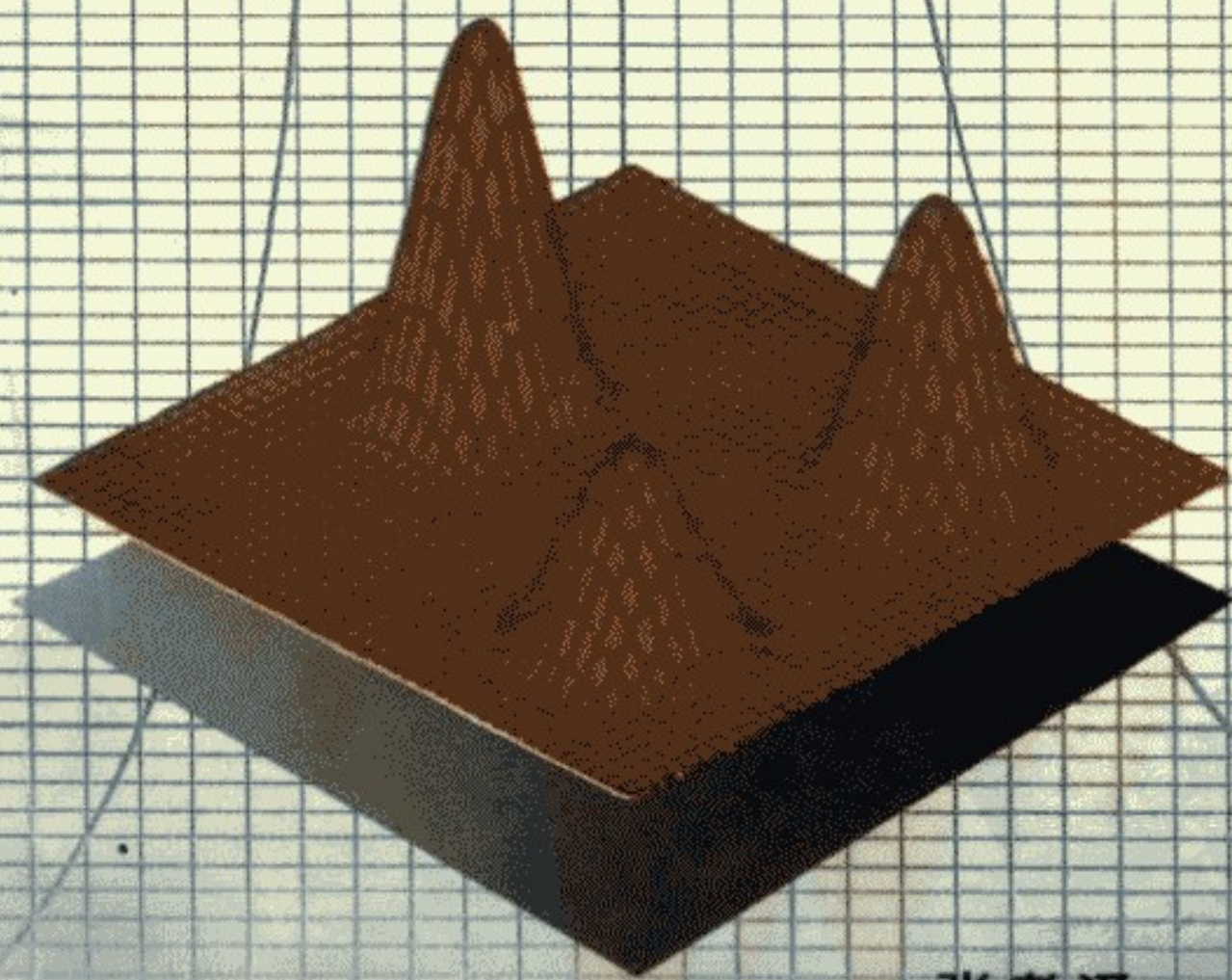


科学与工程计算丛书

蒙特卡罗方法在统计物理中的应用

MENGTEKALUOFANGFAZAITONG
JIWULIZHONGDEYINGYONG



张孝泽 编著
河南科学技术出版社

SECS

51.819

科学与工程计算丛书

蒙特卡罗方法 在统计物理中的应用

张孝泽 编著

河南科学技术出版社

《科学与工程计算丛书》编辑委员会

名誉主编：冯康

名誉编委：（按姓氏笔划为序）

于敏	王仁	冯康	石钟慈	庄逢甘	曲钦岳
朱家骅	李德元	何祚庥	陈能宽	谷超豪	况蕙孙
郑哲敏	周毓麟	秦元勋	黄祖洽	曾庆存	符鸿源
程开甲	裴鹿成				

编委：（按姓氏笔划为序）

于万瑞	王宗皓	王政贤	王宝瑞	王肖钧	冯士桢
孙文心	厉衡隆	石中岳	卢秀球	付德薰	付泽周
纪立人	纪楚群	刘林	刘儒勋	向新民	朱允伦
李荫藩	李作新	吴江航	吴乃龙	吴辉碇	吴其芬
杜书华	杨清建	宋国乡	邱希春	陈健华	何延才
何锦昌	汪翼云	金时懋	郑邦民	周树荃	范新亚
宓国柱	罗吉庭	张立存	张志杰	张若棋	张锁春
胡乃雄	姚凯伦	浣石	顾昌鑫	倪浩清	徐国华
常文蔚	常谦顺	赖定文	蒋伯诚	董绍静	鲍家駉

常务编委：（按姓氏笔划为序）

孙文心	刘儒勋	吴江航	何延才	金时懋	徐国华
蒋伯诚					

执行主编：何延才

编辑部：

蒋伯诚	张锁春	张立存	张志杰	周春生	杜慧娴
陈吉斌					

目 录

第一章	Metropolis抽样方法	(1)
1.1	正则系综平均量的计算	(1)
1.1.1	正则系综的平均观察量	(1)
1.1.2	内能、比热、自由能和熵	(3)
1.1.3	蒙特卡罗计算	(5)
1.2	Metropolis抽样方法	(9)
1.2.1	离散分布的Metropolis抽样方法	(9)
1.2.2	Metropolis抽样方法的收敛性	(11)
1.2.3	Metropolis抽样方法的一般描述	(13)
