Lecture8

Tuesday, October 2, 2018 9:26 AM

MARKOV CHAIN MONTE CARLO:

Def: a collection $\{m_i\}$ is called a Markov Chain if the distribution of m_k depends ONLY on the immediate previous state m_{k-1}

(tomorrow depends on the past only through today).

Def: We call the probability of m_k inside a set A starting from m_{k-1} as the transition probability: $\mathbb{P}(m_{k-1}, A)$. Abuse of notation:

$$\mathbb{P}(m_{k-1}, A) \stackrel{def}{=} \int_A \underbrace{\mathbb{P}(m_{k-1}, m)}_{\rightarrow \text{ density}} dm$$

$$\stackrel{def}{=} \int_A \mathbb{P}(m_{k-1}, dm)$$

Notice: $0 \le \mathbb{P}(m_{k-1}, A) \le 1$

$$\mathbb{P}(m_{k-1}, S) = \int_S \mathbb{P}(m_{k-1}, m)\, dm = 1$$

Example: Define Transition matrix

Example : Define transition matrix

$$\underline{P} = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix}$$

j-th row is the transition probability to all nodes starting from node j

- Let $m_{k-1} = L_1$, then the probabily kernel (or density) $\mathbb{P}[m_{k-1} = L_1, m]$

$$\mathbb{P}[m_{k-1} = L_1, m] = \underline{P}(4,:) = [1/3, 1/3, 0, 0, 1/3]$$

$$\parallel$$

$$[0,0,0,1,0] \times \underline{P}$$

Observation : $\mathbb{P}[m_{k-1}, m_k]$ is the joint density between m_{k-1}, m_k.

$$\mathbb{P}[m_{k-1} = L_1, m_k = 1]$$

$$\parallel \quad A = \{1\}$$

$$\int \underline{P}[m_{k-1} = 4, m] \, dm$$

$$\int_A \underline{P}[m_{k-1}=4, m]\, dm$$

$A = \{1\}$

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & \text{otherwise} \end{cases}$$

$$\| $$

$$\int_S \underline{P}[m_{k-1}=4, m]\, \mathbb{1}_A(m)\, dm$$

$\|$ discrete case

$$\sum_{j=1}^{S} \underline{P}[m_{k-1}=4, m]\, \delta(m-1)$$

$$= \underline{P}[4, 1] = 1/3$$

<u>Def</u>: (Invariant distribution)

we say $\mu(dm) = \pi(m)\, dm$ is an invariant distribution of the transition probability $\underline{P}(m_k, dm)$ if

$$\mu(dm) = \pi(m)\, dm = \int_S \underline{P}(m_k, dm)\, \pi(m_k)\, dm_k$$

$\|$ linear operator

$$\langle \underline{P}(m_k, \ddot{dm}), \pi(m_k)\, dm_k \rangle$$

side note

<u>Def</u>: (Reversibility = detailed balance)

A Marko chain $\{m_i\}$ is reversible w.r.t. $\pi(m)$ if

$$\pi(m)\, \underline{P}(m,p) = \pi(p)\, \underline{P}(p,m)$$

$$\forall\ m, p \in \{m_i\}$$

Digest:
$$\int_S \pi(m)\, \underline{P}(m,p)\, dm = \int_S \pi(p)\, \underline{P}(p,m)\, dm$$

$$= \pi(p) \int_S \underline{P}(p,m)\, dm = \pi(p)$$

$\Rightarrow$ if $\{m_i\}$ is reversible w.r.t. $\pi(m)$
$\Rightarrow$ $\pi(m)$ is an invariant distribution of $\{m_i\}$.

- Reversibility, roughly speaking, means that the likelihood of moving from $m \to p$ is the same as the likelihood of moving from $p \to m$.

Proposition: if $\{m_i\}$ is reversible w.r.t. $\pi(m)$ then $\pi(m)$ is an invariant distribution

$$\pi(m)\, dm = \int \underline{P}(p,m)\, \pi(p)\, dp\, dm$$

$$\pi(m)\,dm = \int \underline{P}(p, m)\,\pi(p)\,dp\,dm$$

Markov chain Monte Carlo (MCMC)

- In general it is hard, if not impossible, to generate i.i.d. samples from $\pi(m)$
- MCMC method is a universal approach to generate **NON** i.i.d samples from $\pi(m)$ while still preserving the LLN-like result.
- MCMC methods generates reversible (w.r.t. π) Markov chain

Algorithm: Metropolis - Hasting MCMC method.

