

# Markov chain Monte Carlo methods for Uncertainty Quantification

## General concept

### Lecture 21

March 31, 2016

# Motivation: Bayesian inference & prediction

Data:  $D = (D_1, \dots, D_N)$

Parameters:  $X = (X_1, \dots, X_d)$

Likelihood:  $\mathcal{L}(D|X)$

Prior model:  $\pi(X)$

Posterior model: The basis for inference about  $X$   
... via Bayes theorem

$$\pi(X|D) = \frac{\mathcal{L}(D|X)\pi(X)}{\int \mathcal{L}(D|Y)\pi(Y)dY}$$

Predictive model: Predictive distribution of a future obs.  $D_{\text{future}}$ .

$$\mathcal{L}(D_{\text{future}}|D) = \int \mathcal{L}(D_{\text{future}}|X)\pi(X|D)dX$$

... expected likelihood where uncertainty of  $X$  is  
constrained w.r.t  $\pi(X|D)$

# The problem: Challenges in Bayesian Inference

For some function  $g(\cdot)$ ,

- ▶ the derivation of any posterior quantity requires the computation of integrals of the form:

$$E_{\pi(X|D)}(g(X)) = \int g(X)\pi(X|D)dX$$

- ▶ the posterior distribution density  $\pi(X|D)$  or  $\pi(g(X)|D)$  is intractable because of

$$\int \mathcal{L}(D|X)\pi(X)dX = ??$$

CODE

CODE

# Application: Bayesian hierarchical model

Likelihood

$$D_i \sim \text{Bernoulli}(p(t_i|\alpha, \beta)), \quad i = 1, \dots, 23$$

$$p(t_i|\alpha, \beta) = \frac{\exp(\alpha + \beta t_i)}{1 + \exp(\alpha + \beta t_i)}$$

Prior

$$\alpha \sim \text{N}(\mu = 0, \sigma^2 = 10^2)$$

$$\beta \sim \text{N}(\mu = 0, \sigma^2 = 10^2)$$

Posterior:

$$\begin{aligned} \pi(X = (\alpha, \beta) | D) &= \prod_{i=1}^{23} \left( \frac{\exp(\alpha + \beta t_i)}{1 + \exp(\alpha + \beta t_i)} \right)^{D_i} \left( \frac{1}{1 + \exp(\alpha + \beta t_i)} \right)^{1-D_i} \\ &\quad \times \exp\left(-\frac{1}{2}\alpha^2/10^2\right) \times \exp\left(-\frac{1}{2}\beta^2/10^2\right) \times \frac{1}{\text{CONST.}} \end{aligned}$$

# The cure: Monte Carlo methods

Due to the 'essential correspondence' between  
density  $\pi(X|D)$  & samples  $\{X^{(n)} \sim \pi(X|D)\}$  :

- ▶ Posterior density could be re-created via
  - ▶ histograms estimators,
  - ▶ kernel density estimators,
  - ▶ Normal mixture models, etc...
- ▶ Expectations could be approx. via

$$E_{\pi(X|D)}(g(X)) \approx \frac{1}{N} \sum_{n=1}^N g(x^{(n)})$$

# Monte Carlo methods (main idea)

- ▶ Generate an i.i.d. sample  $X^{(n)} \sim \pi(dX)$ , for  $n = 1, \dots, N$ 
  - ▶ Inverse probability integral transform
  - ▶ Rejection sampling
  - ▶ Importance sampling

- ▶ Approx. integral  $E_\pi(g(X)) = \int g(X)\pi(X)dX$

$$\text{with } \bar{g}^{(N)} \approx \frac{1}{N} \sum_{n=1}^N g(X^{(n)})$$

... according to the  $\sqrt{N}$ -CLT:

$$\bar{g}^{(N)} \overset{\text{approx}}{\sim} N(E_\pi(g(X)), \frac{1}{N} \text{Var}_\pi(g(X)))$$

if  $E_\pi(g(\theta)^2) < \infty$

Ripley (2001). Stochastic simulation.