1.2.4	预选矩阵 Q 的几种具体形式	(18)
1.2.5	最佳转移概率矩阵的选择	(20)
1.2.6	复合型转移矩阵和热浴方法	(27)
1.2.7	推广到连续分布的情况	(35)
1.2.8	Metropolis抽样方法的一些性质	(44)
1.2.9	非归一分布的迭代抽样方法	(50)
1.3	计算的若干细节	(52)
1.3.1	随机数的使用	(52)
1.3.2	初始分布的影响	(53)
1.3.3	方差的估计	(53)
1.3.4	周期边界条件	(54)
1.3.5	计算例子	(54)
	参考文献	(57)

第二章 Ising模型的蒙特卡罗模拟	(59)
2.1 蒙特卡罗方法解 Ising 模型	(59)
2.1.1 Ising 模型	(59)
2.1.2 主要的物理量	(60)
2.1.3 周期边界条件	(62)
2.1.4 最近邻相互作用	(63)
2.1.5 蒙特卡罗解 Ising 模型	(64)
2.1.6 计算例题	(68)
2.2 研究二元合金系统	(70)
2.3 蒙特卡罗方法研究 XY 模型	(71)
2.3.1 Ising 模型的推广	(71)
2.3.2 蒙特卡罗方法计算 XY 模型	(73)
2.4 能量响应函数系综平均随温度变化曲线的计算方法	(78)
2.4.1 偏移抽样方法	(73)
2.4.2 Ferrenberg—Swendsen 方法	(79)
2.5 求总极值的模拟退火方法	(81)
2.5.1 原理	(82)
2.5.2 Metropolis 抽样方法模拟退火过程	(82)
2.5.3 收敛性问题的说明	(84)
参考文献	(86)
第三章 蒙特卡罗方法解量子多体问题	(89)
3.1 量子多体系统的最低能量	(89)
3.1.1 量子多体问题	(89)
3.1.2 变分原理	(90)
3.2 McMillan—Metropolis 方法	(92)
3.2.1 McMillan—Metropolis 方法	(92)
3.2.2 试探函数族 $\Psi_i(R, a)$	(93)

