Markov Chain Monte Carlo Methods

John Geweke
University of Iowa, USA

2005 Institute on Computational Economics
University of Chicago - Argonne National Laboaratories

July 22, 2005

The problem

$p(\boldsymbol{\theta}, \boldsymbol{\omega} \mid I)$ Distribution of interest

$\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\omega} \in \Omega$; $I =$ "Information"

$$p(\boldsymbol{\theta}, \boldsymbol{\omega} \mid I) = p(\boldsymbol{\theta} \mid I) \cdot p(\boldsymbol{\omega} \mid \boldsymbol{\theta}, I)$$

$p(\boldsymbol{\omega} \mid \boldsymbol{\theta}, I)$ Tractable for simulation

$p(\boldsymbol{\theta} \mid I)$ Not so easy

Leading example:

$$I = \{\text{Model specification}\} \cup \{\text{Data}\}$$

Origins of the problem in econometric inference

Complete model

$$\left.\begin{array}{r} p\left(\mathbf{y} \mid \boldsymbol{\theta}, A\right) \\ p\left(\boldsymbol{\theta} \mid A\right) \end{array}\right\} \Longrightarrow p\left(\boldsymbol{\theta} \mid \mathbf{y}^{o}, A\right)$$

Vector of interest $\boldsymbol{\omega}$

$$p\left(\boldsymbol{\omega} \mid \mathbf{y}^{o}, \boldsymbol{\theta}, A\right)$$

Simulation problem

$$\begin{aligned} \boldsymbol{\theta}^{(m)} &\sim p\left(\boldsymbol{\theta} \mid \mathbf{y}^{o}, A\right) \\ \boldsymbol{\omega}^{(m)} &\sim p\left(\boldsymbol{\omega} \mid \mathbf{y}^{o}, \boldsymbol{\theta}, A\right) \end{aligned}$$

Some background: Random number generation

Mother of all random number generators (Geweke 1996):

$$u^{(m)} \overset{iid}{\sim} uniform\,(0,1)$$

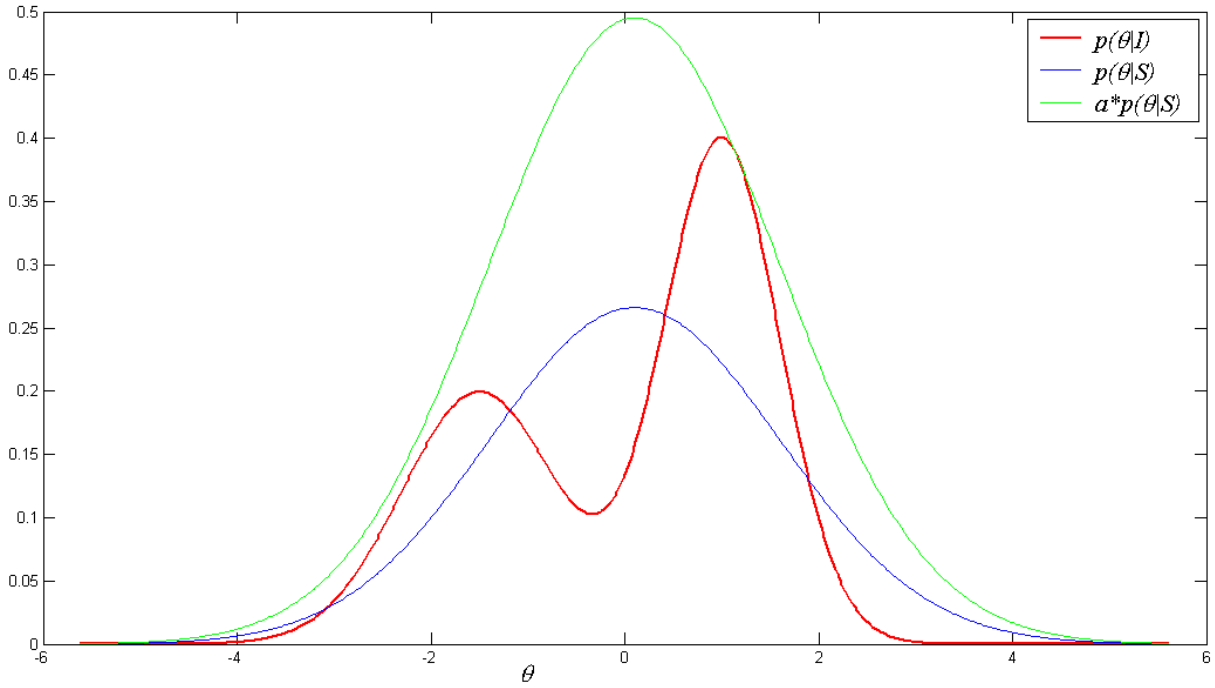
Inverse c.d.f. transformation to simulate $\theta : P(\theta \leq c) = F(c)$:

