Theory for Markov Chain Monte Carlo

Key concepts (for discrete state space Markov Chains)

- Homogeneous Markov Chain: The conditional probabilities do not depend on the time index.

$$P(X_t = j \mid X_{t-1} = i) = P(X_{t+m} = j \mid X_{t+m-1} = i)$$

for all integer m.

- Transition probability (or Kernel) of homogeneous Markov Chain:

$$K(i,j) = P(X_t = j \mid X_{t-1} = i).$$

- Note that we can express

$$P(X_{t+1} = j) = \sum_i P(X_t = i) K_{ij} \tag{1}$$

for discrete state space Markov Chains.

Key concepts (Cont'd)

- A distribution P of X_t is said to be invariant or stationary for a Markov Kernel K if $PK = P$. That is,

$$P(X_t = j) = \sum_i K_{ij} P(X_t = i)$$

- Note that the Kernel matrix for a stationary process is idempotent.
- A stationary stochastic process is said to be reversible if the statistics of the time-reversed version of the process match those of the process in the forward distribution, so that reversing time makes no discernible difference to the sequence of distributions which are obtained, that is the distribution of any collection of future states given any past history must match the conditional distribution of the past conditional upon the future being the reversal of that history.

Key Result

Theorem

If a Markov Kernel satisfies the following condition for some distribution P

$$\forall x, y \in E : P_x K_{xy} = P_y K_{yx} \tag{2}$$

then:

1. *P is the invariant distribution of the chain.*
2. *The chain is reversible with respect to P.*

Condition (2) means that

$$P(X_t = x)P(X_{t+1} = y \mid X_t = x) = P(X_t = y)P(X_{t+1} = x \mid X_t = y)$$

and it is called the detailed balance condition.

Theoretical Properties for M-H algorithm

Lemma

The Metropolis-Hastings kernel satisfies the detailed balance condition

$$K\left(X^{(t-1)}, X^{(t)}\right) f\left(X^{(t-1)}\right) = K\left(X^{(t)}, X^{(t-1)}\right) f\left(X^{(t)}\right).$$

Thus, $f(X)$ is the invariant distribution of the Markov Chain $(X^{(0)}, X^{(1)}, \cdots)$. Furthermore, the Markov chain is reversible.

Ergodic

- X: generic random vector with density $f(X)$
- $f(X)$: difficult to simulate directly
- Goal: construct a Markov chain $\left\{X^{(t)}; t = 1, 2 \cdots\right\}$ with f as its stationary distribution,

  $$P\left(X^{(t)}\right) \to f \quad \text{as } t \to \infty$$

  or

  $$\frac{1}{N}\sum_{t=1}^{N} h\left(X^{(t)}\right) \to E_f\left[h(X)\right] = \int h(x) f(x)\, dx \tag{3}$$

  as $N \to \infty$.
- A Markov chain that satisfies (3) is called ergodic.

Main Properties: An ergodic theorem

Theorem

If the Markov chain generated by the Metropolis-Hastings algorithm is irreducible, then for any integrable function h

$$\lim_n \frac{1}{n} \sum_{t=1}^{n} h\left(X^{(t)}\right) = E_f\{h(X)\}$$

with probability one, for almost every starting value $X^{(0)}$.

Recall: The Markov chain is irreducible if $q(X \mid X^{(t-1)}) > 0$ for all $X, X^{(t-1)} \in \text{supp}(f)$: every state can be reached in a single step.