

第一次 Project

一. 主题介绍

- 本次大作业的主题是——马氏链蒙特卡洛方法
- 马氏链蒙特卡洛 (Markov Chain Monte Carlo, MCMC) 方法, 是马氏链理论的一个重要应用。从 1950 年萌芽, 马氏链蒙特卡洛方法在实践中不断发展, 广泛应用于各种学科领域 (如信息科学、物理、化学、生物学、金融、材料等) 的科学计算, 展示出越来越强大的威力。
- MCMC 的一个简短介绍见文献 [1]——林元烈编著的《应用随机过程》第 3.5 节 P^n 的极限性态与平稳分布。英文的简短介绍可见文献 [2] 的第 29、30 章。更系统的介绍见文献 [3]。

二. 作业题目

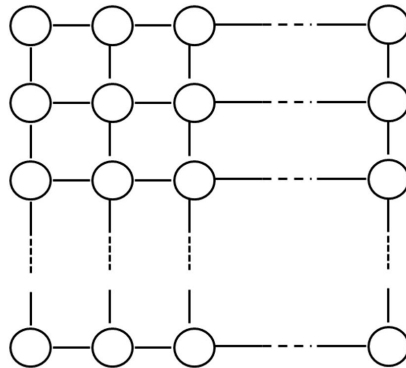
(a) 二维高斯分布的估计。

用 Metropolis-Hastings (MH) 算法, 对下述二维高斯分布进行随机采样

$$\mathcal{N}\left\{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \middle| \begin{pmatrix} 5 \\ 10 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}\right\}$$

- 具体设计出采用的 Metropolis-Hastings 算法, 利用采样结果估计二维高斯分布的相关系数 ρ 。结合实验结果, 分析算法性能, 特别是关注估计的准确性和计算效率。
- 选取不同的举荐分布, 从接受比例、收敛速度等方面对算法进行分析。

(b) 估计 Potts 模型的归一化常数, 绘制能量直方图。



Potts 模型源于统计物理, 是一种重要的概率模型。对 Potts 模型的归一化常数的估计, 代表了一大类科学计算问题, 至今仍是非常活跃的一个研究课题。考虑定义在上图网格上的 $K = 20 \times 20$ 个离散随机变量 $x = (x_1, \dots, x_{(20 \times 20)})$, 每个变量有 q 个可能取值。 K 个变量的联合分布列为:

$$p(x) = \frac{1}{Z(T)} \exp\left[-\frac{u(x)}{T}\right], \quad (1)$$

$$u(x) = - \sum_{\substack{i \leftrightarrow j \\ i, j=1, \dots, K}} 1(x_i = x_j), \quad (2)$$

$i \leftrightarrow j$ 表示格点 i 和 j 相邻。 $1(x_i = x_j)$ 是示性函数, 当 $x_i = x_j$ 时取值为 1, 否则为 0。在统计物理中, $u(x)$ 称为能量, T 表示温度, 归一化常数 $Z(T) = \sum_x \exp\left[-\frac{u(x)}{T}\right]$ 称为配分函数 (Partition Function)。

对 $q = 10$, $T^{(-1)} = 1.4; 1.4065; 1.413; 1.4195; 1.426$, 代表了五个温度下的五个不同的 Potts 模型,

- i. 试使用 MCMC 方法生成的随机样本来估计归一化常数 $Z(T)$;
- ii. 绘制不同温度下的 $\frac{u(x)}{K}$ 的直方图 (称为能量直方图);
- iii. 绘制不同温度下 Potts 模型的典型样本。

注: 请同学们在调研和阅读相关文献的基础上, 设计出有效的采样算法, 结合实验结果, 分析算法性能, 特别是比较采用不同的采样算法下估计的准确性和计算效率。相关文献包括但不限于, [4] 提出的 Nested sampling 方法, [4] 提到的 Swendsen-Wang 方法, [5] 提出的 Stochastic approximation Monte Carlo (SAMS) 方法, [5] 提到的 Parallel Tempering 方法。

三. 具体要求

(a) 希望同学充分调研和阅读相关文献, 积极动脑 + 动手, 取得有**自己见解**的结果, 整理成最终报告。

(b) **最终提交包括:**

i. **报告**

报告的书写要求参见《Project 报告撰写建议》。

ii. **源程序**

不带任何参数运行的主程序命名为 **run.m**, 将输出对上述五个 Potts 模型的归一化常数的估计 (获得的最优估计), 绘制出能量直方图以及典型样本。同时务必包含**自己所有的原始程序**: 请**重点关注**你所写代码的可读性, 务必在代码中包含**充足的注释**。请在**你所有**的 m 文件开头注明本文件中代码实现的功能, 并保证全部代码是可运行的。如果代码的可读性较差或运行出现问题, 会对评分结果产生**较大影响**。

将以上两项一起压缩打包, 命名为 “**学号 _ 姓名.rar**” 进行提交。

(c) 评分标准: 报告书写清晰和规范, 工作新意及深入程度, 工作量及完整程度。

(d) 一旦发现抄袭, 计零分。

(e) 请大家在规定截止时间前提交。晚交的处理方法如下: 按晚交天数, 以 90% 的几何级数进行折扣。晚交时间在 (0, 24 小时], 按 90% 折扣。晚交时间在 (24 小时, 48 小时], 按 90%*90% 折扣。以此类推。

参考文献

- [1] 林元烈, 应用随机过程. 清华大学出版社, 2002.
- [2] MacKay and C. Davidj., *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- [3] J. S. Liu, *Monte Carlo strategies in scientific computing*. Springer Science & Business Media, 2008.
- [4] I. Murray, D. J. C. Mackay, Z. Ghahramani, and J. Skilling, “Nested sampling for potts models,” *Advances in Neural Information Processing Systems*, pp. 947–954, 2006.
- [5] Z. Tan, “Optimally adjusted mixture sampling and locally weighted histogram analysis,” *Journal of Computational and Graphical Statistics*, vol. 26, no. 1, pp. 54–65, 2017.