第一章 Metropolis抽样方法

本章着重讨论在统计物理中广泛应用的Metropolis抽样方法以及它的各种形式。

1.1 正则系综平均量的计算

1.1.1 正则系综的平均观察量

在统计物理中，平衡态下某一物理量 A 的平均观察量 $\langle A \rangle$ 是按某一系综分布来取的：

$$\langle A \rangle = \int_{\Omega} A(\vec{x}) p(\vec{x}) d\vec{x} \quad (1.1)$$

这儿， $\vec{x}$ 表示相空间 Ω 中的一个点（或一个状态，一个组态，一个位形）， $p(\vec{x})$ 是这个物理系统的系综分布， $A(\vec{x})$ 是某个微观观察量。显然， $\langle A \rangle$ 实际上是一个平均值，或叫期望值，因为系统分布 $p(\vec{x})$ 满足条件：

$$p(\vec{x}) \geq 0, \quad \int_{\Omega} p(\vec{x}) d\vec{x} = 1 \quad (1.2)$$

假定，这个系统有 N 个相同性质的粒子，一个粒子的状况通常由一系列动力学参数（位置 $\vec{r}$ ，速度 $\vec{v}$ ，自旋 s 等）来描述，用

$$\alpha = (\vec{r}, \vec{v}, s, \dots)$$

来表示。于是， Ω 中的点 $\vec{x}$ 可表为

$$\vec{x} = (\alpha_1, \alpha_2, \dots, \alpha_N). \quad (1.3)$$

正则系综是最常用的系综。这时，系统的三个参数：粒子数

N , 温度 T 和体积 V 固定, 其分布密度函数为 Boltzmann 分布:

$$p(\vec{x}) = \frac{e^{-H_N(\vec{x})/K_B T}}{\int_0 e^{-H_N(\vec{x})/K_B T} d\vec{x}} \quad (1.4)$$

这里, K_B 为 Boltzmann 常数, T 为系统周围的绝对温度, $H_N(\vec{x})$ 就是通常的哈密顿量。它是由系统中粒子间的相互作用和外部影响所决定, 也就是系统的能量。这时, 系综平均量 (1.1) 改写为

$$\langle A \rangle = \frac{\int_0 A(\vec{x}) e^{-H_N(\vec{x})/K_B T} d\vec{x}}{\int_0 e^{-H_N(\vec{x})/K_B T} d\vec{x}}. \quad (1.5)$$

如果系统的状态 $\vec{x}$ 取离散点上的值, 那么积分 (1.5) 就转变为求和

$$\langle A \rangle = \frac{\sum_i A(\vec{x}_i) e^{-H_N(\vec{x}_i)/K_B T}}{\sum_i e^{-H_N(\vec{x}_i)/K_B T}} \quad (1.6)$$

在平衡态统计物理中, 除了正则系综以外, 还有微正则系综、巨正则系综和等温等压系综。微正则系综是描写孤立的、能量固定的系统的统计平衡下的状态。而在巨正则系综中, 化学势 μ 、体积 V 和温度 T 保持固定, 粒子数 N 是变化的, 它的分布密度函数为

$$p(N, \vec{x}) = \frac{e^{+\mu N/K_B T} e^{-H_N(\vec{x})/K_B T}}{\int_0 \int_0 e^{\mu N/K_B T} e^{-H_N(\vec{x})/K_B T} dN d\vec{x}}. \quad (1.7)$$

至于等温等压系综, 压力 P 、温度 T 和粒子数 N 保持固定, 而体积 V 变化, 其分布密度函数为

$$p(V, \vec{x}) = \frac{e^{-(PV+H_N(\vec{x})/K_B T)}}{\int_0^\infty dV \int_\Omega e^{-(PV+H_N(\vec{x})/K_B T)} d\vec{x}} \quad (1.8)$$

由于计算巨正则系综和等温等压系综的平均观察量所使用的方法，在原理上与计算正则系综的平均量是相同的，因此，下面我们主要讨论正则系综的平均量的计算。

1.1.2 内能、比热、自由能和熵

在统计物理中，系统的内能 U_N ，比热 C_V ，自由能 F_N 和熵 S_N 都是重要的物理量，它们都可以通过计算(1.5)那样的系综平均得到。

实际上，根据定义，内能 U_N 为^[1]：

$$U_N = \langle H_N(\vec{x}) \rangle = \frac{\int_\Omega H_N(\vec{x}) e^{-H_N(\vec{x})/K_B T} d\vec{x}}{\int_\Omega e^{-H_N(\vec{x})/K_B T} d\vec{x}} \quad (1.9)$$