$$\begin{aligned} \theta^{(m)} &= F^{-1}\left(u^{(m)}\right) \\ \theta^{(m)} &\leq c \iff u^{(m)} = F\left(\theta^{(m)}\right) \leq F(c) \\ &\Longrightarrow P\left(\theta^{(m)} \leq c\right) = P\left[u^{(m)} \leq F(c)\right] = F(c) \end{aligned}$$

Examples:

$$\begin{aligned} \text{Sensible:} \quad & \theta^{(m)} = -\log\left(u^{(m)}\right)/\alpha \implies F(\theta) = 1 - \exp(-\alpha\theta) \\ \text{Questionable:} \quad & \theta^{(m)} = \Phi^{-1}\left(u^{(m)}\right) \implies F(\theta) = \Phi(\theta) \end{aligned}$$

Some background: Acceptance sampling (Motivation)



Some background: Acceptance sampling (Algorithm)

$$\begin{aligned} \text{Target density:} \quad & p(\boldsymbol{\theta} \mid I) \\ \text{Source density:} \quad & p(\boldsymbol{\theta} \mid S) \\ \text{Corresponding kernels:} \quad & k(\boldsymbol{\theta} \mid I) \propto p(\boldsymbol{\theta} \mid I),\ k(\boldsymbol{\theta} \mid S) \propto p(\boldsymbol{\theta} \mid S) \\ & r = \sup\left[k(\boldsymbol{\theta} \mid I) / k(\boldsymbol{\theta} \mid S)\right] < \infty \end{aligned}$$

1. $u \overset{iid}{\sim} uniform(0, 1)$
2. $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta} \mid S)$
3. If $u > k(\boldsymbol{\theta}^* \mid I) \,/\, [r \cdot k(\boldsymbol{\theta}^* \mid S)]$ then return to step 1
4. Set $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}^*$

Essence of Markov Chain Monte Carlo (MCMC)

$$\boldsymbol{\theta}^{(m)} \sim p\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m-1)}, C\right)$$

Convergence in distribution:

$$\boldsymbol{\theta}^{(m)} \xrightarrow{d} p\left(\boldsymbol{\theta} \mid I\right)$$

Ergodicity:

$$M^{-1} \sum_{m=1}^{M} g\left(\boldsymbol{\theta}^{(m)}\right) \xrightarrow{a.s.} E\left[g\left(\boldsymbol{\theta}\right) \mid I\right]$$

If $\boldsymbol{\omega}^{(m)} \sim p\left(\boldsymbol{\omega} \mid \boldsymbol{\theta}^{(m)}\right)$, then

$$\left\{\boldsymbol{\theta}^{(m)}, \boldsymbol{\omega}^{(m)}\right\} \xrightarrow{d} p\left(\boldsymbol{\theta}, \boldsymbol{\omega} \mid I\right) \text{ and } M^{-1} \sum_{m=1}^{M} h\left(\boldsymbol{\omega}^{(m)}\right) \xrightarrow{a.s.} E\left[h\left(\boldsymbol{\omega}\right) \mid I\right]$$

Gibbs sampler: Bivariate distribution

$$\boldsymbol{\theta}' = \left(\theta_{(1)}, \theta_{(2)}\right)$$

$$\theta_{(1)}^{(0)} \sim \text{??}$$

$$\theta_{(2)}^{(0)} \sim p\left(\theta_{(2)} \mid \theta_{(1)}^{(0)}, I\right)$$

