Reversible Jump MCMC

Intro to MCMC (Andrieu, et al.)
Trans-dimensional MCMC (Green)
Applications of RJMCMC (Laine)

Qiyam Tung

February 17, 2009

Outline

Outline

Problem Statement

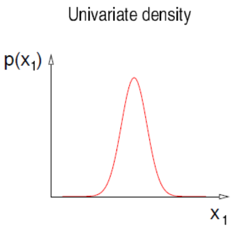
We want to know the probability distribution:

- Exact inference: Sum-product
- Approximation: MCMC
- Approximate for several models: Reverse jump MCMC

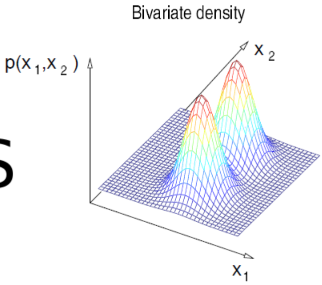
Reverse jump MCMC essentially allows us to jump across different models

"The number of things you don't know is one of things you don't know"

Would a better model be to use 1, 2, or 3 parameters?



VS



Detailed Balance

Recall detailed balance from basic MCMC (discrete)

$$p(x^{(i)})T(x^{(i-1)}|x^{(i)}) = p(x^{(i-1)})T(x^{(i)}|x^{(i-1)})$$

A Markov chain on a continuous general state space

$$\int_{(x,x')\epsilon A\times B} \pi(dx)P(x,dx') = \int_{(x,x')\epsilon A\times B} \pi(dx')P(x',dx).$$

Detailed Balance

If we want to do Metropolis-Hastings on a general space, we replace the transition kernel by the acceptance probabilty $\alpha(x, x')$ and the proposal measure $q(x, dx')$

$$\int_{(x,x')\epsilon A\times B} \pi(dx)P(x, dx') = \int_{(x,x')\epsilon A\times B} \pi(dx')P(x', dx).$$

$$\int_{(x,x')\epsilon A\times B} \pi(dx)q(x, dx')\alpha(x, x') = \int_{(x,x')\epsilon A\times B} \pi(dx')q(x', dx)\alpha(x', x)$$

Outline

Defining a General State Space

The space we're sampling from is:

$$X \triangleq \cup_{m=1}^{M} m \times X_m \tag{2}$$

where

- $x_1 = \{1, 2\}$
- $x_2 = \{(1,1),(1,2),(2,1),(2,2)\}$
- etc.

$$X = \{(1,1),(1,2),(2,(1,1)),(2,(1,2)),\ldots\} \tag{3}$$

To jump between spaces, we need to create:

- Communicating State Spaces
- Bijective function that maps from one space to the other

Communicating State Spaces

Dimensions should be equal.

$$\overline{X_{m,n}} \triangleq X_m \times U_{m,n} \tag{4}$$

and

$$\overline{X_{n,m}} \triangleq X_n \times U_{n,m} \tag{5}$$

U's are random quantities (possibly from a Gaussian distribution).
X's and U's are like vectors (not really)

Bijective Function

The function must satisfy these properties:

$$f_{n\to m}(x_n, u_{n,m}) = (x_m, u_{m,n}) \tag{6}$$

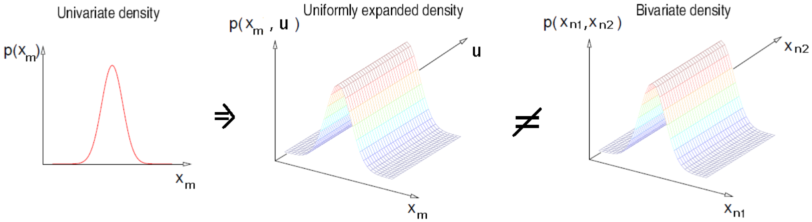
$$f_{m\to n}(x_m, u_{m,n}) = (x_n, u_{n,m}) \tag{7}$$

which says it must be deterministic and invertible.

$$d(x_m) + d(u_{m,n}) = d(x_n) + d(u_{n,m}) \tag{8}$$

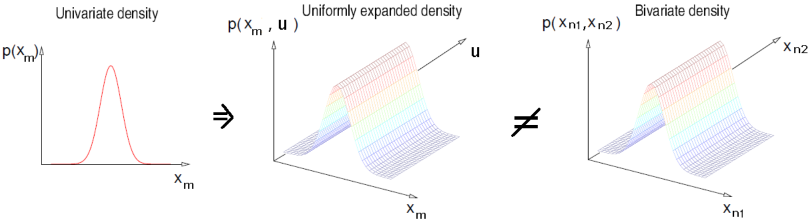
which says the dimensions must be equal.

Example



Univariate distribution's area is 1.
Bivariate distribution's volume is 1.
Expanded distribution's volume is 1 (assuming u = [0,1])

Example



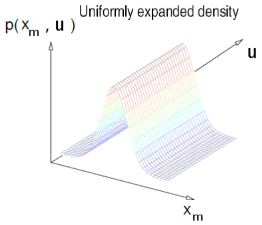
$$(x_m, u) \epsilon X_{m,n}$$

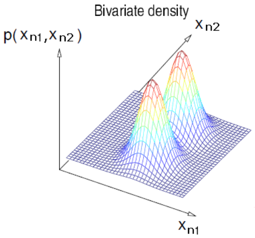
$$(x_{n1}, x_{n2}) \epsilon X_{n,m}$$

$$f_{m \to n}(x_m, u) = (x_{n1}, x_{n2})$$

$$f_{n \to m}(x_{n1}, x_{n2}) = (x_m, u)$$

Example





Most often, we want to capture some information from the other 2 variables into the univariate distribution.

