

Lecture 15 Metropolis algorithm and SSA

Weinan E^{1,2} and Tiejun Li²

¹Department of Mathematics,
Princeton University,
weinan@princeton.edu

²School of Mathematical Sciences,
Peking University,
tieli@pku.edu.cn
No.1 Science Building, 1575

Outline

Metropolis algorithm

Stochastic Simulation Algorithm (SSA)

Metropolis算法

- ▶ 1953年,N. Metropolis等人提出了针对一类特殊形式的高维积分的强有力算法, 由于它能处理的是典型的统计物理中的正则系综平均型积分 (或求和),

$$\langle H(\sigma) \rangle = \sum_{\sigma} H(\sigma) \frac{\exp\{-\beta H(\sigma)\}}{Z_M}, \quad (1)$$

因而具有广泛的应用,后人称为**Metropolis**算法. 这一算法在2000年被评为20世纪十大算法之一.

基本思想

- ▶ 对于形如上式的求和(或积分),自然的想法是采用Monte Carlo方法:

$$\langle H(\sigma) \rangle \approx \frac{1}{N} \sum_{i=1}^N H(\sigma_i)$$

其中 $\{\sigma_i\}_{i=1}^N$ 服从*i.i.d.*概率密度 $\frac{1}{Z_M} \exp\{-\beta H(\sigma)\}$.

- ▶ 问题在于如何生成服从这个概率分布的随机变量. 前面讲到的若干方法都难以奏效!
- ▶ **Metropolis**方法的妙处在于通过一个“迭代法”自发的产生出服从这样分布的随机变量. 这好比我们要求一个非线性方程 $f(x) = 0$ 的根, 一般的直接法是难以实现的, 但是如果采用一个合适的迭代法 $x_{k+1} = g(x_k)$, 使得最后的不动点 x^* 是方程的根, 我们就可以得到非线性方程求根的方法了.

基本思想

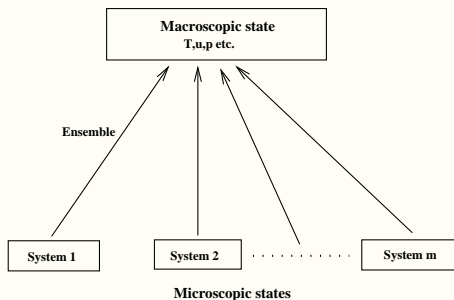
- ▶ Metropolis算法也称为马尔科夫链Monte Carlo方法或动态Monte Carlo方法(Dynamic Monte Carlo). 它建立起一个马氏链, 适当的定义转移概率矩阵 P , 使得概率密度 $\frac{1}{Z_M} e^{-\beta H(\sigma)}$ 恰为其唯一的平稳分布.
- ▶ 如果将 $\frac{1}{Z_M} e^{-\beta H(\sigma)}$ 按不同 σ 排成一个行向量 π , 则 P 矩阵的选取应使得

$$\pi P = \pi,$$

并且对迭代格式 $\nu_{n+1} = \nu_n P$ 应满足与不动点迭代类似的压缩映射性质. 这样从任意的初分布 ν_0 出发产生该马氏链的一个时间序列样本, 经过充分长的时间得到的样本序列应该恰好服从想要的分布 π , 这就是Metropolis算法的巧妙所在!

物理直观

- 假设所有可能的微观态 σ 的数目为 n . 在统计物理中,系综平均表明我们所看到的宏观态实际上是微观状态按一定概率平均的结果.例如我们假想一个容器中有大量分子,并且达到了平衡态,此时我们能测出气体的温度,器壁所受的压力等等.统计物理认为:这一个宏观系统实际上对应了大量的微观系统,它的性质是这些大量系统相应量的统计平均,如下图所示:



物理直观

- ▶ 虽然系统1直至系统m是彼此独立各自发展的系统, 每个系统在任一时刻对应于一个微观态. 但在任何一个时刻, 各种微观状态整体呈现的概率分布是不变的, 这即是平衡态统计系综的观点.
- ▶ **Metropolis**算法对这一问题采取了另一个观点: 既然宏观的系统物理量 T, p, ρ 等随时间不变, 而系统处于一个动态平衡, 这表明时间的平均应该等于系综的平均. 即

$$\langle H(\sigma) \rangle \approx \frac{1}{N} \sum_{i=1}^N H(\sigma_i) \quad (2)$$

- ▶ 如果我们能够找到合适的分子碰撞规则产生状态序列 $\{\sigma_i\}_{i=1}^N$, 就可以在计算中重复这一物理过程, 从而找到求平均的另一种方法. 这种碰撞规则的选取正是前面所述的马氏链的转移概率阵的选取.