# Markov chain Monte Carlo methods (main idea)

- ▶ Generate a Markov chain  $X^{(n)} \sim P(X^{(n-1)}, d\cdot)$ , for  $n = 1, \dots, N$

- ▶ Approx. integral  $E_{\pi}(g(X)) = \int g(X)\pi(X)dX$

$$\text{with } \bar{g}^{(N)} \approx \frac{1}{N} \sum_{n=1}^N g(X^{(n)})$$

...according to the (Markov chain)  $\sqrt{N}$ -CLT:

$$\bar{g}^{(N)} \underset{\text{approx}}{\sim} N(E_{\pi}(g(X)), \frac{1}{N} \tau_g \text{Var}_{\pi}(g(X)))$$

where  $\tau_g \in (0, \infty)$  is the integrated autocorrelation time

Christian P. Robert and George Casella (2004). Monte Carlo Statistical Methods.



# MCMC theory

Main conditions for  $P(\cdot, d\cdot)$

1. Stationarity w.r.t.  $\pi(d\cdot)$ 
  - Reversibility w.r.t.  $\pi(d\cdot)$
2.  $\phi$ -irreducibility
  - Harris recurrent
3. Aperiodicity

Christian P. Robert and George Casella (2004). Monte Carlo Statistical Methods.

# Stationarity

**Definition:** The Markov chain  $\{x^{(n)}; n = 1, \dots, N\}$  simulated by the transition probability  $P(\cdot, d\cdot)$  has stationary (or invariant) distribution  $\pi(\cdot)$  iff

$$\int_{x \in \mathcal{X}} \pi(dx) P(x, dy) = \pi(dy)$$

where  $x, y \in \mathcal{X}$ .

**Explanation:** If  $x_n \sim \pi(d\cdot)$  &  $x_{n+1} \sim P(x_n, d\cdot)$ , then  $x_{n+1} \sim \pi(d\cdot)$

Hopefully, if we run the Markov chain (started from anywhere) for a long time, then for a long  $N$  the distribution of  $X_N$  will be approx. stationary:

$$X_N \stackrel{\text{approx.}}{\sim} \pi(d\cdot).$$

# Reversibility

**Definition:** The Markov chain  $\{x^{(n)}; n = 1, \dots\}$  simulated by the transition probability  $P(\cdot, d\cdot)$  is reversible w.r.t distribution  $\pi(\cdot)$  iff

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx)$$

where  $x, y \in \mathcal{X}$ .

**Explanation:** It expresses an equilibrium in the flow of the Markov chain: The probability of being in  $x$  and moving to  $y$  is the same as the probability of being in  $y$  and moving to  $x$ .

**Property:** Reversibility implies stationarity

**Rational:** It is more conservative assumption, but it is easier to be checked, since no integral is involved

# Reversibility implies stationarity

**Proposition:** If Markov chain  $\{x^{(n)}\}$  with transition probability  $P(\cdot, d\cdot)$  is reversible w.r.t distribution  $\pi(d\cdot)$ , then  $\pi(d\cdot)$  is the stationarity distr.

**Prof:** We compute that

$$\begin{aligned}\int_{x \in \mathcal{X}} \pi(dx) P(x, dy) &= \int_{x \in \mathcal{X}} \pi(dy) P(y, dx) \\ &= \pi(dy) \int_{x \in \mathcal{X}} P(y, dx) \\ &= \pi(dy)\end{aligned}$$

Christian P. Robert and George Casella (2004). Monte Carlo Statistical Methods.

## $\phi$ -irreducibility

**Definition:** The Markov chain  $\{x^{(n)}\}$  with transition probability  $P(\cdot, d\cdot)$  is  $\phi$ -irreducible if for all  $A \subseteq \mathcal{X}$  with  $\phi(A) > 0$ , there exists a positive integer  $n$  s.t.  $P^n(x, A) > 0$ , for all  $x \in \mathcal{X}$ , where  $P^n(x, A)$  is the  $n$ -step transition probability of the Markov chain.

