Markov chain Monte Carlo (MCMC)

Our goal (as before): Estimate $\mu = E_\pi(g(x))$ where expectation is difficult/impossible to do analytically.

~~Most~~ Often:
Cannot use iid Monte Carlo since we are unable to simulate $X_1, \ldots, X_n \overset{iid}{\sim} \pi$.

Can use imp. sampling in principle but if π is complicated, difficult to find good importance function, especially as dimensions increase.

<u>MCMC solution</u>: Construct a Harris ergodic Markov chain $\underset{\sim}{X} = (X_1, X_2, \ldots\ldots)$ s.t. it has stationary distribution π. Simulate $X_1, X_2, X_3, \ldots\ldots X_n$ and, obtain estimate $\hat{\mu}_n = \frac{\sum_{i=1}^{n} g(X_i)}{n}$.

Under similar conditions to iid. case, we can appeal to S.L.L.N. $\hat{\mu}_n \to \mu$.

Conditions for a C.L.T. are more ~~dif~~ complicated.

Note: Importance sampling and MCMC are not mutually exclusive. Can reweight dependent samples above and obtain importance sampling estimators. E.g. tail probabilities, max. likelihood, multiple expectations w/ a single set of samples etc

Metropolis-Hastings algorithm.

Algorithm used to construct a Harris ergodic chain w/ stationary distr. $\pi(x)$, $x \in \Omega$

Let transition kernel, $K(x,y)$, be a generalization of a transition probability matrix, P_{xy} or $P(x,y)$. (P_{ij} before)

As w/ tpm, transition kernel specifies conditional probabilities, i.e., $K(x,y) = k(y|x)$, where $k(y|x)$ is a pdf when x,y are states in a continuous state space M.Chain.

We want $K(x,y)$ to satisfy detailed balance w.r.t. π:

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

We specify such a $K(x,y)$ using the Metropolis-Hastings algorithm: Define $h(x)$ s.t. $\pi(x) = h(x)/c$

Start w/ $X_0 = x_0 \in \Omega$. For $n = 0, 1, 2, \ldots$

If $X_n = x$, X_{n+1} is generated as follows:

① Generate a candidate or proposal $y \sim q(x,y)$, where $q(x,y)$ is really $q(y|x)$ so proposal y may depend on current value x. For cnts state space, $q(y|x)$ is just a conditional pdf.

② Set $X_{n+1} = y$ (accept proposal y) w/ probability

$$\alpha(x,y) = \begin{cases} \min\left(\dfrac{h(y)\, q(y,x)}{h(x)\, q(x,y)}, 1\right) & \text{if } \pi(x)\, q(x,y) > 0 \\ 1 & \text{else} \end{cases}$$

else reject proposal y and set $X_{n+1} = x$.

Requirements on q:

(a) $q(x,y) = 0 \Rightarrow q(y,x) = 0 \quad \forall x, y \in \Omega$

(b) $q(x,y)$ is transition kernel of irreducible M-C.

$X_0, X_1, X_2, \ldots$ is an M.C., Harris ergodic, w/ stationary distr. π.

E.g. Observe $Y_1, \ldots, Y_n$

Model: $Y_i | \theta \sim N(\theta, 1)$ condtl. indep.

$\theta \sim \text{Log-}t\ (\mu, \sigma, r)$

Want posterior distr. of θ:

$$\pi(\theta | \underset{\sim}{Y}) \propto \mathcal{L}(\underset{\sim}{Y} | \theta)\, p(\theta)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(Y_i - \theta)^2\right\} \times \frac{1}{\theta}\left[1 + \frac{1}{r}\left(\frac{\log\theta - \mu}{\sigma}\right)\right]^{-\frac{(r+1)}{2}}$$

$$\propto \frac{1}{\theta} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(Y_i - \theta)^2\right\} \times \left[1 + \frac{1}{r}\left(\frac{\log\theta - \mu}{\sigma}\right)\right]^{-(r+1)/2}$$

Suppose we chose to sample from $\pi(\theta | \underset{\sim}{Y})$ by Metropolis-Hastings and let $\pi(\theta | \underset{\sim}{Y}) = h(\theta)/c$ where c is unknown normalizing constant.

$$h(\theta) = \frac{1}{\theta} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(Y_i - \theta)^2\right\} \times \left[1 + \frac{1}{r}\left(\frac{\log\theta - \mu}{\sigma}\right)\right]^{-(r+1)/2}$$

We now need to find proposal q.