For $m = 1, 2, \ldots$

$$\theta_{(1)}^{(m)} \sim p\left(\theta_{(1)} \mid \theta_{(2)}^{(m-1)}, I\right)$$

$$\theta_{(2)}^{(m)} \sim p\left(\theta_{(2)} \mid \theta_{(1)}^{(m)}, I\right)$$

Gibbs sampler: Bivariate distribution example

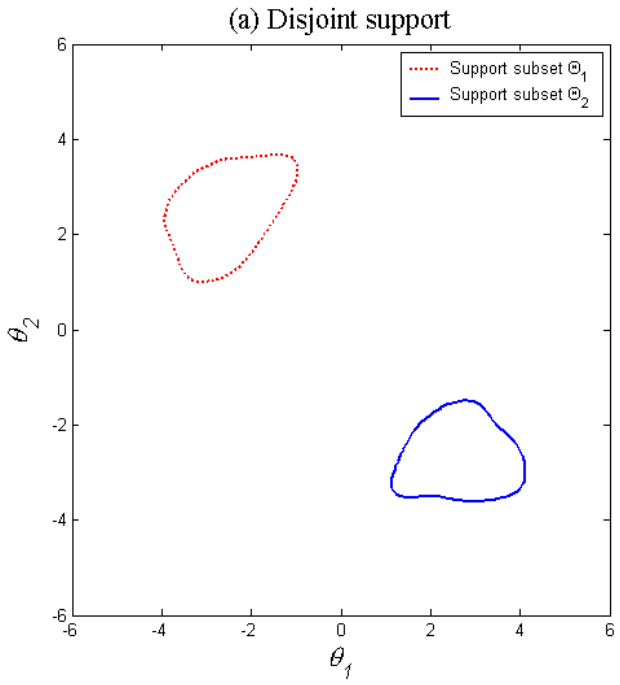
$$\underset{2\times 1}{\boldsymbol{\theta}} \mid I \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

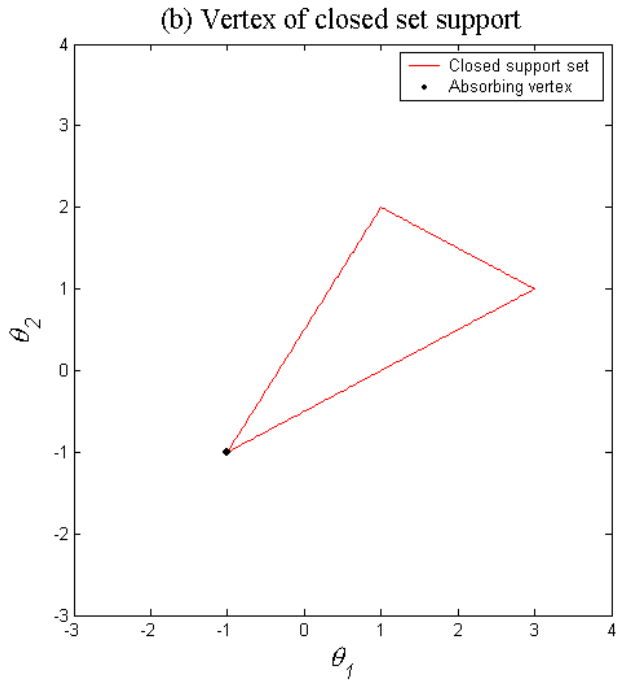
subject to

$$\theta_1 \in \left(\underline{\theta}_1, \overline{\theta}_1\right) \text{ and } \theta_2 \in \left(\underline{\theta}_2, \overline{\theta}_2\right)$$

$$\theta_1 \mid (\theta_2, I) \sim N\left[\mu_1 + \frac{\sigma_{12}\left(\theta_2 - \mu_2\right)}{\sigma_{22}}, \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right] \text{ subject to } \theta_1 \in \left(\underline{\theta}_1, \overline{\theta}_1\right)$$

Gibbs sampler: Potential problems with a bivariate distribution





Gibbs sampler: General case

Blocking of $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}' = \left(\boldsymbol{\theta}'_{(1)}, \ldots, \boldsymbol{\theta}'_{(B)}\right); \;\; \boldsymbol{\theta}_{(b)} \mid \left[\boldsymbol{\theta}_{(a)} \;\; (a \neq b), I\right] \text{ is tractable}$$

