Markov Chain Monte Carlo

Bo Li, Spring 2019

Markov chain Monte Carlo (MCMC) is a method for simulating a given distribution π by generating an irreducible Markov chain $\{X_n\}_{n=0}^{\infty}$. Often the distribution π is only given up to a normalizing constant, which is not known, and hard to evaluate. Under some weak conditions, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} 1_{\{X_k \in A\}} = \pi(A)$$

for any A, a subset of S, a state space. If g is a (~~continuous~~) function on S, then the expectation $\mathbb{E}(g(X))$ with X being a random variable that is π distributed, is

$$\mathbb{E}[g(X)] = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g(X_k).$$

In the continuous distribution case,

$$\mathbb{E}[g(X)] = \lim_{n\to\infty} \int_S g(x)\,\pi(x)\,dx$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g(X_k).$$

The Metropolis Algorithm

Given: S – a discrete state space;
$Q = (q_{ij})$ – a symmetric transition matrix of a (background) Markov chain;
π – a distribution on S in the form $\pi(i) = \frac{1}{\mathcal{Z}} b(i)$, $i \in S$, where all $b(i) > 0$, and $\mathcal{Z} > 0$ is a normalizing constant, and is unknown.

We generate a Markov chain $\{X_n\}_{n=0}^{\infty}$ with π the invariant distribution. The new Markov chain will be termed a Metropolis chain.

The Metropolis algorithm

Given $X_k = i \in S$.

(1) Select $Y \in S$ randomly according to Q, i.e.,
$$\mathbb{P}(Y = j \mid X_k = i) = q_{ij} \quad \forall j \in S.$$
Set $Y = j$.

(2) Let $\alpha_{ij} = \min\left(1, \frac{\pi(j)}{\pi(i)}\right) \in [0,1]$.
Accept Y with the probability α_{ij}:
Generate $U \sim U[0,1]$.
If $U \leq \alpha$, then accept Y: $X_{k+1} \leftarrow Y$;
If $U > \alpha$, then reject Y: $X_{k+1} \leftarrow X_k$.

Remarks

- We call Y a proposal state.
  We call α_{ij} the acceptance probability.

- For the distribution π, we need only to know the ratios $\frac{\pi(i)}{\pi(j)}$ $\forall i,j \in S$, not π itself. This means we do not need to know the normalizing constant — the "partition function".

- We often choose Q to be as simple as possible, e.g., uniform probabilities, i.e., given $X_k = i \in S$. The probability that $X_{k+1} = j \in S$ is uniform on S.

- In many situations, the calculation of the acceptance probability α_{ij} is rather simple, due to the locality, independant of the system size $|S|$.

Theorem (1) The transition probabilities p_{ij} $(i,j \in S)$ of the Metropolis chain $\{X_n\}_{n=0}^{\infty}$ generated above are given by

$$p_{ij} = \begin{cases} q_{ij}\alpha_{ij} & \text{if } j \neq i, \\ 1 - \sum_{k \in S, k \neq i} q_{ik}\alpha_{ik} & \text{if } j = i. \end{cases}$$

(2) If Q is symmetric ~~and irreducible. Then the Metropolis chain $\{X_n\}_{n=0}^{\infty}$ is also irreducible.~~

then the transition matrix $P = (p_{ij}) = (p(i,j))$ and the distribution π satisfy the detailed balance.

$$\pi(i)\, p(i,j) = \pi(j)\, p(j,i) \quad \forall i,j \in S.$$

(3) If all $q_{ij} > 0$ $\forall i,j \in S$, then P is irreducible. If in addition, $p(i,i) > 0$ $\forall i \in S$, then P is also aperiodic. In this case, π is the unique invariant distribution. and

$$\lim_{n\to\infty} \pi_0 P^n = \pi$$

for any initial distribution π_0. Moreover, with probability 1,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k \in A\}} = \pi(A), \quad \forall A \subseteq S.$$

Proof (1) The event $\{Y = j\}$, given $\{X_k = i\}$, and the event $\{X_{k+1} = j\}$ with $j \neq i$, are independent. So, the probability that both events occur is the product of the individual probabilities.

$$p(i,j) = q(i,j)\, \alpha_{ij} \qquad (j \neq i).$$

For $j = i$, we have

$$p(i,i) = 1 - \sum_{k\in S,\, k\neq i} p(i,k)$$

$$= 1 - \sum_{k\in S,\, k\neq i} q(i,k)\, \alpha(i,k).$$

(2) If $i = j$, then $\pi(j)\, p(j,i) = \pi(i)\, p(i,j) = \pi(i)\, p(i,i)$.

If $i \neq j$ then

$$\pi(i)\, p(i,j) = \pi(i)\, q(i,j)\, \alpha_{ij}$$

$$= \pi(i)\, q(i,j) \min\left(1, \frac{\pi(j)}{\pi(i)}\right)$$

$$= q(i,j) \min(\pi(i), \pi(j))$$

$$= q(j,i) \min(\pi(j), \pi(i))$$

$$= \pi(j)\, q(j,i) \min\left(1, \frac{\pi(i)}{\pi(j)}\right)$$

$$= \pi(j)\, p(j,i).$$

(3) Since $p(i,j) = q(i,j)\,\alpha_{ij} > 0$ if $i \neq j$, then P is irreducible.