在体积固定(不变)下的比热 C_V 为：

$$C_V = \frac{\partial U_N}{\partial T} = (\langle H_N^2(\vec{x}) \rangle - \langle H_N(\vec{x}) \rangle^2) / K_B T^2 \quad (1.10)$$

自由能 F_N 和熵 S_N 分别为：

$$F_N = -K_B T \ln Z_N \quad , \quad (1.11)$$

$$S_N = (U_N - F_N) / T \quad , \quad (1.12)$$

其中 Z_N 为：

$$Z_N = \int_\Omega e^{-H_N(\vec{x})/K_B T} d\vec{x} \quad , \quad (1.13)$$

人们称之为配分函数。

稍加变换， Z_N 可写为

$$Z_N = \frac{\Omega}{\langle \exp(H_N(\vec{x})/K_B T) \rangle} \quad (1.14)$$

从以上 U_N , C_V , F_N , S_N 和 Z_N 的表达式(1.9)~(1.14)可知, 它们都可化为形如(1.5)的系综平均。这就表明, 计算系综平均在统计物理的研究中, 确实起着十分重要的作用。

需要指出, 通过计算(1.14)中的 $\langle \exp(H_N(\vec{x})/K_B T) \rangle$ 来得到 Z_N , 进而得到 F_N 和 S_N , 不是一种好办法。因为 $\exp(H_N(\vec{x})/K_B T)$ 变化很大, 计算 $\langle \exp(H_N(\vec{x})/K_B T) \rangle$ 的结果难于稳定。它的方差, 有的时候不是有限, 因而收敛很慢。

计算 S_N 和 F_N 的较好的方法是^[2]:

$$S_N(T) = S_N(\infty) + U_N/T - K_B \int_0^{1/K_B T} U_N d\left(\frac{1}{K_B T}\right) \quad (1.15)$$

$$\frac{F_N}{K_B T} = \frac{S_N(\infty)}{K_B} + \int_0^{1/K_B T} U_N d\left(\frac{1}{K_B T}\right). \quad (1.16)$$

由于 $S_N(\infty) = \lim_{T \rightarrow \infty} S_N(T)$, 在许多情况下已知, 因此, 通过对 U_N 求积分的方法, 可以得到 S_N 和 F_N 。

利用公式(1.9)、(1.11)和(1.12)容易推得

$$\frac{\partial S_N}{\partial T} = \frac{\partial S_N}{\partial T} / T, \quad (1.17)$$

进而在 (T, ∞) 上积分, 就可推得(1.15)和(1.16)。实际上, (1.17)可如下推出:

$$\begin{aligned} \frac{\partial S_N}{\partial T} &= \frac{\partial}{\partial T} \left((U_N - F_N)/T \right) = \frac{\partial U_N}{\partial T} / T - \frac{\partial F_N}{\partial T} / T - \frac{U_N - F_N}{T^2} \\ &= \frac{\partial U_N}{\partial T} / T - \frac{1}{T} \left(\frac{\partial F_N}{\partial T} + S_N \right), \end{aligned}$$

而

$$\frac{\partial F_N}{\partial T} = \frac{\partial}{\partial T} (-K_B T \ln Z_N) = -K_B \ln Z_N - K_B T \frac{\frac{\partial Z_N}{\partial T}}{Z_N}$$

$$\frac{\partial Z_N}{\partial T} = U_N Z_N / K_B T^2.$$

所以,

$$\frac{\partial F_N}{\partial T} = -K_B \ln Z_N - K_B T \frac{U_N}{K_B T^2} = \frac{F_N - U_N}{T} = -S_N,$$

由此, 立即得到 (1.17)。

计算自由能 F_N 的其它有效方法可参见 [3]。

1.1.3 蒙特卡罗计算

显然, 系综平均 (1.5) 是一个高维积分。即使对于离散情形, (1.6) 虽是一个求和, 常常也是难于逐项计算。比如, 对于二维 Ising 模型 (见第二章), 取一个小尺寸的点阵: $N = 10 \times 10$, 系统的一个状态 $\vec{x}$ 表为:

$$\vec{x} = (\sigma_1, \sigma_2, \dots, \sigma_{100}),$$

其中 σ_j 是第 j 个地址上的自旋。每个自旋可取两个值, $\sigma_j = 1$, 自旋向上, $\sigma_j = -1$, 自旋向下。这时, 系综平均 $\langle A \rangle$ 要求在 2^{100} 项上求和, 这有很大的计算量。实际上, $2^{100} \approx 1.3 \times 10^{30}$, 因此, 即使对于一台速度为每秒千万次的计算机来说, 它也要算至少 4×10^{15} 年, 这是根本无法做到的。然而, 由于蒙特卡罗方法适于解决多维问题, 能够用一小部分“代表”状态上的算术平均值近似系综平均值, 因此, 采用蒙特卡罗方法来计算它们便是十分自然的事了。

