Metropolis Hastings within Gibbs Algorithm

Kyurhi Kim

November 12, 2023

HDMT Lab Seminar

Introduction

Markov Chain Monte Carlo

- *Markov Chain Monte Carlo* (MCMC) generates samples from a Markov chain whose invariant distribution is $f(x)$.
- MCMC was introduced to resolve difficulties when it comes to generating samples from a probability distribution ($f(x)$), such as impossible direct generation or high dimensionality of x or complexity of $f(x)$.
- There are two fundamental mechanisms in MCMC algorithm
  - Simplifies the high dimensionality
  - Involves an accept/reject rule to guarantee invariant distribution of $f(x)$

Metropolis-Hastings

Metropolis-Hastings

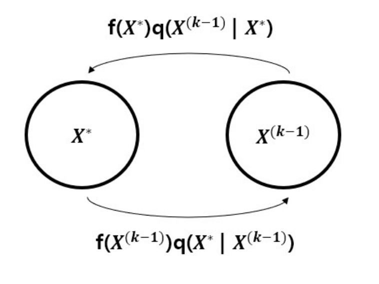
- Metropolis-Hastings (MH) algorithm is one of the most popular MCMC algorithms.
- Basic idea is to simulate Markov chain which has target density as a stationary distribution.
- MH algorithm is defined by two steps
  - Generate a proposal value from a proposal distribution $q(x|x^k)$ given the current value of random variables
  - Accept or Reject the proposed value as a new value with some rule

MH algorithm

Let $f(X)$ be the target density, $X^{(k-1)}$ be a current value, and $q(x|x^{(k)})$ be a proposal density.

- Set up initial value $X^{(0)} = (X_1^{(0)}, ..., X_p^{(0)})$
- For $i = 1, .., M$, repeat:
  1. Draw candidate $X^* \sim q(X|X^{(k-1)})$
  2. Calculate the acceptnace probability $\alpha(X^{(k-1)}, X^*) = min(1, \frac{f(X^*)q(X^{(k-1)}|X^*)}{f(X^{(k-1)})q(X^*|X^{(k-1)})})$
  3. Let $X^{(k)} = \begin{cases} X^* & \text{if} \quad Unif(0,1) \leq \alpha(X^{(k-1)}, X^*) \\ & X^{(k-1)} \text{ otherwise} \end{cases}$
  4. Return the values $X^{(k)} = (X_1^{(k)}, ..., X_p^{(k)})$

Acceptance probability



- Detailed balance condition should be satisfied so that this Markov chain has f as the limiting distribution.
- When the next state is j, let $r_{i,j}$ is the probability of acceptance. Then the Transition probability will be $p_{i,j} = q(i|j)r_{i,j}$, and the restriction for detailed balance condition can be written as $f(X^*)q(x^{(k-1)}|X^*)r_{*,(k-1)} = f(X^{(k-1)})q(x^*|X^{(k-1)})r_{(k-1),*}$

Acceptance probability

- If $f(X^*)q(x^{(k-1)}|X^*) > f(X^{(k-1)})q(x^*|X^{(k-1)})$, then $r_{*,(k-1)}$ should be small and $r_{(k-1),*}$ should be large to satisfy detailed balance condition. Thus, let $r_{*,(k-1)} = 1$ and $r_{(k-1),*} = \frac{1}{\alpha}$. On the other hand, let $r_{*,(k-1)} = \alpha$ and $r_{(k-1),*} = 1$.
- If $\alpha \geq 1$,

$$r_{*,(k-1)} = \frac{1}{\alpha}$$
$$r_{(k-1),*} = 1$$

- If $0 < \alpha < 1$,

$$r_{*,(k-1)} = 1$$
$$r_{(k-1),*} = \alpha$$

Acceptance probability

- Since the next state depends solely on the previous state, $r_{(k-1),*}$ decide whether to accept new value or not, according to the value of α.
- To summarize, If $\alpha \geq 1$,

$$\text{accept } \ X^* \text{ and set } \ X^{(k)} = X^*$$

- If $0 < \alpha < 1$,

$$\begin{gathered} \text{accept } \ X^* \text{ and set } \ X^{(k)} = X^* \text{ with probability } \ \alpha \\ \text{reject } \ X^* \text{ and set } \ X^{(k)} = X^{(k-1)} \text{ with probability } \ 1 - \alpha \end{gathered}$$

Remark

- If the proposal distribution is poorly chosen, it might yields low acceptance rate or too slow movement of Markov chain.
- Also, the choice of proposal distribution is application-dependent, so it may be extremely poor on another even though it works well on one target distribution.