The Metropolis algorithm ('random walk' Metropolis-Hastings)

M-H algorithm where $q(x,y) = q(y,x)$ for all x, y.

That is, M-H algorithm w/ <u>symmetric proposal</u>, so

acceptance prob. $\alpha(x,y) = \min\left\{1, \frac{\pi(y)}{\pi(x)} \frac{q(y,x)}{q(x,y)}\right\}$

$$= \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\}$$

E.g. $q(x,y)$ is a normal density centered at x.

Propose new value $y^* \sim q(x, \cdot)$ so $y^* \sim N(x, \tau^2)$

Variance τ^2 is a tuning parameter: affects how well the algorithm performs.

τ^2 too big: candidates generated far from current value, maybe in tails $\Rightarrow$ low prob. of being accepted.

τ^2 too small: proposals/candidates accepted often but too close to previous value $\Rightarrow$ chain explores state space very slowly and high autocorrelations across sampled values (large variance/ M.C. error)

Metropolis-Hastings : some history

1940s: Monte Carlo invented by physicists working in Los Alamos : S. Ulam, J. von Neumann, N. Metropolis

Metropolis and Ulam (1949) JASA

1953: Metropolis, Rosenbluth², Teller² (J. Chem. Phys.) Los Alamos Natl. Labs

1970: Hastings (Biometrika) rediscovered, generalized

1984: Geman & Geman Gibbs distr. used in image analysis

1987: Tanner & Wong (JASA) Data augmentation alg.

1990: Gelfand & Smith (JASA) Used in Bayesian inference Gibbs sampler

1994: Tierney (Annals) Laid out all theory

1995: Green (Biometrika): Dimension-jumping, reversible-jump M-H alg.

All fall under umbrella of M-H algorithm

Return to our example. Want to simulate from $\pi(\theta|\underline{y})$.

Suppose we use Metropolis algorithm.
$q(\theta, \theta^*)$ is $N(\theta, \tau^2)$. Symmetric since $q(\theta,\theta^*) = q(\theta^*, \theta)$.

Algorithm: When current value of M. chain, say, $\theta^{(n)} = \theta$,
propose new value $\theta^* \sim N(\theta, \tau^2)$.
Accept θ^* w/ prob. $\alpha(\theta, \theta^*) = \min\left\{1, \frac{\pi(\theta^*)}{\pi(\theta)}\right\}$
$= \min\left\{1, \frac{h(\theta^*)}{h(\theta)}\right\}$

M-H recipe:
Start M.C. at $\theta^{(0)} = c$ for some $c > 0$.

For $i = 1, \ldots, n$
Propose $\theta^* \sim N(\theta^{(i-1)}, \tau^2)$
Accept θ^* w/ probability $\alpha(\theta^{(i-1)}, \theta^*) = \min\left\{1, \frac{h(\theta^*)}{h(\theta^{(i-1)})}\right\}$
i.e., $\theta^{(i)} = \theta^*$
else reject, i.e., $\theta^{(i)} = \theta^{(i-1)}$.

For any expectation that exists, $\mu = E_\pi\{g(\theta)\}$ can obtain an estimate $\hat{\mu}_n = \frac{\sum_{i=1}^{n} g(\theta^{(i)})}{n}$.

