(t)}$

Generate I^* from a proposal density, $g(.|i^{(t)})$

$$g(.|i^{(t)}) = \begin{cases} 1 & if\ \left(i^{(t)}\middle|i^*\right) = (m, m-1) \\ \frac{1}{2} & if\ \left|i^* - i^{(t)}\right| = 1 \\ 0 & otherwise \end{cases}$$

Define the Metropolis-Hastings ratio $R_{ST}\left(I^*, i^{(t)}, \boldsymbol{X}^{(t+1)}\right)$,

$$R_{ST}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{z}) = \frac{f_{\boldsymbol{v}}(\boldsymbol{z})p(\boldsymbol{v})g(\boldsymbol{u}|\boldsymbol{v})}{f_{\boldsymbol{u}}(\boldsymbol{z})p(\boldsymbol{u})g(\boldsymbol{v}|\boldsymbol{u})}$$

$$I^{(t+1)} = \begin{cases} I^* & with\ probability\ min\left\{R_{ST}\left(I^*, i^{(t)}, \boldsymbol{X}^{(t+1)}\right), 1\right\} \\ i^{(t)} & otherwise \end{cases}$$

Return to step 1

Example

Slice Sampler

For multimodal distributions,

Univariate case,

$(X, U) \sim f(x, u) = f(x)f(u|x)$

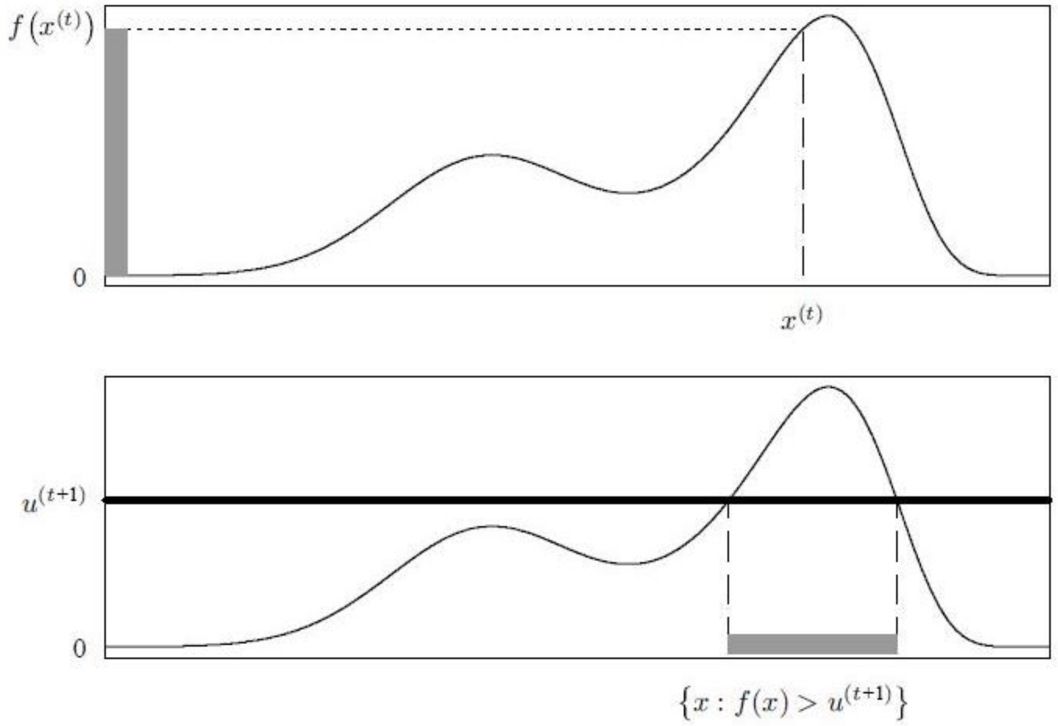
Choose U to speed up the MCMC process

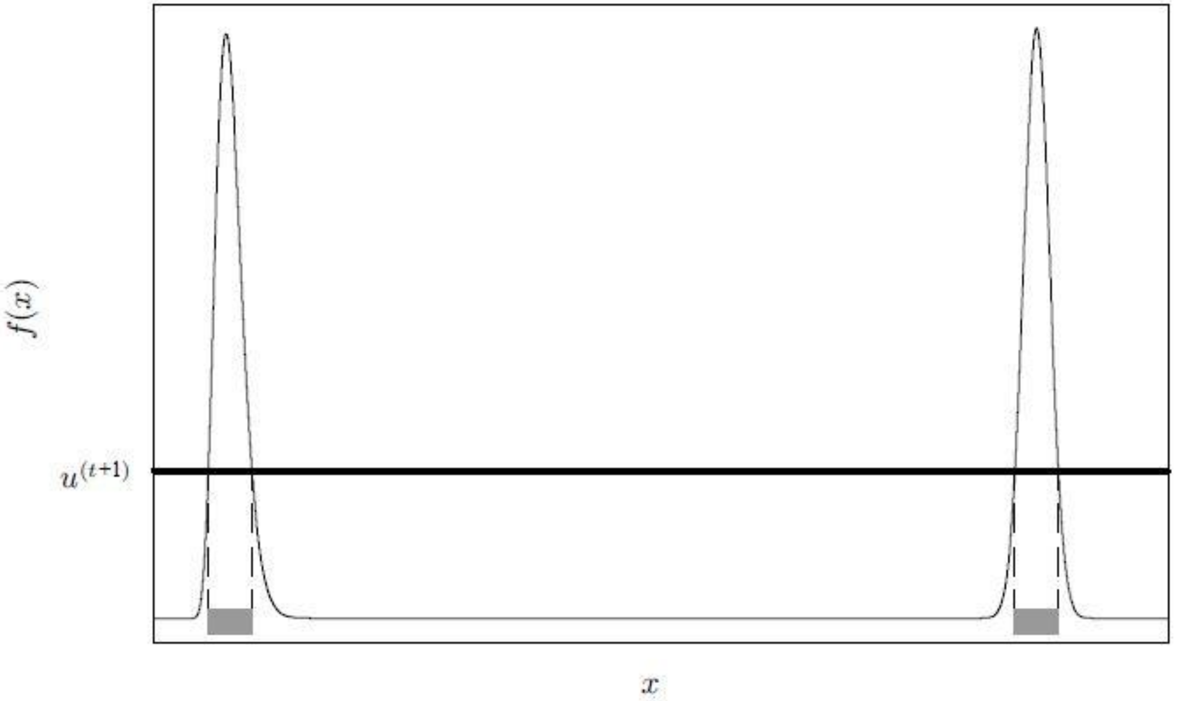
Uses Gibbs sampler

At t+1 iteration, generate

- $U^{(t+1)}|x^{(t)} \sim Unif\left(0, f\left(x^{(t)}\right)\right)$
- $X^{(t+1)}|u^{(t+1)} \sim Unif\left\{x: f(x) \geq u^{(t+1)})\right\}$

Example


$f(x^{(t)})$
0
$x^{(t)}$
$u^{(t+1)}$
0
$\{x : f(x) > u^{(t+1)}\}$


$f(x)$
$u^{(t+1)}$
x

Perfect Sampling

Avoids all drawbacks and generate a chain that has exactly reached the stationary distribution

Challenges in implementing

Coupling from the past (CFTP)

Motivated by the idea that the chain started at $t = -\infty$ and forwarded to $t = 0$

Instead start at an identified time window from $t = \tau < 0$ to $t = 0$

So, chain is in it's stationary distribution by $t = 0$

Impossible to know which state the chain is in at $t = \tau$

Consider multiple chains – every possible chain started from every possible state at $t = \tau$

Coupling / coalescing

Finally algorithm will yield single chain from $t = 0$ – desired stationary distribution

General Idea

Let $X_k^{(t)}$ be the random state at time t of the Markov chain, started in state k with $k = 1, \ldots, K$

