Markov chain Monte Carlo

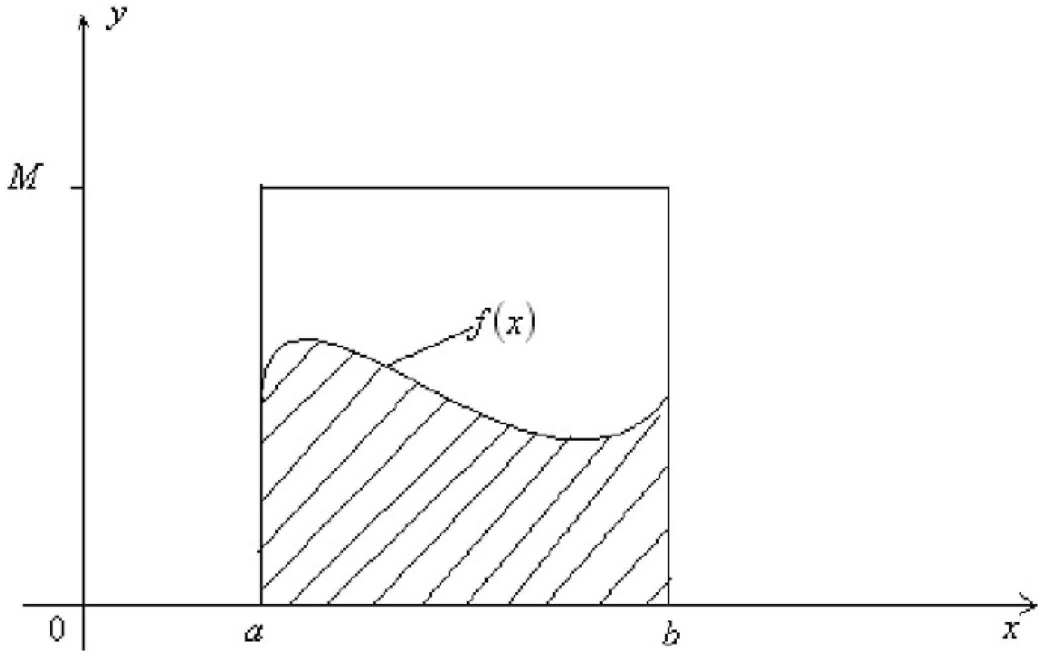
shenchuyun

CONTENTS

- Monte Carlo method
  - Importance Sampling
  - accept-reject sample method
- Markov Chain
- Markov Chain Monte Carlo
  - MH sampling
  - Gibbs sampling

Monte Carlo method

- 假设概率分布的定义已知，通过抽样获得概率分布的随机样本，再通过得到的样本对概率分布的特征进行分析。
- $\theta = \int_a^b f(x)dx$原函数难求
- 随机采样n个点
- $\theta = \int_a^b f(x)dx = \frac{b-a}{n}\sum_{i=0}^{n-1} f(x_i)$



Monte Carlo method

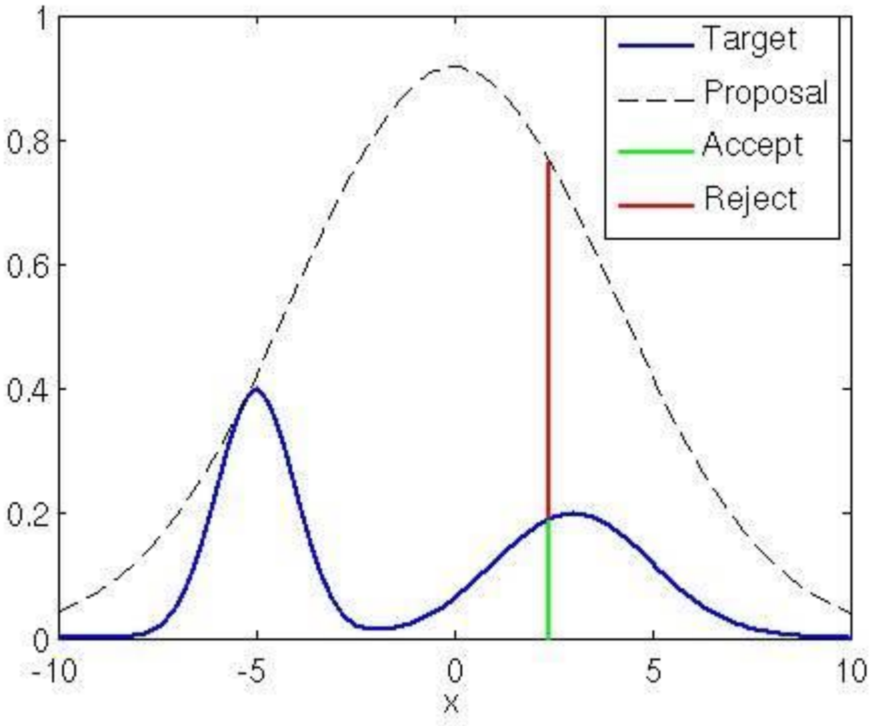
- 求数学期望
- 在p(x)上随机采样n个样本点，然后计算均值$\hat{f}_n = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$
- 作为数学期望$E_{p(x)}[f(x)]$
- $E_{p(x)}[f(x)] = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$, n→ ∞
- 求积分
- $\int_D h(x)\,dx = \int_D f(x)p(x)dx = E_{p(x)}[f(x)]$ 其中$f(x) = \frac{h(x)}{p(x)}$
- $\int_D h(x)\,dx \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$