1 蒙特卡罗方法原理 [4]

现在, 转来叙述蒙特卡罗方法求 (1.1) 和 (1.2) 定义的一般平均值问题。

设 我们得到分布密度函数 $p(\vec{x})$ 的一组代表点:

$$\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M, \quad (1.18)$$

即从分布密度 $p(\vec{x})$ 中取出的容量为 M 的子样, 并组成代表点

(1.18) 上的算术平均值

$$A_M = \frac{1}{M} \sum_{m=1}^M A(\tilde{x}_m), \quad (1.19)$$

那么, 显然有

$$\langle A_M \rangle = \langle A \rangle, \quad (1.20)$$

而且, 只要 $\langle A \rangle$ 有限, 各个 $\tilde{x}_m$ 是独立选取的, 那么 根据 大数定律, 极限

$$\lim_{M \rightarrow \infty} A_M = \langle A \rangle \quad (1.21)$$

以概率1成立. 亦即, 只要 M 充分大, A_M 可以作为 $\langle A \rangle$ 的一个好的估计.

再如果, $A(\tilde{x})$ 的方差有限, 即

$$\sigma^2 = \int_0 A^2(\tilde{x}) p(\tilde{x}) d\tilde{x} - \langle A \rangle^2 < +\infty \quad (1.22)$$

那么, 根据中心极限定理, 误差估计

$$|A_M - \langle A \rangle| < \frac{X_\alpha \sigma}{\sqrt{M}} \quad (1.23)$$

以概率 $1 - \alpha$ 成立:

$$1 - \alpha = \frac{1}{\sqrt{2\pi}} \int_{-X_\alpha}^{X_\alpha} e^{-\frac{1}{2}x^2} dx. \quad (1.24)$$

(1.23) 表明, 用 A_M 来近似 $\langle A \rangle$, 其误差阶为 $O(M^{-\frac{1}{2}})$.

对离散分布 $p_i = p(\tilde{x}_i)$ 上的平均值, 有相同的结论.

分析一下 (1.23) 式, 我们可看出以下两个基本点:

第一, 蒙特卡罗方法的收敛速度是 $O(M^{-\frac{1}{2}})$, 是比较慢的, 因此要想得到十分准确的结果 (误差小于1%) 是困难的.

第二, 但是, 蒙特卡罗方法的误差却与维数无关. 因此, 为得到相同的误差精度, 对于不同维数的问题, M 是相当的, 计算

量只与维数成比例，而不像通常的数值方法，计算量与维数的方幂成比例。这一点就是蒙特卡罗方法适用于解多维问题的原因所在。

如何从各种不同的分布密度函数 $p(\vec{x})$ ，包括 $(0, 1)$ 上的均匀分布和一般离散分布，得到代表点或子样，这是蒙特卡罗方法研究中的基本问题之一。另外，为了减少误差，除了增大代表点数目 M 外，还要设法减少估计量的方差 σ^2 和计算一个估计量所费的运算量的乘积，这也是蒙特卡罗方法研究中的重要问题。关于这些问题的详尽讨论可参阅^[4]。

我们还要指出，上述计算平均值的方法，可用来计算一般积分。实际上，设欲求积分

$$I = \int_{\Omega} G(\vec{x}) d\vec{x} . \quad (1.25)$$

取 Ω 上一个概率密度函数 $p(\vec{x})$ ，满足以下条件：

$$\begin{aligned} p(\vec{x}) &\neq 0, & \text{当 } \vec{x} \in \Omega, & G(\vec{x}) \neq 0 \\ p(x) &\geq 0, & \int_{\Omega} p(\vec{x}) d\vec{x} &= 1 . \end{aligned} \quad (1.26)$$

$$\text{令 } g(\vec{x}) = \begin{cases} G(\vec{x})/p(\vec{x}), & p(\vec{x}) \neq 0 \\ 0, & p(\vec{x}) = 0 \end{cases} \quad (1.27)$$