**Explain:** The Markov chain has positive probability of eventually reaching any state from any other state, in a finite number of iterations.

Christian P. Robert and George Casella (2004). Monte Carlo Statistical Methods.

# Harris recurrent

**Definition:** The Markov chain  $\{x^{(n)}\}$  with transition probability  $P(\cdot, d\cdot)$  is Harris recurrent if for all  $A \subseteq \mathcal{X}$  with  $\pi(A) > 0$  and for all  $x \in \mathcal{X}$ , there exists a positive integer  $n$  s.t.  $P^n(x, A) = 1$ , for all  $x \in \mathcal{X}$ , where  $P^n(x, A)$  is the  $n$ -step transition probability of the Markov chain.

**Explain:** For all  $A \subseteq \mathcal{X}$  with  $\pi(A) > 0$  and for all  $x \in \mathcal{X}$ , the probability that the Markov chain will eventually reach  $A$  from  $x$  is 1.

Christian P. Robert and George Casella (2004). Monte Carlo Statistical Methods.

# Aperiodicity

The Markov chain  $\{x^{(n)}\}$  with transition probability  $P(\cdot, \cdot)$  is aperiodic if there does not exist partition  $\{\mathcal{X}_i; i = 1, \dots, \kappa\}$  where  $\pi(\mathcal{X}_i) > 0$  s.t.

- ▶  $P(x, \mathcal{X}_{i+1}) = 1$  for  $x \in \mathcal{X}_i$  and
- ▶  $P(x, \mathcal{X}_1) = 1$  for  $x \in \mathcal{X}_\kappa$

Christian P. Robert and George Casella (2004). Monte Carlo Statistical Methods.

# Ergodicity

**Theorem:** If Markov chain  $\{x^{(t)}\}$  with transition probability  $P(\cdot, d\cdot)$  is Harris recurrent, aperiodic, and has a stationary distribution  $\pi(d\cdot)$  then for every initial distribution  $\tilde{\pi}$

$$\lim_{n \rightarrow \infty} \left\| \int P^n(x, \cdot) \tilde{\pi}(dx) - \pi(\cdot) \right\| = 0$$

**Explain:** Standard Markov chain theory tells that for any initial seed  $x^{(0)}$ , the realisation of the chain  $\{x^{(1)}, x^{(2)}, x^{(3)}, \dots\}$ , provides via the ergodic theorem, a realisation of the stationary distribution since

$$x^{(n)} \rightarrow \pi(\cdot), \text{ as } n \rightarrow \infty$$



## Markov chain $\sqrt{N}$ -CLT

**Theorem:** If Markov chain  $\{x^{(n)}\}$  with transition probability  $P(\cdot, \cdot)$  is irreducible, aperiodic, and reversible with stationary distribution  $\pi(\cdot)$  then the CLT applies:  
For some function  $g(\cdot)$ , and  $\bar{g}^{(N)} = \frac{1}{N} \sum_{n=1}^N g(x^{(n)})$

$$N^{-1/2}(\bar{g}^{(N)} - E_{\pi}(g(x))) \implies N(0, \tau_g \text{Var}_{\pi}(g(x)))$$

where  $\tau_g = 1 + 2 \sum_{k=1}^{\infty} \text{Cor}(x_n, x_{n+k})$ , if  $\tau_g < \infty$ .

**Explain:** Standard Markov chain theory tells that for any initial seed  $x^{(0)}$ , the realisation of the chain  $\{x^{(1)}, x^{(2)}, x^{(3)}, \dots\}$ , provides, an approx. of the required expectations as

$$\bar{g}^{(N)} \rightarrow E_{\pi}(g(x)), \text{ as } N \rightarrow \infty$$

where  $\bar{g}^{(N)} = \frac{1}{N} \sum_{n=1}^N g(x^{(n)})$ .

