Monte Carlo Methods

for Econometric Inference II

Institute on Computational Economics

July 20, 2006

John Geweke, University of Iowa

Implementing Simulation Methods

Simulation or density evaluation:

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

Have we written the code correctly?

Did we even get the derivations right ??

Density Ratio Tests

$$\underset{n\times 1}{\mathbf{x}^{(m)}} \sim p\left(\mathbf{x}\mid I\right) \quad \text{with}$$

$$M^{-1}\sum_{m=1}^{M} h\left(\mathbf{x}^{(m)}\right) \xrightarrow{a.s.} \int_X h\left(\mathbf{x}\right) p\left(\mathbf{x}\mid I\right) d\mathbf{x}$$

$$k\left(\mathbf{x}\mid I\right) \propto p\left(\mathbf{x}\mid I\right), \;\; c_I = \int_X k\left(\mathbf{x}\mid I\right) d\mathbf{x}.$$

Let f be any p.d.f. with the property: $p\left(\mathbf{x}\mid I\right) > 0 \Longrightarrow f\left(\mathbf{x}\right) > 0$. Then

$$\frac{1}{M}\sum_{m=1}^{M} \frac{f\left(\mathbf{x}^{(m)}\right)}{k\left(\mathbf{x}^{(m)}\mid I\right)} \xrightarrow{a.s.} c_I^{-1}.$$

Claim:

$$\frac{1}{M}\sum_{m=1}^{M} f\left(\mathbf{x}^{(m)}\right)/k\left(\mathbf{x}^{(m)} \mid I\right) \overset{a.s.}{\rightarrow} c_I^{-1}.$$

Proof: For

$$g(\mathbf{x}) = \frac{f(\mathbf{x})}{k(\mathbf{x} \mid I)}$$

we have

$$\begin{aligned} \mathsf{E}\left[g(\mathbf{x}) \mid I\right] &= \int_X g(\mathbf{x})\, p(\mathbf{x} \mid I)\, d\mathbf{x} = c_I^{-1} \int_X g(\mathbf{x})\, k(\mathbf{x} \mid I)\, d\mathbf{x} \\ &= c_I^{-1} \int_X f(\mathbf{x})\, d\nu(\mathbf{x}) = c_I^{-1} \\ &\Longrightarrow \frac{1}{M}\sum_{m=1}^{M} g\left(\mathbf{x}^{(m)}\right) \overset{a.s.}{\rightarrow} c_I^{-1} \end{aligned}$$

$$\begin{aligned} \mathbf{x}^{(m)} &\sim p(\mathbf{x} \mid I), \;\; k(\mathbf{x} \mid I) \propto p(\mathbf{x} \mid I), \;\; c_I = \int_X k(\mathbf{x} \mid I)\, d\mathbf{x} \\ \Longrightarrow & \;\; M^{-1} \sum_{m=1}^{M} \frac{f\left(\mathbf{x}^{(m)}\right)}{p\left(\mathbf{x}^{(m)} \mid I\right)} \xrightarrow{a.s.} c_I^{-1} \end{aligned}$$

Special case $k(\mathbf{x} \mid I) = p(\mathbf{x} \mid I)$:

$$M^{-1} \sum_{m=1}^{M} \frac{f\left(\mathbf{x}^{(m)}\right)}{p\left(\mathbf{x}^{(m)} \mid I\right)} \xrightarrow{a.s.} 1.$$

f/p bounded above $\Longrightarrow$ var $[f(\mathbf{x})/p(\mathbf{x} \mid I)] < \infty$, and we can *test*.

How to construct $f(\mathbf{x})$?

$$\boldsymbol{\mu}^{(M)} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{x}^{(m)}, \quad \boldsymbol{\Sigma}^{(M)} = \frac{1}{M}\sum_{m=1}^{M}\left(\mathbf{x}^{(m)} - \boldsymbol{\mu}^{(M)}\right)\left(\mathbf{x}^{(m)} - \boldsymbol{\mu}^{(M)}\right)'$$