$\therefore$ $\hat{\mu}_n \xrightarrow{P} \mu$ as $n \to \infty$

Variable-at-a-time M-H: (for multivariate distr.)

Main idea: suppose target distr. has k components.

'All-at-once' M-H alg. $\tilde{X}^{(n)} \to \tilde{X}^{(n+1)}$

(already discussed)

$$\begin{pmatrix} X_1^{(n)} \\ X_2^{(n)} \\ \vdots \\ X_k^{(n)} \end{pmatrix} \xrightarrow[\text{via M-H step}]{\text{update}} \begin{pmatrix} X_1^{(n+1)} \\ X_2^{(n+1)} \\ \vdots \\ X_k^{(n+1)} \end{pmatrix}$$

Variable-at-a-time M-H: n^{th} update done in small steps

$$\begin{pmatrix} X_1^{(n)} \\ X_2^{(n)} \\ \vdots \\ X_k^{(n)} \end{pmatrix} \xrightarrow[\text{components } X_1^{(n)} \mid X_2^{(n)}, \ldots, X_k^{(n)}]{\text{M-H update, conditioning on all other}} \begin{pmatrix} X_1^{(n+1)} \\ X_2^{(n)} \\ \vdots \\ X_k^{(n)} \end{pmatrix} \xrightarrow[\text{components } (X_2^{(n)} \mid X_1^{(n+1)}, X_3^{(n)}, \ldots, X_k^{(n)})]{\text{M-H update, given all other}} \begin{pmatrix} X_1^{(n+1)} \\ X_2^{(n+1)} \\ X_3^{(n)} \\ \vdots \\ X_k^{(n)} \end{pmatrix}$$

$$- - - - - - \begin{pmatrix} X_1^{(n+1)} \\ \vdots \\ X_{k-1}^{(n+1)} \\ X_k^{(n)} \end{pmatrix} \xrightarrow{\text{MH update}} \begin{pmatrix} X_1^{(n+1)} \\ \vdots \\ X_k^{(n+1)} \end{pmatrix}$$

Note: each component may itself be multidimensional.

Variable-at-a-time M-H:

Finding transition kernel K (i.e. proposal q) to ~~successiv~~ efficiently sample from π may be difficult when state space is multidimensional.
Similar issues as importance sampling, rejection sampling.

Instead, apply M-H to sub blocks/components.

E.g. if $\underset{\sim}{x} = (x_1, x_2)$

'Full condtl. distr.' of $x_1 = \pi_{1|2}(x_1 | x_2)$

" " " " $x_2 = \pi_{2|1}(x_2 | x_1)$

Let $K_{1|2}(x_1, y_1 | x_2)$ and $K_{2|1}(x_2, y_2 | y_1)$ be transition kernels w/ stationary distr. $\pi_{1|2}$ and $\pi_{2|1}$ respectively.

Two step M-H update:

1. Generate $x_1^{(n+1)}$ using $K_{1|2}(x_1^{(n)}, \cdot \mid x_2^{(n)})$, M-H update
2. Generate $x_2^{(n+1)}$ using $K_{2|1}(x_2^{(n)}, \cdot \mid x_1^{(n+1)})$, M-H update

Equivalent to a single update w/ transition kernel

$$K((x_1, x_2), (y_1, y_2)) = \underbrace{K_{1|2}((x_1, y_1) | x_2)}_{\text{update 1st component}} \underbrace{K_{2|1}(x_2, y_2 / y_1)}_{\text{update 2nd component given 1st update}}$$

Example: suppose $\underset{\sim}{x} = (x_1, x_2, x_3)$ where each block may ~~also~~ be multidimensional.

$$\text{Need: } K_{1|(2,3)} \text{ w/ stationary distr. } \overset{\text{full condtl.}}{\pi_{1|(2,3)}}(x_1 | x_2, x_3)$$
$$K_{2|(1,3)} \quad '' \quad '' \quad '' \quad \pi_{2|(1,3)}(x_2 | x_1, x_3)$$
$$K_{3|(1,2)} \quad '' \quad '' \quad '' \quad \pi_{3|(1,2)}(x_3 | x_1, x_2)$$

Construct $K_{1|2,3}$, $K_{2|1,3}$, $K_{3|1,2}$ by using M-H alg. w/ proposals q_1, q_2, q_3 respectively

If full condtl. distr. is standard distr., can directly sample from it. For e.g. if $\pi_{1|2,3}$ = Normal density simulate update for x_1 from a normal.

