Interacting Multiple-Try Metropolis Sampling

Radu Craiu

Department of Statistics
University of Toronto
web: www.utstat.toronto.edu/craiu/

Joint work with Roberto Casarin (Venice), Antonio di Narzo (Lausanne) and Fabrizio Leisen (Madrid)

JSM - San Diego, 2012
Outline

1. Multiple-Try Metropolis and variations
   - Multi-Distributed-Try Metropolis

2. Interacting MTM
   - Basic principles
   - Practical Issues and Refinements

3. IMTM with Annealing and Subsampling
   - Annealed IMTM
   - Subsampling IMTM

4. Examples
   - Beta Mixture Model
   - Stochastic Volatility Model
Metropolis-Hastings Samplers

- We wish to sample from some distribution for $X \in S$ that has density $\pi$. Obtaining independent draws is too hard.
- We construct and run a Markov chain with transition $K(x_{old}, x_{new})$ that leaves $\pi$ invariant

$$\int_S \pi(x)K(x, y)dx = \pi(y).$$

- The Metropolis-Hastings sampler is one of the most used algorithms in MCMC:
  - Given $x_t$, the current state of the MC, a ”proposed sample” $y$ is drawn from a proposal density $T(y|x_t)$.
  - The proposal $y$ is accepted with probability
    $$\min\{1, \frac{\pi(y)T(x_t|y)}{\pi(x_t)T(y|x_t)}\}.$$  
  - If $y$ is accepted, then $x_{t+1} = y$, otherwise $x_{t+1} = x_t$.  

Suppose $T$ is a proposal density such that

$$T(x|y) > 0 \iff T(y|x) > 0 \quad \text{and} \quad \lambda(x, y)$$

is a symmetric function.
Multiple-Try Metropolis and variations

Interacting MTM

IMTM with Annealing and Subsampling

Examples

Original MTM (Liu, Liang and Wong, JASA 2000)

- Suppose $T$ is a proposal density such that
  $T(x|y) > 0 \iff T(y|x) > 0$ and $\lambda(x, y)$ is a symmetric function.

(i) Draw $K$ independent trial proposals $y_1, \ldots, y_K$ from $T(\cdot|x_t)$.
Sample one with $p_i \propto w(y_i|x_t) = \pi(y_i) T(x_t|y_i) \lambda(x_t, y_i)$. 
Suppose $T$ is a proposal density such that $T(x|y) > 0 \iff T(y|x) > 0$ and $\lambda(x,y)$ is a symmetric function.

(i) Draw $K$ independent trial proposals $y_1, \ldots, y_K$ from $T(\cdot|x_t)$. Sample one with $p_i \propto w(y_i|x_t) = \pi(y_i) T(x_t|y_i) \lambda(x_t, y_i)$.

(ii) Generate $x_1^*, \ldots, x_{k-1}^* \sim T(\cdot|y)$ and put $x_k^* = x_t$. 

Do we better explore the sample space with $K$ proposals? Yes - provided we take advantage of the flexibility offered by the MTM.
Original MTM (Liu, Liang and Wong, JASA 2000)

- Suppose $T$ is a proposal density such that $T(x|y) > 0 \Leftrightarrow T(y|x) > 0$ and $\lambda(x, y)$ is a symmetric function.

(i) Draw $K$ independent trial proposals $y_1, \ldots, y_K$ from $T(\cdot|x_t)$. Sample one with $p_i \propto w(y_i|x_t) = \pi(y_i)T(x_t|y_i)\lambda(x_t, y_i)$.

(ii) Generate $x_1^*, \ldots, x_{k-1}^* \sim T(\cdot|y)$ and put $x_k^* = x_t$.

(iii) Accept $y$ with probability $\min\left\{1, \frac{\sum_{i=1}^K w(y_i|x_t)}{\sum_{i=1}^K w(x_i^*|y)}\right\}$ (generalized MH ratio).
Suppose $T$ is a proposal density such that $T(x|y) > 0 \iff T(y|x) > 0$ and $\lambda(x, y)$ is a symmetric function.

(i) Draw $K$ independent trial proposals $y_1, \ldots, y_K$ from $T(\cdot|x_t)$. Sample one with $p_i \propto w(y_i|x_t) = \pi(y_i) T(x_t|y_i) \lambda(x_t, y_i)$.

(ii) Generate $x^*_1, \ldots, x^*_{k-1} \sim T(\cdot|y)$ and put $x^*_k = x_t$.