马氏链的构造

- ▶ 为了使得我们要求的概率密度

$$\pi(\sigma) = \frac{1}{Z_M} e^{-\beta H(\sigma)}$$

为马氏链的不变分布,自然的马氏链的构造,即 P 矩阵的构造要求满足

$$\pi P = \pi,$$

- ▶ 假设 P 为 $N \times N$ 的矩阵, $\pi P = \pi$ 提供了 N 个方程,加上 P 矩阵行和为 $\mathbf{1}$ 共 N 个方程,而 P 有 N^2 个未知数,尚有非常大的自由度定出 P ,即满足这个条件的马氏链有非常多,我们只需要找出一个即可.

细致平衡条件

- ▶ 定义 如果马氏链满足

$$\pi(\sigma)P(\sigma \rightarrow \sigma') = \pi(\sigma')P(\sigma' \rightarrow \sigma),$$

则称该马链满足细致平衡条件或称该马链为可逆的.

- ▶ 可以证明, 细致平衡条件将推出整体平衡条件

$$\pi P = \pi$$

- ▶ Perron-Frobenius定理告诉我们, 一个满足细致平衡条件的“不可约”马氏链 P , π 是其唯一的不变分布.

细致平衡条件

- ▶ 由细致平衡条件我们得到

$$\frac{P(\sigma \rightarrow \sigma')}{P(\sigma' \rightarrow \sigma)} = \frac{\pi(\sigma')}{\pi(\sigma)} = \exp(-\beta\Delta H),$$

这里

$$\Delta H = H(\sigma') - H(\sigma)$$

为能量差。

- ▶ 由此可以得到确定确定 P 矩阵的一个策略。即当 $\Delta H > 0$ 时，取

$$P(\sigma' \rightarrow \sigma) = 1, \quad P(\sigma \rightarrow \sigma') = \exp(-\beta\Delta H)$$

当 $\Delta H \leq 0$ 时，取

$$P(\sigma \rightarrow \sigma') = 1, \quad P(\sigma' \rightarrow \sigma) = \exp(-\beta\Delta H)$$

此即为MCMC方法的Metropolis策略。

预选过程

- ▶ 上面策略似乎已解决MCMC方法的计算问题，但是其似是而非！注意到

$$P(\sigma' \rightarrow \sigma) = 1, \quad P(\sigma \rightarrow \sigma') = \exp(-\beta\Delta H)$$

使得 P 矩阵的行和一定比1大！这是不可能的。

- ▶ 严格的表述Metropolis算法还需要考虑所谓预选过程。即真正的转移概率是

$$P(\sigma' \rightarrow \sigma) = T(\sigma' \rightarrow \sigma) \times Q(\sigma' \rightarrow \sigma)$$

这里的 Q 是前面的取值为1或 $\exp(-\beta\Delta H)$ 的矩阵。 T 称为预选过程，具体描述如下。

预选过程

- ▶ 所谓预选，即是从当前状态出发选择一个新状态，但是并不是马上从旧状态跳转至新状态，而且在进行决策步骤。我们总是要求预选状态具有对称性，即从A状态选到B状态的概率应该等于从B状态选到A状态的概率。
- ▶ 以一维Ising模型为例，从当前状态 $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ 出发有两种自然可能的预选：
 1. 等可能预选：即对剩下的 $2^n - 1$ 种状态进行同等可能预选，这个预选显然是对称的；
 2. 单点反转预选：即从 σ_k , $k \in \{1, 2, \dots, n\}$ 中随机挑选一个位置，将其状态反转。这个预选显然也是对称的；

Metropolis算法

- 我们以一维Ising模型为例，表述Metropolis算法。

假设当前第 m 步状态 $\sigma^m = (\sigma_1, \sigma_2, \dots, \sigma_n)$ ，这里 $\sigma_i = \pm 1$ 。目标是如何产生下一个状态 σ^{m+1} 。

1. 单点反转预选：

产生均匀分布随机数 $k \in \{1, 2, \dots, n\}$ ，将 σ_k 从1变成-1(或者从-1变成1)，得到新的预选态 σ' 。

2. 计算 $\Delta H = H(\sigma') - H(\sigma^m)$ ，以及

$$R = \min(1, \exp(-\beta\Delta H));$$

3. 生成一均匀分布随机数 $r \sim \mathcal{U}[0, 1]$ ；
4. 如果 $r \leq R$ ，则 $\sigma^{m+1} = \sigma'$ ；否则， $\sigma^{m+1} = \sigma^m$ ，转步1。

Metropolis算法

- ▶ 如果 $\Delta H < 0$, 即预选态 σ' 的能量比旧状态 σ^m 能量低, 于是接受 σ' 作为新的状态; 如果 $\Delta H > 0$, 即预选态 σ' 的能量比旧状态 σ^m 能量高, 则以一定概率接受状态 σ' . 以上操作表明, **Metropolis**算法中状态序列的产生总是倾向于从能量较高的状态跳到能量较低的状态, 即能量较低的状态以更大的几率出现, 而从能量较低的状态也有一定的概率跑到能量较高的状态, 这显然符合概率密度 $\frac{1}{Z} \exp(-\beta H(\sigma))$ 的直观认识.