→ skip below

If current state $= \underset{\sim}{X}^{(n)} = (X_1^{(n)}, X_2^{(n)}, X_3^{(n)})$ produce next state $= \underset{\sim}{X}^{(n+1)} = (X_1^{(n+1)}, X_2^{(n+1)}, X_3^{(n+1)})$ in 3 steps:

① Propose $x_1^* \sim q_1(x_1^{(n)}, x_1^* | x_2^{(n)}, x_3^{(n)})$

Accept x_1^*, ie, set $x_1^{(n+1)} = x_1^*$ w/ prob.

$$\alpha(x_1^{(n)}, x_1^* | x_2^{(n)}, x_3^{(n)}) = \min\left\{1, \frac{\pi_{1|2,3}(x_1^* | x_2^{(n)}, x_3^{(n)}) \, q(x_1^{(n)}, x_1^* | x_2^{(n)}, x_3^{(n)})}{\pi_{1|2,3}(x_1^{(n)} | x_2^{(n)}, x_3^{(n)}) \, q(x_1^*, x_1^{(n)} | x_2^{(n)}, x_3^{(n)})}\right\}$$

else set $x_1^{(n+1)} = x_1^{(n)}$ (reject x^*).

② Propose $x_2^* \sim q_2(x_2^{(n)}, x_2^* \mid x_1^{(n+1)}, x_3^{(n)})$

(note updated value)

(Accept) Set $x_2^{(n+1)} = x_2^*$ w/ prob. $\alpha(x_2^{(n)}, y_2 \mid x_1^{(n+1)}, x_3^{(n)})$

else (Reject): $x_2^{(n+1)} = x_2^{(n)}$.

③ Propose $y_3 \sim q_3(x_3^{(n)}, x_3^* \mid x_1^{(n+1)}, x_2^{(n+1)})$

(Accept): set $x_3^{(n+1)} = x_3^*$ w/ prob. $\alpha(x_3^{(n)}, x_3^* \mid x_1^{(n+1)}, x_2^{(n+1)})$

(Reject) else $x_3^{(n+1)} = x_3^{(n)}$.

The Markov chain constructed by this algorithm is Harris-ergodic w/ stationary distribution π.

Simple example: Poi-Gamma model (C & Louis pg. 143)

$Y_i|\theta_i \sim Poi(\theta_i t_i)$ condtl. indep. $i=1,\dots,k$

Prior $\theta_i|\beta \sim G(a,\beta)$

$t_1,\dots,t_k$ known; a known.

Hyperprior $\beta \sim G(c,d)$. c,d known.

Inference based on posterior distribution

$$\pi(\underset{\sim}{\theta},\beta|\underset{\sim}{y}) \propto L(\underset{\sim}{y}|\underset{\sim}{\theta})\prod_{i=1}^{n} f_1(\theta_i|\beta)\, f_2(\beta)$$

$$= \left\{\prod_{i=1}^{k}\frac{(\theta_i t_i)^{y_i} e^{-\theta_i t_i}}{y_i!}\right\} \times \left\{\prod_{i=1}^{k}\frac{1}{\Gamma(a)\beta^{a}}\theta_i^{a-1}e^{-\theta_i/\beta}\right\} \times \frac{1}{\Gamma(c)d^{c}}\beta^{c-1}e^{-\beta/d}$$

constants

$$\propto \left\{\prod_{i=1}^{k}(\theta_i t_i)^{y_i} e^{-\sum_i^k \theta_i t_i}\right\} \left(\prod_{i=1}^{k}\theta_i^{a-1}\right) e^{-\sum\theta_i/\beta}\, \beta^{c-1-ka} e^{-\beta/d}$$

