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

Metropolis-Hastings Samplers

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

$$\int_{\mathcal{S}} \pi(x) K(x, y) dx = \pi(y).$$

- 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 $\min\{1, \pi(y)T(x_t|y)/\pi(x_t)T(y|x_t)\}$.
  - If y is accepted, then $x_{t+1} = y$, otherwise $x_{t+1} = x_t$.

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.

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 $p_i \propto w(y_i|x_t) = \pi(y_i)T(x_t|y_i)\lambda(x_t, y_i)$.

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 $x_k^* = x_t$.

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).

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 $y_1, \dots, y_K$ 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 $x_1^*, \dots, x_{k-1}^* \sim T(\cdot|y)$ 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 $T(x|y) > 0 \Leftrightarrow T(y|x) > 0$ 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).

- 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 $p_j \propto w(y_j|x_t) = \pi(y_j)T(x_t|y_j)\lambda(x_t, y_j)$.

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 $p_j \propto w(y_j|x_t) = \pi(y_j)T(x_t|y_j)\lambda(x_t, y_j)$.

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 π).
  2. Chains that sample from a distribution not very different from π.

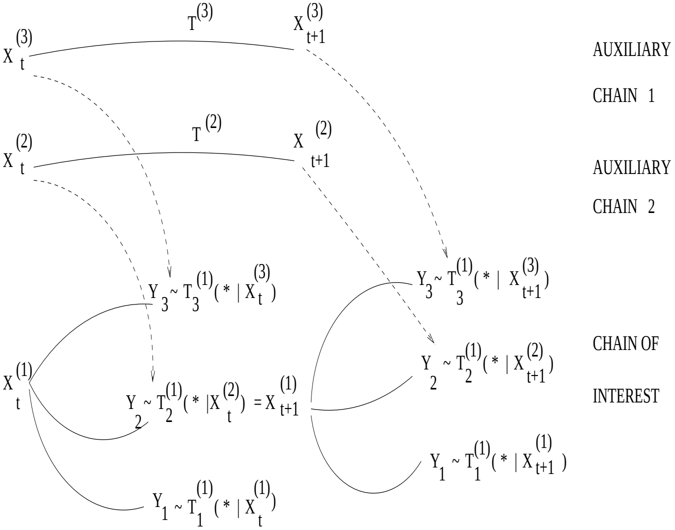
Interacting MTM

- 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 π).
  2. Chains that sample from a distribution not very different from π.
- We need to run many chains!

Interacting MTM

- Consider a population of N chains, $X^{(i)} = \{X_t^{(i)}\}_{n\in\mathbb{N}}$; chain i has MTM transition kernel with M proposal densities $\{T_j^{(i)}\}_{1\leq j\leq M}$.
- Let $\Xi_t = \{x_t^{(i)}\}_{i=1}^N$ 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 Ξ_t.
- The jth proposal for chain i_0 is sampled conditional on $x_t^{(j)}$, $1 \leq j \leq M$ (here we assume $M = N$).

IMTM - A graphical illustration



IMTM

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.

IMTM

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 $\otimes_{i=1}^{N} \pi_i$.

Practical Issues

- 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 $I_M = i$), i.e. $y_j \sim T_j^{(i)}(\cdot|x_{t-1}^{(I_j)})$
- We want to favour contributions from those auxiliary chains that have been "successful" in the previous iteration.
- We suggest using $\tilde{\lambda}_j^{(i)}(x_{t-1},y_j) = \nu_j \lambda_j^{(i)}(x_{t-1},y_j)$, where the factor ν_j is

