

分布式采样理论综述*

凤维明, 尹一通



(计算机软件新技术国家重点实验室(南京大学), 江苏 南京 210023)

通信作者: 尹一通, E-mail: yinyt@nju.edu.cn

摘要: 采样是一类基本的计算问题. 从一个解空间中依特定概率分布进行随机采样, 这一问题在近似计数、概率推断、统计学习等方面都有着诸多重要的应用. 在大数据时代, 采样问题的分布式算法与分布式计算复杂性受到越来越多的关注. 近年来, 有一系列的工作对分布式采样理论展开系统性的研究. 综述了分布式采样的重要结论, 主要包括有严格理论保障的分布式采样算法、采样问题在分布式模型上的计算复杂性以及采样与推断等问题在分布式计算模型中的相互联系.

关键词: 采样; 分布式计算; 计算复杂性; 吉布斯分布; 马尔可夫链
中图法分类号: TP301

中文引用格式: 凤维明, 尹一通. 分布式采样理论综述. 软件学报, 2022, 33(10): 3673–3699. <http://www.jos.org.cn/1000-9825/6372.htm>

英文引用格式: Feng WM, Yin YT. Survey on Theory of Distributed Sampling. Ruan Jian Xue Bao/Journal of Software, 2022, 33(10): 3673–3699 (in Chinese). <http://www.jos.org.cn/1000-9825/6372.htm>

Survey on Theory of Distributed Sampling

FENG Wei-Ming, YIN Yi-Tong

(State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023, China)

Abstract: Sampling is a fundamental class of computational problems. The problem of generating random samples from a solution space according to certain probability distribution has numerous important applications in approximate counting, probability inference, statistical learning, etc. In the big data era, the distributed sampling attracts considerably more attentions. In recent years, there is a line of research works that systematically study the theory of distributed sampling. This study surveys important results on distributed sampling, including distributed sampling algorithms with theoretically provable guarantees, the computational complexity of sampling in the distributed computing model, and the mutual relation between sampling and inference in the distributed computing model.

Key words: sampling; distributed computing; computational complexity; Gibbs distributions; Markov chain

1 引言

采样问题是计算机科学中一类基本的计算任务. 给定一个由问题输入所描述的特定概率分布, 采样问题要求算法返回一个服从或者近似服从该分布的随机样本. 在理论计算机科学中, 采样问题的算法与计算复杂性是一类重要的研究课题, 传统的理论研究致力于给出有严格理论保障的采样算法以及研究采样问题的计算复杂性下界. 在算法研究方面, 人们发展出了一套以马尔可夫链蒙特卡罗(Markov chain Monte Carlo, MCMC)方法^[1]为代表的采样技术; 在复杂性方面, 人们得到了采样问题和计数问题的计算等价性^[2]、采样问题的计算相变(computational phase transition)现象^[3,4]等一系列重要结论.

大数据时代的到来, 为采样理论的研究带来了新的挑战. 随着数据规模的扩大, 计算的一个重要需求是

* 基金项目: 国家重点研发计划重点专项(2018YFB1003202)

收稿时间: 2021-03-24; 修改时间: 2021-05-04; 采用时间: 2021-05-11; jos 在线出版时间: 2021-05-20

算法可以并行以及分布式地处理数据. 在采样问题上, 传统的理论研究大多都聚焦于经典的计算模型, 即计算是串行的、依赖对全局信息的访问以及假设输入为静态数据. 如今, 很多应用场景对分布式采样和并行采样的需求越来越高. 在分布式机器学习等研究领域, 人们已针对很多实际问题提出了分布式采样算法^[5-14]. 这些算法中有很多是启发式算法, 也有些算法有一定程度的理论保障^[15-17]. 尽管分布式采样有很强的现实需求, 但相应的理论研究刚刚起步^[18]. 最近几年, 开始有一系列工作在分布式计算模型上系统性地研究采样问题. 本文将总结这一方向的若干重要结果.

本文考虑分布式计算的经典理论模型——LOCAL 模型^[19]. 在 LOCAL 模型中, 通信网络表示为一张无向图 $G=(V,E)$. 网络中的每个节点 $v \in V$ 对应于一个处理器, 每个处理器有一个唯一身份(UID). 每条无向边 $e=\{u,v\} \in E$ 代表节点 u 和节点 v 之间的双向通信信道. LOCAL 模型是一个理想的同步通信模型, 计算和通信按回合(轮)进行. 初始时, 每个节点仅能访问到自己 and 邻居的 UID 以及本地输入及随机串; 在此后的每一轮中, 每个节点利用当前已收集到的信息进行本地计算, 再发送消息给邻居, 并接受来自邻居的消息. 所有节点可以做不受限计算量的本地计算, 每条消息的长度不设限制, 节点之间的通信也没有延迟. 一个 LOCAL 算法的复杂度为最坏情况下所需通信的总轮数. 显而易见地, 一个 t 轮的 LOCAL 算法可以等价地表示为如下局部信息的函数: 每个节点首先收集距离自己小于等于 t 的所有节点的全部信息, 然后利用这个函数将收集到的信息映射成自己的输出. 因此, 一个 LOCAL 算法的复杂度刻画了解决一个问题所需要的多么局部的信息. 传统的分布式计算理论主要研究如下分布式“构造”问题: 给定一个由局部约束定义的约束满足问题(例如图染色、极大独立集等), 局部地构造一个满足解需要用到多么局部的信息? 这就是局部计算(local computation)问题^[20]. 以图染色问题为例, 给定一张无向图 $G=(V,E)$ 以及 q 种颜色, 问题要求每个点 $v \in V$ 选择一个颜色, 使得所有相邻点的颜色不相同. 在这个问题中, 约束定义在每两个相邻的点上(相邻点颜色不相同), 因此, 这是一个通过一系列局部约束定义的问题. 而在一个 t 轮的 LOCAL 算法中, 每个点 $v \in V$ 只收集以自身为中心、距离不超过 t 的局部信息来计算出一个颜色 X_v , 使得 $(X_v)_{v \in V}$ 构成全图的一种合法染色. 图染色的局部计算问题就是关注当 t 多大时可以构造合法染色, 即解决图染色问题需要多少局部信息. 另一种常见的局部计算模型是 CONGEST 模型, 相对于 LOCAL 模型, CONGEST 模型对通信复杂性有额外限制, 每条消息的大小不超过 $O(\log n)$ 个比特, 其中, n 是网络的总点数. 局部计算领域已有数十年的研究, 并且取得了一系列重要成果.

分布式采样问题是采样理论和分布式计算理论的一次结合. 文献[18]首次严格地从理论上提出了分布式采样问题. 在 LOCAL 模型中, 给定一个由局部约束定义的约束满足问题, 分布式采样问题的目标不仅仅是构造一个满足解, 而是从所有满足解的均匀分布中生成一个随机解. 例如, 对于图染色问题, 分布式采样的目标不仅仅是构造一种合法染色, 而是要求输出的 $(X_v)_{v \in V}$ 是从所有合法染色的均匀分布中采样的一种随机染色. 更一般地, 给定任意一种由局部约束定义的联合分布, 分布式采样问题要求分布式算法采样一个随机样本. 分布式采样的核心是回答如下问题:

由局部约束定义的联合分布是否可以利用局部信息进行采样?

问题提出之后, 一些工作相继展开了对分布式采样问题的研究^[21-26]. 本文将从算法和复杂性两个层面总结该领域目前的成果.

在算法层面, 目前已经出现了若干有严格理论保障的分布式采样算法. 文献[18]设计了两种专门的马尔可夫链采样算法: 卢比-格劳伯(Luby-Glauber)算法和局部梅特罗波利斯(local Metropolis)算法, 两种算法的设计思路分别受到吉布斯采样算法(Gibbs sampling)和梅特罗波利斯算法(Metropolis algorithm)这两个经典串行采样算法的启发. 卢比-格劳伯算法有较好的收敛条件, 但其运行时间依赖于网络图的最大度数; 而局部梅特罗波利斯算法在一些模型上的运行时间可以达到理论最优. 考虑在分布式模型上均匀采样图的合法 q -染色. 假设图的最大度数为 Δ , 点数为 n . 对任意常数 $\delta > 0$, 如果 $q \geq (2+\delta)\Delta$, 卢比-格劳伯算法的收敛时间为 $O(\Delta \log n)$ (定理 5.2); 如果 $q \geq (2+\sqrt{2}+\delta)\Delta$, 则局部梅特罗波利斯算法的收敛时间为 $O(\log n)$. 之后, 两份独立完成的工作^[21,22]给出了改进版的局部梅特罗波利斯算法, 把图染色模型的收敛条件改进成 $q \geq (2+\delta)\Delta$ (定理 5.4). 文献[24]研究在分布式计算模型上模拟串行马尔可夫链——梅特罗波利斯算法, 在一定条件下, 算法只需要

$O(N/n + \log n)$ 轮就可以高概率地模拟串行梅特罗波利斯算法的至少 N 步转移. 对于很多采样问题, 串行梅特罗波利斯算法需要转移 $\Omega(n \log n)$ 步^[27], 这种情况下, 模拟算法可以做到 $\Omega(n)$ 倍的最优加速. 利用这个模拟算法, 串行梅特罗波利斯算法的一些结论可以直接用于分布式采样, 从而很多问题都有了高效的分布式采样算法. 文献[23,26]中的局部拒绝采样算法也可以用于分布式采样. 与之前的基于马尔可夫链的方法不同, 这类算法可以产生精确服从目标分布的随机样本(定理 5.7). 但是这类算法目前只适用于一部分模型的分布式采样, 收敛条件较为苛刻, 算法的性能还有待进一步研究. 本文提到的算法具有较小的通信复杂性, 这些上界结论在 CONGEST 模型上依然成立.