(iii) Accept $y$ with probability $\min \left\{1, \frac{\sum_{i=1}^K w(y_i|x_t)}{\sum_{i=1}^K w(x^*_i|y)} \right\}$ (generalized MH ratio).

Do we better explore the sample space with $K$ proposals?
**Original MTM (Liu, Liang and Wong, JASA 2000)**

- Suppose $T$ is a proposal density such that $T(x|y) > 0 \iff T(y|x) > 0$ and $\lambda(x, y)$ is a symmetric function.

(i) Draw $K$ independent trial proposals $y_1, \ldots, y_K$ from $T(\cdot|x_t)$. Sample one with $p_i \propto w(y_i|x_t) = \pi(y_i) T(x_t|y_i) \lambda(x_t, y_i)$.

(ii) Generate $x_1^*, \ldots, x_{k-1}^* \sim T(\cdot|y)$ and put $x_k^* = x_t$.

(iii) Accept $y$ with probability $\min\left\{1, \frac{\sum_{i=1}^{K} w(y_i|x_t)}{\sum_{i=1}^{K} w(x_i^*|y)}\right\}$ (generalized MH ratio).

- Do we better explore the sample space with $K$ proposals?
  - Yes - provided we take advantage of the flexibility offered by the MTM.
Multi-Distributed-Try Metropolis

- The proposals do not have to be identically distributed.
Multi-Distributed-Try Metropolis

- The proposals do not have to be identically distributed.

Generate $y_j \sim T_j(\cdot|x_t)$ for $1 \leq j \leq k$ and select one with probability $p_j \propto w(y_j|x_t) = \pi(y_j)T(x_t|y_j)\lambda(x_t,y_j)$.
The proposals do not have to be identically distributed.

Generate $y_j \sim T_j(\cdot|x_t)$ for $1 \leq j \leq k$ and select one with probability $p_j \propto w(y_j|x_t) = \pi(y_j) T(x_t|y_j) \lambda(x_t, y_j)$.

If $y = y_{j_0}$ is selected than put $x_{j_0}^* = x_t$ and sample $x_j^* \sim T_j(\cdot|y)$ for all $j \neq j_0$.

Today: Discuss some of the (many) options offered by this general setup.

- Allows the use of two powerful concepts in modern MCMC: interacting chains and adaptive chains.
- Casarin, C. and Leisen (Stat. and Comput., online)
Interacting MTM

- Interacting MCMC uses a *population of chains* to gain insight about the target and improve the mixing properties for the chain(s) of interest.

- Not all chains must have the same stationary distribution and usually they have different convergence properties (e.g. simulated tempering).

- We want to use a population of chains to **guide the generation of multiple proposals**.

- Our population of auxiliary chains includes:
  1. **Chains that mix well** within the state space (usually this means that their stationary distribution is no longer $\pi$).
  2. **Chains that sample from a distribution not very different from** $\pi$. 
Interacting MTM

- Interacting MCMC uses a *population of chains* to gain insight about the target and improve the mixing properties for the chain(s) of interest.

- Not all chains must have the same stationary distribution and usually they have different convergence properties (e.g. simulated tempering).

- We want to use a population of chains to guide the generation of multiple proposals.

- Our population of auxiliary chains includes:
  1. Chains that *mix well* within the state space (usually this means that their stationary distribution is no longer $\pi$).
  2. Chains that *sample from a distribution not very different from* $\pi$.

- We need to run many chains!
Consider a population of $N$ chains, $X^{(i)} = \{X^{(i)}_t\}_{n \in \mathbb{N}}$; chain $i$ has MTM transition kernel with $M$ proposal densities $\{T^{(i)}_j\}_{1 \leq j \leq M}$.

Let $\Xi_t = \{x^{(i)}_t\}_{i=1}^N$ is the vector of values taken at iteration $n \in \mathbb{N}$ by the population of chains.

Each proposal distribution used at iteration $t + 1$ is allowed to depend on $\Xi_t$.