Given $\boldsymbol{\theta}^{(0)}$,

$$\boldsymbol{\theta}^{(1)}_{(b)} \sim p\left(\boldsymbol{\theta}_{(b)} \mid \left[\boldsymbol{\theta}^{(1)}_{(a)} \;\; (a < b), \boldsymbol{\theta}^{(0)}_{(a)} \;\; (a > b), I\right]\right) \quad (b = 1, \ldots, B)$$

defines

$$p\left(\boldsymbol{\theta}^{(1)} \mid \boldsymbol{\theta}^{(0)}, G\right) = \prod_{b=1}^{B} p\left(\boldsymbol{\theta}_{(b)} \mid \left[\boldsymbol{\theta}^{(1)}_{(a)} \;\; (a < b), \boldsymbol{\theta}^{(0)}_{(a)} \;\; (a > b), I\right]\right)$$

Gibbs sampler: Convergence result

Suppose that for every point $\boldsymbol{\theta} \in \Theta$ and every measurable $A \subseteq \Theta$,

$$\int_A p(\boldsymbol{\theta} \mid I)\, d\nu(\theta) > 0 \;\Rightarrow\; \int_A p(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, G)\, d\nu(\theta^*) > 0.$$

Then the transition kernel G of the Gibbs sampler is ergodic: if $E[g(\boldsymbol{\theta}), I]$ exists, then for all $\boldsymbol{\theta}^{(0)} \in \Theta$,

$$M^{-1} \sum_{m=1}^{M} g\left(\boldsymbol{\theta}^{(m)}\right) \overset{a.s.}{\to} \int_\Theta g(\boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid I)\, d\theta.$$

For alternative conditions see Geweke (2005) and for weaker conditions still see Tierney (1994).

Metropolis-Hastings algorithm: Structure

Begin with an arbitrary transition density function

$$q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(m-1)}, H\right)$$

Set $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}^*$ with probability

$$\alpha\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(m-1)}, H\right) \in (0, 1)\,.$$

Otherwise set $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}^{(m-1)}$.

Density kernel of accepted candidates $\boldsymbol{\theta}^*$:

$$q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(m-1)}, H\right) \cdot \alpha\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(m-1)}, H\right)$$

Reversibility

Suppose a transition T is reversible with respect to $p(\boldsymbol{\theta} \mid I)$:

$$p\left(\boldsymbol{\theta}^{(m-1)} \mid I\right) p\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, T\right) = p\left(\boldsymbol{\theta}^{(m)} \mid I\right) p\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{(m)}, T\right).$$

Then $p(\boldsymbol{\theta} \mid I)$ is an invariant density for T:

$$\begin{aligned}
&\int_{\Theta} p\left(\boldsymbol{\theta}^{(m-1)} \mid I\right) p\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, T\right) d\nu\left(\theta^{(m-1)}\right) \\
&= \int_{\Theta} p\left(\boldsymbol{\theta}^{(m)} \mid I\right) p\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{(m)}, T\right) d\nu\left(\boldsymbol{\theta}^{(m-1)}\right) \\
&= p\left(\boldsymbol{\theta}^{(m)} \mid I\right) \int_{\Theta} p\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{(m)}, T\right) d\nu\left(\boldsymbol{\theta}^{(m-1)}\right) = p\left(\boldsymbol{\theta}^{(m)} \mid I\right)
\end{aligned}$$

Note that reversibility holds trivially when $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}^{(m-1)}$.

Metropolis-Hastings algorithm: Making it work

For $\boldsymbol{\theta}^{(m)} \neq \boldsymbol{\theta}^{(m-1)}$ we require

$$p\left(\boldsymbol{\theta}^{(m-1)} \mid I\right) p\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right) = p\left(\boldsymbol{\theta}^{(m)} \mid I\right) p\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{(m)}, H\right).$$

where $p\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right) \propto q\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right) \cdot \alpha\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right)$.
Hence

$$\begin{aligned}
& p\left(\boldsymbol{\theta}^{(m-1)} \mid I\right) q\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right) \alpha\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right) \\
= \quad & p\left(\boldsymbol{\theta}^{(m)} \mid I\right) q\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{(m)}, H\right) \alpha\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{(m)}, H\right) \\
\Longrightarrow \quad & \frac{\alpha\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right)}{\alpha\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{(m)}, H\right)} = \frac{p\left(\boldsymbol{\theta}^{(m)} \mid I\right) q\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{(m)}, H\right)}{p\left(\boldsymbol{\theta}^{(m-1)} \mid I\right) q\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right)}
\end{aligned}$$