在复杂性层面, 若干关于分布式采样的计算复杂性的基本问题得到了回答. 文献[18,26]对于一大类概率分布的采样问题给出了 $\Omega(\log n)$ 下界结论(定理 6.1), 从而证明了 $O(\log n)$ 为理论最优复杂度. 对于有远距离强相关性的概率分布, 例如在最大度数 $\Delta \geq 6$ 的图上所有独立集的均匀分布, 文献[18]证明了分布式计算 $\Omega(\text{diam})$ 的下界结论, 其中, diam 是通信网络的直径(定理 6.2). 粗略来说, 远距离强相关性表示, 在一个概率分布中, 图上不同点取值的相关性不随距离的增加而衰减. 例如, 考虑一个 Δ -正则树上独立集的均匀分布, 当 $\Delta \geq 6$ 时, 第 ℓ 层叶子节点的取值即使在 $\ell \rightarrow \infty$ 时依然与根节点的取值有相关性. 因为在 LOCAL 模型中, 任何计算问题都存在复杂度为 $O(\text{diam})$ 的平凡算法, 这说明此类概率分布不可以被分布式算法采样; 同时, 由于分布式构造独立集问题也是一个平凡的问题, 因为每个点的直接输出不在独立集中即可, 因此这一复杂性下界结论说明: 针对某些特定的约束满足问题, 采样一个满足解的分布式计算难度远高于构造一个满足解. 因为 CONGEST 模型相对 LOCAL 模型有更强的限制, 所以这些下界结论在 CONGEST 模型上直接成立. 文献[24]研究了分布式采样问题和分布式计数问题之间的归约. 分布式版本的计数问题是计算目标分布在一个点上的边缘概率分布, 该问题也被称为推断(inference)问题. 文献[24]证明了如下归约关系.

- 对于任意有自归约性质(自规约性质的严格定义见定义 6.1. 粗略来说, 它是指一个问题在一定条件下可以向同类问题规约. 考虑图上 $G=(V,E)$ 独立集的均匀分布 μ . 对于一个点 $v \in V$, 如果已知 v 不在独立集中, 此时由 μ 导出的条件分布恰好为 G' 上独立集的均匀分布, 其中, G' 是 G 删除 v 得到的子图; 如果已知 v 在独立集中, 此时由 μ 导出的条件分布恰好为 G'' 上独立集的均匀分布, 其中, G'' 是 G 删除 v 以及 v 的邻居后得到的子图)的联合分布, 分布式的近似推断和近似采样问题可以在 $\text{polylog}(n)$ 的时间内相互归约(定理 6.3、定理 6.4);
- 对于有自归约性质且由局部约束定义的联合分布(局部吉布斯分布, 定义 4.1), 分布式的近似推断、近似采样和精确采样问题可以在 $\text{polylog}(n)$ 的时间内相互归约(定理 6.3–定理 6.5); 这些问题可以被分布式算法高效解决当且仅当目标分布满足强空间混合性质(定理 6.7).

在经典的图灵机计算模型上, 著名的 Jerrum-Valiant-Vazirani 定理(JVV 定理, 定理 3.2)证明, 对于有自归约性质的分布, 采样和计数问题可以在多项式时间内相互归约. 因此, 上述结论可以看成分布式版本的 JVJ 定理——采样问题和计数问题的计算等价性在分布式模型上依然成立.

考虑在最大度数 $\Delta \leq 5$ 的图上均匀采样独立集问题, 利用已有的强空间混合性质的结论^[4], 上述结论可以推出一个 $\text{polylog}(n)$ 轮的 LOCAL 算法. 而文献[18]中的下界结论指出: 当 $\Delta \geq 6$ 时, 不存在局部采样算法. 这就得到了第一个分布采样的计算相变现象, 这一结论与经典图灵机模型上的计算相变结论一致^[4,28].

本文第 2 节介绍基本定义和预备知识. 第 3 节回顾一些经典采样的结论以方便读者和分布式采样的结果作比较. 第 4 节定义分布式采样和计数问题. 第 5 节总结已有的分布式采样算法. 第 6 节总结已有的分布式采样计算复杂性结论. 第 7 节介绍本领域目前的公开问题.

2 定义与预备知识

2.1 图和概率分布的记号

- 图的记号

令 $G=(V,E)$ 是一个无向简单图. 对图上任意一个点 $v \in V$, 定义 $\Gamma_G(v)=\Gamma(v) \triangleq \{u \in V | \{u,v\} \in E\}$ 为 v 在图 G 上邻居的集合. 对于图上任意两个点 $u,v \in V$, 定义 $dist_G(u,v)$ 为 u 到 v 在 G 上最短路的距离; 对任意集合 $A \subseteq V$, 定义 $dist_G(v,A)=\min\{dist_G(v,u) | u \in A\}$ 为点 v 到集合 A 的距离.

- 概率分布的记号

令 V 为一个大小为 $n=|V|$ 随机变量的集合, 其中, 每个 $v \in V$ 在一个有限大小的集合 Σ 上取值, Σ 满足 $q=|\Sigma|=poly(n)$. 令 $\Omega=\Sigma^V$ 表示样本空间. 样本空间中的每个元素 $\sigma \in \Omega$ 称为一个配置. 对任意集合 $S \subseteq V$, 我们用 σ_S 或者 $\sigma(S)$ 表示 σ 在变量集合 S 上的配置.

令 μ 是定义在 $\Omega=\Sigma^V$ 上的一个联合分布. 每个随机向量 $Y \sim \mu$ 都是 n 个变量 $Y_v \in \Sigma$ 的联合分布. 对于任意随机变量的子集 $S \subseteq V$, 我们用 μ_S 表示 μ 在集合 S 上诱导出的边缘分布, 即:

$$\forall \tau \in \Sigma^S, \mu_S(\tau) = \Pr_{Y \sim \mu}[Y_S = \tau] = \sum_{\sigma \in \Omega: \sigma_S = \tau} \mu(\sigma).$$

特别地, 当 $S=\{v\}$ 时, 我们把 $\mu_{\{v\}}$ 写成 μ_v . 一个配置 $\sigma \in \Omega$ 相对于 μ 合法当且仅当 $\mu(\sigma) > 0$; 对于任意一个定义在子集 $A \subseteq V$ 上的部分配置 $\tau \in \Sigma^A$, τ 相对于 μ 合法当且仅当 $\mu_A(\tau) > 0$. 规定空集 \emptyset 上面的空配置 \emptyset 是一个合法配置. 给定子集 $A \subseteq V$ 上的合法部分配置 $\tau \in \Sigma^A$, 定义 μ^τ 为在 A 上的取值固定为 τ 的条件下, 分布 μ 诱导出的条件概率分布. 即:

$$\forall \sigma \in \Omega, \mu^\tau(\sigma) = \Pr_{Y \sim \mu}[Y = \sigma | Y_A = \tau].$$

相应地, 我们可以对任意子集 $S \subseteq V$, 定义 μ_S^τ 为 μ^τ 在集合 S 上诱导出的边缘概率分布.

全变差(total variation distance)是衡量两个分布相似程度的重要参数, 本文用 $d_{TV}(\cdot, \cdot)$ 表示两个分布的全变差. 严格地, 对于定义在 Ω 上的两种概率分布 ν 和 μ , 它们的全变差定义为

$$d_{TV}(\nu, \mu) = \frac{1}{2} \sum_{\sigma \in \Omega} |\nu(\sigma) - \mu(\sigma)|.$$

在上下文意义明确时, 本文会混用随机变量和随机变量对应的分布. 例如: 令 $X \in \Omega$ 是一个随机变量, 我们用 $d_{TV}(X, \mu)$ 表示随机变量 X 的分布和 μ 的全变差.