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

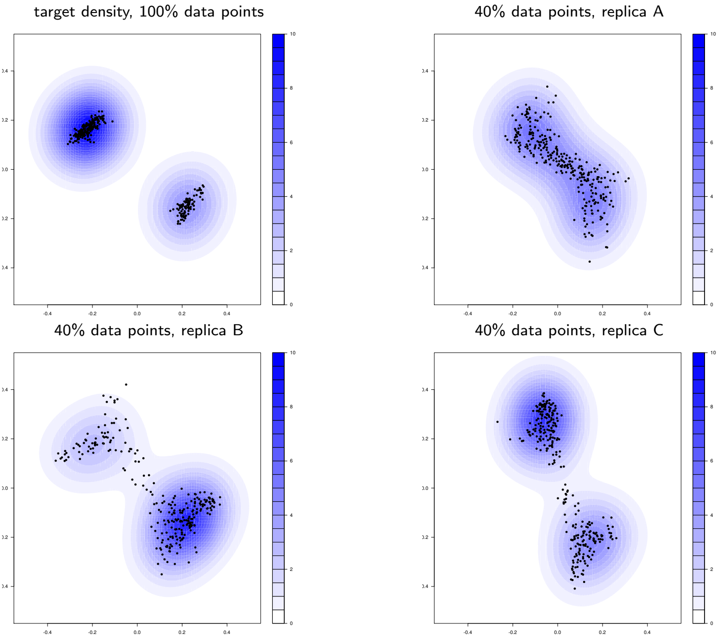
and $\mathbf{1}_c(I_j) = 1$ whenever $y_j \sim T_j^{(c)}(\cdot|x_{t-2}^{(I_j)})$ was selected in the c-th chain update at iteration $t-1$.

Annealed IMTM (AIMTM)

- 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 π has an MTM kernel.

Subsampling IMTM

- Set π_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 π_t may not be "close" to π_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 $N >> M$.
  - With high-volume data it can lead to significant savings.


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 π is $\{x_t^{(1)}\}_t$.

Update for the chain of interest

- Suppose $M = N$; the chain ergodic to π 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 π 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 π 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

Update for the chain of interest

- Suppose $M = N$; the chain ergodic to π is $\{x_t^{(1)}\}_t$.
- For $j = 1, \dots, M$ 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 $J \in \{1, \dots, M\}$ with probability proportional to $w_j^{(1)}(y_j, x_t^{(1)})$, $j = 1, \dots, M$ and set $y = y_J$.
- Let $x_J^* = x_t^{(1)}$ and for $j = 1, \dots, M$, $j \neq J$,
  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
- Compute $w_j^{(i)}(x_j^*, y)$ using the same rule as above.

Update for the chain of interest

- Suppose $M = N$; the chain ergodic to π 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
- Compute $w_j^{(i)}(x_j^*, y)$ using the same rule as above.
- Set $x_{t+1}^{(i)} = y$ with probability ρ_i, where ρ_i is the generalized MH ratio and $x_{t+1}^{(i)} = x_t^{(i)}$ with probability $1 - \rho_i$.

Examples: Beta Mixture Model

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

$$\sum_{h=1}^{k} \tau_h f(y|\mu_h, \eta_h^{-1}) \tag{2}$$

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 $\lambda_j^{(i)}(x, y) = 2\{T_j^{(i)}(x|y) + T_j^{(i)}(y|x)\}^{-1}$ weights. The j-th proposal uses $T_j^{(i)}(y|x) = N(x, \sigma_j^2 \mathbf{I})$ where $\sigma_j = 0.01 + 0.59 * j/M$ for all $1 \leq j \leq M = 10$, $1 \leq i \leq N$.
- IMTM-IS: An IMTM algorithm identical to IMTM-TA but using $\lambda_j^{(i)}(x, y) = \{T_j^{(i)}(x|y) T_j^{(i)}(y|x)\}^{-1}$ 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 $\sigma_j = 0.01 + 0.59 * j/N$ 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.

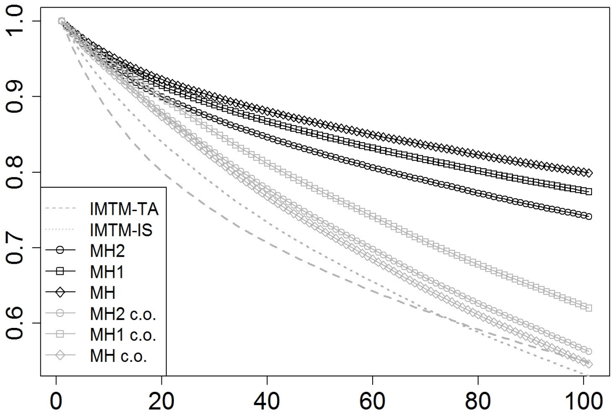
MH2 A population of Monte Carlo algorithms in which each of the N transition kernels is a mixture of four RWMH kernels with same standard deviations as those defined for MH2.

