Reversible Jump MCMC

Sampling from a distribution without a fixed base measure

Hengrui Luo

Department of Statistics
The Ohio State University

STAT 7730 Computational Statistics

Gibbs Sampler Revisited I

The Gibbs sampler can be written as following and interpreted as a special case of Metropolis-Hastings. It creates trajectory of samples of an ***irreducible aperiodic*** *Markov chain that has a stationary distribution.*

1. Initial values for the d-dimensional parameters from some distribution $X^{(0)} = \left(x_1^{(0)}, \cdots x_d^{(0)}\right)$.

Gibbs Sampler Revisited II

2. Given $X^{(t-1)} = \left(x_1^{(t-1)}, \cdots x_d^{(t-1)}\right)$ we update coordinates using samples from following scheme (visitation scheme)

$$
\begin{aligned}
X_1^{(t)} &\sim f\left(x_1 \mid x_2^{(t-1)}, \cdots x_d^{(t-1)}\right) \\
X_2^{(t)} &\sim f\left(x_2 \mid x_1^{(t)}, x_3^{(t-1)}, \cdots x_d^{(t-1)}\right) \\
&\vdots \\
X_d^{(t)} &\sim f\left(x_d \mid x_1^{(t)}, x_2^{(t)}, \cdots x_{d-1}^{(t)}\right)
\end{aligned}
$$

and let $X^{(t)} = \left(x_1^{(t)}, \cdots x_d^{(t)}\right)$

Gibbs Sampler Revisited III

3. $t = t+1$ and go back to step (b) until reasonable many steps or convergence criteria is reached.

In step (b), for each $i = 1, 2, \cdots, d$ we are actually generate $X_i^{(t)} \sim f\left(x_i \mid x_{-i}^{(t-1)}\right)$ and it is equivalent to Metropolis-Hasting sampling with proposal distribution $q_i(y \mid x) := f(y_i \mid x_{-i})\mathbf{1}_{\{y_{-i}=x_{-i}\}}$.
The Markov transition kernel $q(y \mid x)$ should satisfy **detailed balance to ensure that the stationary distribution is π**

$$\pi(x)q(y \mid x) = \pi(y)q(x \mid y)$$

In continuous case we write the Markov transition kernel as $P(x, dx')$ and the **detailed balance** becomes

$$\int_A \int_B \pi(dx)P(x, dx') = \int_B \int_A \pi(dx')P(x', dx)$$

Gibbs Sampler Revisited IV

- What if the stationary distribution π may have different support of different dimension?
- How to check the convergence of the posterior samples (a.k.a. mixing problem)?
- How to sample from an Infinite-dimensional distribution (a.k.a. nonparametric sampling)?

In this lecture, we focus on the first question. The main difficulty of sampling from spaces $\mathscr{X}_1, \mathscr{X}_2, \cdots$ of various dimensions is:
How to design a transition kernel that allows us to move $\mathscr{X}_1$ to $\mathscr{X}_2$ (and also reversely) while maintaining the detailed balance, to ensure that the Markov chain still converges to the stationary distribution.

Existing Methods

For specific models like (mixture of) Dirichlet process/Polya trees, such a dimensionality problem can be well-addressed due to **conjugacy**. However for more general/complicated models we cannot expect the conditionals are still of closed forms therefore we step back and wish we could have **a sequence of dependent posterior samples** from some algorithm.

- Tierney's Hybrid Sampler [5].
- Grenander-Miller's Jump-diffusion Sampler [6].
- Continuous Time Monte Carlo Sampler [7].

The method we are going to introduce in this lecture is the Reversible-jump MCMC Sampler [1, 2].

Idea: Try to modify the proposal in order to move among different spaces $\mathscr{X}_1, \mathscr{X}_2, \cdots$ *of different dimensions. And also adjust the acceptance probability to ensure that the detailed balance holds.*

General Case I

Theorem

[1] For general case, suppose that we consider one move in each proposal step from one space $\mathscr{X}_1$ to $\mathscr{X}_2$ with certain probability. There are D such moves from $\mathscr{X}_1 \mapsto \mathscr{X}_2$ and suppose we choose move $m \in \{1,2,\cdots,D\}$. If we want to ensure that the probability of moving over the combined space $(\mathscr{X}_1 \oplus \mathscr{X}_2)$ from A to B equal the probability that moves from B to A for $A,B \in (\mathscr{X}_1 \oplus \mathscr{X}_2)$ (detailed balance) then it is sufficient to ensure that following detailed balance in Metropolis-Hastings holds

$$\int_A \int_B \pi(dx) q_m(x,dx') \alpha_m(x,x') = \int_B \int_A \pi(dx') q_m(x',dx) \alpha_m(x',x)$$

where we propose a move of type m that would take the state x to dx' with probability $q_m(x,dx')$ with acceptance probability $\alpha_m(x,x')$ in MH.

