Lecture 18:

Metropolis to Metropolis-Hastings

Note Title 5/2/2011

In previous lecture we applied Metropolis to the 1-D Ising model. We took samples $\underline{X}^{50}, \underline{X}^{100}, \ldots, \underline{X}^{50k}, \ldots$ from the Markov chain at $t = 50, 100, \ldots, 50k, \ldots$

Why did we pick 50, 100, ... ?

Two related concepts: Burn-In and Stickiness

Burn-In is the amount of time required for the MCMC to generate samples from $\pi(\underline{x})$

Above we assume burn-in time = 50.

Note: sample $\underline{X}_{51}$ at $t = 51$ will also be a sample from $\pi(\underline{x})$ but it will be correlated to $\underline{X}_{50}$ and so will not be an independent sample from $\pi(\underline{x})$. But, if burn-in = 50, then sample $\underline{X}_{100}$ at $t = 100$ will also be an independent sample from $\pi(\underline{x})$, same for $\underline{X}_{150}, \underline{X}_{200}, \ldots$

(recall, we count i.i.d. samples from $\pi(\underline{x})$)

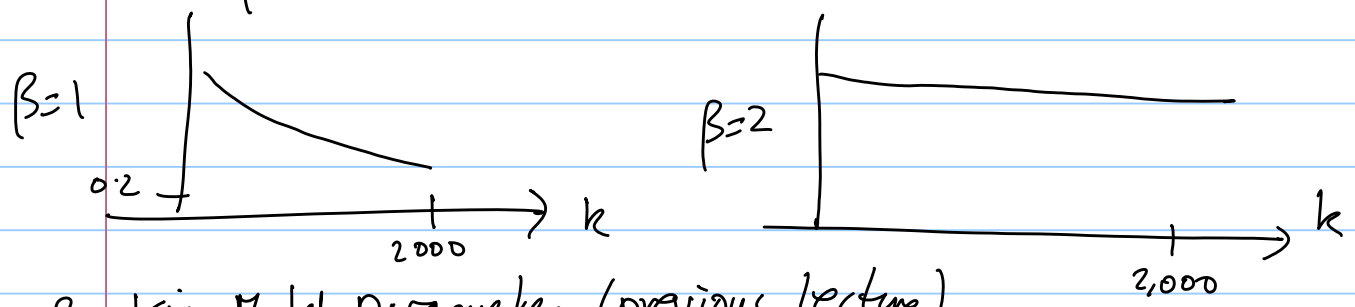
Stickiness is a measure of how slowly the MCMC explores the space of $\underline{x}$. This will affect the burn-in period and how long we need wait to get independent samples.

Page 2.

Note Title 5/10/2006

Autocorrelation is a practical way to measure stickiness.

ρ_k = Correlation $(M^{(1)}, M^{(k+1)})$



β - Ising Model parameter (previous lecture)

So the MCMC is very sticky for $\beta = 2$ because the autocorrelation decreases very slowly with time.

Hence sampling at $t = 50, 100, 150, \ldots$ may be okay for $\pi(\underline{x})$ with $\beta = 1$, but need more burn-in for $\beta = 2$. E.G. $t = 1000, 2000, 3000, 4000, \ldots$

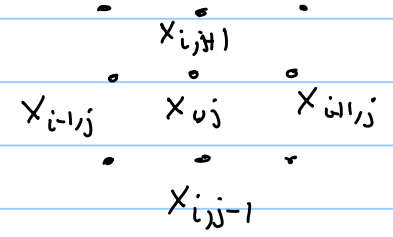
Page 3.

2-D Ising Model.

$X_{ij} \in \{\pm 1\}$

$$\pi(\underline{x}) = \frac{1}{Z} e^{-E[\underline{x}]} \qquad \underline{X} = \{(X_{ij}) : ij \in \Lambda\}$$

$$E[\underline{x}] = \sum_{ij,kl} J_{ij,kl} X_{ij} X_{kl} + \sum_{ij} \theta_{ij} X_{ij}$$



Neighborhood:

$$\underline{x}^t = \{(X^t_{ij}) : ij \in \Lambda\}$$

has neighborhood $\{\underline{x}^t_{/k,l} : k,l \in \Lambda\}$,

where $\underline{x}^t_{/k,l}$ differs from $\underline{x}^t$ by changing $X^t_{k,l}$ to $-X^t_{k,l}$.

Step 1: select k,l at random (unif dist).

