MCMC

Gwangsu Kim

JBNU

Second semester of 2024

Markov Chain with discrete (countable) state space I

- $\{X_t : t = 1, 2, \ldots\}$ sequence of random variables taking values in $S = \{1, \ldots, \}$ (state space) such that
  1. Markovian: $\mathbb{P}(X_{t+1} = x_{t+1} \mid X_t, = x_t, X_{t-1} = x_{t-1}, \ldots, X_1 = x_1) = \mathbb{P}(X_{t+1} \mid X_t = x_t)$
  2. Homogeneous and stationary: $\forall t,\ \mathbb{P}(X_{t+1} = j \mid X_t = i) = p(i, j)$ such that

$$\sum_j p(i, j) = 1,\ p(i, j) \geq 0$$

Markov Chain with discrete (countable) state space II

- Chapman-Kolmogorov equation: Let $P = \{p(i,j)\}_{(i,j)\in S^2}$

$$\mathbb{P}|(X_{t+k} = j \mid X_t = i) = P^K(i,j)$$

- Unique stationary distribution: aperiodic and irreducible Markov chain $\{X_t\}$ have the following:
  1. Unique π such that

$$\pi P = \pi \left(\sum_i \pi(i) = 1, \pi(i) \geq 0 \right)$$

Markov Chain with discrete (countable) state space III

- Aperiodicity and irreducibility

  Irreducible: $\forall i, j \in S,\ \exists k > 0, P^k(i,j) > 0$

  Aperiodic: 최대공약수$\{k : P^k(i,i) > 0\} = 1$, sufficient condition-$\forall i,\ P(i,i) > 0$.

- Reversible Markov chain: under the aperiodicity and irreducibility, if $\exists \pi$ such that

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

Markov Chain with discrete (countable) state space IV

then π should be the stationary distribution.

$$\begin{aligned}\sum_i \pi(i)p(i,j) &= \sum_i \pi(j)p(j,i) \\ &= \pi(j)\sum_i p(j,i) = \pi(j)\end{aligned}$$

* $\pi P = \pi$

Markov Chain with (continuous) state space (real line) I

- $\{X_t : t = 1, 2, \ldots\}$ sequence of random variables taking values in $S = \mathbb{R}$ (state space) such that
  1. $\mathbb{P}(X_{t+1} \in A \mid X_t = x_t, X_{t-1} = x_{t-1}, \ldots, X_1 = x_1) = \mathbb{P}(X_{t+1} \in A \mid X_t = x_t)$
  2. $\forall t,\ \mathbb{P}(X_{t+1} \in A \mid X_t = x) = \int_A p(x, dy)$ such that

$$\int_{\mathbb{R}} p(x, dy) = 1,\ \forall A\ \int_A p(x, dy) \geq 0$$

Markov Chain with (continuous) state space (real line) II

- Under mild conditions for $p(x, dy)$,

$$\int_A \pi(y)dy = \int_A p(x, dy)\pi(x)dx$$

determines the stationary distribution π

$$\mathbb{P}(X_{t+k} \in \cdot \mid X_t = x) \xrightarrow{d} \pi(\cdot) \text{ [a.e. } \pi]$$

as $k \to \infty$.

IDEA of Metropolis('53)-Hastings('70) Algorithm I

- Construct a Markov chain with respect to which a specified π can be stationary distribution.
  - Making the chain be reversible as: $\pi(x)p(x,dy)dx = \pi(y)p(y,dx)dy$
  - Getting π in a iterative manner

$$\int_A \pi^{t+1}(y)dy = \int_A p(x,dy)\pi^t(x)dx$$

$x^t \xrightarrow{d} \pi$ as $t \to \infty$ as long as x is in the support of π.

IDEA of Metropolis('53)-Hastings('70) Algorithm II

- Construct $p(x, dy)$ so that there is always a positive probability of visiting x starting from x ("aperiodicity")

  Transition prob. satisfying the aperiodcity.

  $$p(x, dy) = J(y \mid x)dy \times a(y \mid x) + \delta x(dy) \times \{1 - a(y \mid x)\}$$

  $0 < \int(1 - a(y \mid x))dx < 1$ & $0 < J(y \mid x) < \infty$, and $\delta_x(\cdot)$ is a distribution concentrated on $\{x\}$.

IDEA of Metropolis('53)-Hastings('70) Algorithm III

- Combing the above.

$$\begin{aligned} & \pi(x)J(y \mid x)a(y \mid x)dxdy + \pi(x)(1 - a(y \mid x))\delta_x(dy)dx \\ = \quad & \pi(y)J(x \mid y)a(x \mid y)dydx + \pi(y)(1 - a(x \mid y))\delta_y(dx)dy \end{aligned}$$

The following holds.

$$\pi(x)(1 - a(y \mid x))\delta_x(dy)dx = \pi(y)(1 - a(x \mid y))\delta_y(dx)dy$$

since $\delta_y(dx)$ and $\delta_x(dy)$ have positive probabilities on $x = y$.

IDEA of Metropolis('53)-Hastings('70) Algorithm IV

$$\pi(x)J(y \mid x)a(y \mid x)dxdy \quad = \quad \pi(y)J(x \mid y)a(x \mid y)dydx$$

is sufficient for the reversibility of the transition prob. with stationary prob.