Multivariate normal distribution truncated to highest density region of size $100(1-\alpha)\,\%$, $X_\alpha^{(M)} = \left\{\mathbf{x} : \left(\mathbf{x} - \boldsymbol{\mu}^{(M)}\right)'\left(\boldsymbol{\Sigma}^{(M)}\right)^{-1}\left(\mathbf{x} - \boldsymbol{\mu}^{(M)}\right) < \chi^2_\alpha(n)\right\}$

$$\begin{aligned} f(\mathbf{x}) &= (1-\alpha)^{-1}(2\pi)^{-n/2}\left|\boldsymbol{\Sigma}^{(M)}\right|^{-1/2} \\ &\quad \cdot \exp\left[-\left(\mathbf{x} - \boldsymbol{\mu}^{(M)}\right)'\left(\boldsymbol{\Sigma}^{(M)}\right)^{-1}\left(\mathbf{x} - \boldsymbol{\mu}^{(M)}\right)/2\right] I_{X_\alpha^{(M)}}(\mathbf{x}). \end{aligned}$$

Result:

$$M^{-1}\sum_{m=1}^{M}\frac{f\left(\mathbf{x}^{(m)}\right)}{k\left(\mathbf{x}^{(m)} \mid I\right)} \xrightarrow{a.s.} c_I^{-1}, \quad \text{var}\left(\frac{f\left(\mathbf{x}^{(m)}\right)}{k\left(\mathbf{x}^{(m)} \mid I\right)}\right) < \infty.$$

$p(\mathbf{y} \mid \boldsymbol{\theta}_A)$ example

$$y_t = \rho y_{t-1} + \varepsilon_t, \;\; \varepsilon_t \overset{iid}{\sim} N\left(0, h^{-1}\right) \quad (t = 2, \ldots, T)$$

$$y_1 \sim N\left(0, \frac{1}{h(1-\rho)}\right)$$

The model defines the simulator.

$$p(y_1, \ldots, y_T \mid h, \rho, A) = (2\pi)^{-T/2} h^{T/2} \left(1-\rho^2\right)^{1/2} \cdot \exp\left\{-h\left[y_1^2\left(1-\rho^2\right) + \sum_{t=2}^{T}(y_t - \rho y_{t-1})^2\right]/2\right\}$$

Use

$$\rho = 0.8, \;\; h = 1, \;\; T = 5$$

and $M = 10,000$ simulations.

$p(\mathbf{y} \mid \boldsymbol{\theta}_A)$ example (continued)

$$y_1 \sim N\left(0, \frac{1}{h(1-\rho)}\right)$$

Intentional simulation error: $y_1 \sim N\left(0, h^{-1}\right)$

$$p(y_1, \ldots, y_T \mid h, \rho, A) = (2\pi)^{-T/2} h^{T/2} \left(1-\rho^2\right)^{1/2} \cdot \exp\left\{-h\left[y_1^2\left(1-\rho^2\right) + \sum_{t=2}^{T} (y_t - \rho y_{t-1})^2\right]/2\right\}$$

Intentional p.d.f. evaluation error 1: Omit $h^{T/2}\left(1-\rho^2\right)^{1/2}$

Intentional p.d.f. evaluation error 2: Omit $y_1^2\left(1-\rho^2\right)$

$p(\mathbf{y} \mid \boldsymbol{\theta}_A)$ example (concluded)

Outcomes of tests

$\log\left[M^{-1}\sum_{m=1}^{M} f\left(\mathbf{y}^{(m)}\right)/p\left(\mathbf{y}^{(m)} \mid h, \rho, A\right)\right]$ Standard errors in parentheses			
Density evaluation error:	None	Error 1	Error 2
Simulation error:			
None	-.006 (.010)	.508 (.010)	.259 (.011)
Error	-.342 (.011)	.194 (.011)	-.271 (.011)

Joint Distribution Tests (JASA 2004)