CODE

CODE

## Metropolis-Hastings: the algorithm

Simulate from a Metropolis-Hastings transition probability  $P(\cdot, d\cdot)$ , with target distribution  $\pi(d\cdot)$ , and proposal distribution  $q(\cdot, d\cdot)$ .

i.e.  $X^{(n+1)} \sim P(X^{(n)}, d\cdot)$

Given that the current state of the Markov chain is at  $X^{(n)} = x$ :

1. Generate a 'proposed value'  $x'$  from  $q(x, d\cdot)$
2. Calculate

$$a(x, x') = \min\left(1, \frac{\pi(x')}{\pi(x)} \frac{q(x, x')}{q(x', x)}\right)$$

3. With probability  $a(x, x')$  accept the proposed value and set  $X^{(n+1)} = x'$ ; otherwise reject and set  $X^{(n+1)} = x$ .

Generate  $u \sim U(0, 1)$ , and set  $X^{(n+1)} = \begin{cases} x' & , \text{ if } a(x, x') \geq u \\ x & , \text{ if } a(x, x') < u \end{cases}$

# Metropolis-Hastings: the transition probability

The Metropolis-Hastings transition probability is:

$$P(x, y) = q(x, y)a(x, y) + (1 - r(x))\delta_x(y)$$

where  $r(x) = \int q(x, y)a(x, y)dy$ ,

and  $\delta_x(y)$  is the Dirac mass in  $x$ .

## Metropolis-Hastings: Reversibility

The Metropolis-Hastings (as described above) produces a Markov chain  $\{x^{(t)}\}$  which is reversible w.r.t  $\pi(d\cdot)$ .

Prof: We need to show that

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx)$$

If  $x \neq y$ ,

$$\begin{aligned}\pi(dx)P(x, dy) &= [\pi(x)dx][q(x, y)a(x, y)dy] \\ &= \pi(x)q(x, y) \min(1, \frac{\pi(y)}{\pi(x)} \frac{q(y, x)}{q(x, y)}) dx dy \\ &= \min(\pi(x)q(x, y), \pi(y)q(y, x)) dx dy \\ &= [\pi(y)dy][q(y, x)a(y, x)dx] \\ &= \pi(dy)P(y, dx)\end{aligned}$$

If  $x = y$ , then the equation is trivial.

# Metropolis-Hastings: Main properties

The Metropolis-Hastings (as described above) :

- ▶ is reversible and hence admits stationary distribution  $\pi(d\cdot)$ .
- ▶ is irreducible if

$$q(x, y) > 0, \text{ for every } x \in \mathcal{X}, y \in \mathcal{X}$$

since every set of  $\mathcal{X}$  can be reached in a single step

- ▶ is aperiodic, if it allows events  $\{X^{(t+1)} = X^{(t)}\}$ ,  
i.e. the probability of such an event is not zero

# Metropolis-Hastings: Advantages/Challenges

## Advantages:

We only need to know density  $\pi(\cdot)$  up to a normalisation constant

$$a(x, x') = \min\left(1, \frac{\pi(x')}{\pi(x)} \frac{q(x, x')}{q(x', x)}\right)$$

## Challenges:

If the proposal distribution  $q(\cdot, \cdot)$  is poorly chosen:

- ▶ the number of rejections can be high
- ▶ the exploration of the sampling space will be slow
- ▶ the standard error of the MC estimates will be large; because of high autocorrelations;  $\tau_g = 1 + 2 \sum_{k=1}^{\infty} \text{Cor}(x_h, x_{h+k})$

# Metropolis-Hastings: Special cases

Popular special cases of the Metropolis-Hastings algorithm are:

**IMH:** The independence Metropolis-Hastings sampler

**RWM:** The Random Walk Metropolis algorithm\*

**MALA:** The Langevin adjusted Hastings algorithm



# Independence Metropolis-Hastings algorithm (IMH)

The proposal distribution  $q(\cdot, d\cdot)$  is independent on the current state i.e.  $q(x, x') = q(x')$

$$X^{(n+1)} \sim P^{(\text{IMH})}(X^{(n)}, d\cdot)$$

Given that the current state of the Markov chain is at  $X^{(n)} = x$ :