General Case II

Corollary

If we assume that $\pi(dx)q_m(x,dx')$ *has a finite density* $f_m(x,x')$ *with respect to a symmetric measure on the combined parameter space* $(\mathscr{X}_1 \oplus \mathscr{X}_2) \times (\mathscr{X}_1 \oplus \mathscr{X}_2)$ *then the acceptance probability has a closed form* $\alpha_m(x,x') = min\left\{1, \frac{f_m(x',x)}{f_m(x,x')}\right\}$ *satisfying the detailed balance above.*

Two-Model Example (General) I

Example

Consider $\mathscr{X}_1 = \{1, 2, \cdots\}, \mathscr{X}_2 = \{\theta^{(1)}, \theta^{(2)}, \cdots\}$ of different dimensions and we want to sample from posteriors of different models with different underlying dimensions.
Let $p(\theta^{(1)} \mid \mathscr{X}_1), p(\theta^{(2)} \mid \mathscr{X}_2)$ be two proper densities for these two spaces.
Then the acceptance probability for the proposed transition from $x = \left(1, \theta^{(1)}\right)$ to $x' = \left(2, \theta^{(2)}\right)$ is $\alpha(x, x') = min\left\{1, \frac{f(x',x)}{f(x,x')}\right\}$ where

$$f(x, x') = p(1, \theta^{(1)} \mid y) j(1, \theta^{(1)}) q_1(u^{(1)})$$

$$f(x', x) = p(2, \theta^{(2)} \mid y) j(2, \theta^{(2)}) q_2(u^{(2)}) \left| \frac{\partial(\theta^{(2)}, u^{(2)})}{\partial(\theta^{(1)}, u^{(1)})} \right|$$

Two-Model Example (General) II

Example

So

$$\alpha(x,x') = min\left\{1, \frac{p(2,\theta^{(2)} \mid y)j(2,\theta^{(2)})q_2(u^{(2)})}{p(1,\theta^{(1)} \mid y)j(1,\theta^{(1)})q_1(u^{(1)})}\left|\frac{\partial(\theta^{(2)},u^{(2)})}{\partial(\theta^{(1)},u^{(1)})}\right|\right\}$$

where $u^{(1)}, u^{(2)}$ are from proposal distribution with proper densities q_1, q_2 independent from $\left(\theta^{(1)},\theta^{(2)}\right)$ and
where the probability of choosing certain move is denoted by $j(\bullet)$.
The Jacobian comes from a bijection defined by the move between $(\theta^{(1)},u^{(1)}) \leftrightarrow (\theta^{(2)},u^{(2)})$.

Two-Model Example (General) III

- **Remarks 1.** The joint probability measure on $(\mathscr{X}_1 \oplus \mathscr{X}_2)$ is usually defined as
$\xi(A \times B) = \xi(B \times A) = \lambda\left\{(\theta^{(1)}, u^{(1)}) : \theta^{(1)} \in A, \theta^{(2)}\left(\theta^{(1)}, u^{(1)}\right) \in B)\right\}$
where λ is the Lebesgue measure.
- **Remarks 2.** The bijection is **dependent** on the choice of the random proposals q_1, q_2 and therefore the "recovery from lower dimension to higher dimension" is also affected by such a choice of q_1, q_2. In other words, the performance of reverse jump MCMC is still sensitive to the choice of proposal densities. To get the optimal proposal you have to know some prior information about the bijective correspondence between two models, as we will see in the Poisson-NB example below.

Two-Model Example (General) IV

- **Remarks 3.** The "trap-in-local-optima" problem is still quite bothering. Like general MCMC in Bayesian Gaussian regression, the choice of band width(support) in proposal will lead to very different behavior of the posterior trace plots. Adaptive selection of band width may alleviate the problem but this problem still exists and when dimensions differ drastically, this is a rather serious problem due to my simulation.

Two-Model Examples (Poisson OR Negative Binomial) I

Example

In [3] the author introduced an example of screening the pattern of over-dispersion. For a count data, it is of central important in genetics whether the sampling scheme is Poisson and negative binomial because it involves different marginal assumptions. Reverse jump MCMC can sample with consideration of potential sampling scheme of both. To construct an MCMC which is capable of drawing from both sampling schemes, we need to use an algorithm which is able to “jump” between these two schemes accordingly - Let us try reversible jump MCMC!