Full condtl. ~~only keep~~

$$\pi(\theta_i|\beta,\underset{\sim}{y},\underset{\sim}{\theta}_{-i}) \propto (\theta_i t_i)^{y_i} e^{-\theta_i t_i}\, \theta_i^{a-1} e^{-\theta_i/\beta}$$

$$\propto \theta_i^{y_i+a-1} e^{-\theta_i(t_i+\frac{1}{\beta})} \quad \cancel{\propto \text{Gamma}(y_i+a}$$

$$= \theta_i^{(y_i+a)-1} e^{-\theta_i/(t_i+\frac{1}{\beta})^{-1}} \propto \text{Gamma}\left(y_i+a, \left(t_i+\frac{1}{\beta}\right)^{-1}\right)$$

Recognized density! Simulate from Gamma directly, a $\boxed{\text{Gibbs' update}}$.

In fact priors that result in same type of (posterior) distr. for a given likelihood are called 'conjugate' priors.

Note: conjugate priors are ~~often~~ commonly used to make computation simpler but they should not be used without a good (non-computing) justification — M-H algorithm works even if full condtl. is not of recognizable form.

Above example: Gamma is conjugate prior for Poisson.

Other example: Normal prior for mean of normal; Beta prior for Binomial probability (p) etc.

$$\pi(\beta \mid \underset{\sim}{\theta}, \underline{Y}) \propto e^{-\sum_{i=1}^{k}\theta_i/\beta} \; \beta^{c-1-a} \; e^{-\beta/d} = h(\beta \mid \underset{\sim}{\theta}, \underline{Y}), \text{ say}$$

Not recognizable density.

M-H algorithm/update: e.g. simplest one

Propose $\beta^* \sim N(\beta_{\text{current}}, \underbrace{\tau^2}_{\uparrow \text{ tuning parameter}})$

Accept-reject via M-H prob.

So an M-H algorithm for $\pi(\underset{\sim}{\theta}, \beta \mid Y)$ is:

1) Start M.C. at $(\underset{\sim}{\theta}^{(1)}, \beta^{(1)})$ initial values
   any value that is reasonably likely under π is fine.

2) Nth update of each θ_i for $i=1,\ldots,k$ is according to
   $$\pi(\theta_i \mid \underbrace{\theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_k, \beta, \underline{Y}}_{\text{most recent values}})$$
   For this simple example above is just $\pi(\theta_i \mid \beta, \underline{Y})$
   $= \text{Gamma}\left(Y_i + a, \left[t_i + \frac{1}{\beta}\right]^{-1}\right)$ by condtl. indep. of θ_i's given β.

   Sample $\theta_i^{(n)} \sim \text{Gamma}\left(Y_i + a, \left(t_i + \frac{1}{\beta^{(n)}}\right)^{-1}\right)$ for $i=1,\ldots,k$.
   'Gibbs' update: always accept

3) Nth update of β is according to $\pi(\beta \mid \underset{\sim}{\theta}, \underline{Y})$.
   Propose $\beta^* \sim N(\beta^{(n)}, \tau^2)$
   Accept w/ prob. $\alpha(\beta, \beta^* \mid \underset{\sim}{\theta}, \underline{Y}) = \min\left\{1, \frac{h(\beta^* \mid \underset{\sim}{\theta}, \underline{Y})\, \cancel{q(\beta^*, \beta)}}{h(\beta^{(n)} \mid \underset{\sim}{\theta}, \underline{Y})\, \cancel{q(\beta, \beta^*)}}\right\}$

4) Return to step (2).

