Johns Hopkins Engineering

625.464 Computational Statistics

MCMC and the Metropolis Hastings Algorithm

Module 5 Lecture 5E

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Markov Chain Monte Carlo

$$\hat{\mu}_{mc} = \frac{1}{n}\sum_{t=1}^{n} h(x^{(t)}) \qquad E_f[h(x)]$$

where $x^{(t)} \sim f$

Basic Idea: Create an ergodic MC whose stationary dist is f.

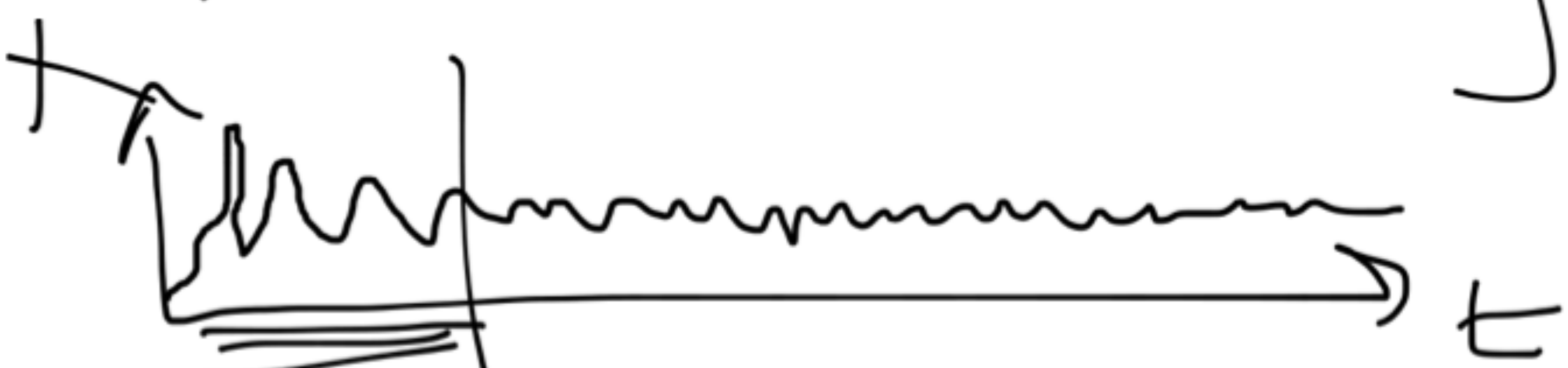
$$\frac{1}{n}\sum_{t=1}^{n} h(x^{(t)}) \longrightarrow E_f[h(x)]$$

Comments on MCMC

① Supports Bayesian inference.

② need t large enough

③ often the initial behavior of the MC is not limiting → burn in period



④ $x^{(0)}, x^{(1)}, \ldots$ dependent

Metropolis Hastings Algorithm

$f(x)$ $\quad g(y|\cdot)$.

M-H Algorithm to generate $\{X^{(t)}\}$

0. For $t=0$ draw a random x_0, with $f(x_0)>0$ & set $X^{(0)}=x_0$

Given $X^{(t)}=x^{(t)}$ compute $X^{(t+1)}$ as follows:

1. Generate a value x^* from the proposal dist $g(\cdot|x^{(t)})$
2. Set r equal to the M-H ratio

$$r = R(x^{(t)}, x^*) = \frac{f(x^*)\, g(x^{(t)}|x^*)}{f(x^{(t)})\, g(x^*|x^{(t)})}$$

Metropolis Hastings Algorithm continued

3. If $r \geq 1$, set $x^{(t+1)} = x^*$ accept

   o.w.
   Generate u from $\mathcal{U}(0,1)$
   if $u < r$, set $x^{(t+1)} = x^*$ accept
   o.w. set $x^{(t+1)} = x^{(t)}$ rejected.

4. increment t by 1 and return to Step 1

Comments on the M-H Algorithm

(0) x^* will come from the support of f since o.w. $f(x^*)=0$

(1) In step (3) we assign $x^{(t+1)}$ as follows

$$x^{(t+1)} = \begin{cases} x^* & \text{w/prob } \min\{R, 1\} \\ x^{(t)} & \text{o.w.} \end{cases}$$

(2) Since f $\underline{\qquad\qquad} = R$

(3) If $g(x^{(t)}|x^*) = g(x^*|x^{(t)})$ we have the Metropolis Alg $\dfrac{f(x^*)}{f(x^{(t)})}$

Comments on M-H Algorithm

(4) Clearly $\{X^{(t)}\}$ is a MC since $X^{(t+1)}$ only depends on $X^{(t)}$.

(5) Is the chain ergodic?
depends on choice of g

$\{X^{(t)}\}$

Proof that the MC's limiting distribution f

The unique stationary dist for the MC is f

pf Suppose $X^{(t)} \sim f(x)$ and consider $x_1 \neq x_2 \in S$ for which $f(x_1), f(x_2) > 0$.

w.l.o.g assume $f(x_2) g(x_1 | x_2) \geq f(x_1) g(x_2 | x_1)$

2 possibilities

1. $X^{(t)} = x_1$ and $X^{(t+1)} = x_2$
2. $X^{(t+1)} = x_1$ and $X^{(t)} = x_2$

Proof continued

Case 1 $X^{(t)} = x_1,\ X^{(t+1)} = x_2$

$$\boxed{f(x_1)\, g(x_2 \mid x_1) \cdot 1} \qquad R \geq 1$$

Case 2 $X^{(t)} = x_2$ & $X^{(t+1)} = x_1$ $\qquad R \leq 1$

$$f(x_2)\, g(x_1 \mid x_2) \cdot R$$

$$f(x_2)\, g(x_1 \mid x_2) \cdot \frac{f(x_1)\, g(x_2 \mid x_1)}{f(x_2)\, g(x_1 \mid x_2)} = f(x_1)\, g(x_2 \mid x_1)$$

joint density of $X^{(t)}$ & $X^{(t+1)}$ is symmetric $X^{(t+1)} \sim f$.