Since $p(i,i) > 0$, P is also aperiodic.

Thus, π is the unique distribution, and the limits hold true. □

The Metropolis-Hastings Algorithm

This is an improvement of the original Metropolis algorithm: the transition matrix Q is allowed to be non symmetric.

Given: – a discrete state space S;

– a proposal transition matrix $Q = (q_{ij}) = (q(j,i))$ which may not be symmetric.

– a target probability distribution π on S with $\pi > 0$ (i.e., $\pi(i) > 0 \ \forall i \in S$).

Generate a Markov chain X_n $(n = 0, 1, \dots)$ with π the invariant distribution.

The Metropolis-Hastings Algorithm

Suppose $X_n = i \in S$.

(1) Generate $Y \in S$ according to Q, i.e.,

$$P(Y = j \mid X_n = i) = q_{ij} \quad \forall j \in S.$$

Let $Y = j \in S$.

(2) Set $\alpha_{ij} = \min\left\{1, \frac{\pi(j)q(j,i)}{\pi(i)q(i,j)}\right\}$

Accept Y with the probability α_{ij}:

Generate $U \sim U[0,1]$.

If $U \le \alpha$, accept Y and $X_{k+1} \leftarrow Y$;

If $U > \alpha$, reject Y and $X_{k+1} \leftarrow X_k$

Remarks

(1) Only $\frac{\pi(i)}{\pi(j)}$ $(i,j \in S)$ are needed. This means we need only to know π up to a normalizing constant.

The following result is similar to that for the Metropolis method

Theorem (1) The transition probabilites of the generated Markov chain $X_n (n = 0, 1, \cdots)$ are given by

$$p(i,j) = \begin{cases} q(i,j)\alpha_{ij} & \text{if } j \neq i, \\ 1 - \sum_{k \in S, k \neq i} q(i,k)\alpha_{ik} & \text{if } j = i, \end{cases}$$

(2) The transition matrix $P = (p(i,j))$ and the distribution π satisfy the detailed balance

$$\pi(i)p(i,j) = \pi(j)p(j,i) \quad \forall i,j \in S.$$

Hence π is the unique invariant measure for P.

(3) If all $q(i,j) > 0$ $\forall i,j \in S$, then P is irreducible and aperiodic. Hence

$$\lim_{n \to \infty} \pi_0 P^n = \pi$$

for any initial distribution, and with the probability 1,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k \in A\}} = \pi(A) \quad \forall A \subseteq S.$$

Markov chain Monte Carlo Methods for the knapsack problem

The knapsack problem

Given: m — a positive integer

$\vec{v} = (v_1, \cdots, v_m) \in \mathbb{R}^m$, each $v_i > 0$

$\vec{w} = (w_1, \cdots, w_m) \in \mathbb{R}^m$, each $w_i > 0$

$b > 0$

Define $S = \{\vec{z} = (z_1, \cdots, z_m) \in \mathbb{R}^m : z_j = 0 \text{ or } 1,\ j = 1, \cdots, m,$

$\vec{w} \cdot \vec{z} = \sum_{j=1}^{m} w_j z_j \le b\}$.

$\vec{z} \in S$: feasible

$$\max_{z \in S} \vec{v} \cdot \vec{z}$$

Remarks

① Modeling: m items, with the ith item with value \$$v_i$, weight w_i kg. Pick up items with maximal value but with total weight $\le b$.

② Usually solved by the dynamic programming method. But, in general, it takes an exponential number of steps.

③ large system. $m = 50$, $|S| > 10^{15}$; $m = 100$, $|S| = 2^m > 10^{30}$.

Design Monte Carlo methods to produce feasible $\vec{z} \in S$ that is close to ~~the~~ an maximizer with high probability. One can generate many such $\vec{z}$'s and compute $\vec{v} \cdot \vec{z}$ to choose the maximal value.

The accept-reject method to

Method 1 Generate a random $\vec{z} \in S$ that is uniformly ~~distributed~~ distributed on S.

① Pick up $\vec{z} = (z_1, \cdots, z_m)$, $z_j = 0$ or 1 $(1 \le j \le m)$ uniformly at random

② If $\vec{z} \in S$ accept it; if $\vec{z} \notin S$, reject it. Go to ①.

Drawback ⓐ For large m, slow, since the acceptance rate can be small.

Example Set $w_1 = w_2 = \cdots = w_m = 1$, $b = \frac{m}{3}$.

$$S = \left\{ \vec{z} = (z_1, \cdots z_m) : z_j = 0 \text{ or } 1,\ 1 \le j \le m,\ \sum_{i=1}^{m} z_i \le m/3 \right\}.$$

$$|S| = \sum_{j=0}^{m/3} \binom{m}{j} \le \left(1 + \frac{m}{3}\right) \binom{m}{\frac{m}{3}}$$

$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$.

By ~~the~~ Stirling's approximation, the probability of acceptance is

$$\frac{|S|}{2^m} \approx (0.83)^m.$$

If $m = 100$, this is around 10^{-8}, very small!

Method 2 Markov chain Monte Carlo method to generate $\vec{z} \in S$ uniformly distributed on S.

