

Monte Carlo Markov Chain methods

Dhruv Patel
Myntra

March 11, 2020

Markov Chains

- Defined by transition probability $T_m(Z^m, Z^{m+1})$
- For time homogeneous chains T_m is same for all m .

Stationary Distribution

A distribution p is stationary with respect to markov chain defined by transition probability T if

$$p(z) = \sum_{z'} T(z', z) p(z')$$

Detailed Balance

Theorem

Given p , if we choose T such that detailed balance equation is satisfied then p will be stationary distribution for this T .

$$p(z)T(z, z') = p(z')T(z', z)$$

Proof.

$$\begin{aligned}\sum_{z'} p(z')T(z', z) &= \sum_{z'} p(z)T(z, z') && \text{by hypothesis} \\ &= p(z) \sum_{z'} T(z, z') \\ &= p(z)\end{aligned}$$



Metropolis-Hastings Algorithm

- Generalizes Metropolis algorithm
- No need for proposal distribution to be symmetric
- Acceptance probability for proposal of z^* from z^τ

$$A(z^*, z^\tau) = \min \left(1, \frac{\tilde{p}(z^*)q(z^\tau|z^*)}{\tilde{p}(z^\tau)q(z^*|z^\tau)} \right)$$

Where $\tilde{p}(z)$ is the unnormalized density at z and $q(z|z')$ is a proposal distribution at z' .

- Note that for symmetric q , this boils down to Metropolis algorithm
- Transition probability is given by

$$T(z^\tau, z^*) = q(z^*|z^\tau)A(z^*, z^\tau)$$

We will prove that our target distribution p (i.e. normalized version of \tilde{p}) is stationary distribution with respect to transition probability defined in last slide by proving that it satisfies detailed balance condition. (detailed balance \implies stationary distribution)

$$\begin{aligned} p(z)T(z, z') &= p(z)q(z'|z)A(z', z) \\ &= p(z)q(z'|z)\min\left(1, \frac{\tilde{p}(z')q(z|z')}{\tilde{p}(z)q(z'|z)}\right) \\ &= \min\left(p(z)q(z'|z), p(z)q(z'|z)\frac{\tilde{p}(z')q(z|z')}{\tilde{p}(z)q(z'|z)}\right) \\ &= \min(p(z)q(z'|z), p(z')q(z|z')) \\ &= \min(p(z')q(z|z'), p(z)q(z'|z)) \\ &= p(z')q(z|z')\min\left(1, \frac{p(z)q(z'|z)}{p(z')q(z|z')}\right) \\ &= p(z')q(z|z')\min\left(1, \frac{\tilde{p}(z)q(z'|z)}{\tilde{p}(z')q(z|z')}\right) \\ &= p(z')T(z', z) \end{aligned}$$



Gibbs Sampling

Metropolis-Hastings algorithm can be inefficient if proposal distribution is not chosen properly, as seen in demo. However if our target distribution is over m variables and we can compute conditional distribution $p(z_m | z_{-m})$ where z_{-m} denotes all variables other than z_m , Gibbs sampling is efficient algorithm.

Gibbs Sampling

Metropolis-Hastings algorithm can be inefficient if proposal distribution is not chosen properly, as seen in demo. However if our target distribution is over m variables and we can compute conditional distribution $p(z_m|z_{-m})$ where z_{-m} denotes all variables other than z_m , Gibbs sampling is efficient algorithm.

Algorithm

- initialize $z_i \quad \forall i \in [1, 2, \dots, m]$
- for next step $\tau + 1$
 - $z_1^{\tau+1} \sim p(z_1|z_2^\tau, z_3^\tau, \dots, z_m^\tau)$
 - $z_2^{\tau+1} \sim p(z_2|z_1^{\tau+1}, z_3^\tau, \dots, z_m^\tau)$
 - \vdots
 - $z_m^{\tau+1} \sim p(z_m|z_1^{\tau+1}, z_2^{\tau+1}, \dots, z_{m-1}^{\tau+1})$

First we will prove that this transitions leads to stationary distribution.

Proof.

$$\begin{aligned}& \sum_{z'} p(z') T(z', z) \\&= \sum_{z'_i} \sum_{z'_{-i}} T(z', z) p(z'_i | z'_{-i}) p(z'_{-i}) \\&= \sum_{z'_i} T(z', z) p(z'_i | z'_{-i}) p(z'_{-i}) \\&= p(z'_{-i}) \sum_{z'_i} p(z_i | z'_{-i}) p(z'_i | z'_{-i}) \\&= p(z'_{-i}) p(z_i | z'_{-i}) \sum_{z'_i} p(z'_i | z'_{-i}) \\&= p(z_{-i}) p(z_i | z_{-i}) && \text{because } z'_{-i} = z_{-i} \\&= p(z)\end{aligned}$$

Gibbs sampling as a special case of Metropolis-Hastings

Gibbs sampling is a special case of Metropolis-Hastings in which proposal distribution is univariate conditional and no proposals are rejected. Thus is more efficient than simple MH algorithm.

Proof.

Suppose we are about to do k th update in this iteration. Our proposal distribution is $q(z^*|z) = p(z_k^*|z_{-k})$. Let's calculate ratio of MH algorithm.

$$\begin{aligned} & \frac{p(z^*)q(z|z^*)}{p(z)q(z^*|z)} \\ &= \frac{p(z_k^*|z_{-k}^*)p(z_{-k}^*)p(z_k|z_{-k}^*)}{p(z_k|z_{-k})p(z_{-k})p(z_k^*|z_{-k})} \\ &= 1 \end{aligned}$$

as $z_{-k}^* = z_{-k}$