The $j$th proposal for chain $i_0$ is sampled conditional on $x^{(j)}_t$, $1 \leq j \leq M$ (here we assume $M = N$).
IMTM - A graphical illustration

\[ X_t^{(3)} \xrightarrow{T^{(3)}} X_{t+1}^{(3)} \]
\[ X_t^{(2)} \xrightarrow{T^{(2)}} X_{t+1}^{(2)} \]
\[ Y_3 \sim T_3^{(1)}(\cdot \mid X_t^{(3)}) \]
\[ Y_2 \sim T_2^{(1)}(\cdot \mid X_t^{(2)}) = X_{t+1}^{(1)} \]
\[ Y_1 \sim T_1^{(1)}(\cdot \mid X_t^{(1)}) \]

AUXILIARY
CHAIN 1
AUXILIARY
CHAIN 2

CHAIN OF
INTEREST
The transition kernel $K_i(x_t^{(i)}, x_{t+1}^{(i)})$ of the $i$-th chain of the IMTM algorithm satisfies the detailed balanced condition.
The transition kernel $K_i(x_t^{(i)}, x_{t+1}^{(i)})$ of the $i$-th chain of the IMTM algorithm satisfies the detailed balanced condition.

The joint transition kernel $K(\Xi_t, \Xi_{t+1})$ is ergodic to $\otimes_{i=1}^{N} \pi_i$. 
If all the chains in the population have an MTM kernel (IMTM):

**Pros**: At each step we choose among a large number of proposals placed in different regions of the sample space.

**Cons**: The computational load increases rapidly.

How to choose M (number of proposals) and N (number of chains)?
IMTM - Practical Issues

- $N$ is generally large so we set $M << N$.

- At $t$-th iterate of the $i$-th chain, we sample at random from the set $\{1, \ldots, N\}$ the indices $I_1, \ldots, I_{M-1}$ of the chains to be used in the transition (always $l_M = i$), i.e. $y_j \sim T_j^{(i)}(\cdot|\mathbf{x}_{t-1}^{(I_j)})$

- We want to favour contributions from those auxiliary chains that have been ”successful” in the previous iteration.

- We suggest using $	ilde{\lambda}_j^{(i)}(x_{t-1}, y_j) = \nu_j \lambda_j^{(i)}(x_{t-1}, y_j)$, where the factor $\nu_j$ is

\[
\nu_j = \frac{1}{N} \left[ 1 + \sum_{c=1}^{N} \mathbf{1}_c(l_j) \right], \quad j = 1, \ldots, M, \quad (1)
\]

and $\mathbf{1}_c(l_j) = 1$ whenever $y_j \sim T_j^{(c)}(\cdot|\mathbf{x}_{t-2}^{(I_j)})$ was selected in the $c$-th chain update at iteration $t - 1$. 

Consider the sequence of annealed distributions \( \pi_t = \pi^t \) with \( t \in \{\xi_1, \xi_2, \ldots, \xi_N\} \), where \( 1 = \xi_1 > \xi_2 > \ldots > \xi_N \), e.g. \( \xi_t = 1/t \).

The Monte Carlo population is made of \( N - 1 \) MH chains having \( \{\pi_2, \ldots, \pi_N\} \) as stationary distributions.

The chain ergodic to \( \pi \) has an MTM kernel.
Subsampling IMTM

- Set \( \pi_t \) to be the posterior obtained with \( t\% \) of the data.
- Sampling from the prior at \( t = 0 \) and from the target at \( t = 1 \).
- Requires proper priors and exchangeable data.
- It is NOT similar to annealing:
  - When \( t \approx s \) then \( \pi_t \) may not be “close” to \( \pi_s \). Even is \( s = t \), \( \pi_t \neq \pi_s \).
  - We may run a few “copies” of the chains corresponding to the same \( t \).
  - Fits into the IMTM setup which can use \( N \gg M \).
  - With high-volume data it can lead to significant savings.
Multiple-Try Metropolis and variations

Interacting MTM

IMTM with Annealing and Subsampling

Examples

- Target density, 100% data points
- 40% data points, replica A
- 40% data points, replica B
- 40% data points, replica C
Update for the chain of interest

- Suppose $M = N$; the chain ergodic to $\pi$ is $\{x_t^{(1)}\}_t$. 

A description of the update for the chain of interest when $M = N$. The chain is ergodic to the target distribution $\pi$, and the notation $x_t^{(1)}$ represents the state of the chain at time $t$. This setup is a key part of the Multiple-Try Metropolis (MTM) algorithm and its variations, which are discussed in the context of Markov Chain Monte Carlo (MCMC) methods.
Update for the chain of interest

- Suppose $M = N$; the chain ergodic to $\pi$ is $\{x_t^{(1)}\}_t$.
- For $j = 1, \ldots, M$ draw independently $y_j \sim T_j^{(1)}(\cdot | x_t^{(j)})$.