那么，(1.25)中的 I 可改写为

$$I = \int_{\Omega} g(\vec{x}) p(\vec{x}) d\vec{x} = \langle g \rangle, \quad (1.28)$$

与(1.1)形式相同。 $p(\vec{x})$ 的最简单形式是取 Ω 上的均匀分布：

$$p(\vec{x}) = \begin{cases} 1/\Omega, & \vec{x} \in \Omega \\ 0, & \text{其它} \end{cases} \quad (1.29)$$

这儿， Ω 也表示相空间的体积。这时，

$$g(\vec{x}) = \begin{cases} \Omega \cdot G(\vec{x}), & \vec{x} \in \Omega \\ 0, & \text{其它} . \end{cases} \quad (1.30)$$

2 蒙特卡罗方法计算系综平均

计算 (1.5) 式的系综平均 $\langle A \rangle$ 的一种方法是, 从某个已知分布密度 $\pi(\vec{x})$

$$\pi(\vec{x}) \geq 0 \quad \int_{\Omega} \pi(\vec{x}) d\vec{x} = 1 \quad (1.31)$$

抽样 $\vec{x}_m, m = 1, 2, \dots, M$, 然后, 分别计算 (1.5) 式中的分子和分母, 那么, $\langle A \rangle$ 的估计为

$$\langle A \rangle \approx \frac{\sum_{m=1}^M A(\vec{x}_m) e^{-H_N(\vec{x}_m)/K_B T} [\pi(\vec{x}_m)]^{-1}}{\sum_{m=1}^M e^{-H_N(\vec{x}_m)/K_B T} [\pi(\vec{x}_m)]^{-1}} \quad (1.32)$$

$\pi(\vec{x})$ 的选法多种多样。它的一种好的选择是取系综分布本身^[5]即取

$$\pi(\vec{x}) = \frac{e^{-H_N(\vec{x})/K_B T}}{\int_{\Omega} e^{-H_N(\vec{x})/K_B T} d\vec{x}}$$

这样一来, (1.32) 式变为 (1.19)

$$\langle A \rangle \approx A_M = \frac{1}{M} \sum_{m=1}^M A(\vec{x}_m)$$

剩下的问题是如何由 $\pi(\vec{x})$ 抽样 $\vec{x}$ 。由于在 $\pi(\vec{x})$ 中有一个未知的归一常数 (分母)

$$\int_{\Omega} e^{-H_N(\vec{x})/K_B T} d\vec{x},$$

因此, 直接法抽样^[4]难以应用; 另一方面, 由于 $H_N(\vec{x})$ 的复杂性, 或者难以求得 $\pi(\vec{x})$ 的最大值 ($H_N(\vec{x})$ 可能取负值); 或者虽能求得 $\pi(\vec{x})$ 的最大值, 但 $\pi(\vec{x})$ 随 $\vec{x}$ 变化值相差很大, 都使挑选法^[4]失效。N. Metropolis 等人在 1953 年提出了一种抽样方法, 现在人们称它为 Metropolis 抽样方法^[6], 巧妙地解决了这类分布的抽样问题。从此, 它在统计物理中得到了广泛的应用。

1.2 Metropolis 抽样方法

1.2.1 离散分布的Metropolis 抽样方法

现在, 我们研究(1.6)的计算问题, 即由离散分布

$$\pi_i = \frac{e^{-H_N(\tilde{x}_i)/K_B T}}{\sum_i e^{-H_N(\tilde{x}_i)/K_B T}}, \quad i=1, 2, \dots, I \quad (1.33)$$

的抽样问题。不失一般性, 设 $\pi_i > 0, i=1, 2, \dots, I$ 。代表点, 或子样

$$\tilde{x}_{i_0}, \tilde{x}_{i_1}, \dots, \tilde{x}_{i_m}, \tilde{x}_{i_{m+1}}, \dots \quad (1.34)$$

可由以下手续产生:

i) 初始点 $\tilde{x}_{i_0}$ 由任意初始分布 $S(\tilde{x}_{i_0})$ 抽样, 这儿, $S(\tilde{x}_{i_0})$

