CS258: Information Theory

Fan Cheng

Spring, 2018. chengfan@sjtu.edu.cn

Metropolis Algorithm

- Start with any initial value θ_0 satisfying $f(\theta_0) > 0$
- Using current θ value, sample a candidate point θ^* from some jumping distribution $q(\theta_1, \theta_2)$, which is the probability of returning a value of θ_2 given a previous value of θ_1. This distribution is also referred to as the proposal or candidate-generating distribution. The only restriction on the jump density in the Metropolis algorithm is that it is symmetric, i.e., $q(\theta_1, \theta_2) = q(\theta_2, \theta_1)$.
- Given the candidate point θ^*, calculate the ratio of the density at the candidate (θ^*) and current (θ_{t-1}) points,

$$\alpha = \frac{p(\theta^*)}{p(\theta_{t-1})} = \frac{f(\theta^*)}{f(\theta_{t-1})}$$

- If the jump increases the density ($\alpha > 1$), accept the candidate point (set $\theta_t = \theta^*$) and return to step 2. If the jump decreases the density ($\alpha < 1$), then with probability α accept the candidate point, else reject it and return to step 2.

Metropolis Algorithm

Q_1: show it is a special case of Metropolis-Hasting algorithms
Q_2: Define

$$Pr(x \to y) = q(x,y)\alpha(x,y),$$

show that

$$Pr(x \to y)p(x) = Pr(y \to x)p(y)$$

Discuss: $q(x,y)p(x) => < q(y,x)p(y)$

Gibbs is a special case of MH

$$\alpha(x_n^{cand}, x_{-n}^{(i-1)} | x_n^{(i-1)}, x_{-n}^{(i-1)}) = 1$$

Proof.

$$\begin{aligned}
&\min\{1, \frac{q(x_n^{(i-1)}, x_{-n}^{(i-1)} | x_n^{cand}, x_{-n}^{(i-1)}) p(x_n^{cand}, x_{-n}^{(i-1)})}{q(x_n^{cand}, x_{-n}^{(i-1)} | x_n^{(i-1)}, x_{-n}^{(i-1)}) p(x_n^{(i-1)}, x_{-n}^{(i-1)})}\} && (1) \\
=&\min\{1, \frac{p(x_n^{(i-1)} | x_{-n}^{(i-1)}) p(x_n^{cand}, x_{-n}^{(i-1)})}{p(x_n^{cand} | x_{-n}^{(i-1)}) p(x_n^{(i-1)}, x_{-n}^{(i-1)})}\} && (2) \\
=&\min\{1, \frac{p(x_n^{(i-1)} | x_{-n}^{(i-1)}) p(x_n^{cand} | x_{-n}^{(i-1)}) p(x_{-n}^{(i-1)})}{p(x_n^{cand} | x_{-n}^{(i-1)}) p(x_n^{(i-1)} | x_{-n}^{(i-1)}) p(x_{-n}^{(i-1)})}\} && (3) \\
=&1 && (4)
\end{aligned}$$

□