Alternative simulators $\left\{\mathbf{x}^{(m)}\right\}$ and $\left\{\tilde{\mathbf{x}}^{(m)}\right\}$

both ergodic, with the same invariant distribution

$$\mathsf{E}\left[|g(\mathbf{x})| \mid I\right] < \infty \implies \frac{1}{M}\sum_{m=1}^{M} g\left(\mathbf{x}^{(m)}\right) - \frac{1}{M}\sum_{m=1}^{M} g\left(\tilde{\mathbf{x}}^{(m)}\right) \stackrel{a.s.}{\rightarrow} 0$$

$$\text{var}\left[g(\mathbf{x}) \mid I\right] \implies \text{Tests possible}$$

We're interested in

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

Marginal-conditional simulator:

$$\boldsymbol{\theta}_A^{(m)} \sim p\left(\boldsymbol{\theta}_A^{(m)} \mid A\right), \mathbf{y}^{(m)} \sim p\left(\mathbf{y} \mid \boldsymbol{\theta}_A^{(m)}, A\right) \quad (m = 1, \ldots, M)$$

Then

$$M^{-1} \sum_{m=1}^{M} g\left(\boldsymbol{\theta}_A^{(m)}, \mathbf{y}^{(m)}\right) \xrightarrow{a.s.} \int_{\Theta_A} \int_Y g\left(\boldsymbol{\theta}_A, \mathbf{y}\right) p\left(\boldsymbol{\theta}_A, \mathbf{y} \mid A\right) d\mathbf{y} d\boldsymbol{\theta}_A.$$

Posterior simulator: $\widetilde{\boldsymbol{\theta}}_{A,\mathbf{y}^o}^{(m)} \sim p\left(\boldsymbol{\theta}_A \mid \widetilde{\boldsymbol{\theta}}_{A,\mathbf{y}^o}^{(m-1)}, \mathbf{y}^o, C\right)$

Successive conditional simulator:

$$\widetilde{\boldsymbol{\theta}}_A^{(0)} \sim p\left(\boldsymbol{\theta}_A \mid A\right)$$

$$\widetilde{\mathbf{y}}^{(m)} \sim p\left(\mathbf{y} \mid \widetilde{\boldsymbol{\theta}}_A^{(m-1)}, A\right),\ \widetilde{\boldsymbol{\theta}}_A^{(m)} \sim p\left(\boldsymbol{\theta}_A \mid \widetilde{\boldsymbol{\theta}}_A^{(m-1)}, \mathbf{y}^{(m)}, C\right)$$
$$(m = 1, ..., M)$$

$$M^{-1}\sum_{m=1}^{M} g\left(\widetilde{\boldsymbol{\theta}}_A^{(m)}, \widetilde{\mathbf{y}}^{(m)}\right) \xrightarrow{a.s.} \int_{\Theta_A}\int_Y g\left(\boldsymbol{\theta}_A, \mathbf{y}\right) p\left(\boldsymbol{\theta}_A, \mathbf{y} \mid A\right) d\mathbf{y}\, d\boldsymbol{\theta}_A.$$

Example: Mixed-t posterior simulator

$$y_t \sim t\left(\mu_1, h_1^{-1}; \nu\right) \text{ with probability } p = p_1,$$

$$y_t \sim t\left(\mu_2, h_2^{-1}; \nu\right) \text{ with probability } 1 - p = p_2.$$

In this example, $\nu = 5$ is assumed (i.e., dogmatic prior)

Priors for other parameters:

$$\mu_j \sim N\left(\underline{\mu}, \underline{h}_\mu^{-1}\right) \quad (j = 1, 2) \quad \left(\underline{\mu} = 0, \underline{h}_\mu = 1\right)$$

$$\underline{s}^2 h_j \overset{i.i.d.}{\sim} \chi^2\left(\underline{\nu}\right) \quad (j = 1, 2) \quad \left(\underline{s}^2 = 3, \underline{\nu} = 3\right)$$

$$p \sim \text{Beta}\left(\underline{r}, \underline{r}\right) \quad \left(\underline{r} = 2\right)$$

Observables simulator (conditional on the parameters):

(1) Latent variables (i.i.d., $t = 1, \ldots, T$):

$$\begin{aligned} P(\widetilde{s}_t = 1) &= p,\ \ P(\widetilde{s}_t = 2) = 1 - p \\ \nu \widetilde{h}_t &\sim \chi^2(\nu) \end{aligned}$$

(2) Then for $t = 1, \ldots, T$,

$$y_t \mid (\widetilde{s}_t = j) \sim N\left[\mu_j, \left(h_j \widetilde{h}_t\right)^{-1}\right]$$