2.2 吉布斯分布

吉布斯分布(Gibbs distribution)是一种重要的联合分布, 在一些文献中也被称为概率图(probabilistic graphical model)或者因子图(factor graph)模型^[29]. 吉布斯分布是经典采样算法研究的关键对象之一. 在第 4 节分布式采样问题的定义中, 吉布斯分布用于对由局部约束定义的联合分布进行建模. 吉布斯分布的定义如下.

定义 2.1(吉布斯分布). 一个吉布斯分布 μ 由三元组 (V, Σ, \mathcal{F}) 定义, 其中, V 是一个变量的集合, Σ 是一个大小为 $q=|\Sigma| \leq poly(n)$ 的取值集合, \mathcal{F} 是一个约束(因子)的集合. 一个约束 $(f, S) \in \mathcal{F}$ 由一个作用域 $S \subseteq V$ 和一个非负约束函数 $f: \Sigma^S \rightarrow \mathbb{R}_{\geq 0}$ 构成. 如果 f 取值始终为正, 则约束 $(f, S) \in \mathcal{F}$ 被称为柔性约束(soft constraint); 否则, 约束 $(f, S) \in \mathcal{F}$ 被称为硬性约束(hard constraint).

每个配置 $\sigma \in \Sigma^V$ 的权重定义为

$$w(\sigma) \triangleq \prod_{(f,S) \in \mathcal{F}} f(\sigma_S).$$

吉布斯分布规定每种配置 $\sigma \in \Sigma^V$ 出现的概率正比于配置的权重, 即:

$$\mu(\sigma) \triangleq \frac{w(\sigma)}{Z},$$

其中, 归一化常数 $Z \triangleq \sum_{\tau \in \Sigma^V} w(\tau)$ 是吉布斯分布的配分函数(partition function).

在定义 2.1 中, 一个特殊情况是所有的约束函数都是取值为 $\{0,1\}$ 的布尔函数, 此时, 三元组 (V, Σ, \mathcal{F}) 定义了一个约束满足问题, 满足解是所有的合法配置. 吉布斯分布是所有满足解的均匀分布. 配分函数是所有满足解的个数.

空间马尔可夫性质(spatial Markovian)是吉布斯分布的重要性质, 它在很多文献中也被称为条件独立性

(conditional independence). 如下命题阐述了空间马尔可夫性质.

命题 2.1(空间马尔可夫性质^[29]). 令 μ 为一个被三元组 (V, Σ, \mathcal{F}) 定义的吉布斯分布. 令 $H=(V, \mathcal{F})$ 为一个超图, 超图的点集是变量集合 V , 边集 $F=\{S|(f, v) \in \mathcal{F}\}$ 是所有约束作用域的集合. 令 $A, B, C \subseteq V$ 是 3 个不相交的变量子集, 且把 C 从 H 中删除会使 A 和 B 不连通. 对于随机样本 $Y \sim \mu$, 当 C 被固定成任何一种合法的配置 $Y_C = \sigma_C$ 时, Y_A 和 Y_B 相互独立. 严格地, 对于任意合法的 $\sigma_C \in \Sigma^C$, 任意 $\sigma_A \in \Sigma^A$ 和 $\sigma_B \in \Sigma^B$, 都有如下结论成立:

$$\Pr_{Y \sim \mu}[Y_A = \sigma_A \wedge Y_B = \sigma_B | Y_C = \sigma_C] = \Pr_{Y \sim \mu}[Y_A = \sigma_A | Y_C = \sigma_C] \cdot \Pr_{Y \sim \mu}[Y_B = \sigma_B | Y_C = \sigma_C].$$

自旋系统 (spin system) 是一类定义吉布斯分布的重要概率模型, 它有时也被称为马尔可夫随机场 (Markov random field, MRF). 在一个自旋系统中, 所有约束要么定义在单个变量上, 要么定义在变量的二元组上. 自旋系统的严格定义如下.

定义 2.2(自旋系统). 令 $G=(V, E)$ 为一个无向简单图. 如果以下两个条件同时成立, 则 (V, Σ, \mathcal{F}) 是一个定义在 G 上的自旋系统.

- (1) 所有的约束定义在 G 的点和边上, $\mathcal{F} = \{(f, v) | v \in V\} \cup \{(f, e) | e \in E\}$;
- (2) 对于任意 $e = \{u, v\} \in E$, 函数 $f_e: \Sigma^e \rightarrow \mathbb{R}_{\geq 0}$ 是一个对称函数, 即: 对任意 $c_u, c_v \in \Sigma, f_e(c_u, c_v) = f_e(c_v, c_u)$.

在自旋系统中, 函数 f_v 称为点 v 的外场, 函数 f_e 称为 e 上的相互作用函数或相互作用矩阵.

自旋系统最早在物理学中用于对微观粒子系统进行建模^[29]. 目前, 它在机器学习和理论计算机等领域^[30,31]都有着重要的应用. 下面是 3 种重要的自旋系统.

- 列表染色/图染色 (list coloring/coloring): 一个列表染色模型由三元组 (V, E, L) 定义, 其中, $G=(V, E)$ 为一个简单无向图, $L=(L_v)_{v \in V}$ 为颜色列表的集合. 一个图的合法列表染色 $\sigma \in \prod_{v \in V} L_v$, 给每个节点颜色 $\sigma \in L_v$, 使得对于任意边 $e = \{u, v\} \in E, \sigma_u \neq \sigma_v$. 吉布斯分布 μ 是所有合法列表染色的均匀分布. 图的 q -染色是列表染色的一个特例, 它满足: 对于任意 $v \in V, L_v = [q] = \{1, 2, \dots, q\}$. 图的 q -染色用三元组 $(V, E, [q])$ 表示;
- 硬核模型 (hardcore model): 硬核模型由三元组 (V, E, λ) 定义, 其中, $G=(V, E)$ 为一张图, $\lambda \in \mathbb{R}_{\geq 0}$ 为逸度参数, 每个变量的取值范围为 $\{0, 1\}$. 硬核模型定义了 G 上所有独立集的加权分布. 一个配置 $\sigma \in \{0, 1\}^V$ 被称为独立集当且仅当 $I_\sigma = \{v \in V | \sigma_v = 1\}$ 是 G 上的一个独立集. 如果配置 $\sigma \in \{0, 1\}^V$ 是独立集, 则它的权重为 $w(\sigma) = \lambda^{|I_\sigma|}$; 如果 $\sigma \in \{0, 1\}^V$ 不是独立集, 则它的权重为 $w(\sigma) = 0$. 吉布斯分布 μ 满足 $\mu(\sigma) \propto w(\sigma)$;
- 伊辛模型 (Ising model): 伊辛模型由三元组 (V, E, β) 定义, 其中, $G=(V, E)$ 为一张简单无向图, $\beta \in \mathbb{R}$ 为温度参数, 每个变量的取值范围为 $\{-1, +1\}$. 对于任意一种配置 $\sigma \in \{-1, +1\}^V$, 它的权重定义为 $w(\sigma) = \exp(\beta \sum_{\{u, v\} \in E} \sigma_u \sigma_v)$. 吉布斯分布 μ 满足 $\mu(\sigma) \propto w(\sigma)$. 当 $\beta > 0$ 时, 此模型被称为铁磁伊辛模型; 当 $\beta < 0$ 时, 此模型被称为反铁磁伊辛模型.

3 经典采样理论回顾

本节回顾一些经典的采样理论, 方便后文与分布式采样理论作对比.

3.1 采样和计数

采样和计数是两个密切相关的重要的计算问题. 我们考虑一般联合分布上的采样和计数问题. 令 V 为一个随机变量的集合, Σ 是每个随机变量的取值范围, \mathbf{x} 是一个编码联合分布的比特串. 三元组 (V, Σ, \mathbf{x}) 定义了一个权重函数 $w = w_{(V, \Sigma, \mathbf{x})}: \Omega \rightarrow \mathbb{R}_{\geq 0}$, 每个配置 $\sigma \in \Omega$ 的权重定义为 $w(\sigma)$. 三元组 (V, Σ, \mathbf{x}) 确定了一个唯一的联合分布 $\mu = \mu_{(V, \Sigma, \mathbf{x})}$, 对于任意 $\sigma \in \Omega$:

$$\mu(\sigma) = \frac{w(\sigma)}{Z},$$

其中, 归一化因子 $Z = Z_{(V, \Sigma, \mathbf{x})}$ 称为联合分布的配分函数. 特别地, 当 (V, Σ, \mathbf{x}) 编码了一个吉布斯分布 (定义 2.1) 时, 比特串 \mathbf{x} 编码了吉布斯分布所有的约束.