Random walk Metropolis-Hastings

- If the proposals are formed as $X^* = X^{(k-1)} + \epsilon_k$ where $\{\epsilon_k\}$ is a sequence of independent draws from $q(\cdot)$, then the algorithm is called Random walk Metropolis-Hastings.
- Note that $q(X^*|X^{(k-1)}) = q(X^* - X^{(k-1)}) = q(\epsilon_k)$ and $q(X^{(k-1)}|X^*) = q(X^{(k-1)} - X^* = q(-\epsilon_k)$
- So, the acceptance probability is $\alpha = \frac{f(x^{(k-1)} + \epsilon_k)q(-\epsilon_k)}{f(x^{(k-1)})q(\epsilon_k)}$

Independent Metropolis-Hastings

- If the proposals are independent of the previous state, that is $q(X^*|X^{(k-1)}) = q(X^*)$, then the algorithm is called Independent Metropolis-Hastings.
- The the acceptance probability is $\alpha = \frac{f(x^*)q(X^*)}{f(x^{(k-1)})q(X^{(k-1)})}$
- When $f(X) = q(X)$ for any X, the proposal is always accepted since the acceptance probability is 1.
- On the contrary, the more the proposal distribution differs from the target distribution, the lower the acceptance probabilities are, and the more often the proposals are rejected.

Gibbs Sampling

Gibbs Sampling

- Given the target distribution $\pi(\mathbf{x}) = \pi(x_1, ..., x_p)$, Gibbs samplers generates samples for each of the random variable X_i from the full conditional distribution $(X_i|X_1, ..., X_{(i-1)}, X_{(i+1)}, ..., X_p)$.
- it reduces the dimensionality but converges slowly and might suffer from high correlation when X exhibit heavy dependence.
- Gibbs sampling is a special case of Metropolis-Hastings where the proposed moves are always accepted (the acceptance probability is 1).

Gibbs sampling Algorithm: Random-scan

- Set up initial value $X^{(0)} = (X_1^{(0)}, ..., X_p^{(0)})$
- For $i = 1, .., M$, repeat:
    1. Randomly select random variable i from $\{1, ..., p\}$,
    2. Draw a sample $X_i^{(k+1)}$ from $p(X_i | X_{-i}^{(k)})$ where $X_{-i}^{(k)} = (X_1^{(k)}, ..., X_{i-1}^{(k)}, X_{i+1}^{(k)}, ..., X_p^{(k)})$.
    3. Let $X^{(k+1)} = (X_1^{(k)}, ..., X_{i-1}^{(k)}, X_i^{(k)}, X_{i+1}^{(k)}, ..., X_p^{(k)})$.

Gibbs sampling Algorithm: Systemic-scan

- Set up initial value $X^{(0)} = (X_1^{(0)}, ..., X_p^{(0)})$
- For $i = 1, .., M$, repeat:
  1. Draw a sample $X_1^{(k+1)}$ from $p(X_1|X_2^{(k)}, ...X_p^{(k)})$.
  2. Draw a sample $X_2^{(k+1)}$ from $p(X_2|X_1^{(k+1)}, X_3^{(k)}, ...X_p^{(k)})$.

     $\vdots$
  3. Draw a sample $X_p^{(k+1)}$ from $p(X_p|X_1^{(k+1)}, ...X_{p-1}^{(k+1)})$.
  4. Let $X^{(k+1)} = (X_1^{(k+1)}, ..., X_p^{(k+!)})$.

Property regarding acceptance probability

- Gibbs sampler satisfy detailed balance condition by setting the proposed density as a full conditional distribution.
- Thus, proposed value is accepted as a new value always, without the need to correct the transition probability through alpha.
- To prove, suppose two random variable X_1 and X_2.
- Let current value $X^{(k)} = (X_1^{(k)}, X_2^{(k)})$ and proposed value $X^* = (X_1^*, X_2^{(k)})$.
- We know that $q(X^*|X^{(k)}) = p(X_1^*, X_2^{(k)}|X_2^{(k)}) = p(X_1^*|X_2^{(k)})$ and $q(X^{(k)}|X^*) = p(X_1^{(k)}, X_2^{(k)}|X_2^{(k)}) = p(X_1^{(k)}|X_2^{(k)})$.