Given $X_k = \vec{z} = (z_1, \cdots, z_m) \in S$

① Choose $J \in \{1, \cdots, m\}$ uniformly at random

② "Flip" ~~$\vec{z}$~~ z_J, i.e., set $\vec{y} = (z_1, \cdots z_{J-1}, 1 - z_J, z_{J+1}, \cdots, z_m)$

③ If $\vec{y} \in S$ then $X_{k+1} \leftarrow \vec{y}$.
If $\vec{y} \notin S$ then $X_{k+1} \leftarrow X_k$.

The Markov chain $\{X_n\}_{n=0}^{\infty}$ satisfies:

① symmetric, i.e., the transition matrix $P = (p_{ij}) = (p_{\vec{y}\vec{z}})$ is symmetric.

$p_{\vec{y},\vec{z}} = p_{\vec{z},\vec{y}} \quad \forall \vec{z}, \vec{y} \in S$

So, $\vec{z} \to \vec{0}$ Wlg. P is symmetric, so, $\vec{0} \to \vec{z}$.

① Irreducible: every state $\vec{z} \in S$ is connected to $\vec{0} \in S$. (One can always remove items!)

① It's aperiodic: (assuming $\sum_{i=1}^{n} w_i > b$), since $\exists \vec{z} \in S$ such that no items can be added, so, $p_{\vec{z},\vec{z}} \geq \frac{1}{m} > 0$.

For a finite-state Markov chain, if it is irreducible, aperiodic, and symmetric, then the unique invariant (or stationary) dist. is the uniform distribution.

$$\pi(\vec{z}) = \frac{1}{|S|} \quad \forall \vec{z} \in S.$$

$\pi(i) = \frac{1}{|S|} \quad \forall i \in S$

Remark ① We sample π even $|S|$ is unknown.

② Uniform distribution is not good, as there may exist only a few states close to optimal.

Suggest: try $\pi(\vec{z}) = e^{\vec{v}\cdot\vec{z}}/C \quad \forall \vec{z} \in S$?

Method 3 The Metropolis Alg. — a Markov chain Monte Carlo method to generate $\vec{z} \in S$ with a new distribution

Given: S — same as before

$Q = (q_{ij})$ — transition matrix: uniform!

$\pi(\vec{z}) = e^{\beta(\vec{v}\cdot\vec{z})}/C_\beta$ where $\beta > 0$ (fixed)

$C_\beta = \sum_{\vec{y}\in S} e^{\beta(\vec{v}\cdot\vec{y})}$ — a normalizing constant.

~~Do not need to know C_β.~~

Given $X_k = \vec{z} \in S$.

(1) Choose $J \in \{1, \dots, m\}$ uniformly at random.

(2) Set $\vec{y} = (z_1, \dots, z_{J-1}, 1 - z_J, z_{J+1}, \dots, z_m)$.

(3) If $\vec{y} \notin S$ then $X_{k+1} \leftarrow X_k$.
If $\vec{y} \in S$ then accept $\vec{y}$ and $X_{k+1} \leftarrow \vec{y}$ with the prob. $\alpha = \min\left\{1, \frac{\pi^{\beta}(\vec{y})}{\pi^{\beta}(\vec{z})}\right\}$.

i.e., generate $u \sim U[0,1]$.
If $u \le \alpha$ accept $\vec{y}$ and $X_{k+1} \leftarrow \vec{y}$,
If $u > \alpha$ reject $\vec{y}$ and $X_{k+1} \leftarrow X_k$.

Remarks (1) Do not need to know C_β.

(2) Simple calc of α. $\frac{\pi^{\beta}(\vec{y})}{\pi^{\beta}(\vec{z})} = e^{\beta\vec{v}\cdot(\vec{y}-\vec{z})} = e^{\beta\vec{v}\cdot(0,\dots,0,1-2z_J,0,\dots,0)}$

$$\Rightarrow \alpha = \begin{cases} e^{-\beta v_J} & \text{if } z_J = 1 \\ 1 & \text{if } z_J = 0 \end{cases}$$

(3) Trapped in a local state.

Example $v_i = w_i = 1$ if $i = 1, \cdots, m-1$,

$v_m = w_m = m$ 并

$b = m$.

$$S = \{\vec{z} = (z_1, \cdots, z_m) : z_j = 0 \text{ or } 1,\ 1 \le j \le m,\ \sum_{j=1}^{m-1} z_j + z_m m \le m\}$$

$$= \{(z_1, \cdots, z_{m-1}, 0) : z_i = 0 \text{ or } 1,\ 1 \le i \le m-1\} \cup \{(0, \cdots, 0, 1)\}$$

$$C_\beta = \sum_{\vec{y} \in S} e^{\beta \vec{v} \cdot \vec{y}} = \sum_{\vec{y} \in S,\ y_m = 0} e^{\beta \sum_{i=1}^{m-1} y_i} + e^{\beta m}$$

$$= e^{\beta \cdot 0} \binom{m-1}{0} + e^{\beta \cdot 1} \binom{m-1}{1} + \cdots + e^{\beta(m-1)} \binom{m-1}{1} + e^{\beta m}$$

$$= (1 + e^\beta)^{m-1} + e^{\beta m}.$$

Let $\vec{e} = (0, \cdots, 0, 1) \in S$. The only feasible solution adjacent to $\vec{e}$ is $\vec{0}$. [Flip one component of $\vec{e}$, still in S.]