满足:

$$S_{i_0} = S(\tilde{x}_{i_0}) \geq 0, \quad \sum_{i_0=1}^I S(\tilde{x}_{i_0}) = 1 \quad (1.35)$$

ii) 给定一个对称的转移概率矩阵 $P^* = (p^*_{ij})$,

即满足以下条件:

$$p^*_{ij} = p^*(\tilde{x}_j | \tilde{x}_i) = p^*(\tilde{x}_i | \tilde{x}_j) = p^*_{ji} > 0. \quad (*)$$

$$\sum_{j=1}^I p^*_{ij} = 1, \quad i=1, 2, \dots, I \quad (1.36)$$

然后, 对任何整数 $m \geq 0$, 在 $\tilde{x}_{i_m}$ 已知条件下, $\tilde{x}_{i_{m+1}}$ 可如下产生:

* $p^*_{ij} > 0$ 这条件可以放宽见1.2.3。

从对称转移矩阵 P^* 的第 i_m 行

$$p_{i_m i'} = p^*(\vec{x}_{i'} | \vec{x}_{i_m}), \quad i' = 1, 2, \dots, I \text{ 抽样产生 } \vec{x}_{i'}$$

↓
计算能量差

$$\Delta H_N(\vec{x}_{i_m}, \vec{x}_{i'}) = H_N(\vec{x}_{i'}) - H_N(\vec{x}_{i_m})$$

$$\Delta H_N \leq 0? \quad \begin{array}{l} \leq (+) \\ > (-) \end{array}$$

$$\xi \leq e^{-\Delta H_N / K_B T} \quad ? \quad \begin{array}{l} (+) \\ (-) \end{array} \rightarrow \vec{x}_{i_{m+1}} = \vec{x}_{i'}$$

$$\vec{x}_{i_{m+1}} = \vec{x}_{i_m} \quad (1.37)$$

这儿, ξ 为 $(0, 1)$ 上均匀分布的随机数 (下同)。

注意到

$$e^{-\Delta H_N / K_B T} = \frac{\pi_{i'}}{\pi_{i_m}},$$

$$\Delta H_N \leq 0 \iff e^{-\Delta H_N / K_B T} \geq 1 \iff \frac{\pi_{i'}}{\pi_{i_m}} \geq 1,$$

(1.37) 可改写为一般形式:

从对称转移矩阵 P^* 的第 i_m 行

$$p_{i_m i'} = p^*(\vec{x}_{i'} | \vec{x}_{i_m}), \quad i' = 1, 2, \dots, I \text{ 抽样产生 } \vec{x}_{i'}$$

$$\xi \leq \frac{\pi(\vec{x}_{i'})}{\pi(\vec{x}_{i_m})} = \frac{\pi_{i'}}{\pi_{i_m}}? \quad \begin{array}{l} \leq (+) \\ > (-) \end{array}$$

$$\vec{x}_{i_{m+1}} = \vec{x}_{i_m} \quad \vec{x}_{i_{m+1}} = \vec{x}_{i'} \quad (1.38)$$

这就是著名的Metropolis 抽样方法。不断重复 (1.37) 或

(1.38), 就可得到(1.34)。

分析一下(1.37)和(1.38)这种抽样过程, 可以看到, 当能量 $H_N(\vec{x}_i')$ 比 $H_N(\vec{x}_{i_m})$ 减小时, 即 $\pi(\vec{x}_i')$ 增大时, 状态 $\vec{x}_i'$, 总被接受为 $\vec{x}_{i_{m+1}}$; 而当 $H_N(\vec{x}_i')$ 比 $H_N(\vec{x}_{i_m})$ 大时, 即 $\pi(\vec{x}_i')$ 变小时, $\vec{x}_i'$ 被接受的概率与 $\pi(\vec{x}_i')$ 成比例。这是符合概率密度 $\pi(\vec{x}_i')$ 大时, 抽 $\vec{x}_i'$ 的机会多这一特点的。其次, 还可看到, 已知 $\vec{x}_{i_m}$ 后, $\vec{x}_{i_{m+1}}$ 的确定只与 $\vec{x}_{i_m}$ 有关, 而与 $\vec{x}_{i_0}, \vec{x}_{i_1}, \dots, \vec{x}_{i_{m-1}}$ 无关, 因此 $\{\vec{x}_{i_m}\}_{m=0}^{\infty}$ 组成一个有限状态的均匀马尔科夫链。最后, 可以导出, 这种抽样过程与概率分布中的未知归一常数无关。因而避开了分布(1.33)中有一未知归一常数这一困难。