- Metropolis Alg.
  - Using J such that $J(y \mid x) = J(x \mid y)$.
  - $\pi(x)a(y \mid x) = \pi(y)a(x \mid y)$ where $a(x \mid y) = \min\{\frac{\pi(x)}{\pi(y)}, 1\}$.
  - So that $\pi(x)a(y \mid x) = \pi(y)a(x \mid y) = \min\{\pi(y), \pi(x)\}$.

IDEA of Metropolis('53)-Hastings('70) Algorithm V

- Hastings Alg.
  - $\pi(x)J(y \mid x)a(y \mid x) = \pi(y)J(x \mid y)a(x \mid y)$ where $a(y \mid x) = \min\{\frac{\pi(y)J(x|y)}{\pi(x)J(y|x)}, 1\}$.
  - So that $\pi(x)J(y \mid x)a(y \mid x) = \pi(y)J(x \mid y)a(x \mid y) =$ $\min\{\pi(y)J(x \mid y), \pi(x)J(y \mid x)\}$.

  - Irreducibility is satisfied if $\sup(J(\cdot \mid x)) \subset \sup(\pi),\ \forall x \in \{x : \pi(x) > 0\}$

Sampling Algorithms I

- Goal of **MH algorithm**: generate random variables from $p(\theta \mid y)$, using $q(\theta \mid y) = p(\theta \mid y)\nu$.

1. Initial values: generating θ_0 from $p_0(\theta)$ with $q(\theta_0 \mid y) > 0$.
2. Generate $\theta^* \sim J(\cdot \mid \theta_0)$ and compute $r(\theta^* \mid \theta_0) = \begin{cases} \frac{q^*(\theta^*|y)}{q^*(\theta_0|y)} \\ \frac{q^*(\theta^*|y)J(\theta_0|\theta^*)}{q^*(\theta_0|y)J(\theta^*|\theta_0)} \end{cases}$
3. Generate $U \sim Unif[0, 1]$ and
$$\theta_1 = \begin{cases} \theta^* & U \leq \min\{r(\theta^* \mid \theta_0), 1\} \\ \theta_0 & \text{o.w.} \end{cases}$$
4. Repeat 2–4 so that $\theta_t \xrightarrow{d} p(\cdot \mid y)$ as $t \to \infty$.

Sampling Algorithms II

- Idea of **Gibbs sampling**: generation of $(x_1, x_2) \sim p_{1,2}(x_1, x_2)$ from $p_{2|1}(x_2 \mid x_1)$ and $p_{1|2}(x_1 \mid x_2)$.
  - Basic equation

$$
\begin{aligned}
p_2(x_2) &= \int p_{2|1}(x_2 \mid x) p_1(x) dx \\
p_1(x) &= \int p_{1|2}(x \mid y) p_2(y) dy \\
p_2(x_2) &= \int \left\{ p_{2|1}(x_2 \mid x) p_{1|2}(x \mid y) dx \right\} p_2(y) dy \\
&= \int J(x_2 \mid y) p_2(y) dy
\end{aligned}
$$

Sampling Algorithms III

- Basic Gibbs samples

$$
\begin{aligned}
p_2^{t+1}(x_2) &= \int \underbrace{J(x_2 \mid y)}\,\underbrace{p_2(y)^t}_{start}\, dy \\
J(x_2 \mid y) &= \int p_{2|1}(x_2 \mid x) p_{1|2}(x \mid y) dx
\end{aligned}
$$

- Note that the $r = 1$.

Sampling Algorithms IV

- Algorithm
  1. Generate x_2^0 from $p_0(x2$ where the p_0 approximates p_2.
  2. (x_1^1, x_2^2) from x_2^0
     - $x_1^1 \sim p_{1|2}(x \mid x_2^0)$
     - $x_2^1 \sim p_{2|1}(x \mid x_1^1)$
     - Repeat the above
  3. $(x_1^t, x_2^t) \xrightarrow{d} p_{1,2}(x_1, x_2)$ as $t \to \infty$.

Sampling Algorithms V

- Gibbs sampler (alternating conditional sampling)
- Goal: sampling $p(\theta \mid y)$ where $\theta = (\theta_1, \ldots, \theta_d)$.
- Algorithm
  1. Generate $\theta_1^1 \sim p(\theta_1 \mid \theta_2^0, \ldots, \theta_d^0)$
  2. Generate $\theta_2^1 \sim p(\theta_2 \mid \theta_1^1, \theta_3^0, \ldots, \theta_d^0)$

     $\vdots$ $\ldots$

  3. Generate $\theta_d^1 \sim p(\theta_d \mid \theta_1^1, \ldots, \theta_{d-1}^1)$

  - $(x_1^t, \ldots, x_d^t) \xrightarrow{d} p(x_1, \ldots, x_d)$ as $t \to \infty$.

Sampling Algorithms VI

- Remarks
  - Use the 2nd half of the generated sequence to diminish the effect of initial values.
  - Use multiple sequences of random #'s (vectors)
  - Use starting random numbers from over-dispersed distribution so that

$$\text{sup}(J(\cdot \mid x)) \subset \text{supp}(\pi), \forall x : \pi(x) > 0.$$

  - Monitoring scalar estimands.
  - All these methods are valid under "propriety – proper posterior".