Hard for the Markov chain to get out of $\vec{e}$.

Show:

$$P_{\vec{e},\vec{0}} = \underbrace{Q_{\vec{e},\vec{0}}}_{\text{unif.}} \min\left(1, \frac{\pi(\vec{0})}{\pi(\vec{e})}\right)$$

$$= \frac{1}{m} e^{\beta(\vec{0} - \vec{e})} = \frac{1}{m} e^{-\beta m}.$$

$P_{\vec{e},\vec{0}} + P_{\vec{e},\vec{e}} = 1$.

So, $P^{(k)}_{\vec{e},\vec{e}} \ge \left(1 - \frac{e^{-\beta m}}{m}\right)^k \ge 1 - \frac{k}{m} e^{-\beta m}$

As $k \to \infty$, $P_k(i,j) \to \pi(j)$ $\forall j \in S$.

Comparing this with

$$\pi^\beta(\vec{e}) = \frac{e^{\beta m}}{C_\beta} \approx \left(\frac{e^\beta}{1 + e^\beta}\right)^{m-1} e^\beta$$

we see $P^{(k)}_{\vec{e},\vec{e}} = P_k(\vec{e},\vec{e}) >> \pi^\beta(\vec{e})$ unless $k = O(e^{\beta m})$.

Later: the method of simulated annealing.

Sampling Methods for Generating the Transition Matrix Q

- **Independence Sampler.** $q(i,j) = g(j) \quad \forall i,j \in S$.

- **Uniform Sampler.** This is for sampling a uniform distribution. Define a neighbood structure on S. Let $n_i = \#$ neighbors of $i \in S$. Choose
$$q(i,j) = \frac{1}{n_i} \qquad \forall i,j \in S.$$
Note: With such a Q, and a uniform distribution π as the target distribution, the acceptance probability in the Metropolis-Hastings algorithm is $\alpha(i,j) = \min\left\{1, \frac{\pi(j)}{\pi(i)}\right\} \quad \forall i,j \in S$.
The limiting invariant distribution of the Metropolis-Hastings chain is the uniform distribution.

- **Random Walk Sampler.** Given $X_k = i \in S$. Generate $Y = i + \sigma Z$, where Z is generated from a spherically symmetrical distribution (in the continuous case), e.g. $N(\underline{0}, \sigma^2 \underline{I})$. In this case, Q is symmetric and $\alpha(i,j) = \min\left\{1, \frac{\pi(j)}{\pi(i)}\right\}$ in the Metropolis-Hastings algorithm.

In the continuous case with $f(x)$ the PDF, one can use the Langevin diffusion to sample X_k with the PDF $f(x)$: solving the SDE
$$dX_t = \frac{1}{2} \nabla \ln f(X_t)\, dt + dW_t,$$
say, using the Euler time-discretization. This is the Langevin Metropolis-Hastings algorithm.

The Simulated Annealing Method

This is a general stochastic optimization method. It generalizes the Metropolis-Hastings method to allow the acceptance probabilities to depend on the steps. This method is closedly related to a simple observation of a physical system, where increasing the temperature will increase the chance for an underlying system to get out of a local minimum. Since the generated Markov chain will no longer be time homogeneous. The mathematical theory for convergence becomes more complicated.

The Simulated Annealing Algorithm for the Knapsack Problem

We define as before

$$S = \{\vec{z} = (z_1, \cdots, z_m) : z_j = 0 \text{ or } 1,\ 1 \le j \le m,\ \vec{w} \cdot \vec{z} = \sum_{j=1}^{m} w_j z_j \le b\}$$

$\vec{w} = (w_1, \cdots, w_m) > 0$ is given

$b > 0$ is given

$\vec{v} = (v_1, \cdots, v_m) > 0$ is given.

The original knapsack problem is: $\max_{\vec{z} \in S} \vec{v} \cdot \vec{z}$.

We use Markov chain Monte Carlo methods to

Sample feasible solutions $\vec{z} \in S$ with certain distribution π on S. As seen in the Metropolis algorithm applied to the knapsack problem, one can design a target distribution so that the chain is distributed around states with the large values of the objective function in the optimization problem, and then evaluate the function values of such states to get approximate solutions to the original problem. Here, we will not fix such a distribution, but we change it dynamically (depending on steps).

We define $\pi^{\beta(t)}(\vec{z}) = \frac{1}{C_{\beta(t)}} e^{\beta \vec{v} \cdot \vec{z}} \quad \forall \vec{z} \in S$

where $\beta(t) \uparrow \infty$ as $t \uparrow \infty$! For instance, we can choose

$$\beta(t) = \log t \quad \text{or} \quad (1.0001)^t.$$

We assume $Q = (q_{ij}) = (q(i,j))$ is the transition matrix of a background Markov chain. Usually this is a simple matrix of transition probabilities, e.g., uniform. In general, Q may not be symmetric.

Given $X_k = \vec{z} \in S \qquad \vec{z} = (z_1, \cdots, z_m)$.

(1) Select $J \in \{1, \cdots, m\}$ uniform at random.