Prior simulator:

$$\mu_j \overset{i.i.d.}{\sim} N\left(\underline{\mu}, \underline{h}_{\mu}^{-1}\right) \quad (j = 1, 2) \quad \left(\underline{\mu} = 0, \underline{h}_{\mu} = 1\right)$$

$$\underline{s}^2 h_j \overset{i.i.d.}{\sim} \chi^2\left(\underline{\nu}\right) \quad (j = 1, 2) \quad \left(\underline{s}^2 = 3, \underline{\nu} = 3\right)$$

$$p \sim \text{Beta}\left(\underline{r}, \underline{r}\right) \quad \left(\underline{r} = 2\right)$$

Error 1:

$$p \sim \text{Beta}(1, 1)$$

... We now have the marginal-conditional simulator.

Next: Successive-conditional simulator

Gibbs sampling algorithm:

$$\mu_j \sim N\left(\overline{\mu}_j, \overline{h}_j^{-1}\right) \quad (j = 1, 2)$$

where

$$\overline{h}_j = \underline{h}_\mu + h_j \sum_{t:s_t=j} \widetilde{h}_t \; \overline{\mu}_j = \overline{h}_j^{-1} \left(\underline{h}_\mu \underline{\mu} + \sum_{t:s_t=j} h_j \widetilde{h}_t y_t \right)$$

Error 3:

$$\mu_j = \overline{\mu}_j \;\; (j = 1, 2)$$

$$\overline{s}_j^2 h_j \sim \chi^2 \left(\overline{\nu}_j\right) \quad (j = 1, 2)$$

where

$$\overline{\nu}_j = \underline{\nu} + T_j, \overline{s}_j^2 = \underline{s}_j^2 + \sum_{t:s_t=j} \widetilde{h}_t \left(y_t - \mu_j\right)^2$$

$$p \sim Beta\left(T_1 + \underline{r}, T_2 + \underline{r}\right).$$

In algorithm MCMC1, $\left(s_t, \widetilde{h}_t\right)$ is drawn jointly:

$$p\left(s_t = j\right) \propto p_j h_j^{1/2} \left[1 + \nu^{-1} h_j \left(y_t - \mu_j\right)^2\right]^{-(\nu+1)/2} \quad (\nu = 5) \qquad (1)$$

$$\left[h_j \left(y_t - \mu_{s_t}\right)^2 + \nu\right] \widetilde{h}_t \sim \chi^2 \left(\nu + 1\right) \quad (\nu = 5) \qquad (2)$$

Error 4:

$$\nu \widetilde{h}_t \sim \chi^2 \left(\nu\right) \quad (\nu = 5)$$

Error 5: Use (2), but not right after (1)

In algorithm MCMC2 $\widetilde{s}_t$ is drawn separately:

$$p\left(\widetilde{s}_t = j\right) \propto p_j h_j^{1/2} \exp\left[-h_j \widetilde{h}_t \left(y_t - \mu_j\right)^2 / 2\right] \quad (j = 1, 2)$$

$$\left[h_{\widetilde{s}_t}\left(y_t - \mu_{s_t}\right)^2 + \nu\right] \widetilde{h}_t \sim \chi^2\left(\nu + 1\right) \quad (\nu = 5)$$

(same as (2).

Conditional on the parameters **and** $\widetilde{h}_t$ $(t = 1, \ldots, T)$

$$y_t \mid \left(\widetilde{s}_t = j\right) \sim N\left[\mu_j, \left(h_j \widetilde{h}_t\right)^{-1}\right]$$

Error 2:

$$y_t \mid \left(\widetilde{s}_t = j\right) \sim t\left(\mu_j, h_j^{-1}, \nu\right) \quad (\nu = 5)$$

The tests

Marginal-conditional simulator: $\left\{\boldsymbol{\theta}_A^{(m)}, \mathbf{y}^{(m)}\right\}$

Successive-conditional simulator: $\left\{\widetilde{\boldsymbol{\theta}}_A^{(m)}, \widetilde{\mathbf{y}}^{(m)}\right\}$

