Probability and Random Process

COSE382

Wonzoo Chung

Metropolis-Hastings

Algorithm 12.1.1 (Metropolis-Hastings). Let $X_0, X_1, X_2, \cdots$ be a Markov chain on state space $\{1, \cdots, M\}$ with transition matrix P.

Let $\mathbf{s} = (s_1, \cdots, s_M)$ be a desired stationary distribution on state space ($s_i > 0$ for all i)

Our goal is to modify P to construct a Markov chain $X_0, X_1, \cdots$ so that new transition matrix Q has stationary distribution $\mathbf{s}$. Starting from X_0;

1. For $X_n = i$, propose a new transition probability q_{ij} from p_{ij}.
2. Compute the *acceptance probability*

$$a_{ij} = \min\left(\frac{s_j p_{ji}}{s_i p_{ij}}, 1\right)$$

3. Let

$$q_{ij} = a_{ij} p_{ij}$$

   This can be implemented by

   - Flip a coin that lands Heads with probability a_{ij}.
   - If the coin lands Heads, accept the transition with respect to p_{ij}, setting $X_{n+1} = j$. Otherwise, reject the proposal (i.e., stay at i), setting $X_{n+1} = i$.

Proof of the Metropolis-Hastings Algorithm:

To check is the reversibility $s_i q_{ij} = s_j q_{ji}$ for all i and j: from $q_{ij} = \min\left(\frac{s_j p_{ji}}{s_i p_{ij}}, 1\right) p_{ij}$,

If $q_{ij} > 0 \Longrightarrow p_{ij} > 0$ and $p_{ji} > 0$ (otherwise, the a_{ij} would be 0).

If $p_{ij} > 0$ and $p_{ji} > 0, \Longrightarrow q_{ji} > 0$.

So q_{ij} and q_{ji} are either both zero or both nonzero. Assume they are both nonzero:

$$q_{ij} = \min\left(\frac{s_j p_{ji}}{s_i p_{ij}}, 1\right) p_{ij}$$

If $s_j p_{ji} \le s_i p_{ij}$, we have $a_{ij} = \frac{s_j p_{ji}}{s_i p_{ij}}$ and $a_{ji} = 1$, so

$$s_i q_{ij} = s_i a_{ij} p_{ij} = s_i \frac{s_j p_{ji}}{s_i p_{ij}} p_{ij} = s_j p_{ji} = s_j p_{ji} a_{ji} = s_j q_{ji}.$$

Symmetrically, if $s_j p_{ji} > s_i p_{ij}$, we again have $a_{ji} = \frac{s_i p_{ij}}{s_j p_{ji}}$ and $a_{ij} = 1$,

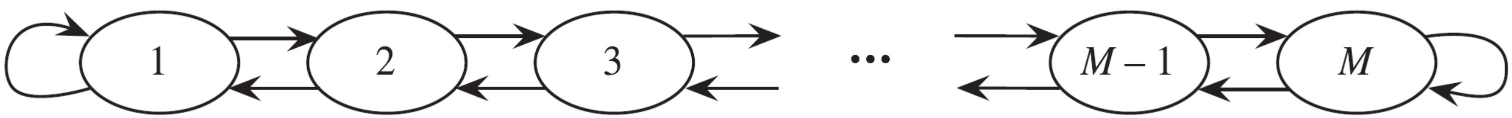
$$s_j q_{ji} = s_j a_{ji} p_{ji} = s_j \frac{s_i p_{ij}}{s_j p_{ji}} p_{ji} = s_i p_{ij} = s_i p_{ij} a_{ij} = s_i q_{ij}.$$

Example 12.1.3 (Zipf distribution simulation). An r.v. X has the *Zipf* distribution with parameter $a > 0$ if its PMF is

$$P(X = k) = \frac{1}{\sum_{j=1}^{M} j^{-a}} k^{-a}, \text{ for } k = 1, 2, \ldots, M(M \geq 2)$$

Create a Markov chain $X_0, X_1, \ldots$ whose stationary distribution is the Zipf(a) with $|X_{n+1} - X_n| \leq 1$ for all n, i.e. $s_i = \frac{1}{\sum_{j=1}^{M} j^{-a}} i^{-a}$.