(2) Flip to define

$$\vec{y} = (z_1, \cdots, z_{J-1}, 1 - z_J, z_{J+1}, \cdots, z_m).$$

(3) If $\vec{y} \notin S$, then reject $\vec{y}$ and $X_{k+1} \leftarrow X_k$

(4) If $\vec{y} \in S$, accept it with the probability

$$\alpha^{(k)}_{\vec{z},\vec{y}} := \min\left\{1, \frac{\pi^{\beta(k)}(\vec{y})\, q(\vec{y},\vec{z})}{\pi^{\beta(k)}(\vec{z})\, q(\vec{z},\vec{y})}\right\}$$

This means that: Generate $U \sim U[0,1]$.

If $U \le \alpha^{(k)}_{\vec{z},\vec{y}}$, accept $\vec{y}$ and $X_{k+1} \leftarrow \vec{y}$.

If $U > \alpha^{(k)}_{\vec{z},\vec{y}}$, reject $\vec{y}$ and set $X_{k+1} \leftarrow X_k$.

Remarks ① The transition probabilities for the generated chain X_k $(k=0,1,2,\dots)$ are

$$P_{ij} = P(X_{k+1}=j \mid X_k = i)$$
$$= \begin{cases} q(i,j)\,\alpha^{(k)}_{ij} & \text{if } i \neq j, \\ 1 - \sum_{k \in S, k \neq i} q(i,k)\,\alpha^{(k)}_{ki} & \text{if } i = j. \end{cases}$$

② We do not need to know $C_{\beta(k)}$, the normalizing constant in each step k.

③ Suppose $M = \max_{\vec{z} \in S} \vec{v}\cdot\vec{z} > 0$. Let

$$S_{max} = \{\vec{z} \in S : \vec{v}\cdot\vec{z} = M\}.$$

Then, we have

$$\lim_{\beta\to\infty} e^{-\beta M} \sum_{\vec{z}\in S} e^{\beta(\vec{v}\cdot\vec{z})} = |S_{max}|.$$

Hence, $\lim_{\beta\to\infty} \pi^{\beta}(\vec{z}) = \begin{cases} \frac{1}{|S_{max}|} & \text{if } \vec{z} \in S_{max}, \\ 0 & \text{if } \vec{z} \notin S_{max}. \end{cases}$

That is, the limiting distribution is concentrated on S.

The Simulated Annealing Method for a General Optimization Problem

Given: $S \neq \emptyset$, a discrete set.

$G : S \to \mathbb{R}$ bounded above.

The optimization problem: $\max_{\vec{z} \in S} G(\vec{z})$.

Let Q be a stochastic matrix on S. (We often assume it is irreducible and aperiodic.)

Let $\beta(t) : [1, \infty) \to [0, \infty)$ be an increasing function such that $\lim_{t \to \infty} \beta(t) = +\infty$.

Define $\pi^{\beta(t)}(\vec{z}) = e^{\beta(t) G(\vec{z})} \quad \forall \vec{z} \in S$.

The algorithm

Given $X_k = \vec{z} \in S$.

(1) Generate $Y \in S$ according to Q:

$$\mathbb{P}(Y = \vec{y} \mid X_k = \vec{z}) = q(\vec{z}, \vec{y}).$$

Let $Y = \vec{y} \in S$.

(2) Set $\alpha^{(k)}_{\vec{z}, \vec{y}} = \min\left\{1, \dfrac{\pi^{\beta(k)}(\vec{y})\, q(\vec{y}, \vec{z})}{\pi^{\beta(k)}(\vec{z})\, q(\vec{z}, \vec{y})}\right\}$

(3) Accept Y with the probability $\alpha^{(k)}_{\vec{z}, \vec{y}}$:

Generate $U \sim \mathcal{U}[0,1]$.

If $U \leq \alpha$ then accept Y and set $X_{k+1} \leftarrow Y$.

If $U > \alpha$ then reject Y and set $X_{k+1} \leftarrow X_k$.

The Gibbs Sampler

This is a Markov chain Monte Carlo method for sampling random vectors in $\mathbb{R}^d$ ($d \geq 2$). It constructs a Markov chain from a sequence of conditional distributions. This method is therefore different from the Metropolis, Metropolis-Hastings, and simulated annealing method. It is though possible to combine the Gibbs sampling method with the Metropolis-Hastings, or other, method.

Example Gibbs Sampler on the Discrete Hypercubes.

Let $S = \{ \vec{z} = (z_1, \cdots, z_m) \in \mathbb{R}^m : z_j = 0 \text{ or } 1, \ 1 \leq j \leq m \}$.

Let $\pi > 0$ be a probability distribution on S. The Gibbs sampler is the Markov chain $X_0, X_1, \cdots$ on S defined as follows.

Given $X_k = \vec{z} = (z_1, \cdots, z_m) \in S$.

(1) Pick $J \in \{1, \cdots, m\}$ uniformly at random.