$m = 1, \ldots, M$, $M = 250,000$ in each case

Moments for testing: 5 first moments and 15 second moments of

$$\boldsymbol{\theta}'_A = (\mu_1, \mu_2, h_1, h_2, p)'$$

		Rejections (out of 20) at			
Algorithm	*Error*	$p = .05$	$p = .01$	$p = .005$	$p = .001$
MCMC1	*0. None*	0	0	0	0
MCMC2	*0. None*	0	0	0	0
MCMC1	*1. Prior simulation of* p	4	3	3	2
MCMC1	*2. Simulation of* $\mathbf{y}$	10	9	9	9
MCMC1	*3.* μ *variance*	11	10	10	9
MCMC1	*4.* $\widetilde{h}_t$ *degrees of freedom*	5	3	3	3
MCMC1	*5.* $(\widetilde{s}_t, \widetilde{h}_t)$ *draw*	7	6	6	6

Variance Reduction

The main idea

$$\overline{h}^{(M)} = \sum_{m=1}^{M} w\left(\boldsymbol{\theta}_A^{(m)}\right) h\left(\omega^{(m)}\right) / \sum_{m=1}^{M} w\left(\boldsymbol{\theta}_A^{(m)}\right) \stackrel{a.s.}{\rightarrow} \mathsf{E}\left[h\left(\boldsymbol{\omega}\right) \mid I\right].$$

Can we find $h^*\left(\boldsymbol{\theta}_A, \boldsymbol{\omega}\right)$ such that:

$$\mathsf{E}\left[h^*\left(\boldsymbol{\theta}_A, \boldsymbol{\omega}\right) \mid I\right] = \mathsf{E}\left[h\left(\boldsymbol{\omega}\right) \mid I\right],$$

$$\mathsf{var}\left[h^*\left(\boldsymbol{\theta}_A, \boldsymbol{\omega}\right) \mid I\right] < \mathsf{var}\left[h\left(\boldsymbol{\omega}\right) \mid I\right],$$

$$\overline{h}^{*(M)} = \frac{\sum_{m=1}^{M} w\left(\boldsymbol{\theta}_A^{(m)}\right) h^*\left(\boldsymbol{\theta}_A^{(m)}, \boldsymbol{\omega}^{(m)}\right)}{\sum_{m=1}^{M} w\left(\boldsymbol{\theta}_A^{(m)}\right)} \stackrel{a.s.}{\rightarrow} \mathsf{E}\left[h\left(\boldsymbol{\omega}\right) \mid I\right]?$$

Concentrated Expectations

The principle: Suppose we need to evaluate $\int\int f(x,y)\,p(x,y)\,dxdy$.

Direct sampling:

$$\left(x^{(m)},y^{(m)}\right)\overset{i.i.d.}{\sim}p(x,y)\,,$$

$$M^{-1}\sum_{m=1}^{M}f\left(x^{(m)},y^{(m)}\right)\overset{a.s.}{\rightarrow}\int\int f(x,y)\,p(x,y)\,dxdy$$

$$\int\int f(x,y)\, p(x,y)\, dx dy$$

Suppose we have an analytical evaluation of

$$g(x) = \mathsf{E}[f(x,y) \mid x] = \int f(x,y)\, p(x,y)\, dy / \int p(x,y)\, dy$$

By the law of iterated expectations $\mathsf{E}[g(x)] = \mathsf{E}[f(x,y)]$ and so

$$M^{-1} \sum_{m=1}^{M} g\left(x^{(m)}\right) \xrightarrow{a.s.} \mathsf{E}[f(x,y)] = \int\int f(x,y)\, p(x,y)\, dx dy.$$

By the Rao-Blackwell theorem

$$\mathsf{var}[g(x)] \leq \mathsf{var}[f(x,y)].$$

Theorem 4.4.1 Concentrated expectations in posterior simulation

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\omega} \mid I) &= p(\boldsymbol{\theta} \mid I)\, p(\boldsymbol{\omega} \mid \boldsymbol{\theta}, I), \\ \overline{\omega} &= \mathsf{E}(\omega \mid I), \\ \boldsymbol{\theta}' &= \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right) \end{aligned}$$