给定一个联合分布的输入 (V, Σ, \mathbf{x}) , 采样问题要求从 μ 中生成随机样本. 经典的采样理论考虑如下问题.

- 近似采样问题: $ApproxSample(V, \Sigma, \mathbf{x}, \varepsilon)$:
 - 输入: 一个描述联合分布的输入 (V, Σ, \mathbf{x}) , 一个误差参数 $\varepsilon > 0$;
 - 输出: 一个随机样本 $X \in \Sigma^V$, 它满足 $d_{TV}(X, \mu) \leq \varepsilon$;
- 精确采样问题: $ExactSample(V, \Sigma, \mathbf{x}, \delta)$:
 - 输入: 一个描述联合分布的输入 (V, Σ, \mathbf{x}) , 一个误差参数 $\delta > 0$;
 - 输出: 一个随机样本 $X \in \Sigma^V$ 以及一个指示变量 $F \in \{0, 1\}$, 在 $F=0$ 的条件下, $X \sim \mu$ 且 $\Pr[F=0] \geq 1 - \delta$;
- 零误差精确采样问题: $ZeroErrorSample(V, \Sigma, \mathbf{x})$:
 - 输入: 一个描述联合分布的输入 (V, Σ, \mathbf{x}) ;
 - 输出: 一个随机样本 $X \sim \mu$.

经典的采样理论有一系列算法, 近似采样算法有马尔可夫链蒙特卡洛(Markov chain Monte Carlo, MCMC)算法^[32], MCMC 算法是研究最为深刻、应用最为广泛的算法, 我们会在第 3.2 节简要介绍 MCMC 算法的背景知识. 相对而言, 精确采样算法的结论较为有限. 对于精确采样问题 $ExactSample$, 常见的算法有 JVV 采样算法(定理 3.1)、Fill 算法^[33]、随机性回收器(randomness recycler)^[34]以及局部拒绝采样算法^[23,25,26,35,36], 独立重复这些算法多次, 它们都能用于解决 $ZeroErrorSample$ 问题. 另一类重要的精确采样算法是 Coupling from the past (CFTP)^[37], 例如著名的 bounding chain^[16,38]. CFTP 算法有随机的运行时间, 可以用于解决零误差精确采样问题 $ZeroErrorSample$. 文献[39]系统地介绍了精确采样算法.

计数问题是一个与采样密切相关的问题. 给定一个描述联合分布的输入 (V, Σ, \mathbf{x}) , 计数问题要求计算配分函数 Z . 很多精确的计数问题是 #P 难问题^[40]. 经典计数理论主要考虑如下定义的两类近似计数问题.

- 确定近似计数问题: $DetCount(V, \Sigma, \mathbf{x}, \varepsilon)$:
 - 输入: 一个描述联合分布的输入 (V, Σ, \mathbf{x}) , 一个近似参数 $\varepsilon > 0$;
 - 输出: 一个配分函数的近似值 \tilde{Z} , 它满足 $\exp(-\varepsilon)Z \leq \tilde{Z} \leq \exp(\varepsilon)Z$;
- 随机近似计数问题: $RanCount(V, \Sigma, \mathbf{x}, \varepsilon, \delta)$:
 - 输入: 一个描述联合分布的输入 (V, Σ, \mathbf{x}) , 一个近似参数 $\varepsilon > 0$ 以及一个误差参数 $\delta > 0$;
 - 输出: 一个配分函数的随机近似值 \tilde{Z} , 它满足 $\Pr[\exp(-\varepsilon)Z \leq \tilde{Z} \leq \exp(\varepsilon)Z] \geq 1 - \delta$.

在一些文献中, 随机的近似计数问题直接定义为 $RanCount(V, \Sigma, \mathbf{x}, \varepsilon, 1/4)$ ^[41]. 这是因为, 如果存在一个能解决 $RanCount(V, \Sigma, \mathbf{x}, \varepsilon, 1/4)$ 的算法 \mathcal{A} , 只需要把 \mathcal{A} 独立重复地运行 $\text{poly}(|V|, |\Sigma|, |\mathbf{x}|) \log 1/\delta$ 次, 再取出所有输出的中位数就能解决 $RanCount(V, \Sigma, \mathbf{x}, \varepsilon, \delta)$.

经典的 Jerrum-Valiant-Vazirani 定理(JVV 定理)^[2]指出, 在图灵机模型上, 如果一类联合分布有自归约性质, 则在这类分布上的采样与计数问题在多项式时间内可以相互归约.

定义 3.1(多项式时间自归约性). 令 \mathfrak{M} 为一个联合分布类. 如果 \mathfrak{M} 满足如下两个性质, 则 \mathfrak{M} 具有多项式时间自归约性质.

- (1) 对于任意联合分布 $\mu_{(V, \Sigma, \mathbf{x})} \in \mathfrak{M}$ 、任意集合 $A \subseteq V$ 、任意合法配置 $\tau \in \Sigma^A$, 都存在一个联合分布 $\mu_{(V, \Sigma, \mathbf{y})} \in \mathfrak{M}$ 且满足 $\mu_{(V, \Sigma, \mathbf{y})}$ 是条件概率分布 $\mu_{(V, \Sigma, \mathbf{x})}^\tau$, 即在 A 上的取值固定为 τ 时, $\mu_{(V, \Sigma, \mathbf{x})}$ 在 Σ^V 上诱导出的条件分布;
- (2) 给定任意联合分布 $\mu_{(V, \Sigma, \mathbf{x})} \in \mathfrak{M}$ 、任意集合 $A \subseteq V$ 、任意合法配置 $\tau \in \Sigma^A$, 性质(1)中描述的输入实例 (V, Σ, \mathbf{y}) 可以在 $\text{poly}(|V|, |\Sigma|, |\mathbf{x}|)$ 的时间内计算.

给定一个联合分布类后, 可以定义在联合分布类上的高效采样和计数算法. 根据采样和计数问题的不同要求, 可以定义出以下 4 类多项式时间采样/计数算法.

定义 3.2(联合分布类上的多项式采样/计数算法). 令 \mathfrak{M} 为一个联合分布类:

- (1) 全多项式时间近似计数算法: 给定任意 $\mu_{(V, \Sigma, \mathbf{x})} \in \mathfrak{M}$ 、任意 $\varepsilon > 0$, 算法以 $\text{poly}(|V|, |\Sigma|, |\mathbf{x}|, 1/\varepsilon)$ 的时间复杂度解决确定近似计数问题 $DetCount(V, \Sigma, \mathbf{x}, \varepsilon)$;
- (2) 全多项式时间随机近似计数算法: 给定任意 $\mu_{(V, \Sigma, \mathbf{x})} \in \mathfrak{M}$ 、任意 $\varepsilon, \delta > 0$, 算法以 $\text{poly}(|V|, |\Sigma|, |\mathbf{x}|, 1/\varepsilon, \log 1/\delta)$ 的时间复杂度解决随机近似计数问题 $RanCount(V, \Sigma, \mathbf{x}, \varepsilon, \delta)$;

- (3) 全多项式时间近似采样算法: 给定任意 $\mu_{(V, \Sigma, \mathbf{x})} \in \mathfrak{M}$ 、任意 $\varepsilon > 0$, 算法以 $\text{poly}(|V|, |\Sigma|, |\mathbf{x}|, 1/\varepsilon)$ 的时间复杂度解决确定近似采样问题 $\text{ApproxSample}(V, \Sigma, \mathbf{x}, \varepsilon)$;
- (4) 全多项式时间精确采样算法: 给定任意 $\mu_{(V, \Sigma, \mathbf{x})} \in \mathfrak{M}$ 、任意 $\delta > 0$, 算法以 $\text{poly}(|V|, |\Sigma|, |\mathbf{x}|, 1/\delta)$ 的时间复杂度解决确定近似采样问题 $\text{ExactSample}(V, \Sigma, \mathbf{x}, \delta)$.

在很多英文文献中, 全多项式时间近似计数算法被称为(fully polynomial-time approximation scheme, FPTAS), 全多项式时间随机近似计数算法被称为(fully polynomial-time randomized approximation scheme, FPRAS); 当联合分布为均匀分布时, 全多项式时间近似采样算法被称为(fully polynomial-time almost uniform sampler, FPAUS)^[41].

JVV 定理指出了采样和计数问题的著名等价关系.