1. Generate  $x'$  from  $q(d\cdot)$
2. Calculate

$$\begin{aligned} a(x, x') &= \min\left(1, \frac{\pi(x')}{\pi(x)} \frac{q(x, x')}{q(x', x)}\right) \\ &= \min\left(1, \frac{\pi(x')}{\pi(x)} \frac{q(x)}{q(x')}\right) \end{aligned}$$

3. With probability  $a(x, x')$  accept and set  $X^{(n+1)} = x'$ ; otherwise reject and set  $X^{(n+1)} = x$ .

# Independence Metropolis-Hastings: notes

1. The proposal distribution  $q(\cdot, d\cdot)$  is independent on the current state; i.e.  $q(x, x') = q(x')$
2. The proposal distribution should be as close as possible to the target (stationary) distribution; i.e.  $q(\cdot) \approx \pi(\cdot)$
3. Ideally, if  $q(\cdot) = \pi(\cdot)$ , then  $a(x, x') = 1$ , and the algorithm reduces to .... i.i.d. sampling from  $\pi(d\cdot)$
4. It is a little bit .... difficult to find 'good'  $q(\cdot)$  s.t.  $q(\cdot) \approx \pi(\cdot)$ , however possible if you try hard...

E.g.: If  $\pi(\cdot)$  is uni-modal,  $q(\cdot)$  can be a multivariate Normal distribution  $N(\mu_\pi, \Sigma_\pi)$  where  $\mu_\pi, \Sigma_\pi$  are properly chosen.

## Random walk Metropolis algorithm (RWM)

The proposal distribution  $q(\cdot, d\cdot)$  is s.t.

$$q(x, x') = \mathcal{N}(x'|x, \sigma^2 \mathbb{I}), \text{ for } \sigma^2 > 0$$

$$X^{(n+1)} \sim P^{(\text{RWM})}(X^{(n)}, d\cdot)$$

Given that the current state of the Markov chain is at  $X^{(n)} = x$ :

1. Generate  $x' \sim \mathcal{N}(x, \sigma^2 \mathbb{I})$
2. Calculate

$$\begin{aligned} a(x, x') &= \min\left(1, \frac{\pi(x')}{\pi(x)} \frac{q(x', x)}{q(x, x')}\right) = \min\left(1, \frac{\pi(x')}{\pi(x)} \frac{\cancel{\mathcal{N}(x|x', \sigma^2 \mathbb{I})}}{\cancel{\mathcal{N}(x'|x, \sigma^2 \mathbb{I})}}\right) \\ &= \min\left(1, \frac{\pi(x')}{\pi(x)}\right) \end{aligned}$$

3. With probability  $a(x, x')$  accept and set  $X^{(n+1)} = x'$ ; otherwise reject and set  $X^{(n+1)} = x$ .

# Random walk Metropolis: notes 1

- ▶ Rational: “Local” exploration of the sampling space, around the neighbourhood of  $X_n = x$ .

$$q(\cdot, d\cdot) : x' = x + \sigma z \quad ; \quad z \sim N(0, 1)$$

- ▶ Move towards modes of  $\pi(\cdot)$  more often than moving away from them
  - ▶ “Uphill moves” are all accepted w.p.  $a(x, x') = 1$
  - ▶ “Downhill moves” may be accepted w.p.  $a(x, x') < 1$ , or rejected w.p.  $1 - a(x, x')$
- ▶ Advantages:
  - ▶ RWM is flexible: the choice of  $q(\cdot, d\cdot)$  is simple
  - ▶ RWM uses the previously simulated value  $x$  at stage  $X^{(n)}$  to generate the proposed value  $x'$  for stage  $X^{(n+1)}$ .

## Random walk Metropolis: notes 2

- ▶ RWM achieves optimal performance, if the proposal scale  $\sigma^2$  is s.t.  $\bar{\alpha}_{\text{opt}} \approx 0.234$

...if the components of  $x := (x_1, \dots, x_d)$  are independent.