$$\boldsymbol{\theta}^{(m)} \overset{iid}{\sim} p\left(\boldsymbol{\theta}^{(m)} \mid I\right), \quad \begin{aligned} \omega_1^{(m)} &\sim p\left(\omega \mid \boldsymbol{\theta}^{(m)}, I\right) \\ \omega_2^{(m)} &= \mathsf{E}\left(\omega \mid \boldsymbol{\theta}^{(m)}, I\right) \\ \omega_3^{(m)} &= \mathsf{E}\left(\omega \mid \boldsymbol{\theta}_1^{(m)}, I\right) \end{aligned}$$

$$\boldsymbol{\theta}^{(m)} \overset{iid}{\sim} p\left(\boldsymbol{\theta}^{(m)} \mid I\right), \quad \begin{aligned} \omega_1^{(m)} &\sim p\left(\omega \mid \boldsymbol{\theta}^{(m)}, I\right) \\ \omega_2^{(m)} &= \mathsf{E}\left(\omega \mid \boldsymbol{\theta}^{(m)}, I\right) \\ \omega_3^{(m)} &= \mathsf{E}\left(\omega \mid \boldsymbol{\theta}_1^{(m)}, I\right) \end{aligned}$$

$$\overline{\omega}_j^{(M)} = M^{-1} \sum_{m=1}^{M} \omega_j^{(m)} \, (j = 1, 2, 3)$$

Then

$$M^{1/2}\left(\overline{\omega}_j^{(M)} - \overline{\omega}\right) \xrightarrow{d} N\left(0, \tau_j^2\right) \quad (j = 1, 2, 3)$$
$$\tau_3^2 \leq \tau_2^2 \leq \tau_1^2$$

Note: Theorem 4.4.1 is not known to hold for importance sampling.

This is not important.

What is important:

(1) Principle of concentrated expectations;

(2) We can still get the numerical standard error.

Same points apply to MCMC.

Antithetic Sampling

The principle: Suppose we need to evaluate $\mathsf{E}(x)$.

There is an *antithetic variate* y with the (defining) properties

$$\mathsf{E}(y) = \mathsf{E}(x), \quad \mathsf{var}(y) = \mathsf{var}(x), \quad \mathsf{cov}(y, x) < 0.$$

Then

$$\mathsf{E}\left(\frac{x+y}{2}\right) = \mathsf{E}(x),$$
$$\mathsf{var}\left(\frac{x+y}{2}\right) = \frac{1}{2}\mathsf{var}(x) + \frac{1}{2}\mathsf{cov}(x, y) < \frac{1}{2}\mathsf{var}(x).$$

Application to simulation

$$
\begin{aligned}
\left(\omega^{(1,m)}, \omega^{(2,m)}\right)' &\sim p\left(\boldsymbol{\omega} \mid R\right) \\
\omega^{(j,m)} &\sim p\left(\omega \mid I\right) \quad (j = 1, 2) \\
\text{cov}\left(\omega^{(1,m)}, \omega^{(2,m)} \mid R\right) &< 0
\end{aligned}
$$

Then

$$
\text{var}\left[\sum_{m=1}^{M}\left(\omega^{(1,m)} + \omega^{(2,m)}\right)/2M \mid R\right] < (1/2)\,\text{var}\left(\sum_{m=1}^{M}\omega^{(1,m)}/M \mid I\right).
$$

(Remember the extra time taken to get $\omega^{(2,m)}$.)

To approximate

$$\mathsf{E}\left[h\left(\omega\right) \mid I\right],$$

use

$$\sum_{m=1}^{M}\left[h\left(\omega^{(1,m)}\right)+h\left(\omega^{(2,m)}\right)\right]/2M.$$

Guaranteed improvement? No.

But: You can try it and look at the numerical standard error.

Transition Mixtures

In the Metropolis-Hastings algorithm:

$$q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H) = \sum_{j=1}^{J} \pi_j q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right)$$

(1) Select $q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right)$ with probability π_j $(j = 1, \ldots, J)$
(2) Draw $\boldsymbol{\theta}^* \sim q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right)$
(3) Accept with probability

$$\alpha(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H) = \min\left[\frac{p(\boldsymbol{\theta}^* \mid I)/q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H)}{p(\boldsymbol{\theta} \mid I)/q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*, H)}, 1\right].$$

Suppose instead:

(1) Select $q\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}, H_{j}\right)$ with probability π_{j} $(j=1, \ldots, J)$
(2) Draw $\boldsymbol{\theta}^{*} \sim q\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}, H_{j}\right)$
(3) Accept with probability

$$\alpha\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}, H_{j}\right)=\min \left[\frac{p\left(\boldsymbol{\theta}^{*} \mid I\right) / q\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}, H_{j}\right)}{p(\boldsymbol{\theta} \mid I) / q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{*}, H_{j}\right)}, 1\right]$$

This works.