Metropolis算法

- ▶ 根据上述Metropolis算法从任何一个初始状态出发就可以生成随机状态序列

$$\sigma^1, \sigma^2, \dots, \sigma^m, \dots$$

- ▶ 进行时间平均（即经验平均）

$$\langle H \rangle \approx \frac{1}{N} \sum_{i=1}^N H(\sigma^i)$$

在概率论中有所谓遍历定理(ergodic theorem)保证其收敛性，这一点是与前面简单的大数定律是不一样的。

Outline

Metropolis algorithm

Stochastic Simulation Algorithm (SSA)

Stochastic simulations

- ▶ Biological network

Suppose there are N species of molecules S_i , $i = 1, \dots, N$, and M reaction channels R_j , $j = 1, \dots, M$. x_i is the number of molecules of species S_i . Then the state of the system is given by

$$\mathbf{x} = (x_1, x_2, \dots, x_N).$$

Each reaction R_j is characterized by a rate function $a_j(\mathbf{x})$ and a vector ν_j that describes the change of state due to reaction (after the j -th reaction, $\mathbf{x} \rightarrow \mathbf{x} + \nu_j$). In shorthand denote

$$R_j = (a_j, \nu_j)$$

How to simulate this biological process?

SSA

- ▶ The mathematical formulation of this SSA is based on the so-called **chemical master equation**:

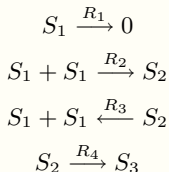
$$\frac{\partial}{\partial t} P(x, t | x_0, t_0) = \sum_{j=1}^M \left[a_j(x - \nu_j) P(x - \nu_j, t | x_0, t_0) - a_j(x) P(x, t | x_0, t_0) \right]$$

where $p(x, t | x_0, t_0)$ is the probability that random variable $X_t = x$ given that $X_{t_0} = x_0$ ($t \geq t_0$).

- ▶ This equations is easily understood from the transition of probability.

SSA: Example

- Consider the following chemical reactions:



The propensity functions are given by

$$a_1 = x_1, \quad a_2 = 5x_1(1 - x_1), \quad a_3 = 1000x_2, \quad a_4 = 0.1x_2$$

and

$$\nu_1 = (-1, 0, 0), \quad \nu_2 = (-2, 1, 0), \quad \nu_3 = (2, -1, 0), \quad \nu_4 = (-1, 0, 1).$$

Initial state

$$X_1(0) = 400, \quad X_2(0) = 798, \quad X_3(0) = 0,$$

i.e. $\mathbf{x}(0) = (400, 798, 0)$ and final time $T = 0.2$

SSA

- ▶ The direct simulation method of SSA is as follows

Suppose the current state is x , define

$$a_0(x) = \sum_{j=1}^M a_j(x)$$

1. Draw two independent samples r_1 and r_2 of $\mathcal{U}[0, 1]$;
2. Take

$$\tau = \frac{1}{a_0(x)} \ln\left(\frac{1}{r_1}\right),$$

3. Take

$$j = \text{the smallest integer satisfying } \sum_{k=1}^j a_k(x) > r_2 a_0(x)$$

4. Then the reaction R_j happens and take $x = x + \nu_j$.

SSA

- ▶ The essence of SSA is that the waiting time of the j -th reaction takes the form

$$p(\tau, j|x, t) = a_j(x) \exp(-a_0(x)\tau)$$

- ▶ Decompose $p(\tau, j|x, t)$ as

$$p(\tau, j|x, t) = p_1(\tau|x, t)p_2(j|\tau, x, t)$$

and

$$p_1(\tau|x, t) = a_0(x) \exp(-a_0(x)\tau)$$

$$p_2(j|\tau, x, t) = \frac{a_j(x)}{a_0(x)}.$$

- ▶ To sample random variables according to $p_1(\tau|x, t)$ and $p_2(j|\tau, x, t)$ is the direct method of SSA.

References

- ▶ G. Winkler, Image Analysis, Random Fields and Dynamic Monte Carlo Methods, Springer-Verlag, Berlin and New York, 1995.
- ▶ D.P. Landau and K. Binder, A guide to Monte Carlo Simulations in Statistical Physics, World Publishing House, 2004.
- ▶ Jun S. Liu, Monte Carlo strategies in scientific computing, Springer-Verlag, 2001.
- ▶ D.T. Gillespie, J. Phys. Chem. 81, 403(1977).