Solution: Let us start with a simple Markov chain with $|X_{n+1} - X_n| \leq 1$ for all n as the following (all transition probabilities are 1/2)



Let P be the transition matrix of this chain (The stationary distribution for P is uniform).

Let X_0 be any starting state, and generate a chain $X_0, X_1, \ldots$ as follows.

1. Generate a proposal state j according to the proposal chain P.
2. Accept the proposal with probability

$$\min\left(\frac{s_j \cdot 1/2}{s_i \cdot 1/2}, 1\right) = \min(i^a/j^a, 1)$$

   If the proposal is accepted, go to j; otherwise, stay at i.

Note that the normalizing constant $\sum_{j=1}^{M} j^{-a}$ is *not* needed to run the chain.

Example 12.1.4 (Beta simulation). We want to generate $W \sim \text{Beta}(a, b)$ with i.i.d. $U_n \sim \text{Unif}(0, 1)$ r.v.s.

Solution: Note that U_n can be viewed as a Markov Chain with the transition probability

$$f_{U_{n+1}|U_n}(U_{n+1} = u | U_n = v) = 1$$

Let W_0 be any starting state, and generate a chain $W_0, W_1, \ldots$ as follows. If the chain is currently at state w (a real number in (0, 1)), then:

1. Generate a proposal u by drawing a Unif(0, 1) r.v.
2. Accept the proposal with probability $\min\left(\frac{w^{a-1}(1-u)^{b-1}}{w^{a-1}(1-w)^{b-1}}, 1\right)$ If the proposal is accepted, go to u; otherwise, stay at w.

The normalizing constant $\beta(a, b)$ was not needed in order to run the chain.
Running the Markov chain, we have that $W_n, W_{n+1}, W_{n+2}, \ldots$ are approximately $\text{Beta}(a, b)$ for n large. Note that $\{W_n\}$ are correlated r.v.s., not i.i.d.

Example 12.1.7 (Knapsack problem). B finds M treasures and wishes to maximize the total worth of the treasure he takes since the maximum weight he can carry is W;

$$\text{maximize } V(x) = \sum_{j=1}^{M} x_j g_j, \text{ subject to } \sum_{j=1}^{M} x_j w_j \leq W$$

w_j and g_j denote the weight and the value of the jth treasure, respectively, and x_j is 1 if he takes the jth treasure and 0 otherwise. Let's define $C := \left\{x = (x_1, \cdots, x_M) \in \{0,1\}^M \mid \sum x_j w_j < W\right\}$.

(a) Consider the following Markov chain over C. Starting at $(0, 0, \ldots, 0)$, suppose the current state is $x = (x_1, \ldots, x_M)$. Choose a uniformly random $j \in \{1, 2, \ldots, M\}$, and obtain y from x by replacing x_j with $1 - x_j$. If $y \notin C$ stay at x, else move to y. Show that the uniform distribution over C is stationary for this chain.

(b) Show that the chain from (a) is irreducible, and that it may or may not be aperiodic (depending on $W, w_1, \ldots, w_M$).

(c) B is interested in finding solutions where the value $V(x)$ is high. Specifically, suppose that we want to simulate from the distribution

$$s(x) \propto e^{\beta V(x)},$$

Create a Markov chain whose stationary distribution is as desired.

Solution:

(a) symmetric $q_{xy} = q_{yx}$: for $x \neq y$, either $q_{xy} = q_{yx} = 1/M$ or $q_{xy} = q_{yx} = 0$. So the stationary distribution is uniform over C.

(b) irreducible: For $x, y \in C$, we can $x \to (0, 0, \ldots, 0)$ and $(0, 0, \ldots, 0) \to y$ one at a time. periodicity:

- $(0, \cdots, 0)$ has period 2, if $w_1 + \cdots + w_M < W$.
- Every state has period 1, if $w_i > W$ for an $i \in \{1, \cdots, M\}$

(c) We apply Metropolis-Hastings using the chain from (a) to make proposals. Start at $(0, 0, \ldots, 0)$. Suppose the current state is $x = (x_1, \ldots, x_M)$. Then:

1. Choose a uniformly random j in $\{1, 2, \ldots, M\}$, and obtain y from x by replacing x_j with $1 - x_j$.
2. If y is not in C, stay at x. If y is in C, flip a coin that lands Heads with probability

$$\min\left(\frac{s_y \cdot 1/M}{s_x \cdot 1/M}, 1\right) = \min(e^{\beta(V(y)-V(x))}, 1)$$

If the coin lands Heads, go to y; otherwise, stay at x.