定理 3.2(Jerrum-Valiant-Vazirani 定理^[2]). 令 \mathfrak{M} 为以一个有多项式时间自归约性质的联合分布类. 假设对于任意 $\mu_{(V, \Sigma, \mathbf{x})} \in \mathfrak{M}$, 其所对应的权重函数 $w_{(V, \Sigma, \mathbf{x})}$ 可以在多项式时间内计算, 则有:

- (1) 如果 \mathfrak{M} 上存在全多项式时间近似计数算法, 则 \mathfrak{M} 上存在全多项式时间精确采样算法, 因此也存在全多项式时间近似采样算法;
- (2) 如果 \mathfrak{M} 上存在全多项式时间随机近似计数算法, 则 \mathfrak{M} 上存在全多项式时间近似采样算法;
- (3) 如果 \mathfrak{M} 上存在全多项式时间近似采样算法或全多项式时间精确采样算法, 则 \mathfrak{M} 上存在全多项式时间随机近似计数算法.

根据 JVJ 定理, 解决近似计数问题的一个有力工具是高效的采样算法. 基于这个技术, 人们成功地解决了很多著名的近似计数问题. 例如, 近似非负矩阵积和式(permanent)^[42]、求解组合数学的经典计数问题^[43]、近似物理模型的配分函数^[44]等. 在 JVJ 定理提出之后, 人们又发现了基于模拟退火的归约算法^[45-48], 相对于原始 JVJ 定理中的归约, 模拟退火算法可以更高效地把采样算法转化为计数算法.

确定近似计数问题也有一些专门的技术, 例如, 基于相关性衰减的递归算法^[4, 49-51]、多项式插值算法^[52, 53]以及线性规划算法^[54]. 利用 JVJ 定理, 这些算法也可以得出相应模型的精确采样算法.

3.2 马尔可夫链蒙特卡洛算法

马尔可夫链蒙特卡洛算法(MCMC 算法)是研究最为广泛的采样算法. 给定一个输入 (V, Σ, \mathbf{x}) , 它定义了样本空间 \mathcal{S}^V 上的联合分布 $\mu = \mu_{(V, \Sigma, \mathbf{x})}$. MCMC 算法设计了一个 \mathcal{S}^V 上的马尔可夫链 $(X_t)_{t \geq 0}$, 使得 μ 是马尔可夫链的唯一平稳分布; 然后模拟这个马尔可夫链 T 步并输出 X_T , 当 T 足够大时, X_T 的分布就能足够接近目标分布 μ . 马尔可夫链蒙特卡洛方法在很多问题上取得了成功的应用, 著名的例子有估计凸体体积^[55]、估计非负矩阵行列式^[42]、计算伊辛模型配分函数^[44]、采样图染色^[17, 56, 57]和采样硬核模型^[58]等.

本节简要介绍一些马尔可夫链的基本概念, 更多马尔可夫链的背景知识可参考相关教材^[32]. 令 $(X_t)_{t \geq 0}$ 是一个空间 Ω 上的马尔可夫链, 令 $P: \Omega \times \Omega \rightarrow \mathbb{R}_{\geq 0}$ 为马尔可夫链的转移矩阵, 它满足 $\forall \sigma, \tau \in \Omega, \Pr[X_t = \tau | X_{t-1} = \sigma] = P(\sigma, \tau)$. 本文经常用转移矩阵 P 指代其所对应的马尔可夫链. 一个马尔可夫链 P 是不可约的(irreducible)当且仅当对任意 $x, y \in \Omega$, 存在 $t \geq 0$, 使得 $P^t(x, y) > 0$; 不可约的链 P 是非周期的(aperiodic)当且仅当对任意 $x \in \Omega, \gcd\{t > 0 | P^t(x, x) > 0\} = 1$. 一个 Ω 上的分布 π 视为一个行向量是马尔可夫链 P 的平稳分布当且仅当 $\pi P = \pi$. 如果马尔可夫链 P 是不可约且非周期的, 那么 P 有唯一平稳分布. 马尔可夫链 P 相对于 Ω 上的分布 π 可逆(reversible)当且仅当细致平衡方程(detailed balance equation)成立:

$$\forall \sigma, \tau \in \Omega, \pi(\sigma)P(\sigma, \tau) = \pi(\tau)P(\tau, \sigma).$$

可逆性质可以推出 π 是 P 的一个平稳分布.

本文介绍两种经典的 MCMC 算法.

第 1 种是吉布斯采样(Gibbs sampling)算法, 它有时又称为格劳伯动态(Glauber dynamics). 给定一个由输入 (V, Σ, \mathbf{x}) 定义的联合分布 $\mu = \mu_{(V, \Sigma, \mathbf{x})}$. 吉布斯采样算法从任意一个合法配置 $X \in \mathcal{S}^V$ 开始, 每步转移执行如下操作.

1. 随机等概率选择一个点 $v \in V$;

- 重新采样 $X_v \sim \mu_v^{X_{V \setminus \{v\}}}$, 其中, $\mu_v^{X_{V \setminus \{v\}}}$ 表示在 $V \setminus \{v\}$ 上所有变量的取值固定为 $X_{V \setminus \{v\}}$ 的条件下, μ 在 v 上诱导出的条件边缘概率分布.

吉布斯采样算法重复 T 步转移之后输出当前随机配置 X .

第 2 种重要的 MCMC 算法是梅特罗波利斯算法(Metropolis 算法). 实际上, 梅特罗波利斯算法是一类 MCMC 算法的统称, 这类算法在每次转移时, 先从某个概率分布中提出一个随机候选值, 再以一定概率接受或者拒绝候选值. 本节介绍一种常见的梅特罗波利斯算法. 给定一个由输入 (V, Σ, \mathbf{x}) 定义的联合分布 $\mu = \mu_{(V, \Sigma, \mathbf{x})}$. 梅特罗波利斯算法从任意一个合法配置 $X \in \Sigma^V$ 开始, 每步转移执行如下操作.

- 随机等概率选择一个点 $v \in V$;
- 随机等概率地从 Σ 中采样一个候选值 c , 令 $X' \in \Sigma^V$ 为把 X 在 v 上的取值改成 c 之后得到的配置;
- 以概率 $\min\{1, \mu(X')/\mu(X)\}$ 接受候选值并令 $X_v \leftarrow c$; 以概率 $1 - \min\{1, \mu(X')/\mu(X)\}$ 拒绝候选值并保持 X_v 不变.

梅特罗波利斯算法重复 T 步转移之后输出当前随机配置 X .

如下定理说明吉布斯采样算法和梅特罗波利斯算法的正确性.

定理 3.3(吉布斯采样算法和梅特罗波利斯算法正确性^[32]). 令 P_G 为吉布斯采样算法定义的马尔可夫链, P_M 为梅特罗波利斯算法定义的马尔可夫链. P_G 和 P_M 都相对于目标分布 μ 可逆, 且对于任意合法配置 σ , 都有 $P_G(\sigma, \sigma) > 0, P_M(\sigma, \sigma) > 0$. 因此, 对于 P_G 和 P_M , 只要马尔可夫链不可约, 则 μ 是马尔可夫链的唯一平稳分布.

对于很多采样问题, 很容易证明马尔可夫链最终可以收敛到目标分布, 但是分析收敛速度却需要非常艰深的理论分析工具. 假设马尔可夫链 $(X_t)_{t \geq 0}$ 有唯一平稳分布 μ , 马尔可夫链的混合时间(mixing time)定义为

$$T(\varepsilon) \triangleq \max_{X_0} \min\{t \mid d_{TV}(X_t, \mu) \leq \varepsilon\}.$$

混合时间表示从最坏的初始状态开始, 马尔可夫链需要转移多少步才与平稳分布足够接近. 混合时间的分析是经典采样理论和分布式采样理论中的重要课题.

4 分布式采样和计数问题

4.1 一般联合分布的分布式采样和计数问题

我们考虑 LOCAL 模型上的分布式采样和计数问题. LOCAL 模型的定义见第 1 节. 我们用 (G, Σ, \mathbf{x}) 表示分布式采样/计数问题实例, 其中, $G=(V, E)$ 是一个 $n=|V|$ 个点的简单无向图, 它表示 LOCAL 模型的通信网络; Σ 是一个大小为 $q=|\Sigma|=poly(n)$ 的取值范围; $\mathbf{x}=(x_v)_{v \in V}$ 是一个 n 维向量, 每一个 x_v 是一个比特串. 实例 (G, Σ, \mathbf{x}) 唯一编码一个样本空间 Σ^V 上的联合分布 $\mu = \mu_{(G, \Sigma, \mathbf{x})}$. 我们称 μ 为实例 (G, Σ, \mathbf{x}) 的目标分布.