(2) Set

$$\vec{y}[0] = (z_1, \cdots, z_{J-1}, 0, z_{J+1}, \cdots, z_m)$$

$$\vec{y}[1] = (z_1, \cdots, z_{J-1}, 1, z_{J+1}, \cdots, z_m)$$

$$\alpha_0 = \frac{\pi(\vec{y}[0])}{\pi(\vec{y}[0]) + \pi(\vec{y}[1])}$$

$$\alpha_1 = 1 - \alpha_0 = \frac{\pi(\vec{y}[1])}{\pi(\vec{y}[0]) + \pi(\vec{y}[1])}$$

(3) Assign

$$X_{k+1} = \begin{cases} \vec{y}[0] & \text{with probability } \alpha_0, \\ \vec{y}[1] & \text{with probability } \alpha_1. \end{cases}$$

That is. generate $U \sim \mathcal{U}[0,1]$.

If $U \le \alpha_0$ then $X_{k+1} \longleftarrow \vec{y}[0]$;

If $U > \alpha_0$ then $X_{k+1} \longleftarrow \vec{y}[1]$.

This method erases the j-th coordinate of X_k and replaces it by a number drawn from the conditional (equilibrium) distribution of the j-th coordinate given the other $m-1$ coordinates.

To be more precise, let us denote

$$\vec{u}_{-k} = (u_1, \cdots, u_{k-1}, u_{k+1}, \cdots, u_m) \in \mathbb{R}^{m-1}$$

for a given $\vec{u} = (u_1, \cdots, u_{k-1}, u_k, u_{k+1}, \cdots, u_m) \in \mathbb{R}^m$ and a given $k \in \{1, \cdots, m\}$. Then, with $\vec{z}$, $\vec{y}[0]$, and $\vec{y}[1]$ as above, $\vec{u} = \vec{y}[i] \Leftrightarrow \vec{u}_{-j} = \vec{z}_{-j}$ and $u_j = i$ ($i = 0$ or 1).

Let $\vec{V} = (V_1, \cdots, V_m) \in \mathbb{R}^m$ be a random vector π-distributed

Then

$$\pi(\vec{y}[0]) = \mathbb{P}(\vec{V}_{-j} = \vec{z}_{-j} \text{ and } V_j = 0),$$

$$\pi(\vec{y}[1]) = \mathbb{P}(\vec{V}_{-j} = \vec{z}_{-j} \text{ and } V_j = 1),$$

and $$\pi(\vec{y}[0]) + \pi(\vec{y}[1]) = \mathbb{P}(\vec{V}_{-j} = \vec{z}_{-j}).$$

Therefore

$$\alpha_0 = \frac{\pi(\vec{y}[0])}{\pi(\vec{y}[0]) + \pi(\vec{y}[1])} = \mathbb{P}(V_j = 0 \mid \vec{V}_{-j} = \vec{z}_{-j}).$$

This means that the probabilities in Step (2) above are indeed the corresponding conditional probabilities of the equilibrium distribution π.

The general discrete case

Let $S \subseteq \mathbb{R}^m$ be a discrete set, π a probability distribution on S, and $\vec{V} \in S$ a random variable with distribution π. For each $j \in \{1, \cdots, m\}$, define the matrix P_j by

$$P_j(\vec{z}, \vec{w}) = \mathbb{P}\left(\vec{V} = \vec{w} \,\middle|\, \vec{V}_{-j} = \vec{z}_{-j}\right)$$

$$= \begin{cases} 0 & \text{if } \vec{w}_{-j} \neq \vec{z}_{-j}, \\ \dfrac{\pi(\vec{w})}{\pi_{-j}(\vec{z}_{-j})} & \text{if } \vec{w}_{-j} = \vec{z}_{-j}, \end{cases}$$

(P_j is of order $|S| \times |S|$.)

where $\pi_{-j}(\vec{z}_{-j}) = \mathbb{P}\left(\vec{V}_{-j} = \vec{z}_{-j}\right) = \sum_{\vec{u} \in S,\, \vec{u}_{-j} = \vec{z}_{-j}} \pi(\vec{u})$.

Define the matrix

$$P_{RS} = \frac{1}{m} \sum_{j=1}^{m} P_j.$$

This is the transition matrix for the random scan Gibbs sampler. Define

$$P_{SS} = P_1 P_2 \cdots P_m.$$

This is the transition matrix for the systematic scan Gibbs sampler.

Example Let $\{\vec{S}_k\}_{k=0}^{\infty}$ be the Markov chain with transition matrix P_{SS} for some distribution $\pi > 0$ on $\mathbb{R}^3$. Then one transition of P_{SS} corresponds to the following:

If $\vec{S}_k = (z_1, z_2, z_3)$, then:

(1) Generate Y_1 from the conditional distribution V_1 given $\vec{V}_{-1} = (z_2, z_3)$;

(2) Generate Y_2 from the conditional distribution of V_2 given $\vec{V}_{-2} = (Y_1, z_3)$;

(3) Generate Y_3 from the conditional distribution of V_3 given $\vec{V}_{-3} = (Y_1, Y_2)$.

Then set $\vec{S}_{n+1} \leftarrow (Y_1, Y_2, Y_3)$.

Proposition With the above definitions:

(1) Each of $P_1, \cdots, P_m$ and P_{RS} satisfies the detailed balance with respect to π;

(2) π is an invariant distribution for P_{SS}, even though P_{SS} need not be reversible (i.e., P_{SS} and π may not satisfy the detailed balance.