Choose initial state m_0

For $k = 0, \ldots$ do

1) Draw a sample p from a Given/chosen proposal density $q(m_k, p)$
2) Compute $\pi(p)$, $\pi(m_k)$
   $q(m_k, p)$, $q(p, m_k)$

3) Compute the acceptance probability
(Metropolization)

$$\alpha(m_k, p) = \min\left\{1, \frac{\pi(p)\, q(p, m_k)}{\pi(m_k)\, q(m_k, p)}\right\}$$

4) * accept p with probability $\alpha(m_k, p)$
set $m_{k+1} = p$
* otherwise reject p: set $m_{k+1} = m_k$

Question: is $\pi(m)$ an invariant distribution of $\{m_0, m_1, \dots\}$

- we have seen that if the reversibility holds then then $\pi(m)$ is an invariant distribution. What remains is to show that MCMC algorithm generate transition probability that satisfies the detailed balance equation

Prop: the Markov chain generated from the above algorithm is reversible w.r.t $\pi(m)$.

Proof: we need to identify $P(m_k, m_{k+1})$ and then show

$$\pi(m_k)\, P(m_k, m_{k+1}) \overset{??}{=} \pi(m_{k+1})\, P(m_{k+1}, m_k)$$

Two cases:

1) Assume p is accepted: $m_{k+1} = m_k$

2) p is rejected: $m_{k+1} = m_k$

Let's work with the first case: $\mathbb{P}(m_k, m_{k+1})$ is joint probability of m_k, $m_{k+1} = p$ where $p = m_{k+1}$ is drawn from $q(m_k, p)$ and $p = m_{k+1}$ is accepted.

- define B as the acceptance event:

$$\mathbb{P}(m_k, p) = \pi(B, p)$$

$$= \pi(B \mid p)\, \pi(p)$$

$$\pi(p) = q(m_k, p)$$

$$\pi(B \mid p) = \alpha(m_k, p)$$

$$\alpha(m_k, p)\; q(m_k, p)$$

Question: $\pi(m_k)\, \mathbb{P}(m_k, p) \overset{??}{=} \pi(p)\, \mathbb{P}(p, m_k)$

$$\pi(m_k)\, q(m_k, p) \min\left\{1, \frac{\pi(p)\, q(p, m_k)}{\pi(m_k)\, q(m_k, p)}\right\}$$

$$\pi(p)\, q(p, m_k) \times \min\left\{1, \frac{\pi(m_k)\, q(m_k, p)}{\pi(p)\, q(p, m_k)}\right\}$$

we have $\pi(m_k)\, \mathbb{P}(m_k, p)$

$$\pi(m_k)\,\alpha(m_k,p)\,q(m_k,p) \quad \| \quad \underline{P}(m_k,p) = \alpha(m_k,p)\,q(m_k,p)$$

$$\| \text{ def } \alpha$$

$$\pi(m_k)\,q(m_k,p)\,\min\left\{1, \frac{\pi(p)\,q(p,m_k)}{\pi(m_k)\,q(m_k,p)}\right\}$$

$$a\min\{c,d\} \quad \| = \min\{ac, ad\}$$

$$\min\{\pi(m_k)\,q(m_k,p),\ \pi(p)\,q(p,m_k)\}$$

$$\| \quad \min\{ac,ad\} = a\min\{c,d\}$$

$$\min\left\{\frac{\pi(m_k)\,q(m_k,p)}{\pi(p)\,q(p,m_k)}, 1\right\}\pi(p)\,q(p,m_k)$$

$$\| \text{ def } \alpha$$

$$\alpha(p,m_k)\,\pi(p)\,q(p,m_k)$$

$$\| \text{ def transition probability } \underline{P}(\cdot,\cdot)$$

$$\pi(p)\,\underline{P}(p,m_k)$$

$\Rightarrow$ the resulting Markov chain is reversible w.r.t $\pi(m)$

$\Downarrow$

$\pi(m)$ is AN invariant measure for $\{m_i\}$

??? $\Downarrow$ understood

$\{m_i\}$ is eventually distributed by $\pi(m)$

Is $\{m_i\}$ distributed by $\pi(m)$?

Def: (Irreducibility). If

1) $\pi(m)$ is finite everywhere.

2) $q(\cdot,\cdot)$ is positive and continuous

then $\{m_i\}$ is