给定一个问题实例 (G, Σ, \mathbf{x}) , 令 $G=(V, E)$ 为一个 LOCAL 模型的通信网络, 每个点 $v \in V$ 的输入包含取值范围 Σ 、输入的比特串 x_v . 除去目标分布的信息, 每个 x_v 中还编码了一个 v 的唯一身份 UID_v , 一个全局所有点数 n 的上界 $n \leq N \leq poly(n)$, 如果是近似采样/计数问题, 每个点的输入还额外包含了一个全局误差上界.

分布式算法的运行时间为通信的总轮数. 本文主要考虑有固定运行时间上界的分布式算法. 我们要求分布式算法以高概率成功, 且所有的失败都可以被算法局部地发现. 严格地, 在分布式算法运行结束之后, 每个点 $v \in V$ 输出一个随机值 $Y_v \in \Sigma$ 以及一个随机比特 $F_v \in \{0, 1\}$, 指示点 v 处的算法是否失败. 如果对于所有 $v \in V$ 都有 $F_v = 0$, 则分布式算法运行成功; 否则算法失败, 所有随机比特满足 $\mathbb{E}(\sum_{v \in V} F_v) = O(1/n)$.

本文主要考虑如下两类分布式采样问题.

- 分布式精确采样: 给定任意问题实例 (G, Σ, \mathbf{x}) , 算法在成功时输出随机向量 $Y=(Y_v)_{v \in V}$, 每个点 $v \in V$ 输出 Y_v 或者算法失败. 在所有点都成功的条件下, Y 的分布恰好是 $\mu_{(G, \Sigma, \mathbf{x})}$;
- 分布式近似采样: 给定任意问题实例 (G, Σ, \mathbf{x}) 以及任意 $\varepsilon > 0$, 算法在成功时输出随机向量 $Y=(Y_v)_{v \in V}$. μ_{alg} 为在所有点都成功的条件下 Y 的概率分布, 问题要求满足 $d_{TV}(\mu_{alg}, \mu_{(G, \Sigma, \mathbf{x})}) \leq \varepsilon$.

计数问题是一个全局性的问题, 不适合用分布式算法计算. 分布式版本的计数问题是推断(inference), 即

计算变量的边缘概率分布. 一个变量的边缘分布实际上包含了一定的整个分布的信息. 在 JVV 定理从计数问题到采样问题的归约中, 计数问题被分解成一系列估计边缘概率的问题. 对于存在非零远距离相关性的概率分布, 由于分布式算法信息的局部性, 精确推断不可能被完成. 我们考虑如下近似推断问题.

- 分布式近似推断

给定任意问题实例 (G, Σ, \mathbf{x}) 以及任意误差 $\varepsilon > 0$, 算法结束时, 每个点输出一个 Σ 上的边缘分布 $\tilde{\mu}_v$, 或者算法失败, 如果点 v 上的算法成功, 则有 $d_{TV}(\tilde{\mu}_v, \mu_{v,(G,\Sigma,\mathbf{x})}) \leq \varepsilon$, 其中, $\mu_{v,(G,\Sigma,\mathbf{x})}$ 表示目标分布 $\mu_{(G,\Sigma,\mathbf{x})}$ 投影到 v 上的边缘分布.

4.2 由局部约束定义的联合分布

由局部约束定义的联合分布是分布式采样和计数问题的重要研究对象. 这类联合分布一般用局部吉布斯分布进行建模. 根据定义 2.1, 一个吉布斯分布由 $\mu = \mu_{(V,\Sigma,\mathcal{F})}$ 三元组 (V, Σ, \mathcal{F}) 确定, \mathcal{F} 是一系列约束 (S, f) 的集合, $S \subseteq V$ 是作用域, $f: \Sigma^S \rightarrow \mathbb{R}_{\geq 0}$ 是约束函数. 每种配置出现的概率为 $\mu(\sigma) \propto \prod_{(f,S) \in \mathcal{F}} f(\sigma_S)$.

定义 4.1(局部吉布斯分布). 令 $G=(V,E)$ 为一个 LOCAL 模型的通信网络, 三元组 (V, Σ, \mathcal{F}) 定义了一个吉布斯分布 $\mu = \mu_{(V,\Sigma,\mathcal{F})}$. 如果对于任意 $(S, f) \in \mathcal{F}$, 集合 S 在 G 上的直径 $\max_{u,v} \text{dist}_G(u,v) = O(1)$, 则 $\mu_{(V,\Sigma,\mathcal{F})}$ 是通信网络 $G=(V,E)$ 上的局部吉布斯分布.

考虑局部吉布斯分布的分布式采样和计数问题. 如果问题实例 (G, Σ, \mathbf{x}) 编码的分布是一个局部吉布斯分布 (V, Σ, \mathcal{F}) , 则对于所有 $v \in V$, 比特串 x_v 中包含了所有的满足 $v \in S$ 的约束 $(S, f) \in \mathcal{F}$.

吉布斯分布的一类重要特例是自旋系统(定义 2.2). 令三元组 (V, Σ, \mathcal{F}) 为一个 $G=(V,E)$ 上的自旋系统, 则它的约束集合为 $\mathcal{F} = \{(f_v, v) | v \in V\} \cup \{(f_e, e) | e \in E\}$. 对于自旋系统上的分布式采样和计数问题, 很多文献考虑如下自然的定义: 令分布式计算网络 G 为自旋系统所在的图 G . 每个点 $v \in V$ 输入的比特串 x_v 中包含函数 f_v 以及满足 $v \in e$ 的所有函数 f_e .

例如, 考虑列表染色的分布式采样和计数问题. 每个点 $v \in V$ 输入的比特串 x_v 中包含 v 可以使用的颜色列表以及所有与 v 相关的边 $e \ni v$ 上的约束, 即 e 的两个端点不能取相同的颜色.

5 分布式采样算法

在最近几年的研究中, 针对第 4.1 节提出的分布式采样问题, 人们提出了一些有严格理论保障的分布式采样算法. 文献[18]首次研究了这类采样问题, 并针对自旋系统的分布式采样提出了卢比-格劳伯算法和局部梅特罗波利斯算法. 之后, 文献[22]和文献[21]通过引入分布式破对称性操作(distributed symmetry breaking), 改进了原始的局部梅特罗波利斯算法. 文献[24]研究在分布式计算模型上正确并高效地模拟串行梅特罗波利斯采样算法, 从而证明了一些串行采样算法的结论对分布式采样依然成立. 文献[26]在研究洛瓦兹局部引理(Lovász local lemma)采样问题时, 提出了局部拒绝采样算法(partial rejection sampling, PRS), PRS 类算法^[23,26]也可用于一类局部吉布斯分布的分布式采样. 本节的所有算法不仅在 LOCAL 模型上有较好的复杂度, 而且应用到很多具体模型上时, 算法不需要使用 LOCAL 模型不受限制的本地计算能力.

5.1 卢比-格劳伯算法

令三元组 (V, Σ, \mathcal{F}) 为一个 $G=(V,E)$ 上的自旋系统, 则它的约束集合为 $\mathcal{F} = \{(f_v, v) | v \in V\} \cup \{(f_e, e) | e \in E\}$, 其中, 所有 f_e 是对称函数. 经典的串行马尔可夫链蒙特卡罗算法每次随机选择一个点 $v \in V$, 每次转移基于 v 邻居的状态更新 v 的取值. 一个自然的并行策略就是每次更新图 G 上的一个独立集.

卢比-格劳伯算法就是一种对吉布斯采样算法(又称为格劳伯动态)的并行化方法. 一般联合分布的吉布斯采样算法描述在第 3.2 节, 把吉布斯采样算法应用到自旋系统上可以得到算法 1.

算法 1. 自旋系统的吉布斯采样算法.