Monte Carlo method (Importance Sampling)

- $\mu = \int_D f(x)p(x)dx = \int_D \frac{f(x)p(x)}{q(x)} q(x)dx = E_q\left(\frac{f(X)p(X)}{q(X)}\right)$
- $p(x)$不好直接采样，在q上采样n个点
- $\mu \approx \frac{1}{n}\sum_{i=1}^{n} \frac{f(x_i)p(x_i)}{q(x_i)}$

Monte Carlo method（accept-reject sample method）

- P(x)不可以直接抽样，找一个可以直接抽样的分布（proposal distribution）
- 重复以下步骤，直到采样到m个样本:
  - 从q（x）采样一个x^*，满足$c * q(x) \geq p(x)$
  - 从$U(0,1)$均匀分布中得到u
  - if u $\leq \frac{p(x^*)}{cq(x^*)}$,则将$x^*$作为抽样结果
  - else 拒绝该样本



Markov Chain

- $P(X_t|X_0, X_1, \dots, X_{t-1}) = P(X_t|X_{t-1})$
- $P(X_t|X_{t-1})$称为马尔科夫链的转移概率分布
- 离散状态马尔可夫转移矩阵P=$\begin{bmatrix} P_{11} & P_{12} & P_{13} & \dots \\ P_{21} & P_{22} & P_{23} & \dots \\ P_{31} & P_{32} & P_{33} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$
- X在时刻t的概率分布$\pi(t) = \begin{bmatrix} \pi_1(t) \\ \pi_2(t) \\ \dots \end{bmatrix}$
- $\pi_i(t) = P(X_t = i), i = 1,2, \dots$

Markov Chain

- X在t时刻的状态分布可以由时刻(t-1)的状态分布和转移矩阵决定
- $\pi(t) = P\pi(t-1)$
- 平稳分布$\pi = P\pi$
- 遍历定理：直观解释：满足一定条件的马尔科夫链，当t趋于无穷，马尔科夫链的状态分布趋于平稳分布

Markov Chain Monte Carlo

- 基本思想：我们的目标分布是p(x)，定义一个马尔科夫链$X = \{X_0, X_1, X_2, \ldots, X_t, \ldots\}$，使其平稳分布就是抽样的目标p(x)，然后在马尔科夫链上随机游走，每个时刻就可以得到一个样本
- **Detailed Balance Condition：** $\pi(i)P_{ij} = \pi(j)P_{ji}\ for\ all\ i, j$
- 充分性证明：
  - $\sum_i \pi(i)P_{ij} = \sum_i \pi(j)P_{ji} = \pi(j)\sum_i P_{ji} = \pi(j)$
  - $\pi P = \pi$

Metropolis-Hastings

- $p(x^*|x) = q(x^*|x)\alpha(x^*|x)(proposal\ distribution) * (acceptance\ distribution)$ $q(x^*|x)$是容易抽样的分布
- $\alpha(x^*|x) = \min\left\{1, \frac{\pi(x^*)q(x^*|x)}{\pi(x)q(x|x^*)}\right\}$
- *MH*
  - $u \sim U(0,1)$
  - $x^* \sim q\left(x^* \middle| x^i\right)$
  - $if\ u < \alpha(x^*|x) \quad x^{i+1} = x^*$
  - $else\ x^{i+1} = x^i$

Gibbs sampling

- Given a starting sample $(x_1, y_1, z_1)^T$
- You want to sample
  - $(x_2, y_2, z_2)^T, (x_3, y_3, z_3)^T, \dots, (x_N, y_N, z_N)^T \sim P(x, y, z)$
- Sampling:
  - $x_2 \sim P(x|y_1, z_1)$
  - $y_2 \sim P(y|x_2, z_1)$
  - $z_2 \sim P(z|x_2, y_2)$
  - $\cdots$

Gibbs sampling

- A special case of M-H
  - Let $\mathbf{x} = x_1, \cdots, x_D$
  - When sampling k^{th} component, $q_k(\boldsymbol{x}^*|\boldsymbol{x}) = \pi(\boldsymbol{x}_k^*|\boldsymbol{x}_{-k})$
  - When sampling k^{th} component, $\boldsymbol{x}_{-k}^* = \boldsymbol{x}_{-\mathrm{k}}$

$$\frac{\pi(\mathbf{x}^*)q(\mathbf{x}|\mathbf{x}^*)}{\pi(\mathbf{x})q(\mathbf{x}^*|\mathbf{x})} = \frac{\pi(\mathbf{x}^*)\pi(x_k|\mathbf{x}^*{}_{-k})}{\pi(\mathbf{x})\pi(x_k^*|\mathbf{x}_{-k})} = \frac{\pi(x_k^*|\mathbf{x}_{-k}^*)\pi(x_k|\mathbf{x}^*{}_{-k})}{\pi(x_k|\mathbf{x}_{-k})\pi(x_k^*|\mathbf{x}_{-k})} = 1$$