Why does this work? The reversibility condition for

$$q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H\right) = \sum_{j=1}^{J} \pi_j q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right)$$

is

$$\begin{aligned} & p\left(\boldsymbol{\theta} \mid I\right) \sum_{j=1}^{J} \pi_j q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right) \alpha\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right) \\ = \; & p\left(\boldsymbol{\theta}^* \mid I\right) \sum_{j=1}^{J} \pi_j q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*, H_j\right) \alpha\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*, H_j\right). \end{aligned}$$

$$
\begin{aligned}
& p(\boldsymbol{\theta} \mid I) \sum_{j=1}^{J} \pi_j q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right) \alpha\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right) \\
= \; & p(\boldsymbol{\theta}^* \mid I) \sum_{j=1}^{J} \pi_j q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*, H_j\right) \alpha\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*, H_j\right).
\end{aligned}
$$

holds if

$$
\begin{aligned}
& p(\boldsymbol{\theta} \mid I) \, q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right) \alpha\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right) \\
= \; & p(\boldsymbol{\theta}^* \mid I) \, q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*, H_j\right) \alpha\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*, H_j\right) \quad (j = 1, \ldots, J)
\end{aligned}
$$

$$
\Longleftarrow \; \alpha\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right) = \min\left[\frac{p(\boldsymbol{\theta}^* \mid I) / q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}, H_j\right)}{p(\boldsymbol{\theta} \mid I) / q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*, H_j\right)}, 1\right].
$$

Transition mixtures can guarantee the ergodicity of MCMC.

Example:

$$q\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}, H_{1}\right)=p\left(\boldsymbol{\theta}^{*} \mid I\right)$$

Metropolis within Gibbs

Motivation: Nice Gibbs sampler except that we cannot perform

$$\boldsymbol{\theta}_{(b)} \sim p\left(\boldsymbol{\theta}_{(b)} \mid \boldsymbol{\theta}_{-(b)}, I\right).$$

Metropolis within Gibbs: Draw

$$\boldsymbol{\theta}^*_{(b)} \sim q\left(\boldsymbol{\theta}^*_{(b)} \mid \boldsymbol{\theta}^{(m)}_{<(b)}, \boldsymbol{\theta}^{(m-1)}_{>(b-1)}, H_b\right)$$

and set $\boldsymbol{\theta}^{(m)}_{(b)} = \boldsymbol{\theta}^*_{(b)}$ with probability...

$$\alpha\left(\boldsymbol{\theta}^*_{(b)} \mid \boldsymbol{\theta}^{(m)}_{<(b)}, \boldsymbol{\theta}^{(m-1)}_{>(b-1)}, H_b\right)$$

$$= \min\left\{\frac{p\left(\boldsymbol{\theta}^{(m)}_{<(b)}, \boldsymbol{\theta}^*_{(b)}, \boldsymbol{\theta}^{(m-1)}_{>(b)} \mid I\right) / q\left(\boldsymbol{\theta}^*_{(b)} \mid \boldsymbol{\theta}^{(m)}_{<(b)}, \boldsymbol{\theta}^{(m-1)}_{>(b-1)}, H_b\right)}{p\left(\boldsymbol{\theta}^{(m)}_{<(b)}, \boldsymbol{\theta}^{(m-1)}_{>(b-1)} \mid I\right) / q\left(\boldsymbol{\theta}^{(m-1)}_{(b)} \mid \boldsymbol{\theta}^{(m)}_{<(b)}, \boldsymbol{\theta}^*_{(b)}, \boldsymbol{\theta}^{(m-1)}_{>(b)}, H_b\right)}, \quad 1\right\}$$

(Otherwise, $\boldsymbol{\theta}^{(m)}_{(b)} = \boldsymbol{\theta}^{(m-1)}_{(b)}$.)