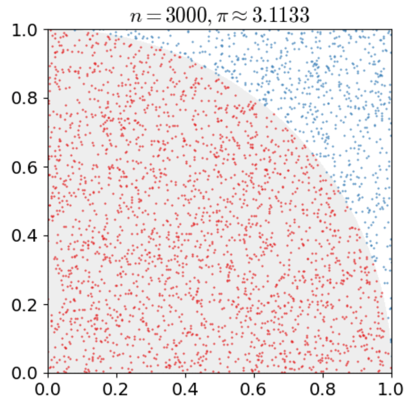
Monte Carlo Simulation

- Named after Monte Carlo Casino in 1940s by a group of men working on the nuclear bomb John von Neumann, Stanislaw Ulam and Nicholas Metropolis: "The Monte Carlo Method," *Journal of the American Statistical Association*, 44 (247), 335-341, 1949.

- Using a large number of simulated trials in order to approximate a solution to a problem;
  - **Problem 1:** to generate samples $\{x_n\}$ from a given probability distribution $P(x)$.
  - **Problem 2:** to estimate expectations of functions under this distribution, for example

$$\hat{\phi}(\mathbf{x}) = \int \phi(\mathbf{x}) P(\mathbf{x}) d\mathbf{x}$$

- The key is to generate good samples of $P(\mathbf{x})$, especially for the high dimensional $\mathbf{x}$.

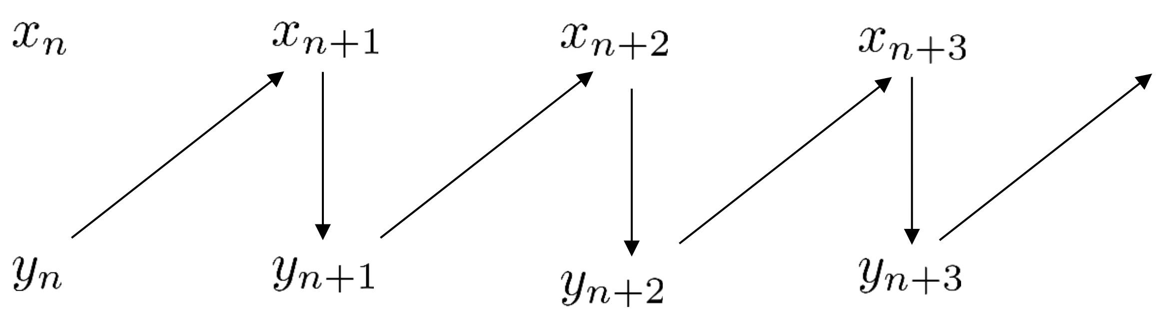


Gibbs Sampler

Algorithm 12.2.1 (Systematic scan Gibbs sampler). Let X and Y be discrete r.v.s with joint PMF $p_{X,Y}(x, y) = P(X = x, Y = y)$. We wish to construct a two-dimensional Markov chain (X_n, Y_n) whose stationary distribution is $p_{X,Y}$ with known conditional distributions $f_{X|Y}$ and $f_{Y|X}$:

1. Draw a value x_{n+1} from $f_{X|Y}(x|Y = y_n)$, and set $X_{n+1} = x_{n+1}$.
2. Draw a value y_{n+1} from $f_{Y|X}(y|X = x_{n+1})$, and set $Y_{n+1} = y_{n+1}$.

Repeating steps 1 and 2 over and over, the stationary distribution of the chain (X_0, Y_0), (X_1, Y_1) , $(X_2, Y_2), \ldots$ is $p_{X,Y}$.



Algorithm 12.2.2 (Random scan Gibbs sampler). Let X and Y be discrete r.v.s with joint PMF $p_{X,Y}(x, y)$. We wish to construct a two-dimensional Markov chain (X_n, Y_n) whose stationary distribution is $p_{X,Y}$ with known conditional distributions $f_{X|Y}$ and $f_{Y|X}$:

1. Choose which component to update, with equal probabilities.
2. If the X-component was chosen,
   - draw a value x_{n+1} from $f_{X|Y}(x|Y = y_n)$,
   - set $X_{n+1} = x_{n+1}, Y_{n+1} = y_n$.