... and the condition

$$\frac{\alpha\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right)}{\alpha\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{(m)}, H\right)} = \frac{p\left(\boldsymbol{\theta}^{(m)} \mid I\right) q\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{(m)}, H\right)}{p\left(\boldsymbol{\theta}^{(m-1)} \mid I\right) q\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right)}$$

is satisfied if

$$\alpha\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right) = \min\left\{\frac{p\left(\boldsymbol{\theta}^{(m)} \mid I\right) q\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{(m)}, H\right)}{p\left(\boldsymbol{\theta}^{(m-1)} \mid I\right) q\left(\boldsymbol{\theta}^{(m)} \mid \boldsymbol{\theta}^{(m-1)}, H\right)}, 1\right\}.$$

Metropolis-Hastings algorithm: Convergence result

Suppose that for every point $\boldsymbol{\theta} \in \Theta$ and every measurable $A \subseteq \Theta$,

$$\int_A p(\boldsymbol{\theta} \mid I)\, d\nu(\theta) > 0 \;\Rightarrow\; \int_A q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H)\, d\nu(\theta^*) > 0.$$

Then the transition kernel H of the Metropolis-Hastings algoirthm is ergodic: if $E[g(\boldsymbol{\theta}), I]$ exists, then for all $\boldsymbol{\theta}^{(0)} \in \Theta$,

$$M^{-1} \sum_{m=1}^{M} g\left(\boldsymbol{\theta}^{(m)}\right) \overset{a.s.}{\to} \int_\Theta g(\boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid I)\, d\theta.$$

For alternative conditions see Geweke (2005) and for weaker conditions see Tierney (1994).

Metropolis independence chain

$$q\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}, H\right) = q\left(\boldsymbol{\theta}^{*} \mid H\right)$$

If $q\left(\boldsymbol{\theta}^{*} \mid H\right) > 0$ for all $\boldsymbol{\theta}^{*} \in \Theta$, then this chain must be ergodic. Note

$$\begin{aligned}
\alpha\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}^{(m-1)}, H\right) &= \alpha\left(\boldsymbol{\theta}^{*} \mid H\right) = \min\left\{\frac{p\left(\boldsymbol{\theta}^{*} \mid I\right) q\left(\boldsymbol{\theta}^{(m-1)} \mid H\right)}{p\left(\boldsymbol{\theta}^{(m-1)} \mid I\right) q\left(\boldsymbol{\theta}^{*} \mid H\right)}, 1\right\} \\
&= \min\left\{\frac{p\left(\boldsymbol{\theta}^{*} \mid I\right) / q\left(\boldsymbol{\theta}^{*} \mid H\right)}{p\left(\boldsymbol{\theta}^{(m-1)} \mid I\right) / q\left(\boldsymbol{\theta}^{(m-1)} \mid H\right)}, 1\right\}
\end{aligned}$$

Metropolis random walk chain

$$q\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}, H\right) = q\left(\boldsymbol{\theta}^{*} - \boldsymbol{\theta}, H\right) = q\left(\boldsymbol{\theta} - \boldsymbol{\theta}^{*}, H\right)$$

Leading example: $\boldsymbol{\theta}^{*} \mid H \sim N\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}\right)$

This is an instance of a Metropolils chain because $q\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}, H\right) = q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{*}, H\right)$ implying

$$\begin{aligned} \alpha\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}^{(m-1)}, H\right) &= \min\left\{\frac{p\left(\boldsymbol{\theta}^{*} \mid I\right) q\left(\boldsymbol{\theta}^{(m-1)} \mid \boldsymbol{\theta}^{*}, H\right)}{p\left(\boldsymbol{\theta}^{(m-1)} \mid I\right) q\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}^{(m-1)}, H\right)}, 1\right\} \\ &= \min\left\{p\left(\boldsymbol{\theta}^{*} \mid I\right) / p\left(\boldsymbol{\theta}^{(m-1)} \mid I\right), 1\right\} \end{aligned}$$

Metropolis within Gibbs

Suppose that we block a problem for the Gibbs sampler

$$\boldsymbol{\theta}' = \left(\boldsymbol{\theta}'_{(1)}, \ldots, \boldsymbol{\theta}'_{(B)}\right)$$

but $\boldsymbol{\theta}_{(b)} \mid \left[\boldsymbol{\theta}_{(a)} \ \ (a \neq b), I\right]$ is intractable for some b.