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.

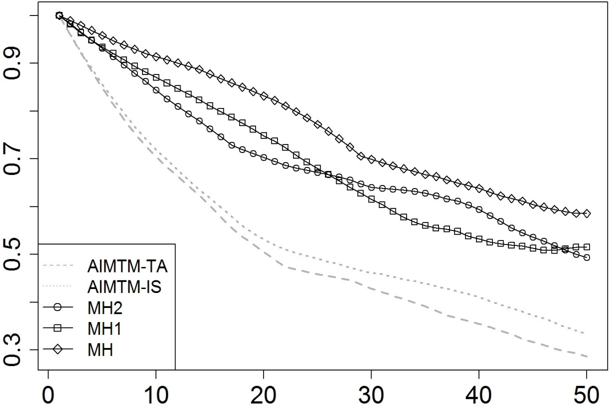
ACF Comparison



Error estimates

	N=100					N=20				
	1	2	3	4	MSE	1	2	3	4	MSE
MH	0.81	0.42	2.08	1.06	18.83	0.39	0.69	0.67	2.28	26.76
	(4.22)	(4.37)	(4.39)	(4.10)		(5.35)	(5.16)	(6.02)	(3.15)	
MH1	0.72	0.21	0.62	0.91	5.42	0.10	0.17	0.66	0.78	7.35
	(2.12)	(2.09)	(2.14)	(2.19)		(2.47)	(1.89)	(2.49)	(2.91)	
MH2	0.99	1.89	1.47	1.01	3.30	0.11	2.80	0.42	0.37	5.09
	(1.57)	(1.73)	(1.87)	(1.89)		(1.99)	(1.71)	(1.98)	(1.85)	
MH c.o.	1.87	1.09	1.91	1.66	7.89	1.74	1.11	1.01	1.75	11.02
	(2.52)	(2.79)	(2.88)	(2.92)		(3.14)	(3.12)	(3.58)	(3.33)	
MH1 c.o.	0.65	0.21	1.59	1.46	2.77	0.51	0.22	1.83	1.12	3.51
	(1.86)	(1.35)	(1.24)	(1.35)		(1.48)	(1.91)	(1.27)	(1.91)	
MH2 c.o.	1.11	1.69	1.27	1.26	2.17	0.59	1.68	0.97	1.14	2.26
	(1.33)	(1.34)	(1.76)	(1.29)		(1.43)	(1.16)	(1.36)	(1.58)	
IMTM-IS	1.40	1.52	1.37	1.42	1.05	1.36	1.39	1.61	1.69	1.42
	(1.01)	(0.98)	(1.22)	(0.87)		(0.98)	(1.20)	(1.12)	(1.42)	
IMTM-IS-a	1.37	1.44	1.58	1.54	0.49	1.31	1.71	1.35	1.72	1.18
	(0.83)	(0.56)	(0.71)	(0.64)		(0.81)	(0.97)	(1.23)	(1.24)	
IMTM-TA	1.31	1.46	1.53	1.61	0.52	1.29	1.21	1.70	1.32	0.89
	(0.38)	(1.06)	(0.48)	(0.73)		(1.34)	(1.05)	(0.31)	(0.59)	
IMTM-TA-a	1.56	1.39	1.60	1.37	0.47	1.63	1.75	1.61	1.44	0.85
	(0.48)	(0.91)	(0.76)	(0.42)		(0.76)	(0.86)	(1.02)	(0.97)	

ACF for AIMTM



Subsampling IMTM

Data Generating Process:

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

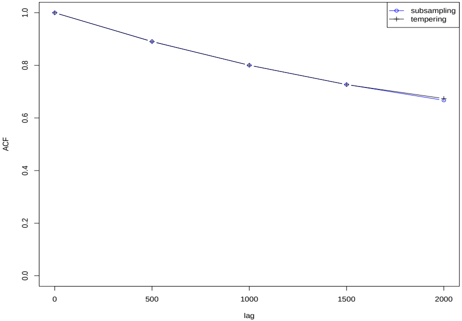
Priors:

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

MCMC settings:

- $40k$ samples, $N = 10$ parallel chains, $M = 10$
- temperatures $\in [0.4, 1]$ equally spaced.
- sampling proportion for subsampling: 40%

ACF Comparison



Relative Reduction in running time: 11% for sample size $n = 1000$ and 28% when $n = 10K$.