...however this rule leads to satisfactory performance in general cases

- ▶ Variations:  $q(\cdot, d\cdot)$  can be any symmetric dist. s.t.

$$q(x, x') = q(|x - x'|)$$

E.g.  $q(\cdot, d\cdot)$  :

propose  $x' \sim U(x - \sigma, x + \sigma)$ .

propose  $x' = x + \sigma z$  ;  $z \sim U(-1, 1)$

## Langevin adjusted Hastings algorithm (MALA)

The proposal distribution  $q(\cdot, d\cdot)$  is s.t.

$$q(x, x') = \mathcal{N}(x' | x + \frac{\sigma^2}{2} \nabla \log(\pi(x)), \sigma^2 \mathbb{I}), \text{ for } \sigma^2 > 0$$

$$X^{(n+1)} \sim P^{(\text{MALA})}(X^{(n)}, d\cdot)$$

Given that the current state of the Markov chain is at  $X^{(n)} = x$ :

1. Generate  $x' \sim \mathcal{N}(x + \frac{\sigma^2}{2} \nabla \log(\pi(x)), \sigma^2 \mathbb{I})$
2. Calculate

$$\begin{aligned} a(x, x') &= \min\left(1, \frac{\pi(x')}{\pi(x)} \frac{q(x', x)}{q(x, x')}\right) \\ &= \min\left(1, \frac{\pi(x')}{\pi(x)} \frac{\mathcal{N}(x | x' + \frac{\sigma^2}{2} \nabla \log(\pi(x')), \sigma^2 \mathbb{I})}{\mathcal{N}(x' | x + \frac{\sigma^2}{2} \nabla \log(\pi(x)), \sigma^2 \mathbb{I})}\right) \end{aligned}$$

3. With probability  $a(x, x')$  accept and set  $X^{(n+1)} = x'$ ; otherwise reject and set  $X^{(n+1)} = x$ .

# Langevin adjusted Hastings: notes

1. Goal: Direct the proposed values toward areas where density  $\pi(\cdot)$  is likely to be larger by using information from  $\pi(\cdot)$ .
2. Rational: the inclusion of  $\nabla \log(\pi(\cdot))$  in the proposal centre encourages moves towards the modes of  $\pi(\cdot)$

$$q(\cdot, d\cdot) : x' = x + \frac{\sigma^2}{2} \nabla \log(\pi(x)) + \sigma z ; z \sim N(0, 1)$$

3. In difficult settings, exact gradients  $\nabla \log(\pi(\cdot))$  can be replaced by numerical derivatives
4. MALA achieves optimal performance, if the proposal scale  $\sigma^2$  is s.t.  $\bar{\alpha}_{\text{opt}} \approx 0.57$

...if the components of  $x := (x_1, \dots, x_d)$  are independent.

CODE

CODE



# Tuning Metropolis-Hastings algorithms

The issue:

- ▶ How do we select a satisfactory proposal distr.  $q(\cdot, d\cdot)$  ?
- ▶ Well, ... this is not easy in general Metropolis-Hastings algorithm
- ▶ But, ... for some special cases, it is possible by adjusting the proposals

Recall that :

- ▶ About RWM, the  $\sigma^2$  is unknown
  - ▶ RWM can achieve satisfactory performance, if the proposal scale  $\sigma^2$  is s.t.  $a_{\text{opt}} \approx 0.234$
- ▶ About MALA, the  $\sigma^2$  is unknown
  - ▶ MALA can achieve satisfactory performance, if the proposal scale  $\sigma^2$  is s.t.  $a_{\text{opt}} \approx 0.57$

# An adaptive scheme for RWM, MALA

Goal: Adjust the proposal scaling  $\sigma^2$  in RWM or MALA algorithms

For  $n = 0, 1, 2, \dots$ , iterate:

1. Simulate  $X^{(n+1)}$  from  $P_{\sigma_n^2}^{(\text{RWM})}(X^{(n)}, d\cdot)$
2. Adjust  $\sigma^2$  s.t.  $\log(\sigma_{n+1}^2) = \log(\sigma_n^2) + \gamma_{n+1}(a_{n+1}^{\text{RWM}} - \bar{a}_{\text{opt}})$

Acceptance prob. of RWM at  $n$ -th iteration:  $a_n^{\text{RWM}}$

Optimal acc. prob. value: 
$$a_{\text{opt}} = \begin{cases} 0.234 & , \text{ for RWM} \\ 0.57 & , \text{ for MALA} \end{cases}$$

Gain sequence:  $\gamma_n : \mathbb{N} \rightarrow \mathbb{R}_+$ , a decreasing function  $\gamma_n \searrow 0$

E.g.  $\gamma_n = C/n^\varsigma$ ,  $C > 0$ ,  $\varsigma \in (0.5, 1)$

# An adaptive scheme for RWM, MALA: notes 1

Rational: At state  $n$ ,

- ▶ if  $a_{n+1}^{\text{RWM}} < \bar{a}_{\text{opt}}$ ,
  - $\implies \log(\sigma_n^2) < \log(\sigma_{n-1}^2)$
  - $\implies \sigma_n^2$  decreases
  - $\implies$  RWM/MALA will perform smaller steps at stage  $n + 1$
- ▶ if  $a_{n+1}^{\text{RWM}} > \bar{a}_{\text{opt}}$ ,
  - $\implies \log(\sigma_n^2) > \log(\sigma_{n-1}^2)$
  - $\implies \sigma_n^2$  increases
  - $\implies$  RWM/MALA will perform larger steps at stage  $n + 1$

Christophe Andrieu and Johannes Thoms (2008). A tutorial on adaptive MCMC

# Adaptive RWM and MALA algorithms: notes 2

Gain sequence  $\gamma_n$

- ▶  $\gamma_n$  must present a smooth slow decay
- ▶ As  $n \uparrow$ ,  $\gamma_n \downarrow$ , and the influence of adaptation vanishes
- ▶ A reasonable choice is

$$\gamma_n = C/n^\varsigma, \quad C > 0, \quad \varsigma \in (0.5, 1)$$

- ▶  $\varsigma$  controls the speed that  $\gamma_n$  decays to 0

Christophe Andrieu and Johannes Thoms (2008). A tutorial on adaptive MCMC

CODE

CODE

# Blockwise MCMC samplers

Consider r.v.  $X := (X_1, \dots, X_d)$  that follows  $X \sim \pi(dX)$ .

**Challenge:** In many cases, it is difficult to select appropriate proposals to construct an efficient Metropolis-Hastings algorithm that ‘updates’ simultaneously the whole  $X := (X_1, \dots, X_d)$

**Reasons:**  $X$  can be high dimensional.

Different  $X_i$  may have different ranges, types, etc.

etc...

**Cure:** ‘Break’ sampling of  $X$  by combining M-H algorithms targeting the conditional distributions of  $\pi(d\cdot)$

# Blockwise MCMC sampler

Consider r.v.  $X := (X_1, \dots, X_d)$  that follows  $X \sim \pi(dX)$ .

How to generate  $X^{(n)} \sim P^{(\text{blockwise})}(X^{(n-1)}, d\cdot)$  targeting  $\pi(dX)$  ??

At iteration  $n$ :

- ▶ Simulate  $X_1^{(n)} \sim \pi(dX_1^{(n)} | X_2^{(n-1)}, \dots, X_d^{(n-1)})$  via M-H alg  
⋮
- ▶ Simulate  $X_i^{(n)} \sim \pi(dX_i^{(n)} | X_1^{(n)}, \dots, X_{i-1}^{(n)}, X_{i+1}^{(n-1)}, \dots, X_d^{(n-1)})$  via M-H alg  
⋮
- ▶ Simulate  $X_d^{(n)} \sim \pi(dX_d^{(n)} | X_1^{(n)}, X_3^{(n)}, \dots, X_{d-1}^{(n)})$  via M-H alg