输入: 一个 $G=(V,E)$ 上的吉布斯分布为 $\mu = \mu_{(V,\Sigma,\mathcal{F})}$ 的自旋系统 (V, Σ, \mathcal{F}) , 一个整数 T ;

- 1 初始 X 取 Σ^V 上的任意一种配置(可能不合法);
- 2 **for** $t=1$ 到 T **do**
- 3 随机等概率选择一个点 $v \in V$;
- 4 重新采样 $X_v \sim \mu_v^{X_{\Gamma(v)}}$
- 5 **return** X ;

算法 1 允许初始配置不合法. 每次转移时, 利用吉布斯分布的条件独立性(命题 2.1), 每次转移只需在 v 所有的邻居 Γ_v 取当前值的条件下, 利用条件分布更新 v 的取值. 严格地, 算法 1 第 4 行的条件概率为

$$\forall v \in V, \sigma \in \Sigma^{\Gamma_v}, c \in \Sigma, \mu_v^\sigma(c) = \frac{f_v(c) \prod_{u \in \Gamma_v} f_{uv}(c, \sigma_u)}{\sum_{b \in \Sigma} f_v(b) \prod_{u \in \Gamma_v} f_{uv}(b, \sigma_u)} \quad (1.1)$$

因为允许初始配置不合法, 我们对吉布斯分布做如下假设:

$$\forall v \in V, \sigma \in \Sigma^{\Gamma_v}, \sum_{b \in \Sigma} f_v(b) \prod_{u \in \Gamma_v} f_{uv}(b, \sigma_u) > 0,$$

则算法 1 中的转移概率良定义, 且算法 1 以概率 1 收敛到合法配置. 上述假设对所有由柔性约束定义的吉布斯分布都成立; 对很多由硬性约束定义的吉布斯分布也都成立, 例如硬核模型以及 $q \geq \Delta + 1$ 时的图的 q -染色问题(其中, Δ 是图的最大度数)等.

卢比-格劳伯算法利用如下方法并行吉布斯采样算法, 在每次转移的过程中,

1. 构造图 G 上的一个随机独立集 $I \subseteq V$;
2. 对于独立集上的每个点 $v \in I$, 重新采样 $X_v \sim \mu_v^{X_{\Gamma(v)}}$.

卢比-格劳伯算法借用卢比算法(Luby algorithm)的思想构造独立集 $I \subseteq V$. 每个点 $v \in V$ 随机均匀地采样一个随机实数 $\beta_v \in [0, 1]$, 如果一个点 v 的 β_v 大于所有邻居 $u \in \Gamma_v$ 的 β_u , 则 v 进入独立集 I . 之后, I 中的每个点并行地按照格劳伯动态的规则更新取值. 卢比-格劳伯算法的伪代码在算法 2 中给出.

算法 2. 卢比-格劳伯算法.

输入: 每个点 $v \in V$ 收到值域 Σ , 函数 f_v 以及满足 $v \in e$ 的所有函数 f_e , 一个整数参数 T ;

- 1 每个点 $v \in V$ 把 X_v 设置为 Σ 中任意一个值;
- 2 **for** $t=1$ 到 T **do**
- 3 每个点 v 独立均匀地采样一个随机实数 $\beta_v \in [0, 1]$;
- 4 **foreach** 点 v 满足 $\beta_v \geq \max_{u \in \Gamma(v)} \beta_u$ **do**
- 5 重新采样 $X_v \sim \mu_v^{X_{\Gamma(v)}}$;
- 6 **return** X ;

算法 2 第 5 行的概率分布 $\mu_v^{X_{\Gamma(v)}}$ 可以按照公式(1.1)收集局部信息计算, 所以算法 2 中每次 for-循环可以用 $O(1)$ 轮的代价实现. 文献[18]分析了卢比-格劳伯算法的正确性和收敛性. 卢比-格劳伯算法定义了一个 Σ^V 上的马尔可夫链 $(X_t)_{t \geq 0}$, 每个 $X_t \in \Sigma^V$ 表示第 t 次循环后产生的状态.

定理 5.1(卢比-格劳伯算法正确性^[18]). 对于任意自旋系统 (V, Σ, \mathcal{F}) , 如果串行吉布斯采样算法(算法 1)从任意初始状态出发, 最终一定收敛到合法配置且在合法配置的空间上不可约, 则卢比-格劳伯算法 $(X_t)_{t \geq 0}$ 满足对于任意初始状态 $X_0 \in \Sigma^V$, 当 $t \rightarrow \infty$ 时, X_t 的概率分布一定收敛到目标吉布斯分布 $\mu_{(V, \Sigma, \mathcal{F})}$.

定义 5.1(多布鲁申(Dobrushin)条件^[59]). 令 (V, Σ, \mathcal{F}) 为一个自旋系统, $\mu = \mu_{(V, \Sigma, \mathcal{F})}$ 为其所定义的吉布斯分布. 对任意 $u, v \in V$, 定义 v 对 u 的影响为

$$\rho(u, v) \triangleq \max_{(\sigma, \tau) \in S_v} d_{TV}(\mu_v^\sigma, \mu_v^\tau).$$

集合 S_v 包含了所有的配置二元组 $(\sigma, \tau) \in \Sigma^V \times \Sigma^V$ 满足 σ 和 τ 只在 v 一处取值不同. 定义全影响 α 为

$$\alpha = \max_{v \in V} \sum_{u \in \Gamma(v)} \rho(u, v).$$

如果 $\alpha < 1$, 则自旋系统满足多布鲁申条件.

定理 5.2(卢比-格劳伯算法收敛性^[18]). 在定理 5.1 的假设下, 如果输入的自旋系统 (V, Σ, \mathcal{F}) 满足多布鲁申条件, 则卢比-格劳伯算法以 $O\left(\frac{\Delta}{1-\alpha} \log \frac{1}{\varepsilon}\right)$ 轮的时间返回一个随机样本 $X \in \Sigma^V$ 满足 $d_{TV}(X, \mu_{(V, \Sigma, \mathcal{F})}) \leq \varepsilon$, 其中, $n=|V|$, Δ 是自旋系统所在图的最大度数, $O(\cdot)$ 记号隐藏了一个绝对常数因子.

多布鲁申条件是串行吉布斯采样算法的重要收敛条件^[59,60], 定理 5.2 说明, 分布式的卢比-格劳伯算法也能在同样的条件下收敛. 例如, 对对应到图的 q 染色问题, 如果 $q \geq (2+\delta)\Delta$, 其中, $\delta > 0$ 是一个常数, 卢比-格劳伯算法的运行时间为 $O(\Delta \log n)$. 卢比-格劳伯算法的主要缺陷是运行时间正比于最大度数 Δ , 所以它在稠密图上效率偏低. 运行时间里的 Δ 因子是不可避免的, 这是因为算法每次只能更新一个独立集, 所以算法需要 $O(\Delta)$ 轮才能把图上所有点都更新一遍. 在 $G=(V, E)$ 是一个完全图的极端情况下, 卢比-格劳伯算法就直接退化成串行的吉布斯采样算法.

卢比-格劳伯算法的思想也被一些分布式机器学习的研究使用过^[10], 它与吉布斯采样算法的系统性扫描 (systematic scan)^[61] 也有一定联系, 但是这类技术无法在一般图上获得高效的分布式算法.

5.2 局部梅特罗波利斯算法

令三元组 (V, Σ, \mathcal{F}) 为一个 $G=(V, E)$ 上的自旋系统, 则它的约束集合为 $\mathcal{F} = \{(f_v, v) | v \in V\} \cup \{(f_e, e) | e \in E\}$, 其中, 所有 f_e 是对称函数. 我们可以对所有约束函数做如下归一化操作.

- 对所有 $v \in V$, 定义:

$$\forall c \in \Sigma, \tilde{f}_v(c) = \frac{f_v(c)}{\sum_{b \in \Sigma} f_v(b)} \tag{1.2}$$

- 对所有 $e \in E$, 定义:

$$\forall c, c' \in \Sigma, \tilde{f}_e(c, c') = \frac{f_e(c, c')}{\max_{b, b' \in \Sigma} f_e(b, b')} \tag{1.3}$$

根据定义 2.2, 三元组 $(V, \Sigma, \tilde{\mathcal{F}})$ 和三元组 (V, Σ, \mathcal{F}) 定义了同一个吉布斯分布, 其中,

$$\tilde{\mathcal{F}} = \{(\tilde{f}_v, v) | v \in V\} \cup \{(\tilde{f}_e, e) | e \in E\}.$$

我们把这个自旋系统定义的吉布斯分布记为 $\mu = \mu_{(V, \Sigma, \mathcal{F})} = \mu_{(V, \Sigma, \tilde{\mathcal{F}})}$.

算法 3 是一种自旋系统上自然的梅特罗波利斯采样算法.

算法 3 从任意一种可能不合法的初始状态 $X \in \Sigma^V$ 出发, 依次执行 T 步更新, 每一步更新随机、等概率地选择一个变量 $v \in V$, 按照公式(1.2)中定义的概率分布 \tilde{f}_v 采样一个候选值; 然后, 点 v 以概率 $\prod_{u: \{u, v\} \in E} \tilde{f}_e(X_u, c_v)$ 接受候选值并执行更新 $X_v \leftarrow c$; 以剩下 $1 - \prod_{u: \{u, v\} \in E} \tilde{f}_e(X_u, c_v)$ 的概率拒绝候选值并保持 X_v 不变. 例如, 对于图染色模型, 被选中的点 $v \in V$ 从 $[q]$ 中随机、等概率地选择一个候选颜色 c_v , 如果对于 v 的所有邻居 $u \in \Gamma_v$ 都有 $c_v \neq X_u$, 则 v 接受候选颜色并执行更新 $X_v \leftarrow c_v$; 否则, v 拒绝候选颜色并保持 X_v 不变.