不难导出, 这种均匀马尔科夫链的转移概率矩阵 $P^M = (p_{ij}^M)$ 具有以下形式:

$$p_{ij}^M = \begin{cases} p_{ij}^* \min\left(1, \frac{\pi_j}{\pi_i}\right) & j \neq i \\ 1 - \sum_{j \neq i} p_{ij}^M, & j = i \end{cases} \quad (1.39)$$

其中, $P^* = (p_{ij}^*)$ 满足条件(1.36)。(1.39)就是最早由Metropolis等人提出和使用的转移概率矩阵^[7]。

因此, 概括起来, Metropolis 抽样方法产生子样(或代表点)(1.34)的步骤如下:

- i) $\vec{x}_{i_0}$ 由任意分布 $S'(\vec{x}_{i_0})$ 抽样确定;
- ii) 对任何整数 $m \geq 0$, $\vec{x}_{i_{m+1}}$ 由转移概率矩阵第 i_m 行 $p_{i_m i_{m+1}}^M = P(\vec{x}_{i_{m+1}} | \vec{x}_{i_m})$ 抽样确定;
- iii) $m = m + 1$, 重复 ii)。

1.2.2 Metropolis抽样方法的收敛性

首先, 考虑子样(1.34)中元素 $\vec{x}_{i_m}$ 的分布。

容易验证, 关系式

$$\pi_i p_{ij}^M = \pi_j p_{ji}^M, \quad i, j = 1, \dots, I \quad (1.40)$$

成立, 实际上, 当 $i = j$ 时, (1.40) 成立. 当 $i \neq j$ 时, 不妨设 $\pi_i \geq \pi_j$, 这时

$$\text{左端} = \pi_i p_{ij}^M = \pi_i p_{ij}^* \min(1, \pi_j / \pi_i)$$

$$= \pi_i p_{ij}^* \cdot \frac{\pi_j}{\pi_i} = \pi_j p_{ij}^* = \pi_j p_{ji}^*,$$

$$\text{右端} = \pi_j p_{ji}^M = \pi_j p_{ji}^* \min(1, \pi_i / \pi_j)$$

$$= \pi_j p_{ji}^* = \text{左端}.$$

当 $\pi_i < \pi_j$ 时, 同样可证结论成立. 关系式 (1.40) 称为细致平衡 (detail balance). 由 (1.40) 立即得到:

$$\pi_j = \sum_{i=1}^I \pi_i p_{ij}^M, \quad j = 1, 2, \dots, I. \quad (1.41)$$

由于 $p_{ij}^* > 0$, $i, j = 1, 2, \dots, I$, 再由 (1.39) 式可以得出, $p_{ij}^M > 0$, $i, j = 1, \dots, I$. 因此, 从有限马尔科夫链的理论立即得知 (见附录), 由转移概率矩阵 P^M 所定义的马尔科夫链是各态历经的 (遍历的), 且有

$$\lim_{m \rightarrow \infty} p_{ij}^M(m) = p_j^M = \pi_j > 0,$$

$$j = 1, 2, \dots, I. \quad (1.42)$$

这儿, $p_{ij}^M(m)$ 是由状态 $\tilde{x}_i$ 到状态 $\tilde{x}_j$ 的 m 步转移概率. $p_j^M = \pi_j$, $j = 1, \dots, I$, 这是由 p_j^M 是满足方程 (1.41) 的唯一解而得出的.

进一步, $\tilde{x}_{i_m}$ 的分布 $\pi_i^{(m)}$, 就是马尔科夫链经 m 步后处于状态 $\tilde{x}_i$ 的概率, 也称全概率, 它可表示为

$$\pi_i^{(m)} = P(\tilde{x}_{i_m} = \tilde{x}_i) = \sum_{i_0} S(\tilde{x}_{i_0}) P(\tilde{x}_i | \tilde{x}_{i_0}, m)$$

$$= \sum_{i_0} S_{i_0} \cdot p_{i_0 i}^M(m),$$

$$m = 1, 2, \dots, i = 1, 2, \dots, I \quad (1.43)$$

当 $m \rightarrow \infty$, 对 (1.43) 两边取极限, 利用 (1.42) 立即得到