M.C. produced has stationary distr. $\pi(\underset{\sim}{\theta}, \beta \mid \underline{Y})$ and is Harris ergodic.

Some other options: $\Rightarrow$ depend on current value of β

(1) Propose $\beta^* \sim \text{Gamma}(\gamma_1(\beta), \gamma_2(\beta))$

w/ mean = current value and variance τ^2.

so, $\gamma_1(\beta)\gamma_2(\beta) = \beta$ and $\gamma_1(\beta)\gamma_2(\beta)^2 = \tau^2$

$\Rightarrow \gamma_2(\beta) = \tau^2/\beta$ and $\gamma_1(\beta) = \beta/\gamma_2(\beta) = \beta^2/\tau^2$.

$q(\beta,\beta^*) \neq q(\beta^*,\beta)$ so M-H accept prob.

$$= \alpha(\beta,\beta^*) = \min\left\{1, \frac{h(\beta^*|\underset{\sim}{\theta},\underset{\sim}{Y})}{h(\beta|\underset{\sim}{\theta},\underset{\sim}{Y})}\frac{q(\beta^*,\beta)}{q(\beta,\beta^*)}\right\}$$

where $q(\beta,\beta^*) = \text{Gamma}(\gamma_1(\beta),\gamma_2(\beta))$ pdf evaluated at β^*.

(2) Log-transform β, i.e., set $\psi = \log\beta \in (-\infty,\infty)$

Now use random-walk M-H update to sample from $\psi \mid \underset{\sim}{\theta}, Y$.

Can transform to get β draws, i.e., $\beta = \exp(\psi)$.

(3) Laplace approx. for $\pi(\beta \mid \underset{\sim}{\theta}^{(n)}, \underset{\sim}{Y})$ as proposal $q(\beta,\beta^*)$.

Some basic M.C. theory for discrete time, cntns. state spaces. (Borrowing from Jones & Hobert, 2001)

M.C.: $X_0, X_1, X_2, \ldots$ $\quad X_i \in \Omega$

Discrete state space: t.p.m. $\{P_{ij}\}$ where $P_{ij} = Pr$(move to state j from state i) $= P(X_n = j \mid X_{n-1} = i)$ $\quad i, j \in \Omega$

Cntns. state space: transition density (more generally, the 'transition kernel') is a condtl pdf, $K(x,y) = k(y|x)$ s.t.

$$P(X_n \in A \mid X_{n-1} = x) = \int_A k(y|x)\,dy \quad \forall x \in \Omega, \text{ all intervals } A.$$

Technically: $\forall x \in \Omega$, $\forall A \in \mathcal{B}(\Omega)$

(Borel σ-algebra generated by Ω ≈ collection of subsets of Ω)

$K^n(x,y)$ is n-step transition kernel

$$P(X_{i+n} \in A \mid X_i = x) = \int_A k^n(y|x)\,dx$$

This is an M.C. so

$$P(X_n \in A \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots) = P(X_n \in A \mid X_{n-1} = x_{n-1})$$
$$= \int_A k(y|x_{n-1})\,dy$$

E.g. of M.C. on cntns. state space: AR(1) model

$$X_n = \theta X_{n-1} + \varepsilon_n \qquad \theta \in \mathbb{R}$$

$$\varepsilon_1, \varepsilon_2, \dots \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

$$P(X_n \in A \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots)$$

$$= P(X_n \in A \mid X_{n-1} = x_{n-1}) = \int_A f(x_n = x \mid x_{n-1})\, dx$$

where $f(x_n \mid x_{n-1})$ = pdf of ~~xxxxx~~ $N(\theta x_{n-1}, \sigma^2)$

<u>Stationarity</u>: if π is a density s.t.

$$\pi(y) = \int K(y \mid x)\, \pi(x)\, dx$$

then π is the stationary density for the M.C. defined by K.

If the current state of the chain is drawn from π, then marginal density of next state is also π. (Analogous to discrete state space M.C.'s).