Let $\boldsymbol{X}^{(0)}$ be a draw from the desired stationary distribution f

Let $\tau = -1$, generate $\boldsymbol{U}^{(0)}$

Start a chain from each state namely, $x_1^{(-1)}, x_2^{(-1)}, \ldots, x_K^{(-1)}$

Forward each chain to t=0 via the update $X_k^{(0)} = \mathrm{q}(x_K^{(-1)}, \boldsymbol{U}^{(0)})$

If all the chains are in the same state at t=0 then chains are coupled – algorithm stopped

If not, Let $\tau = -2$, generate $\boldsymbol{U}^{(-1)}$

Start a chain from each state namely, $x_1^{(-2)}, x_2^{(-2)}, \dots, x_K^{(-2)}$

Forward each chain to t=0 via the updates $X_k^{(-1)} = \mathrm{q}(x_K^{(-2)}, \boldsymbol{U}^{(-1)})$

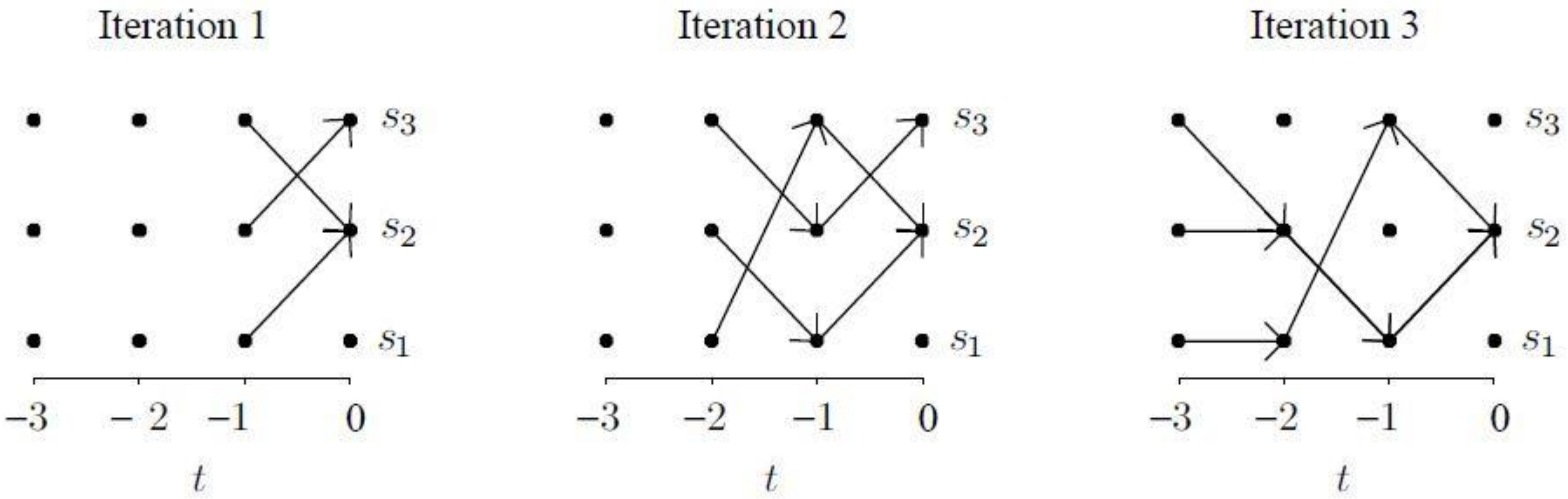
then reusing $\boldsymbol{U}^{(0)}$, generate $X_k^{(0)} = \mathrm{q}(x_K^{(-1)}, \boldsymbol{U}^{(0)})$

If all the chains are in the same state at t=0 then chains are coupled – algorithm stopped

Otherwise – move the starting time back another step, $\tau = -3$

Continue restarting the chains one step further back till all the chains are coalesced by t=0

Example



References

Computational Statistics, 2nd edition; Givens and Hoeting,2013