1. If $j \neq 1$ set $w_j^{(1)}(y_j, x_t^{(1)}) = \pi(y_j) T_j^{(1)}(x_t^{(1)}| x_t^{(j)}) \lambda_j^{(1)}(y_j, x_t^{(1)})$.
2. If $j = 1$ set $w_1^{(1)}(y_1, x_t^{(1)}) = \pi(y_1) T_1^{(1)}(x_t^{(1)}| y_1) \lambda_1^{(1)}(y_1, x_t^{(1)})$.

When $j \neq 1 \rightarrow$ Independent Metropolis.
Update for the chain of interest

- Suppose $M = N$; the chain ergodic to $\pi$ is $\{x_t^{(1)}\}_t$.
- For $j = 1, \ldots, M$ draw independently $y_j \sim T_j^{(1)}(\cdot|x_t^{(j)})$.
  1. If $j \neq 1$ set $w_j^{(1)}(y_j, x_t^{(1)}) = \pi(y_j) T_j^{(1)}(x_t^{(1)}|x_t^{(j)}) \lambda_j^{(1)}(y_j, x_t^{(1)})$.
  2. If $j = 1$ set $w_1^{(1)}(y_1, x_t^{(1)}) = \pi(y_1) T_1^{(1)}(x_t^{(1)}|y_1) \lambda_1^{(1)}(y_1, x_t^{(1)})$.

When $j \neq 1 \rightarrow$ Independent Metropolis.

- Select $J \in \{1, \ldots, M\}$ with probability proportional to $w_j^{(1)}(y_j, x_t^{(1)})$, $j = 1, \ldots, M$ and set $y = y_J$. 
Update for the chain of interest

- Suppose $M = N$; the chain ergodic to $\pi$ is $\{x_t^{(1)}\}_t$.
- For $j = 1, \ldots, M$ draw independently $y_j \sim T_j^{(1)}(\cdot|x_t^{(j)})$.
  
  1. If $j \neq 1$ set $w_j^{(1)}(y_j, x_t^{(1)}) = \pi(y_j) T_j^{(1)}(x_t^{(1)}|x_t^{(j)}) \lambda_j^{(1)}(y_j, x_t^{(1)})$.
  2. If $j = 1$ set $w_1^{(1)}(y_1, x_t^{(1)}) = \pi(y_1) T_1^{(1)}(x_t^{(1)}|y_1) \lambda_1^{(1)}(y_1, x_t^{(1)})$.

When $j \neq 1 \rightarrow$ Independent Metropolis.

- Select $J \in \{1, \ldots, M\}$ with probability proportional to $w_j^{(1)}(y_j, x_t^{(1)}), j = 1, \ldots, M$ and set $y = y_J$.
- Let $x^*_j = x_t^{(1)}$ and for $j = 1, \ldots, M, j \neq J$,
  
  1. If $j \neq 1$ draw $x^*_j \sim T_j^{(1)}(\cdot|x_t^{(j)}) \leftarrow$ independent Metropolis
  2. If $j = 1$ draw $x^*_1 \sim T_1^{(i)}(\cdot|y) \leftarrow$ Metropolis-Hastings
Suppose \( M = N \); the chain ergodic to \( \pi \) is \( \{x_t^{(1)}\}_t \).

For \( j = 1, \ldots, M \) draw independently \( y_j \sim T_j^{(1)}(\cdot|x_t^{(j)}) \).

1. If \( j \neq 1 \) set \( w_j^{(1)}(y_j, x_t^{(1)}) = \pi(y_j) T_j^{(1)}(x_t^{(1)}|x_t^{(j)}) \lambda_j^{(1)}(y_j, x_t^{(1)}) \).

2. If \( j = 1 \) set \( w_1^{(1)}(y_1, x_t^{(1)}) = \pi(y_1) T_1^{(1)}(x_t^{(1)}|y_1) \lambda_1^{(1)}(y_1, x_t^{(1)}) \).

When \( j \neq 1 \rightarrow \text{Independent Metropolis} \).

Select \( J \in \{1, \ldots, M\} \) with probability proportional to \( w_j^{(1)}(y_j, x_t^{(1)}), j = 1, \ldots, M \) and set \( y = y_J \).

Let \( x_J^* = x_t^{(1)} \) and for \( j = 1, \ldots, M, j \neq J \),

1. If \( j \neq 1 \) draw \( x_j^* \sim T_j^{(1)}(\cdot|x_t^{(j)}) \leftarrow \text{independent Metropolis} \)

2. If \( j = 1 \) draw \( x_1^* \sim T_1^{(i)}(\cdot|y) \leftarrow \text{Metropolis-Hastings} \)

Compute \( w_j^{(i)}(x_j^*, y) \) using the same rule as above.
Suppose $M = N$; the chain ergodic to $\pi$ is $\{x_t^{(1)}\}_t$.