Irreducibility

M.C. can reach all interesting (positive prob.) regions (sets/intervals) in the state space.

Discrete case: $\forall i, j \in \Omega, \exists n$ s.t. $P_{ij}^{n} > 0$.

Continuous state space: π-irreducibility

Let $\pi(A) = \int_A \pi(x)\, dx$ (slight abuse of notation π)

M.C. is π-irreducible if $\forall x \in \Omega$ and all A s.t. $\pi(A) > 0$, $\exists n$ s.t. $P^n(x, A) > 0$.

That is, any set w/ positive prob. under π is accessible from every pt. in state space.

Recurrence

Discrete M.C.: $\forall i \in \Omega$, $P(X_n = i | X_0 = i) = 1$ for some $n < \infty$.

M.C. will return to state i after leaving state i, in a finite # of steps w/ prob. 1.

Equivalently all states will be visited infinitely often by M.C.

More general state spaces:

All interesting states will be visited infinitely often from 'almost all' starting values.

A π-irreducible M.C. is <u>recurrent</u> if for all A s.t. $\pi(A) > 0$:

(1) $P(X_n \in A$ infinitely often $| X_0 = x) > 0$ $\forall x \in \Omega$ and

(2) $P(X_n \in A$ " " " $) = 1$ for π-almost all x, i.e. $\forall x \in \Omega$ except possibly on a set N s.t. $\pi(N) = 0$.

π-irreducible M.C. is <u>positive recurrent</u> if π is a prob. distr.

Allowing for a set of starting values from where we may not reach all A infinitely often is a nuisance

Positive Harris recurrence: A π-irreducible M.C. is positive Harris recurrent if for all A with $\pi(A) > 0$, $P(X_n \in A \text{ infinitely often} \mid X_0 = x) = 1 \ \forall x \in \Omega$, and π is a prob. distr.

E.g. of pos-recurrent but non-Harris chain (Roberts & Rosenthal).

Aperiodicity:

Discrete case: Recall that period of state i, $d(i)$, is gcd of all $n \geq 1$ s.t. $P_{ii}^n > 0$.

If M.C. is aperiodic $d(i) = 1 \ \forall i \in \Omega$

General case (e.g. cntns) case : definition is more technical

Intuition: An M.C. is aperiodic if we cannot partition Ω s.t. M.C. makes a regular tour through partition.

Regular tour would mean it only visits certain blocks at certain ~~intervals~~ times, say $n = 2, 4, 6, 8, \ldots$

Markov chains for MCMC.

If M.C. satisfies the following regularity conditions ('well behaved' for MCMC):
① Aperiodicity ② Positive Harris recurrence.
then M.C. is Harris-ergodic.

Thm. 1: If an M.C. $(X_1, X_2, \dots)$ has stationary distr. π and is Harris-ergodic,
SLLN holds:
If $\mu = E_\pi(g(x)) < \infty$ then
$$\hat{\mu}_n = \frac{\sum_{i=1}^{n} g(X_i)}{n} \to \mu \quad \text{w/ prob. } 1.$$

Note ① This looks exactly like iid M.Carlo case
except before: $X_1, \dots, X_n \overset{iid}{\sim} \pi$
now: $X_1, \dots, X_n$ are dependent and
X_i's are not distr. according to π.
$E(\hat{\mu}_n) \neq \mu$ unless M.C. is stationary,
so $\hat{\mu}_n$ is usually a $\boxed{\text{biased}}$ estimate of μ.
② Not easy to estimate $Var(\hat{\mu}_n)$ and C.L.T. may not hold even if $Var_\pi g(X) < \infty$

MCMC II

③ Although estimate is biased unless $X_0 \sim \pi$,
SLLN holds <u>regardless of initial value distribution.</u>

M-H algorithm lets us construct an Harris-ergodic M.C. w/ stationary distr. π, so SLLN ~~satisfied~~ holds.