Ex: Stochastic Volatility Model

$$
\begin{aligned}
y_t | h_t &\sim \mathcal{N}\left(0, e^{h_t}\right) \\
h_t | h_{t-1}, \boldsymbol{\theta} &\sim \mathcal{N}\left(\alpha + \phi h_{t-1}, \sigma^2\right) \\
h_0 | \boldsymbol{\theta} &\sim \mathcal{N}\left(0, \sigma^2/(1-\phi^2)\right)
\end{aligned}
$$

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

$$
\pi(\phi | \sigma^2, \mathbf{h}, \mathbf{y}) \propto (1-\phi^2)^{1/2} \exp\left(-\frac{\phi^2}{2\sigma^2}\sum_{t=2}^{T-1} h_t^2 - \frac{\phi}{\sigma^2}\sum_{t=2}^{T} h_t h_{t-1}\right)\mathbb{I}_{(-1,+1)}(\phi)
$$

$$
\pi(h_t | \alpha, \phi, \sigma^2, \mathbf{h}, \mathbf{y}) \propto \exp\left(-\frac{1}{2\sigma^2}\left[(h_t - \alpha - \phi h_{t-1})^2 - \right.\right.
$$

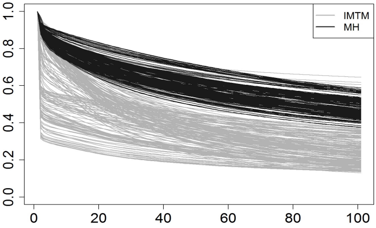
$$
\left.\left.(h_{t+1} - \alpha - \phi h_t)^2\right] - \frac{1}{2}\left(h_t + y_t^2 \exp\{-h_t\}\right)\right).
$$

Stochastic Volatility Model

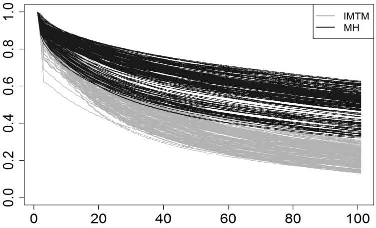
- $(\alpha, \phi, \sigma^2) = (0, 0.99, 0.01)$ corresponds to daily frequency data.
- $(\alpha, \phi, \sigma^2) = (0, 0.9, 0.1)$ corresponds to weekly frequency data.
- $\{h_t\}_{1 \leq t \leq 200}$ are latent variables.
- Compare MH samplers ($N = 20$, 50K iterations) and IMTM ($N = 20$, $M = 5$, 10K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

Stochastic Volatility Model

Daily data

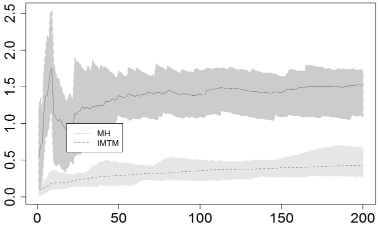


Weekly data

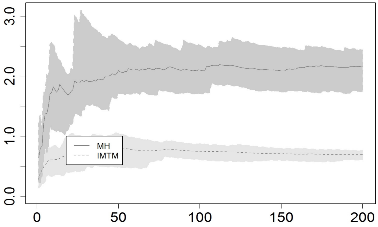


Stochastic Volatility Model - Cumulated RMSE

Daily data



Weekly data



Stochastic Volatility Model

	Daily Data				Weekly Data		
θ	Value	MSE IMTM-IS	MSE MH	θ	Value	MSE IMTM-IS	MSE MH
α	0	0.03018 (0.00583)	0.07392 (0.00201)	α	0	0.00202 (0.00179)	0.00597 (0.00139)
ϕ	0.99	0.19853 (0.02038)	0.29871 (0.04423)	ϕ	0.9	0.01512 (0.03920)	0.08183 (0.04011)
σ^2	0.01	0.00204 (0.00241)	0.01373 (0.00191)	σ^2	0.1	0.00892 (0.00201)	0.07405 (0.00293)

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