Proof (1) Fix $k \in \{1, \cdots, m\}$. We need to show

(*) $$\pi(\vec{z}) P_k(\vec{z}, \vec{w}) = \pi(\vec{w}) P_k(\vec{w}, \vec{z}) \quad \forall \vec{z}, \vec{w} \in S.$$

Note: if $\vec{z}_{-k} \neq \vec{w}_{-k}$ then $P_k(\vec{w}, \vec{z}) = 0 = P_k(\vec{z}, \vec{w})$. So (*) is true for $\vec{z}, \vec{w}$.

If $\vec{z}_{-k} = \vec{w}_{-k}$ then

$$\pi(\vec{z}) P_k(\vec{z}, w) = \pi(\vec{z}) \frac{\pi(\vec{w})}{\pi_{-k}(\vec{z}_{-k})} = \frac{\pi(\vec{w})\pi(\vec{z})}{\pi_{-k}(\vec{w}_{-k})}$$

$$= \pi(\vec{w}) P_k(\vec{w}, \vec{z}).$$

So, (*) is true.

(2) By Part (1), π is an invariant distribution for each P_k, $\pi P_k = \pi$. Hence.

$$\pi P_{SS} = (\pi P_1) P_2 \cdots P_m = (\pi P_2) P_3 \cdots P_m = \pi P_m = \pi. \quad \square$$

The general continuous / discrete case.

Let $\vec{X} = (x_1, \cdots x_d)$ be a random vector with the probability distribution function $f(\vec{x})$ $(\vec{x} \in \mathbb{R}^d)$.
Let $f(x_i | x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_d)$ represent the conditional PDF of the ith component x_i, given the other components, $x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_d$.

The Gibbs Sampler

Suppose $\vec{X}_k$ is generated. $\vec{X}_k = (x_{k,1} \cdots x_{k,d})$

(1) Draw y_1 from the conditional PDF $f(y_1 | x_{k,2}, \cdots, x_{k,d})$.

(2) For $i = 2$ to d

Draw y_i from the conditional PDF $f(y_i | y_1, \cdots y_{i-1}, x_{k,i+1}, \cdots x_{k,d})$.

~~End~~ ~~Set x_{k+1}~~ for.

(3) Set $\vec{X}_{k+1} \leftarrow \vec{Y} = (y_1, \cdots y_d)$.

The transition PDF is given by

$$p_{1 \to d}(\vec{y} | \vec{x}) = \prod_{i=1}^{d} f(y_i | y_1, \cdots y_{i-1}, x_{i+1}, \cdots x_d)$$

where $1 \to d$ means that the components of $\vec{x}$ are updated in the order $1 \to 2 \to 3 \to \cdots \to d$. The transition PDF of the reverse move $\vec{y} \to \vec{x}$ in which $\vec{y}$ is updated in the order ~~d~~ $\to$ ~~$d-1$~~ $\to \cdots \to 1$ is

$$p_{d \to 1}(\vec{x} | \vec{y}) = \prod_{i=1}^{d} f(x_i | y_1, \cdots y_{i-1}, x_{i+1}, \cdots x_d).$$

Let $f_i(x_i)$ be the ith marginal density of the PDF $f(x)$.

Definition A PDF f on $\mathbb{R}^d$ satisfies the positivity condition, if for any $\vec{y} \in = (y_1, \cdots, y_d) \in \mathbb{R}^d$,
$f(\vec{y}) > 0$ provided that $f_i(y_i) > 0$, $i = 1, \cdots, d$.

Theorem If f is a PDF on $\mathbb{R}^d$ that satisfies the positivity condition, then

$$f(\vec{y})\, p_{d \to 1}(\vec{x} \mid \vec{y}) = f(\vec{x})\, p_{1 \to d}(\vec{y} \mid \vec{x}).$$

Proof

$$\frac{p_{1 \to d}(\vec{y} \mid \vec{x})}{p_{d \to 1}(\vec{x} \mid \vec{y})} = \prod_{i=1}^{d} \frac{f(y_i \mid y_1, \cdots, y_{i-1}, x_{i+1}, \cdots, x_d)}{f(x_i \mid y_1, \cdots, y_{i-1}, x_{i+1}, \cdots, x_d)}$$

$$= \prod_{i=1}^{d} \frac{f(y_1, \cdots, y_i, x_{i+1}, \cdots, x_d)}{f(y_1, \cdots, y_{i-1}, x_i, \cdots, x_d)}$$

$$= \frac{f(\vec{y})}{f(\vec{x})} \frac{\prod_{i=1}^{d-1} f(y_1, \cdots, y_i, x_{i+1}, \cdots, x_d)}{\prod_{j=2}^{d} f(y_1, \cdots, y_{j-1}, x_j, \cdots, x_d)}$$

$$= \frac{f(\vec{y})}{f(\vec{x})} \cdot \frac{\prod_{i=1}^{d-1} f(y_1, \cdots, y_i, x_{i+1}, \cdots, x_d)}{\prod_{j=1}^{d-1} f(y_1, \cdots, y_j, x_{j+1}, \cdots, x_d)}$$

$$= \frac{f(\vec{y})}{f(\vec{x})}. \qquad \square$$

The theorem provides a generalized detailed balance.