   If the Y-component was chosen,
   - draw a value y_{n+1} from $f_{Y|X}(y|X = x_n)$,
   - set $X_{n+1} = x_n, Y_{n+1} = y_{n+1}$.

Repeating steps 1 and 2 over and over, the stationary distribution of the chain (X_0, Y_0), (X_1, Y_1),(X_2, Y_2),... is $p_{X,Y}$.

Theorem 12.2.3 (Random scan Gibbs as Metropolis-Hastings). The random scan Gibbs sampler is a special case of the Metropolis-Hastings algorithm.

Proof. Let X and Y be discrete r.v.s with joint PMF $p(x, y)$. Let's consider a Markov chain generated by Random scan Gibbs sampling with the transition probability

$$p_{(x,y),(x',y')} = \begin{cases} 0 & x \neq x', y \neq y' \\ \frac{1}{2}p(y'|x) & x = x', y \neq y' \\ \frac{1}{2}p(x'|y) & x \neq x', y = y' \end{cases}$$

The desired stationary distribution is the joint PMF of X and Y, $p(x, y)$. Let's compute the Metropolis-Hastings acceptance probability for going from (x, y) (state i) to (x', y') (state j), assuming that $x = x'$ (the case $y = y'$ can be handled symmetrically):

$$\frac{s_j p_{ji}}{s_i p_{ij}} = \frac{p(x, y')p(y|x)\frac{1}{2}}{p(x, y)p(y'|x)\frac{1}{2}} = \frac{p(x)p(y'|x)p(y|x)}{p(x)p(y|x)p(y'|x)} = 1.$$

Thus, this Metropolis-Hastings algorithm always accepts the proposal! So it's just running the random scan Gibbs sampler without modifying it.

Example 12.2.6 (Chicken-egg with unknown parameters). Every day a chicken hatches $X = x$ eggs. We model this as:

- A chicken lays N eggs, where $N \sim$Pois(λ).
- Each egg hatches with probability p, where p is unknown but $p \sim$ Beta(a, b).
- The constants λ, a, b are known.

We want to know the probability of hatching p, but...

- We cannot directly access N
- We only observe the number of eggs that hatch, X.

Find $E(p|X = x)$, the posterior mean of p for observed x hatched eggs

Solution.

The distribution of X given p is Pois(λp). The posterior PDF of p is

$$f(p|X=x) = \frac{P(X=x|p)f(p)}{P(X=x)} = \frac{e^{-\lambda p}(\lambda p)^x p^{a-1} q^{b-1}}{\int_0^1 e^{-\lambda p}(\lambda p)^x p^{a-1} q^{b-1} dp},$$

This PDF is not named, nor integrable.

Hence, we will generate samples $p_k \sim f(p|X=x)$ and compute the sample mean

$$\hat{p} = \frac{1}{M}\sum_{k=1}^{M} p_k$$

Although we cannot generate samples of

$$f(p|X=x) \propto P(X=x|p)f(p) \propto e^{-\lambda p}(\lambda p)^x p^{a-1}(1-p)^{b-1},$$

we have

$$\begin{aligned} f(p|X=x, N=n) &\sim \text{Beta}(x+a, n-x+b) \\ f(N|X=x, P=p) &\sim \text{Pois}(\lambda(1-p)) + x \end{aligned}$$

We make an initial guess for p and N, then iterate the following systematic Gibb's sampling:

1. Conditional on $N = n$ and $X = x$, draw a new guess for p from the Beta$(x+a, n-x+b)$ distribution.
2. Conditional on p and $X = x$, the number of unhatched eggs is $Y \sim$Pois$(\lambda(1-p))$ by the chicken-egg story, so we can draw Y from the Pois$(\lambda(1-p))$ distribution and set the new guess for N to be $N = x + Y$.

After many iterations, we have draws for both p and N.

With

- $\lambda = 10$,
- $a = b = 1$
- $X = 7$

we have $\hat{p} = 0.68$