Property regarding acceptance probability

- Detailed balance equation is as follows:

$$
\begin{aligned}
&p(X^{(k)})q(X^*|X^{(k)}) \\
&= p(X_1^{(k)}, X_2^{(k)})p(X_1^*|X_2^{(k)}) \\
&= p(X_1^{(k)}|X_2^{(k)})p(X_2^{(k)})p(X_1^*|X_2^{(k)}) \\
&= p(X_1^{(k)}|X_2^{(k)})p(X_1^*, X_2^{(k)}) \\
&= q(X^{(k)}|X^*)p(X^*)
\end{aligned}
$$

- Therefore, $\alpha = \frac{q(X^{(k)}|X^*)p(X^*)}{q(X^*|X^{(k)})p(X^{(k)})} = 1$

Metropolis-within-gibbs sampler

- Gibbs sampler needs full conditional distribution, so it is restricted when some components cannot be simulated easily.
- Müller(1991) suggested a compromised Gibbs algorithm - the Metropolis-within-Gibbs sampler.
- For any step of the Gibbs sampler that has difficulty in sampling from full conditional distribution, substitue a MH simulation.

Steps

- For $i = 1, ..., p$, given $(x_1^{(k+1)}, ..., x_{i-1}^{(k+1)}, x_i^{(k)} ...., x_p^{(k)})$:
    1. Generate $x_i^* \sim q_i(x_i | x_1^{(k+1)}, ..., x_{i-1}^{(k+1)}, x_i^{(k)}, ..., x_p^{(k)})$.
    2. Calculate $\alpha = \frac{f_i(x_i^* | x_1^{(k+1)}, ..., x_{i-1}^{(k+1)}, x_{i+1}^{(k)}, ..., x_p^{(k)})}{f_i(x_i^{(k)} | x_1^{(k+1)}, ..., x_{i-1}^{(k+1)}, x_{i+1}^{(k)}, ..., x_p^{(k)})} \times \frac{q_i(x_i^{(k)} | x_1^{(k+1)}, ..., x_{i-1}^{(k+1)}, x_i^*, x_{i+1}^{(k)}, ..., x_p^{(k)})}{q_i(x_i^* | x_1^{(k+1)}, ..., x_{i-1}^{(k+1)}, x_i^{(k)}, x_{i+1}^{(k)}, ..., x_p^{(k)})}$
    3. Set $x_i^{(k+1)} = x_i^*$ with probability $min(1, \alpha)$ and $x_i^{(k+1)} = x_i^{(k)}$ with the remaining probability.
- In this algorithm, the MH step is performed only once at each iteration and $f(x_1, ..., x_p)$ is still admitted as the stationary distribution.

Ex: graph-GPA model

- Chung et al. (2017) used metropolis within Gibbs algorithm with hidden Markov random field architecture to integrate a large number of phenotypes.
- Observed p_{it}: p-value of association testing of SNP $t = 1, .., T$ with phenotype $i = 1, ..., n$
- Transform p_{it} as $y_{it} = \Phi^{-1}(1 - p_{it})$, and Let $e_{it} = 1$ if SNP t is associated with phenotype i and $e_{it} = 0$ otherwise.
- Model the density of y_{it} given the latent status e_{it} by a normal mixture:

$$p(y_{it}|e_{it}, \mu_i, \sigma_i^2) = e_{it} LN(y_{it}; \mu_i, \sigma_i^2) + (1 - e_{it}) N(y_{it}; 0, 1)$$

Ex: graph-GPA model

- To integrate multiple GWAS datasets for genetically related phenotypes, graphical model based on an MRF was suggested.
- Conditional distribution of $\mathbf{e}_t = (e_{1t}, ..., e_{nt})$ is:

$$p(\mathbf{e}_t|\alpha, \beta, G) = C(\alpha, \beta, G) exp(\sum_{i=1}^{n} \alpha_i e_{it} + \sum_{i \sim j} \beta_{ij} e_{it} e_{jt}),$$

where $C(\alpha, \beta, G)$ is an normalizing constant.

Ex: graph-gpa model

- For Bayesian inference, conjugate prior distribution for density of y_{it} were introduced as follows:

$$\mu_i \sim N(\theta_\mu, \tau_\mu^2),\ \sigma_i^2 \sim IG(a_\sigma, b_\sigma)$$

- For coefficients in the MRF model, prior distributions are assumed as follows:

$$\alpha_i \sim N(\theta_\alpha, \tau_\alpha^2),\ \beta_{ij} \sim E(i,j)\Gamma(\beta_{ij}; a_\beta, b_\beta) + \{1 - E(i,j)\}\delta_0(\beta_i j)$$

Ex: graph-gpa model

- Given p-values from GWAS datasets, parameters of the model were estimated based on posterior samples from a MCMC sampler.
- Specifically, Metropolis-Hastings within Gibbs algorithm and reversible jump MCMC was implemented.

