Bayesian

Jian Qian

2017 年 10 月 25 日

Definition 1. *Markov Chain* $(X_t)_t$ *on* $\mathcal{X}$ *with transition kernel* K*:*
$P(X_{t+1} \in A|X_t = x) = P(X_1 \in A|X_0 = x) = K(x, A)$ *for* $\forall x, \forall t, \forall A$

Definition 2. *Stationary distribution* π*:*

$$\pi(A) = \int_{\mathcal{X}} K(s, A)\pi(ds)$$
$$\Leftrightarrow \pi(y) = \int_{\mathcal{X}} K(x, y)\pi(x)dx$$

Because the stationary condition is hard to verify, so we introduce a more verifiable condition reversibility.
Meaning if we want to prove a distribution is stationary, we first try with reversible.
And moreover, this condition is intuition, so if we have to choose one to remember, I guess, this condition should be the one to go for.

Definition 3. *Reversibility:*

$$\int_{s\in A}\int_{x\in B} K(s, dx)\pi(ds) = \int_{x\in B}\int_{s\in A} K(x, ds)\pi(dx)$$

Proposition 4. K *reversible for* $\pi \Rightarrow \pi$ *is stationary*

Definition 5. *Irreducibility:*

$$\pi(A) > 0 \Rightarrow P(x\ s.t.\ \exists m\ X_m \in A\ given\ X_0 = x) > 0$$

Definition 6. *Aperiodic:* $\bar{\exists} d$ *s.t.* $\exists$ *disjoint subsets* $\mathcal{X}_1, ..., \mathcal{X}_d$ *such that*

$$K(x, \mathcal{X}_{(i+1) \mod d}) = 1 \qquad \forall i, \forall x \in \mathcal{X}_{i \mod d}$$

Theorem 7.

$$d(K^n(x,\cdot), \pi(\cdot))_{TV} \to 0 \quad a.s.$$

Theorem 8.

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} g(X_i) \to \mathbb{E}_\pi[g]$$

Metropolis Testing algo
choose $Q(\theta, \cdot)$ a kernel with density $q(\theta, \theta')$

- generate $\theta^* \sim Q(\theta_t, \cdot)$
- compute "acceptance ratio" $\alpha(\theta_t, \theta^*)$ with

$$\alpha(s,t) = \min(1, \frac{\widetilde{\pi}(t)q(t,s)}{\widetilde{\pi}(s)q(s,t)})$$

An interesting point about the construction is the idea behind the ratio: If going back is easy, then we don't worry moving forward, while when it is not easy, then we consider for sometime and decide that we want move forward with some possibility. This is kind of trust system, because we want to be stable, so if we are guaranteed that we will be back anyway, we go without hesitation, otherwise we move will the degree of trust.
Well, there is a proof.
Then we find a clever way of choosing Q.
Examples

- $q(\theta, s) = \mathcal{N}(s|\mu = \theta, s^2 = \epsilon)$ random walk MH(Metropolis-Hasting). If $\epsilon << 1$, then the moves will be too cautious. However if $\epsilon >> 1$, then it gets stuck for a while and jump suddenly.

Classical Q's

- independent M.H.: $Q(x, \cdot)$ does not depend on x.
  - target: Beta on [0,1]
  - proposal: $Q = \mathcal{U}_{[0,1]}$
- $\theta^* = \theta_t + \epsilon$ where $\epsilon \sim f$, f is any centered density, thus $q(\theta_t, s) = f(s - \theta_t)$, if f is symmetric, then we have:

$$\alpha(\theta_t, \theta^*) = \min(1, \frac{\widetilde{\pi}(\theta^*)}{\widetilde{\pi}(\theta_t)})$$

Rule of thumb
acceptance rate $\simeq 1/4$ (high dimension d=10),1/2 (low dimension d=2)
Convergence diagnostics
one cannot "accept" the null hypothesis $H_0 : (\theta_t)_t$ is stationary
many test, none of which is absolute
R package: coda
main idea:run several chains in parallel
same behavior of different chains with different starting values
after some Burning period.