The Gibbs sampler and the Metropolis-Hastings method can be combined together to sample a random vector in $\mathbb{R}^d$ with a given PDF $f(x)$ by generating a Markov chain with $f(x)$ the invariant distribution.

Given $\vec{X}_n = \vec{x} = (x_1, \cdots, x_d)$.

(1) Generate Y_i from $f(y_i \mid \vec{x}_{-i})$ conditionally on $Y_i \neq x_i$; i.e., generate Y_i from the PDF

$$\frac{f(y_i \mid \vec{x}_{-i})}{1 - f(y_i \mid \vec{x}_{-i})}, \quad y_i \neq i_i.$$

Set $\vec{Y} = (x_1, \cdots, x_{i-1}, Y_i, x_{i+1}, \cdots, x_d)$.

(2) Apply the Metropolis-Hastings acceptance criterion to accept $\vec{Y}$ with the acceptance probability

$$\alpha(\vec{x}, \vec{Y}) = \min\left\{1, \frac{1 - f(x_i \mid \vec{x}_i)}{1 - f(Y_i \mid \vec{x}_{-i})}\right\}.$$

The Slice Sampler

Suppose we would like to sample a random vector in a subset S of $\mathbb{R}^d$ according to a given PDF $f(\vec{x})$, $x \in S$. In a large class of methods, auxiliary variable methods, one views f as the marginal density

$$f(\vec{x}) = \int \tilde{f}(\vec{x}, \vec{y})\, d\vec{y},$$

where $\tilde{f}$ is the joint probability density for the random variables $\vec{x}$ and $\vec{y}$. Here, $\vec{y}$ is called an auxiliary or latent variable. The slice sampler is one of such auxiliary variable methods.

Given a PDF f on $S \subseteq \mathbb{R}^d$:

$$f(\vec{x}) = b \sum_{k=1}^{m} p_k(\vec{x}),$$

where $p_k : S \to (0, \infty)$ $(k = 1, \cdots, m)$ are given functions, not necessary PDFs, and $b > 0$ is a normalizing constant, which may or may not be known, and $m \geq 1$ is an integer.

We introduce the auxiliary variable $\vec{y} = (y_1, \cdots, y_m)$ such that the joint density of $\vec{x}$ and $\vec{y}$ is

$$\tilde{f}(\vec{x},\vec{y}) = \tilde{b} \prod_{k=1}^{m} \mathbb{I}\{0 \le y_k \le p_k(\vec{x})\},$$

where $\tilde{b} > 0$ is a normalizing constant. (No need to know the value of $\tilde{b}$.) Given $\vec{x}$, each y_k has uniform density on $[0, p_k(\vec{x})]$, and the marginal density

$$\int \tilde{f}(\vec{x},\vec{y})\,d\vec{y}$$

is exactly $f(\vec{x})$. Note that all $y_1, \cdots, y_m$ are independent. The idea of the slice sampler here is to apply a grouped Gibbs sampler on the augmented space by iteratively sampling from the conditional densities $\tilde{f}(\vec{x}|\vec{y})$ and $\tilde{f}(\vec{y}|\vec{x})$

Algorithm of the Slice Sampler

Given $\vec{X}_k \in S$

(1) Generate $\vec{Y}$ from the conditional density $\tilde{f}(\vec{y}|\vec{X}_k)$, i.e.,
generate $U_1, \cdots, U_m \overset{\text{i.i.d.}}{\sim} \mathcal{U}[0,1]$.
Set $Y_j = U_j p_j(\vec{X}_k)$, $j = 1, \cdots, m$

(2) Generate $\vec{X}_{k+1}$ from the conditional density $\tilde{f}(\vec{x}|\vec{Y})$, i.e., draw $\vec{X}_{k+1}$ uniformly from the set $\{\vec{x} \in S : p_j(\vec{x}) \ge Y_j, j = 1, \cdots, m\}$.

Remarks

- The second step is often more involved.
- If $\operatorname{supp} f$ is compact and f is bounded, then the chain produced by the slice sampler is ergodic (i.e., irreducible).

Example A 3D slice sampler. Let $f = f(x)$ be a PDF on $\mathbb{R}^1$, given by

$$f(x) = b \underbrace{[1+\sin^2(3x)]}_{p_1(x)} \underbrace{[1+\cos^4(5x)]}_{p_2(x)} \underbrace{e^{-\frac{x^2}{2}}}_{p_3(x)}.$$

We generate X_k ($k=0,1,2,\dots$) so that in the large k limit, X_k will have the PDF f.

Given $X_k \in \mathbb{R}$

We generate $U_1, U_2, U_3 \overset{iid}{\sim} \mathcal{U}[0,1]$, and set $Y_j = U_j p_j(X_k)$, $j=1,2,3$.

Then, we generate $X_{k+1} \in \mathbb{R}$ uniformly on the set $\bigcap_{j=1}^{3} \{x \in \mathbb{R} : p_j(x) \geq Y_j\}$

$$= \{x \in \mathbb{R} : \sin^2(3x) \geq 1 - Y_1\} \cap \{x \in \mathbb{R} : \cos^4(5x) \geq 1 - Y_2\}$$

$$\cap \{x \in \mathbb{R} : |x| \leq \sqrt{-2 \log Y_3}\}.$$