Two-Model Examples (Poisson OR Negative Binomial) II

Example

The joint likelihood of Poisson is

$$L_{(1,\theta^{(1)})}(y) = \prod_{i=1}^{n} \frac{\left(\theta^{(1)}\right)^{y_i}}{y_i!} exp\left(-\theta^{(1)}\right)$$

where if we use the usual parameterization of $Poi(\lambda)$ then $\theta^{(1)} = \lambda$.
while the joint likelihood of negative binomial is

$$L_{(2,\theta^{(2)})}(y) = \prod_{i=1}^{n} \frac{\left(\theta_1^{(2)}\right)^{y_i}}{y_i!} \cdot \frac{\Gamma\left(\frac{1}{\theta_2^{(2)}} + y_i\right)}{\Gamma\left(\frac{1}{\theta_2^{(2)}}\right)\left(\frac{1}{\theta_2^{(2)}} + \theta_1^{(2)}\right)^{y_i}} \left(1 + \theta_1^{(2)}\theta_2^{(2)}\right)^{-\frac{1}{\theta_2^{(2)}}}$$

where if we use the usual parameterization of $NB(\lambda, \kappa)$ then $\theta^{(2)} = (\lambda, \kappa)$.

Two-Model Examples (Poisson OR Negative Binomial) III

Example

Consider following natural bijection induced by the "jump $1 \to 2$" $\left(\theta^{(1)}, u^{(1)}\right) \mapsto \left(\theta^{(1)}, \mu exp\left(u^{(1)}\right)\right)$ and the "reverse jump $2 \to 1$" $\left(\theta^{(2)}, u^{(2)}\right) \mapsto \left(\theta_1^{(2)}, \log\left(\frac{\theta_2^{(2)}}{\mu}\right)\right)$ (μ being a known constant) and random proposal $q_1, q_2 \sim N_1(0, \sigma^2)$. According to our definition of proposal and with Jacobian

$$\left|\frac{\partial(\theta^{(2)}, u^{(2)})}{\partial(\theta^{(1)}, u^{(1)})}\right| = \begin{vmatrix} 1 & 0 \\ 0 & \mu exp(u^{(1)}) \end{vmatrix} = \mu exp(u^{(1)})$$

$$\left|\frac{\partial(\theta^{(1)}, u^{(1)})}{\partial(\theta^{(2)}, u^{(2)})}\right| = \begin{vmatrix} \frac{1}{\mu} & 0 \\ 0 & \frac{\mu}{\theta_2^{(2)}} \end{vmatrix} = \frac{1}{\theta_2^{(2)}}$$

Two-Model Examples (Poisson OR Negative Binomial) IV

Example

the acceptance probability from Poisson to NB is

$$\alpha_{1\to 2} = min\left\{1, \frac{p(2,\theta^{(2)} \mid y)}{p(1,\theta^{(1)} \mid y)}\left\{\frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{u^{(1)2}}{2\sigma^2}\right]\right\}^{-1} \cdot \mu exp(u^{(1)})\right\}$$

the acceptance probability from NB to Poisson is

$$\alpha_{2\to 1} = min\left\{1, \frac{p(1,\theta^{(1)} \mid y)}{p(2,\theta^{(2)} \mid y)}\left\{\frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{\left(\log\left(\frac{\theta_2^{(2)}}{\mu}\right)\right)^2}{2\sigma^2}\right]\right\}^{-1} \cdot \frac{1}{\theta_2^{(2)}}\right\}$$

Two-Model Examples (Multinomial OR Binomial) I

Example

In [4] the author introduced an example of how to switch between multinomial (dim=2) and binomial example (dim=1), which is a naive model of change-point detection: If we only see response of two kinds in a limited amount of observations, and there could be a third category, are there two or three categories? (OR more? But let us discuss only two models now) Can we sample from such a model with the prior information/consideration of the possibility of having different number of categories - Let us try reversible jump MCMC!

Two-Model Examples (Multinomial OR Binomial) II

Example

The joint likelihood of multinomial (three categories) is

$$L_{(1,\theta^{(1)})}(y) \propto \prod_{i=1}^{n} \left(\theta_1^{(1)}\right)^{I(y_i \in 1)} \left(\theta_2^{(1)}\right)^{I(y_i \in 2)} \left(1-\theta_1^{(1)}-\theta_2^{(1)}\right)^{I(y_i \in 3)}$$

where if we use the usual parameterization of $Multi(p_1, p_2)$ then $\theta^{(1)} = (p_1, p_2)$.
while the joint likelihood of binomial is

$$L_{(2,\theta^{(2)})}(y) \propto \prod_{i=1}^{n} \left(\theta_1^{(2)}\right)^{I(y_i \in 1)} \left(\theta_2^{(2)}\right)^{I(y_i \in 2)}$$

where if we use the usual parameterization of $Binom(p)$ then $\theta^{(2)} = p$.