The posterior inferences are implemented by the following MCMC steps:

S1. For each i and t, draw $e_{it} \sim \text{Bernoulli}(p_1^*)$ where

$$p_1^* = \frac{\exp\left(\alpha_i + \sum_{j\sim i} \beta_{ij} e_{jt}\right) \cdot p(y_{it} | e_{it} = 1, \mu_i, \sigma_i^2)}{\sum_{e^* \in \{0,1\}} \exp\left(\alpha_i e^* + \sum_{j\sim i} \beta_{ij}\ e^*\ e_{jt}\right) \cdot p(y_{it} | e_{it} = e^*, \mu_i, \sigma_i^2)}.$$

S2. For each i, draw μ_i from its full conditional distribution,

$$\mu_i \sim \mathrm{N}\left(\frac{\sigma_i^2 \theta_\mu + \tau_\mu^2 \sum_{\{t:e_{it}=1\}} \log y_{it}}{\sigma_i^2 + \tau_\mu^2 n_i}, \frac{\sigma_i^2 \tau_\mu^2}{\sigma_i^2 + \tau_\mu^2 n_i}\right)$$

where $n_i = \#\{t : e_{it} = 1\}$.

S3. For each i, draw σ_i^2 from its full conditional distribution,

$$\sigma_i^2 \sim \mathrm{IG}\left(a_\sigma + \frac{n_i}{2}, b_\sigma + \frac{\sum_{\{t:e_{it}=1\}} (\log y_{it} - \mu_i)^2}{2}\right)$$

where $n_i = \#\{t : e_{it} = 1\}$

Ex: graph-GPA model

S4. For each i, update α_i with the Metropolis-Hastings:

1. Propose α_i^q from $q(\alpha_i^q|\alpha_i) = \mathrm{N}(\alpha_i^q; \alpha_i, s_\alpha^2)$. We set $s_\alpha = 0.1$.
2. Update $\alpha_i = \alpha_i^q$ with the acceptance probability

$$\min\left[1, \left\{\prod_{t=1}^{T} \frac{\mathrm{C}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{G})}{\mathrm{C}(\boldsymbol{\alpha}^q, \boldsymbol{\beta}, \boldsymbol{G})} \frac{\exp(\alpha_i^q e_{it})}{\exp(\alpha_i e_{it})}\right\} \frac{p(\alpha_i^q)}{p(\alpha_i)}\right]$$

where $\boldsymbol{\alpha}^q = (\alpha_1, \ldots, \alpha_{i-1}, \alpha_i^q, \alpha_{i+1}, \ldots, \alpha_n)$.

S5. For each (i, j) such that $E(i, j) = 1$, update β_{ij} with the Metropolis-Hastings:

1. Propose β_{ij}^q from the truncated normal proposal distribution bounded above zero, $q(\beta_{ij}^q|\beta_{ij}) = \mathrm{N}_+(\beta_{ij}^q; \beta_{ij}, s_\beta^2)$. We set $s_\beta = 0.1$.
2. Update $\beta_{ij} = \beta_{ij}^q$ with the acceptance probability

$$\min\left[1, \left\{\prod_{t=1}^{T} \frac{\mathrm{C}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{G})}{\mathrm{C}(\boldsymbol{\alpha}, \boldsymbol{\beta}^q, \boldsymbol{G})} \frac{\exp(\beta_{ij}^q e_{it} e_{jt})}{\exp(\beta_{ij} e_{it} e_{jt})}\right\} \frac{p(\beta_{ij}^q|E(i,j))}{p(\beta_{ij}|E(i,j))} \frac{q(\beta_{ij}|\beta_{ij}^q)}{q(\beta_{ij}^q|\beta_{ij})}\right]$$

where $\boldsymbol{\beta}^q = (\beta_{12}, \beta_{13}, \ldots, \beta_{i,j-1}, \beta_{ij}^q, \beta_{i,j+1}, \ldots, \beta_{n-1,n-2}, \beta_{n-1,n})$.

References

References:

- Liang, F., Liu, C., Carroll, R. (2011). Advanced Markov chain Monte Carlo methods: learning from past samples (Vol. 714). John Wiley Sons.
- Chung D, Kim HJ, Zhao H (2017) graph-GPA: A graphical model for prioritizing GWAS results and investigating pleiotropic architecture. PLOS Computational Biology 13(2): e1005388. https://doi.org/10.1371/journal.pcbi.1005388
- https://www.statlect.com/fundamentals-of-statistics/Metropolis-Hastings-algorithm
- https://georglsm.r-forge.r-project.org/site-projects/pdf/Hastings$_{w}$*ithin*$_{G}$*ibbs.pdf*