For $j = 1, \ldots, M$ draw independently $y_j \sim T_j^{(1)}(\cdot|x_t^{(j)})$.

1. If $j \neq 1$ set $w_j^{(1)}(y_j, x_t^{(1)}) = \pi(y_j) T_j^{(1)}(x_t^{(1)}|x_t^{(j)}) \lambda_j^{(1)}(y_j, x_t^{(1)})$.
2. If $j = 1$ set $w_1^{(1)}(y_1, x_t^{(1)}) = \pi(y_1) T_1^{(1)}(x_t^{(1)}|y_1) \lambda_1^{(1)}(y_1, x_t^{(1)})$.

When $j \neq 1 \rightarrow$ Independent Metropolis.

Select $J \in \{1, \ldots, M\}$ with probability proportional to $w_j^{(1)}(y_j, x_t^{(1)}), j = 1, \ldots, M$ and set $y = y_J$.

Let $x_J^* = x_t^{(1)}$ and for $j = 1, \ldots, M, j \neq J$,

1. If $j \neq 1$ draw $x_j^* \sim T_j^{(1)}(\cdot|x_t^{(j)}) \leftarrow$ independent Metropolis
2. If $j = 1$ draw $x_1^* \sim T_1^{(1)}(\cdot|y) \leftarrow$ Metropolis-Hastings

Compute $w_j^{(i)}(x_j^*, y)$ using the same rule as above.

Set $x_t^{(i)} = y$ with probability $\rho_i$, where $\rho_i$ is the generalized MH ratio and $x_t^{(i)} = x_t^{(i)}$ with probability $1 - \rho_i$. 
Examples: Beta Mixture Model

Let $y_1, \ldots, y_n$ be $n$ i.i.d. samples with density

$$
\sum_{h=1}^{k} \tau_h f(y | \mu_h, \eta_h^{-1})
$$

(2)

We use: $n = 100$, $k = 4$, $(\mu_1, \mu_2, \mu_3, \mu_4)^T = (-3, 0, 3, 6)^T$, $\tau_h = 0.25$, $\eta_h^{-1/2} = 0.55$, $1 \leq k \leq 4$.

- **IMTM-TA**: An IMTM algorithm with $N = 100$ chains and using $\lambda^{(i)}_j(x, y) = 2\{T^{(i)}_j(x | y) + T^{(i)}_j(y | x)\}^{-1}$ weights. The $j$-th proposal uses $T^{(i)}_j(y | x) = N(x, \sigma^2_j I)$ where $\sigma_j = 0.01 + 0.59 \times j/M$ for all $1 \leq j \leq M = 10$, $1 \leq i \leq N$.

- **IMTM-IS**: An IMTM algorithm identical to IMTM-TA but using $\lambda^{(i)}_j(x, y) = \{T^{(i)}_j(x | y) T^{(i)}_j(y | x)\}^{-1}$ weights.
Competing algorithms

These chains were run 10 times longer.

**MH** A population of $N$ parallel RWMH samplers in which the $j$-th Gaussian proposal distribution has covariance $\sigma_j^2 \mathbf{I}$ where $\sigma_j = 0.01 + 0.59 \times j/N$ for all $1 \leq j \leq N$ (the acceptance rates are between 10-60%).

**MH1** A population $N$ parallel RWMH algorithms whose proposal distribution is a mixture of 4 normal densities. The standard deviations of the proposals are divided equally between 0.01 and 0.3.

**MH2** A population of Monte Carlo algorithms in which each of the $N$ transition kernels is a mixture of four RWMH kernels with same standard deviations as those defined for MH2.