Two-Model Examples (Multinomial OR Binomial) III

Example

Consider following natural bijection induced by the "jump $1 \to 2$"
$\left(\theta^{(1)}, u^{(1)}\right) \mapsto \left(\frac{n_1\theta_1^{(1)} + n_2\theta_2^{(1)}}{n_1+n_2}, \theta_2^{(1)}\right)$ and the "reverse jump $2 \to 1$"
$\left(\theta^{(2)}, u^{(2)}\right) \mapsto \left(\frac{n_1+n_2}{n_1}\theta^{(2)} - \frac{n_2}{n_1}u^{(2)}, u^{(2)}\right)$ (μ being a known constant) where $n_j = \sum_{i=1}^{n} I(y_i \in j), j = 1,2,3$ and random proposal $q_1, q_2 \sim N_1(0, \sigma^2)$. According to our definition of proposal and with Jacobian

$$\left|\frac{\partial(\theta^{(2)}, u^{(2)})}{\partial(\theta^{(1)}, u^{(1)})}\right| = \begin{vmatrix} \frac{n_1+n_2}{n_1} & -\frac{n_1}{n_2} \\ 0 & 1 \end{vmatrix} = 1 + \frac{n_2}{n_1}$$

$$\left|\frac{\partial(\theta^{(1)}, u^{(1)})}{\partial(\theta^{(2)}, u^{(2)})}\right| = \begin{vmatrix} \frac{n_1}{n_1+n_2} & \frac{n_2}{n_1+n_2} \\ 0 & 1 \end{vmatrix} = \frac{n_1}{n_1+n_2}$$

Two-Model Examples (Multinomial OR Binomial) IV

Example

the acceptance probability from Multinomial to Binomial is

$$\alpha_{1\to 2} = min\left\{1, \frac{p(2,\theta^{(2)} \mid y)}{p(1,\theta^{(1)} \mid y)}\left\{\frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{\theta_2^{(1)2}}{2\sigma^2}\right]\right\}^{-1} \cdot \left(1+\frac{n_2}{n_1}\right)\right\}$$

the acceptance probability from Binomial to Multinomial is

$$\alpha_{2\to 1} = min\left\{1, \frac{p(1,\theta^{(1)} \mid y)}{p(2,\theta^{(2)} \mid y)}\left\{\frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{u^{(2)2}}{2\sigma^2}\right]\right\}^{-1} \cdot \left(\frac{n_1}{n_1+n_2}\right)\right\}$$

Using R

- rjmcmc package available in CRAN. It realizes default bijection and user-specified bijection and provides sample-based posterior as well as functional posterior. See our example in rjmcmc_Tutorial.R
- However, it is based on birth-death process which converges to reversible jump MCMC according to a theoretic support [7].
- You need to know the model and parameter prior.
- You need to know the model likelihood or able to sample from it.
- You need to specify the bijection
  - Construct a reasonable bijection like we saw above. rjmcmcpost
  - Use the default bijection derived by [9], usually it works pretty well. defaultpost
- You need to fit the model using correct sample in format consistent with coda, which can be provided by rjags (external Gibbs sampler with possible MH step). This means that reverse jump MCMC is not limited to the conjugate case we discussed above!

Summary

- Summary
  - We construct a jump/move proposal $j(\bullet)$ to decide to which space we are moving to in this step.
  - We construct random proposals q_1, q_2 to aid us moving on the combined space $(\mathscr{X}_1 \oplus \mathscr{X}_2)$ with probability measure ξ.
  - We adjust the acceptance probability α_m by a Jacobian in order to maintain the detailed balance in the move m.
- Applications

  In what kind of scenarios do we want to consider reversible MCMC?

  The most commonly seen cases are model selection; the most general case is when the model priors are supported on parameter spaces of different dimensions. Reversible MCMC is widely used in following scenarios

  - Clustering
  - Change-point detection
  - Non-stationary modeling

- Green, Peter J. "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination." Biometrika 82.4 (1995): 711-732.
- Green, Peter J. “Statistical problems where the parameter dimension varies: MCMC theory and some applications” IMS/ENAR, Memphis, March 1997
- Green, Peter J., and David I. Hastie. "Reversible jump MCMC." Genetics 155.3 (2009): 1391-1403.
- Barker, Richard J., and William A. Link. "Bayesian multimodel inference by RJMCMC: A Gibbs sampling approach." The American Statistician 67.3 (2013): 150-156.
- Tierney, Luke. "Markov chains for exploring posterior distributions." the Annals of Statistics (1994): 1701-1728.

- Grenander, Ulf, and Michael I. Miller. "Representations of knowledge in complex systems." Journal of the Royal Statistical Society. Series B (Methodological) (1994): 549-603.
- Cappé, Olivier, Christian P. Robert, and Tobias Rydén. "Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65.3 (2003): 679-700.
- Besag, Julian, et al. "Bayesian computation and stochastic systems." Statistical science (1995): 3-41.
- Carlin, Bradley P., and Siddhartha Chib. "Bayesian model choice via Markov chain Monte Carlo methods." Journal of the Royal Statistical Society. Series B (Methodological) (1995): 473-484.