11. Hastings-Metropolis Algorithm – Lecture 11 – October 8, 2002

11.1 The Hastings-Metropolis Algorithm

1. Begin with an irreducible Markov matrix Q_{ij}, with $i, j = 1, 2, 3, ...n$ (which need not be symmetric).

what is m ?

2. Let $n = 0$ and $X_0 = k$, for some $1 \leq k \leq m$.

3. Generate a new random variable X such that $Prob\{X = j\} = Q_{X_n,j}$.

4. Generate a random number U uniformly distributed on $(0, 1)$. If $U < \frac{[b(X)Q_{X,X_n}]}{[b(X_n)Q_{X_n,X}]}$ then $NS = X$; otherwise $NS = X_n$.

5. Let $n = n + 1$; set $X_n = NS$. what is NS ?

6. Go to step 3.

11.2 Application – Example 10a in Ross

We begin with a large set L of all permutations of $\{1, 2, ..., n\}$ for which $\sum_{j=1}^{n} jx_j > a$ for a given constant a. Another example is the set of all tree subgraphs of a given graph H. We want a limiting probability distriubtion which is uniform.

– First define a Markov chain graph G whose vertices are the elements of L. We will need a notion of neighbors in L, and we join node i to j by an arc if j is accessible from i in one move.

– For $L = S_n$ we put the arcs of G between states or permutations which differ by a transposition.

– Let $N(s) = \{$ the neighbors of a node $s\}$, and let $|N(x)|$ equal the cardinality of $N(s)$. Let $Q_{s,t} = frac1|N(s)|$ if $t \in N(s)$.

– Since we are interested in sampling uniformly from L, we want $\Pi(s) = \Pi(t) = K = \frac{b_s}{\sum b_j}$. Therefore, by setting

$$min(1, \frac{b_t Q_{ts}}{b_s Q_{st}} = min(\frac{|N_s|}{|N_t|}, 1) \tag{11.1}$$

we get an ergodic Markov chain which is reversible:

$$P_{s,t} = \left\{ \begin{array}{c} Q_{s,t} min(1, \frac{b_t S_{ts}}{b_s Q_{st}}) \\ Q_{s,s} + \sum_{r \neq s} Q_{s,r}(1 - min(1, \frac{b_r Q_{rs}}{b_s Q_{sr}})) \end{array} \right\} \tag{11.2}$$

by a theorem in the last lecture.

– The justification for using $min(1, \frac{b_t Q_{ts}}{b_s Q_{st}} = min(1, \frac{|N_s|}{|N_t|}$ as opposed to the earlier formulation $min(1, \frac{b_s}{b_t}$ is based on relaxing the condition that $Q_{st} = Q_{ts}$.

We check that $\sum_{t=1}^{m} P_{st} = 1$; and $P_{ss} > 0$ for some s, which is a generic property that leads to aperiodicity.

– Suppose the current state is $X_n = S$. Choose one of its neighbors randomly (make a random transposition).

If $|N_t| \leq |N_s|$, then accept $X_{n+1} = t$.

If $|N_t| > |N_s|$, then set $X_{n+1} = t$ with probability $q = \frac{|N_s|}{|N_t|}$. Set $X_{n+1} = s$ with probability $1 - q = 1 - \frac{|N_s|}{|N_t|}$.

11.3 Gibbs Sampler (Vector form of Hastings-Metropolis)

Let $\vec{X} = (X_1, x_2, ..., x_n)$ be a random vector with probability mass function $p(\vec{X}) = cg(\vec{x})$. We want to sample from such a distribution of random vectors, where $g(\vec{X})$ is known but c is not.

Consider a Markov chain where the states are $\vec{x} = (x_1, x_2, ..., x_n)$. Let $\vec{x}$ be the curent vector state. Choose $i = 1, 2, ..., n$ randomly and set the random variable $X = x$ with:

$$Prob\{X = x\} = Prob\{X_i = x | X_j = x_j, j \neq i\} \tag{11.3}$$

(which is given a priori by $p(\vec{x})$), since

$$\begin{aligned} Prob\{X_i = x | X_j = x_j, j \neq i\} &= \frac{Prob\{X_i = x, X_j = x_j, j \neq i\}}{Prob\{X_j = x_j, j \neq i\}} &(11.4) \\ &= \frac{Prob\{X_i = x, X_j = x_j, j \neq i\}}{\sum_k Prob\{X_i = k, X_j = x_j, j \neq i\}} &(11.5) \\ &= \frac{p(x_1, x_2, ..., x_{i-1}, x_i, x_{i+1}, ..., x_n)}{\sum_k p(x_1, x_2, ..., x_{i-1}, k, x_{i+1}, ..., x_n)} &(11.6) \end{aligned}$$

Next, if the random variable $X = x$, then $\vec{y} = (x_1, x_2, ..., x_{i-1}, x, x_{i+1}, ..., x_n)$ is a possible next state vector.