Step 2: evaluate $E[\underline{x}^t_{/k,l}] - E[\underline{x}^t]$

$$= -2\theta_{kl} X_{kl} - 2X_{kl} \sum_{ij} J_{ij,kl} X_{ij}.$$

Step: set $\underline{x}^{t+1} = \underline{x}^t_{/k,l}$ with prob $\min\{1, e^{-E[\underline{x}^t_{/k,l}] + E[\underline{x}^t]}\}$,

otherwise set $\underline{x}^{t+1} = \underline{x}^t$.

Note: this 2D Ising model is represented by a graph with many closed loops — so we cannot do sequential sampling.

Metropolis-Hastings.

This improves the basic Metropolis algorithm by adding a proposal probability.

$T(\underline{y}|\underline{x})$ for $\underline{y} \in N(\underline{x})$

Note: if $T(\underline{y}|\underline{x})$ is uniform then we get Metropolis.

At time t: state $\underline{x}^t$

(i) propose $\underline{y} \in N(\underline{x})$ with probability $T(\underline{y}|\underline{x}^t)$.

(ii) accept this proposal with probability $\min\left\{1, \frac{\pi(\underline{y})\, T(\underline{x}^t|\underline{y})}{\pi(\underline{x}^t)\, T(\underline{y}|\underline{x}^t)}\right\}$ ← modify Metropolis acceptance prob to prevent bias from proposals $T(\underline{y}|\underline{x})$

If accepted, set $\underline{x}^{t+1} = \underline{y}$

otherwise, set $\underline{x}^{t+1} = \underline{x}^t$.

Metropolis Hastings has kernel:

$$K(\underline{y}|\underline{x}) = T(\underline{y}|\underline{x}) \min\left\{1, \frac{\pi(\underline{y})T(\underline{x}|\underline{y})}{\pi(\underline{x})T(\underline{y}|\underline{x})}\right\}, \text{ for } \underline{y} \neq \underline{x}$$

$$K(\underline{x}|\underline{x}) = 1 - \sum_{\underline{y} \neq \underline{x}} K(\underline{y}|\underline{x}).$$

Check Detailed Balance:

$$K(\underline{y}|\underline{x})\pi(\underline{x}) = \min\{\pi(\underline{x})\pi(\underline{y}|\underline{x}), \pi(\underline{y})\pi(\underline{x}|\underline{y})\} = K(\underline{x}|\underline{y})\pi(\underline{y})$$

(symmetric in $\underline{x}$ & $\underline{y}$)

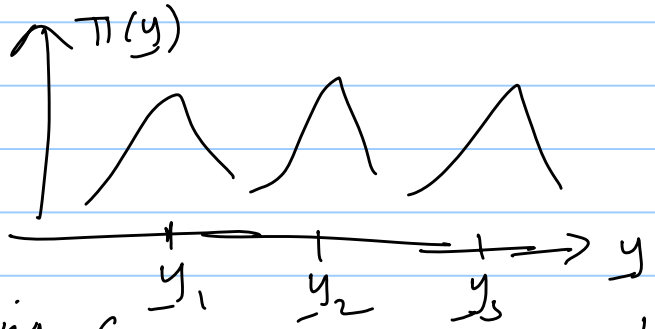
so detailed balance holds.

Metropolis-Hastings is one of the most popular MCMCs.

The speed of Metropolis-Hastings — e.g., the burn in period — depends greatly on the choice of proposal probability $T(\underline{y}|\underline{x})$.

Most practical implementations of Metropolis-Hasting use well-designed proposal probabilities. But selecting proposal probabilities is problem specific. It depends on the structure of $\Pi(\underline{x})$.

Example:



Ideally, we would have proposals that allow us to sample from $\Pi(\underline{y})$ by jumping from y_1 to y_2, y_2 to y_3, etc

More practically, suppose we have a 1-D Ising model $\Pi(\underline{x}) = \frac{1}{Z} e^{-E[\underline{x}]}$, with $E(\underline{x}) = \beta \sum_{i=0}^{d-1} x_i x_{i+1} + \sum_{i=1}^{d} c_i x_i$.

Then define neighborhood (as before):

$N(\underline{x}) = \{\underline{x}_{/k} : k = 0, \ldots d\}$, where $\underline{x}_{/k} = (x_0, \ldots, -x_k, \ldots x_d)$

and proposal $T(\underline{x}_{/k}|\underline{x}) = \frac{e^{|c_k|}}{\sum_i e^{|c_i|}}$ i.e. change variables where c_k is big.

Designing a good MCMC is an art not a science.