Let $q\left(\boldsymbol{\theta}^*_{(b)} \mid \left[\boldsymbol{\theta}_{(a)} (a \neq b)\right], H_b\right)$ be an arbitrary transition density.

At step m of the MCMC algorithm, draw

$$\boldsymbol{\theta}^*_{(b)} \sim q\left(\boldsymbol{\theta}^*_{(b)} \mid \boldsymbol{\theta}^m_{(1)}, \ldots, \boldsymbol{\theta}^{(m)}_{(b-1)}, \boldsymbol{\theta}^{(m-1)}_{(b)}, \boldsymbol{\theta}^{(m-1)}_{(b+1)}, \ldots, \boldsymbol{\theta}^{(m-1)}_{(B)}, H_b\right)$$

... and accept with probability

$$
\begin{aligned}
& \alpha\left(\boldsymbol{\theta}^{*}_{(b)} \mid \boldsymbol{\theta}^{m}_{(1)}, \ldots, \boldsymbol{\theta}^{(m)}_{(b-1)}, \boldsymbol{\theta}^{(m-1)}_{(b+1)}, \ldots, \boldsymbol{\theta}^{(-1)}_{(B)}, H_b\right) \\
= \quad & \min\left\{ \frac{p\left(\boldsymbol{\theta}^{m}_{(1)}, \ldots, \boldsymbol{\theta}^{(m)}_{(b-1)}, \boldsymbol{\theta}^{*}_{(b)}, \boldsymbol{\theta}^{(m-1)}_{(b+1)}, \ldots, \boldsymbol{\theta}^{(-1)}_{(B)}, I\right)}{p\left(\boldsymbol{\theta}^{m}_{(1)}, \ldots, \boldsymbol{\theta}^{(m)}_{(b-1)}, \boldsymbol{\theta}^{(m-1)}_{(b)}, \boldsymbol{\theta}^{(m-1)}_{(b+1)}, \ldots, \boldsymbol{\theta}^{(-1)}_{(B)}, I\right)} \right. \\
& \left. \cdot \frac{q\left(\boldsymbol{\theta}^{(m-1)}_{(b)} \mid \boldsymbol{\theta}^{m}_{(1)}, \ldots, \boldsymbol{\theta}^{(m)}_{(b-1)}, \boldsymbol{\theta}^{(m-1)}_{(b+1)}, \ldots, \boldsymbol{\theta}^{(m-1)}_{(B)}, H_b\right)}{q\left(\boldsymbol{\theta}^{*}_{(b)} \mid \boldsymbol{\theta}^{m}_{(1)}, \ldots, \boldsymbol{\theta}^{(m)}_{(b-1)}, \boldsymbol{\theta}^{(m-1)}_{(b)}, \boldsymbol{\theta}^{(m-1)}_{(b+1)}, \ldots, \boldsymbol{\theta}^{(m-1)}_{(B)}, H_b\right)}, 1 \right\}
\end{aligned}
$$

Chib and Greenberg (1995); Geweke (2005, Section 4.6.2)

Application to inference

Model

$$p(\boldsymbol{\theta} \mid A), \quad p(\mathbf{y} \mid \boldsymbol{\theta}, A)$$

$\boldsymbol{\theta}$	Vector of unobservables
$\mathbf{y}$	Vector of observables

For observed values $\mathbf{y} = \mathbf{y}^o$,

$$p(\mathbf{y}^o \mid \boldsymbol{\theta}, A) \quad \text{Likelihood function}$$

$$p(\boldsymbol{\theta} \mid A)\, p(\mathbf{y}^o \mid \boldsymbol{\theta}, A) \quad \text{Kernel of posterior density function}$$

$$p(\boldsymbol{\theta} \mid \mathbf{y}^o, A) \propto p(\boldsymbol{\theta} \mid A)\, p(\mathbf{y}^o \mid \boldsymbol{\theta}, A)$$