**MH.c.o** The MH algorithm described above with cross-over moves.

**MH1.c.o** The MH1 algorithm described above with cross-over moves.

**MH2.c.o** The MH2 algorithm described above with cross-over moves.
ACF Comparison
### Error estimates

<table>
<thead>
<tr>
<th>Method</th>
<th>N=100</th>
<th>N=20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>MH</td>
<td>0.81</td>
<td>0.42</td>
</tr>
<tr>
<td>MH1</td>
<td>0.72</td>
<td>0.21</td>
</tr>
<tr>
<td>MH2</td>
<td>0.99</td>
<td>1.89</td>
</tr>
<tr>
<td>MH c.o.</td>
<td>1.87</td>
<td>1.09</td>
</tr>
<tr>
<td>MH1 c.o.</td>
<td>0.65</td>
<td>0.21</td>
</tr>
<tr>
<td>MH2 c.o.</td>
<td>1.11</td>
<td>1.69</td>
</tr>
<tr>
<td>IMTM-IS</td>
<td>1.40</td>
<td>1.52</td>
</tr>
<tr>
<td>IMTM-IS-a</td>
<td>1.37</td>
<td>1.44</td>
</tr>
<tr>
<td>IMTM-TA</td>
<td>1.31</td>
<td>1.46</td>
</tr>
<tr>
<td>IMTM-TA-a</td>
<td>1.56</td>
<td>1.39</td>
</tr>
</tbody>
</table>
ACF for AIMTM
Subsampling IMTM

Data Generating Process:

\[ y_i \sim \tau_1 N(\mu_1, \eta_1) + \tau_2 N(\mu_2, \eta_2), \quad i = 1, \ldots, N = 1000 \]
\[ \mu = \{-0.2, 0.2\} \quad \eta = \{0.2, 0.2\} \quad \tau = \{0.5, 0.5\} \]

Priors:

- \( p(\mu, \log(\eta)) \propto 1 \)
- \( \log(\tau_1/(1 - \tau_1)) \sim N(0, 1.2) \)

MCMC settings:

- 40k samples, \( N = 10 \) parallel chains, \( M = 10 \)
- temperatures \( \in [0.4, 1] \) equally spaced.
- sampling proportion for subsampling: 40%
Relative Reduction in running time: 11% for sample size $n = 1000$ and 28% when $n = 10K$. 
**Ex: Stochastic Volatility Model**

\[ y_t | h_t \sim \mathcal{N}(0, e^{h_t}) \]
\[ h_t | h_{t-1}, \theta \sim \mathcal{N}(\alpha + \phi h_{t-1}, \sigma^2) \]
\[ h_0 | \theta \sim \mathcal{N}(0, \sigma^2/(1 - \phi^2)) \]

- \( \pi(\theta) \propto 1/(\sigma \beta) \Pi_{(-1,1)}(\phi) \) where \( \beta^2 = \exp(\alpha) \)
- \( \phi \) and the latent variables have non-standard full conditionals

\[ \pi(\phi | \sigma^2, h, y) \propto (1 - \phi^2)^{1/2} \exp \left( -\frac{\phi^2}{2\sigma^2} \sum_{t=2}^{T-1} h_t^2 - \frac{\phi}{\sigma^2} \sum_{t=2}^{T} h_t h_{t-1} \right) \Pi_{(-1,+1)}(\phi) \]

\[ \pi(h_t | \alpha, \phi, \sigma^2, h, y) \propto \exp \left( -\frac{1}{2\sigma^2} \left[ (h_t - \alpha - \phi h_{t-1})^2 - (h_{t+1} - \alpha - \phi h_t)^2 \right] - \frac{1}{2} \left( h_t + y_t^2 \exp\{-h_t\} \right) \right). \]
Stochastic Volatility Model

- \((\alpha, \phi, \sigma^2) = (0, 0.99, 0.01)\) corresponds to daily frequency data.
- \((\alpha, \phi, \sigma^2) = (0, 0.9, 0.1)\) corresponds to weekly frequency data.
- \(\{h_t\}_{1 \leq t \leq 200}\) are latent variables.
- Compare MH samplers (\(N = 20, 50K\) iterations) and IMTM (\(N = 20, M = 5, 10K\) iterations)
Stochastic Volatility Model

Daily data

Weekly data
Stochastic Volatility Model - Cumulated RMSE

Daily data

Weekly data
## Stochastic Volatility Model

<table>
<thead>
<tr>
<th></th>
<th>Daily Data</th>
<th>Weekly Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IMTM-IS</td>
<td>MH</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Value</td>
<td>MSE</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0</td>
<td>0.03018 (0.00583)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.99</td>
<td>0.19853 (0.02038)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.01</td>
<td>0.00204 (0.00241)</td>
</tr>
</tbody>
</table>
Conclusions

- MTM with different proposals is a flexible instrument.
- It integrates well auxiliary information brought by a population of chains.
- Emphasizes the importance of building a reasonable set of chains: tempering and subsampling.
- Central is also the tuning of the $M$ proposal distributions $\Leftrightarrow$ Adaptive MCMC methods.
- Allows mixing of different kernels (RWM, IM, etc).

The paper related to the talk can be downloaded at www.utstat.toronto.edu/craiu/Papers/index.html