Set  $X^{(n)} = (X_1^{(n)}, X_3^{(n)}, \dots, X_d^{(n)})$

For example, for the  $i$ -th block

Simulate  $X_i^{(n)} \sim \pi(dX_i^{(n)} | X_1^{(n)}, X_{i-1}^{(n)}, \dots, X_{i+1}^{(n-1)}, X_d^{(n-1)})$  via RWM

1. Generate  $x' \sim N(x_i^{(n-1)}, \sigma^2 \mathbb{I})$

2. Calculate

$$a(x, x') = \min\left(1, \frac{\pi(x' | x_1^{(n)}, \dots, x_{i-1}^{(n)}, x_{i+1}^{(n-1)}, \dots, x_d^{(n-1)})}{\pi(x_i^{(n-1)} | x_1^{(n)}, \dots, x_{i-1}^{(n)}, x_{i+1}^{(n-1)}, \dots, x_d^{(n-1)})}\right)$$

3. With probability  $a(x_i^{(n-1)}, x')$  accept and set  $X_i^{(n)} = x'$ ; otherwise reject and set  $X_i^{(n)} = x_i^{(n-1)}$ .



# Gibbs sampler (special case of Blockwise MCMC sampler)

Consider r.v.  $X := (X_1, \dots, X_d)$  that follows  $X \sim \pi(dX)$ .

If all the full conditional dist of  $\pi(dX)$  can be sampled directly

How to  $X^{(n)} \sim P^{(\text{Gibbs})}(X^{(n-1)}, d\cdot)$  targeting  $\pi(dX)$  ??

At iteration  $n$ :

- ▶ Simulate  $X_1^{(n)} \sim \pi(dX_1^{(n)} | X_2^{(n-1)}, \dots, X_d^{(n-1)})$   
⋮
- ▶ Simulate  $X_i^{(n)} \sim \pi(dX_i^{(n)} | X_1^{(n)}, \dots, X_{i-1}^{(n)}, X_{i+1}^{(n-1)}, \dots, X_d^{(n-1)})$   
⋮
- ▶ Simulate  $X_d^{(n)} \sim \pi(dX_d^{(n)} | X_1^{(n)}, X_3^{(n)}, \dots, X_{d-1}^{(n)})$

Set  $X^{(n)} = (X_1^{(n)}, X_3^{(n)}, \dots, X_d^{(n)})$

CODE

CODE

# Blockwise MCMC sampler: notes

The blockwise MCMC sampler admits  $\pi(dX)$  as stationary distribution

Systematic sweep: (described above)

- ▶ The blocks are updated in a fix order: 1-st  $\longrightarrow$   $d$ -th
- ▶ The Markov chain is NOT reversible (why ??)

Permutation sweep:

- ▶ The blocks are updated in a random permutation order
- ▶ The Markov chain is reversible (why ??)

Random sweep:

- ▶ At each iteration, randomly select and update ONLY one block
- ▶ The Markov chain is reversible (why ??)

# Blockwise MCMC sampler (Random sweep)

Consider r.v.  $X := (X_1, \dots, X_d)$  that follows  $X \sim \pi(dX)$ .

Generate  $X^{(n)} \sim P^{(\text{blockwise})}(X^{(n-1)}, d\cdot)$  targeting  $\pi(dX)$

At iteration  $n$ :

Select randomly block  $i \sim U\{1, \dots, d\}$

► Simulate

$X_i^{(n)} \sim \pi(dX_i^{(n)} | X_1^{(n-1)}, \dots, X_{i-1}^{(n-1)}, X_{i+1}^{(n-1)}, \dots, X_d^{(n-1)})$  via  
M-H alg

Set  $X^{(n)} = (X_1^{(n-1)}, \dots, X_{i-1}^{(n-1)}, X_i^{(n)}, X_{i+1}^{(n-1)}, \dots, X_d^{(n-1)})$