$$f_{m \to n}(x_m, u) = (x_m + u, x_m - u)$$

$$f_{n \to m}(x_{n1}, x_{n2}) = (\frac{x_{n1} + x_{n2}}{2}, \frac{x_{n1} - x_{n2}}{2})$$

Example

$$f_{m\to n}(x_m, u) = (x_m + u, x_m - u)$$

$$f_{n\to m}(x_{n1}, x_{n2}) = (\tfrac{x_{n1}+x_{n2}}{2}, \tfrac{x_{n1}-x_{n2}}{2})$$

$$\begin{aligned} f_{m\to n}(x_m, u) &= (x_m + u, x_m - u) \\ &= (\frac{x_{n1} + x_{n2}}{2} + \frac{x_{n1} + x_{n2}}{2}, \frac{x_{n1} + x_{n2}}{2} - (\frac{x_{n1} - x_{n2}}{2})) \\ &= (\frac{2x_{n1}}{2}, \frac{2x_{n2}}{2}) \\ &= (x_{n1}, x_{n2}) \end{aligned}$$

Outline

With our previous Metropolis-Hastings on the general state space

$$\int_{(x,x')\epsilon A\times B} \pi(dx)q(x,dx')\alpha(x,x') = \int_{(x,x')\epsilon A\times B} \pi(dx')q(x',dx)\alpha(x',x)$$

$$\alpha(x,x') = min\{1, \frac{\pi(dx')q(x',dx)}{\pi(dx)q(x,dx')}\}$$

Substitutes q and π with f with respect to u

$$\int_{(x,x')\epsilon A\times B} \alpha(x,x')f(x,x')\mu(dx,dx') = \int_{(x,x')\epsilon A\times B} \alpha(x',x)f(x',x)\mu(dx',$$

$$\alpha(x,x') = min\{1, \frac{f(x',x)}{f(x,x')}\} \quad (9)$$

If we were to make random variables u, our equation looks like

$$\int \pi(x)g(u)\alpha(x,x')dxdu = \int \pi(dx')g(u')\alpha(x',x)dx'du' \quad (10)$$

$g(u)$: random numbers sampled from known density g
$x' = h(x,u)$ and $x = h'(x',u')$
We're essentially moving from one state space to another (but with differing number of u's).

$$\alpha(x,x') = min\{1, \frac{\pi(x')}{\pi(x)} \times \frac{g'(u')}{g(u)} \times |\frac{\partial(x',u')}{\partial(x,u)}|\}$$

$$J_{f_{n\to m}} = |det \frac{\partial f_{n\to m}(x,u)}{\partial(x,u)}| \tag{11}$$

$$= det \begin{pmatrix} \frac{\partial x}{\partial x'} & \frac{\partial x}{\partial u'} \\ \frac{\partial u}{\partial x'} & \frac{\partial u}{\partial u'} \end{pmatrix}$$

Describes a stretching factor or magnification factor from one coordinate system (x,y) to (u,v).
Recall that $x = h(x', u')$

Modifying the equation to deal with multiple models, we change it to

$$\int \pi(x)q_m(x,x')\alpha_m(x,x')dxdu = \int \pi(x')q_m(x',dx)\alpha_m(x',x)dx'du'$$

$$\alpha_m = min\{1, \frac{\pi(x')}{\pi(x)} \times \frac{j_m(x')}{j_m(x)} \times \frac{g'_m(u')}{g_m(u)} \times |\frac{\partial(x',u')}{\partial(x,u)}|\}$$

$j_m(x)$: probability of choosing to move type m when at x
$g_m(u)$: random number sampled from known density g

Outline

Idea

The Metropolis Hastings algorithm

1. Initialise $x^{(0)}$.
2. For $i = 0$ to $N - 1$
   - Sample $u \sim \mathcal{U}_{[0,1]}$.
   - Sample $x^\star \sim q(x^\star | x^{(i)})$.
   - If $u < \mathcal{A}(x^{(i)}, x^\star) = \min\left\{1, \frac{p(x^\star)q(x^{(i)}|x^\star)}{p(x^{(i)})q(x^\star|x^{(i)})}\right\}$

     $$x^{(i+1)} = x^\star$$

     else

     $$x^{(i+1)} = x^{(i)}$$

Figure 5. Metropolis-Hastings algorithm.

Update *x* if the acceptance probability is high enough.

1. Initialisation: set $(k^{(0)}, \mu^{(0)})$.
2. For $i = 0$ to $N - 1$
   - Sample $u \sim \mathcal{U}_{[0,1]}$.
   - If $(u \leq b_k)$
     - then "birth" move.
     - else if $(u \leq b_k + d_k)$ then "death" move.
     - else if $(u \leq b_k + d_k + s_k)$ then "split" move.
     - else if $(u \leq b_k + d_k + s_k + m_k)$ then "merge" move.
     - else update.

     End If.
   - Sample other parameters.

Figure 18. Generic reversible jump MCMC.

Unlike normal MCMC, we can make several decisions depending on our acceptance probability
A generic MJMCMC algorithm has 5 moves:

- Birth: simply add a new component(s)
- Death: simply remove a component(s)
- Split: add a new component by splitting an old one
- Merge: remove 2 components by merging them into one
- Update: Simply update the component

All except update are trans-dimensional

RJMCMC allows us to sample from models of different dimensions.
Issues with RJMCMC include:

- If models were very different, better to do within-model
- Poor performance (?)
- Cumbersome to construct and difficult to tune

Discussion

Target distribution:

$$p(k, dx) = \sum_{m=1}^{M} p(m, dx_m) I_{m \times x_m}(k, x) \tag{12}$$