MCMC and the posterior distribution

Some perspective:

$$\begin{aligned} \boldsymbol{\theta} \sim & p\left(\boldsymbol{\theta} \mid A\right) \quad \textit{Easy} \\ \mathbf{y} \sim & p\left(\mathbf{y} \mid \boldsymbol{\theta}, A\right) \quad \textit{Easy} \end{aligned}$$

We seek $p\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m-1)}, \mathbf{y}^o, C\right)$ with the property

$$M^{-1} \sum_{m=1}^{M} g\left(\boldsymbol{\theta}^{(m)}\right) \stackrel{a.s.}{\rightarrow} E\left[g\left(\boldsymbol{\theta}\right) \mid \mathbf{y}^o, A\right]$$

which implies

$$\boldsymbol{\theta}^{(m)} \stackrel{d}{\rightarrow} p\left(\boldsymbol{\theta} \mid \mathbf{y}^o, A\right)$$

Getting it right: the problem

Errors in derivations and/or coding still produce a Markov chain

$$\boldsymbol{\theta}^{(m)} \sim p\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m-1)}, C\right)$$

that will produce reasonable answers.

Most students and many researchers "check" their code by

$$\boldsymbol{\theta} = \boldsymbol{\theta}^0$$
$$\mathbf{y} \sim p\left(\mathbf{y} \mid \boldsymbol{\theta}^0, A\right)$$

and observing whether or not $\left\{\boldsymbol{\theta}^{(m)} \ (m = 1, \ldots, M)\right\}$ is concentrated near $\boldsymbol{\theta}^0$.

This does not work.

Getting it right: the solution

Note there exists a joint distribution for $\boldsymbol{\theta}$ and $\mathbf{y}$,

$$p\left(\boldsymbol{\theta}, \mathbf{y} \mid A\right) = p\left(\boldsymbol{\theta} \mid A\right) p\left(\mathbf{y} \mid \boldsymbol{\theta}, A\right).$$

There is an easy way to simulate from this distribution:

$$\begin{aligned} \boldsymbol{\theta}^{(m)} &\sim p\left(\boldsymbol{\theta} \mid A\right) \\ \mathbf{y}^{(m)} &\sim p\left(\mathbf{y} \mid \boldsymbol{\theta}^{(m)}, A\right) \end{aligned}$$

There is a hard way to simulate from this distribution:

$$\begin{aligned} \widetilde{\boldsymbol{\theta}}^{(m)} &\sim p\left(\boldsymbol{\theta} \mid \widetilde{\boldsymbol{\theta}}^{(m-1)}, \widetilde{\mathbf{y}}^{(m-1)}, C\right) \\ \widetilde{\mathbf{y}}^{(m)} &\sim p\left(\mathbf{y} \mid \widetilde{\boldsymbol{\theta}}^{(m)}, A\right) \end{aligned}$$

... if all derivations and coding are correct, then a standard HAC (Newey-West) test will show that

$$M^{-1} \sum_{m=1}^{M} g\left(\boldsymbol{\theta}^{(m)}\right) - M^{-1} \sum_{m=1}^{M} g\left(\widetilde{\boldsymbol{\theta}}^{(m)}\right)$$

has mean zero.

... if there are errors in derivations (even conceptual) or coding, then the test will fail.

Geweke (2004; 2005, Section 8.1)

References

Chib, S., and E. Greenberg (1995), "Understanding the Metropolis-Hastings algorithm", The American Statistician 49: 327-335.

Geweke, J. (1996), "Monte Carlo simulation and numerical integration", in: H. Amman, D. Kendrick and J. Rust, eds., Handbook of Computational Economics (North-Holland, Amsterdam) 731-800.

Geweke, J. (2004), "Getting It Right: Joint Distribution Tests of Posterior Simulators," Journal of the American Statistical Association 99: 799-804.

Geweke, J. (2005), Contemporary Bayesian Econometrics and Statistics (Wiley, New York; forthcoming in October).

Tierney, L. (1994), "Markov chains for exploring posterior distributions", Annals of Statistics 22: 1701-1762.