This is equivalent to Hastings-Metropolis with:

$$\begin{aligned} Q(\vec{x},\vec{y}) &= \frac{1}{n} Prob\{X_i = x | X_j = x_j, j \neq i\} && (11.7) \\ &= \frac{p(\vec{y})}{n Prob\{X_j = x_j, j \neq i\}} && (11.8) \\ &= \frac{p(\vec{y})}{n \sum_k p(x_1, x_2, ..., x_{i-1}, k, x_{i+1}, ..., x_n)} && (11.9) \end{aligned}$$

and

$$\begin{aligned} P_{\vec{x},\vec{y}} &= \left\{ \begin{array}{c} Q(\vec{x},\vec{y}) min(1, \frac{p(\vec{y})Q(\vec{y},\vec{x})}{p(\vec{x})Q(\vec{x},\vec{y})}) \\ Q(\vec{x},\vec{x}) + \sum_{\vec{z} \neq \vec{x}} Q(\vec{x},\vec{z})(1 - \frac{p(\vec{z})Q(\vec{z},\vec{x})}{p(\vec{x})Q(\vec{x},\vec{z})}) \end{array} \right\} && (11.10) \\ &= \left\{ \begin{array}{cc} Q(\vec{x},\vec{y}) & \vec{y} \neq \vec{x} \\ Q(\vec{x},\vec{x}) & \vec{y} = \vec{x} \end{array} \right\} && (11.11) \end{aligned}$$

11.4 Application (Example 10b)

the problem is to generate n random points on the unit circle, such that no two points are within distance d of each other, where

$$\beta = Prob\{\text{no two points are within distance } d \text{ of each other}\} \quad (11.12)$$

is assumed to be small.

Do this by applying the Gibbs sampler, starting with n points on the unit sphere $x_1, x_2, ..., x_n$ such that no two are within distance d of each other.

Generate a random number U and let $I = int(nU)+1$. this step picks randomly from $i = 1, 2, ..., n$.

Next, generate a random point on the circle, $X = x$ and if $|x - x_j| > d, j \neq I$, then set $\vec{y} = (x_1, x_2, ..., x_{I-1}, x, x_{I+1}, ..., x_n)$. Otherwise generate another point $X = x'$ and repeat.

11.5 The Metropolis Algorithm for Statistical Mechanics

Construct Aafinal probability distribution

$$\Pi_{\vec{x}} = P_{\vec{x}} = \frac{e^{-\beta E(\vec{x})}}{Z_N} \equiv \frac{b(\vec{x})}{\sum_{\vec{r}} b(\vec{r})} \tag{11.13}$$

where N denotes the number of particles or the number of lattice sites. To be precise let us suppose that the domain is the unit square (with periodic boundary conditions), and the N particles can take any position on a uniform (MxM) grid – let us say $(100x100)$ for this demonstration.

a. Compute the number of states in the problem. For $N = 100$ this number is $(10^4)^M = (10^4)^{100}$ which is huge!

b. So Q is an $M^{2N}xM^{2N}$ matrix $Q(\vec{x}, \vec{y}) = \frac{1}{N} Prob\{X_i = x | X_j = x_j, j \neq i\}$.

c. The important point is that in the Metropolis algorithm,

$$P_{\vec{x},\vec{y}} = \left\{ \begin{array}{c} Q(\vec{x},\vec{y}) min(1, \frac{b_{\vec{y}} Q(\vec{y},\vec{x})}{b_{\vec{x}} Q(\vec{x},\vec{y})}) \\ Q(\vec{x},\vec{x}) + \sum_{\vec{z} \neq \vec{x}, b_j < b_i} Q(\vec{x},\vec{z})(1 - \frac{b_{\vec{y}} Q(\vec{y},\vec{x})}{b_{\vec{x}} Q(\vec{x},\vec{y})}) \end{array} \right\} \tag{11.14}$$

d. Finally, we need to analyze what

$$Prob\{X_i = X | X_j = x_j, j \neq i\}; i, j = 1, 2, ...N \tag{11.15}$$

$x_j \in L$, the MxM grid.

This can be treated as a graph G where the nodes are all possible states $\vec{x}$ of which there are M^{2N}. Arcs connect "neighbors" in G of $\vec{x}$ which are defined by states $\vec{y}$ that can be reached from $\vec{x}$ by a Gibbs sampler move, i.e, randomly choose $i = 1, 2, ..., N$ and then change $X_i = x$, leaving $X_j = x_j$ for $j \neq i$ with

$$\begin{aligned} Prob X_i = x | X_j = x_j &= \frac{Prob\{X_i = x, X_j = x_j\}}{Prob\{X_j = x_j, j \neq i\}} && (11.16) \\ &= \frac{e^{-\beta E(\vec{y})}}{\sum_k e^{-\beta E(X_i = k)}} && (11.17) \end{aligned}$$

e. Putting it together we get

$$Q(\vec{x},\vec{y}) = \frac{e^{-\beta E(\vec{y})}}{N\sum_k e^{-\beta E(X_i=k)}} \tag{11.18}$$

and

$$P_{\vec{x};\vec{\cdot}} = \left\{ \begin{array}{ll} Q(\vec{x},\vec{y}) & \vec{y} \neq \vec{x} \\ Q(\vec{x},\vec{x}) & \vec{y} = \vec{x} \end{array} \right\} \tag{11.19}$$