算法 3. 一种自旋系统的梅特罗波利斯算法.

输入: 一个 $G=(V, E)$ 上的吉布斯分布为 $\mu = \mu_{(V, \Sigma, \mathcal{F})}$ 的自旋系统 (V, Σ, \mathcal{F}) ;

- 1 初始 X 取 Σ^V 上的任意一种配置(可能不合法);
- 2 **for** $t=1$ 到 T **do**
- 3 随机、等概率地选择一个点 $v \in V$;
- 4 从概率分布 \tilde{f}_v 中随机采样一个候选值 $c_v \in \Sigma$;
- 5 以概率 $\prod_{u: \{u, v\} \in E} \tilde{f}_e(X_u, c_v)$ 执行更新 $X_v \leftarrow c_v$; 否则, 保持 X_v 不变;
- 6 **return** X ;

算法 3 中的梅特罗波利斯采样算法是一个串行算法. 文献[18]给出了第一个分布式版本的梅特罗波利斯采样算法——local-Metropolis 算法. 之后, 文献[22]和文献[21]以图染色模型为例, 对文献中的算法引入了一

步分布式破对称性操作,从而得到了更加高效的采样算法.实际上,文献[21,22]中的算法可以推广到一般自旋系统.本文把这一类算法统称为局部梅特罗波利斯采样算法.注意:在文献[21]中,这个算法被称为 lazy local-Metropolis 算法;而在文献[22]中,这个算法被称为 local Glauber dynamics 算法.

算法 4 给出了局部梅特罗波利斯算法的描述.

算法 4. 局部梅特罗波利斯算法.

输入: 每个点 $v \in V$ 收到值域 Σ , 函数 f_v 以及满足 $v \in e$ 的所有函数 f_e , 一个实数参数 $0 \leq p \leq 1$, 一个整数参数 T ;

```

1 每个点  $v \in V$  把  $X_v$  设置为  $\Sigma$  中任意一个值;
2 for  $t=1$  到  $T$  do
3   foreach  $v \in V$  do
4     以概率  $p$  变成活跃, 否则变成休眠;
5   foreach 活跃的点  $v \in V$  do
6     从概率分布  $\tilde{f}_v$  中采样一个随机值  $c_v \in Q$ ;
7   foreach  $u$  和  $v$  都活跃的边  $\{u, v\} \in E$  do
8     以概率  $\tilde{f}_e(c_u, c_v), \tilde{f}_e(c_u, X_v), \tilde{f}_e(X_u, c_v)$  通过测试;
9   foreach  $u$  活跃且  $v$  休眠的边  $\{u, v\} \in E$  do
10    以概率  $\tilde{f}_e(c_u, X_v)$  通过测试;
11  foreach 活跃的点  $v \in V$  do
12    if 所有与  $v$  关联的边都通过测试 then
13       $X_v \leftarrow c_v$ ;
14 每个  $v \in V$  输出  $X_v$ ;
```

给定一个图 $G=(V, E)$ 上的自旋系统 (V, Σ, \mathcal{F}) 以及实数参数 $0 \leq p \leq 1$ 和整数参数 $T \geq 0$, 局部梅特罗波利斯算法初始每个点 $v \in V$ 中任意取一个值 $X_v \in \Sigma$, 之后, 算法执行 T 次如下步骤.

1. 每个点 $v \in V$ 独立地以概率 p 变为活跃状态, 否则, 点 v 变成休眠状态;
2. 所有活跃的点 $v \in V$ 独立地从概率分布 \tilde{f}_v (定义见等式(1.2))中采样一个随机取值 $c_v \in \Sigma$;
3. 每条边 $e=\{u, v\} \in E$ 变为活跃状态当且仅当 u 和 v 中至少一个点处于活跃状态, 所有活跃的边 $e \in E$ 独立地抛一枚硬币, 正面向上的概率 p_e 为

$$p_e = \begin{cases} \tilde{f}_e(c_u, c_v), \tilde{f}_e(c_u, X_v), \tilde{f}_e(X_u, c_v), & \text{如果 } u, v \text{ 都活跃} \\ \tilde{f}_e(c_u, X_v), & \text{如果 } u \text{ 活跃, } v \text{ 休眠.} \\ \tilde{f}_e(X_u, c_v), & \text{如果 } u \text{ 休眠, } v \text{ 活跃} \end{cases}$$

上面函数 \tilde{f}_e 的定义在等式(1.3)中;

4. 对于所有点 $v \in V$, 如果点 v 活跃且与 v 相关的所有边的抛硬币结果都是正面向上, 则点 v 执行更新 $X_v \leftarrow c_v$; 否则, 点 v 保持 X_v 不变.

参数 $0 \leq p \leq 1$ 控制了局部梅特罗波利斯算法每一步尝试更新的节点个数. 特别地, 当 $p=1$ 时, 算法 4 变成了文献[18]中的原始算法. 当 $0 < p < 1$ 时, 算法 4 成为文献[21,22]中的改进算法.

我们用 $(X_t)_{t \geq 0}$ 表示局部梅特罗波利斯算法, 每个 $X_t \in \Sigma^V$ 表示第 t 次循环后产生的状态. 如下定理说明, 对于任意自旋系统, 如果串行梅特罗波利斯算法(算法 3)可以收敛到目标分布, 那么局部梅特罗波利斯算法也能收敛到目标分布.

定理 5.3(局部梅特罗波利斯算法的正确性^[18,21,22]). 对于任意自旋系统 (V, Σ, \mathcal{F}) , 如果串行梅特罗波利斯算法(算法 3)从任意初始状态出发, 最终一定收敛到合法配置且在合法配置的空间上不可约, 则对于任意 $0 < p < 1$, 以 p 为参数的局部梅特罗波利斯算法 $(X_t)_{t \geq 0}$ 满足: 对于任意初始状态 $X_0 \in \Sigma^V$, 当 $t \rightarrow \infty$ 时, X_t 的概率分布一定收

敛到目标吉布斯分布 $\mu_{(V, \mathcal{E}, \mathcal{F})}$.

文献[18]提出了最原始的梅特罗波利斯算法(参数 $p=1$ 时的算法 4), 并把算法应用在图染色问题上, 得出在 $q \geq (2 + \sqrt{2} + \delta)\Delta$ 时, 算法可以在 $O(\log n)$ 的时间内近似均匀采样合法图染色. 文献[21,22]进一步把收敛条件改进到 $q \geq (2 + \delta)\Delta$. 条件 $q \geq (2 + \delta)\Delta$ 是图染色问题的多布鲁申条件^[59], 是串行采样算法收敛的重要条件^[43].

定理 5.4(局部梅特罗波利斯算法在图染色问题上的应用^[21,22]). 令 $\delta > 0$ 为一个常数. 给定任意 n 个点最大度数为 Δ 的图 G , 如果 $q \geq (2 + \delta)\Delta$, 则参数为 $p = \min\{\delta/3, 1/2\}$ 的局部梅特罗波利斯算法可以用 $O\left(\log \frac{n}{\varepsilon}\right)$ 轮的时间返回一个随机样本 $X \in [q]^V$, 满足 $d_{TV}(X, \mu) \leq \varepsilon$, 其中, μ 是所有合法 q -染色的均匀分布, $O(\cdot)$ 记号隐藏了一个只与 δ 相关的常数因子.

对于特殊图上的图染色问题, 文献[21]得到如下结论.

定理 5.5(局部梅特罗波利斯算法在特殊图的图染色问题上的应用^[21]). 对于任意常数 $\delta > 0$, 存在一个常数 $\Delta_0 = \Delta_0(\delta)$, 使得对于任意 n 个点, 最大度数为 Δ , 围长为 $g = g(G)$ 的图, 如果如下两个条件同时成立.

- (1) $\Delta \geq \Delta_0$ 且 $g \geq 9$;
- (2) $q \geq (\alpha^* + \delta)\Delta$, 其中, $\alpha^* \approx 1.763 \dots$ 满足 $\alpha^* = \exp(1/\alpha^*)$,

则参数 $p = p(\delta)$ 的局部梅特罗波利斯算法可以用 $O\left(\log \frac{n}{\varepsilon}\right)$ 轮的时间返回一个随机样本 $X \in [q]^V$, 满足 $d_{TV}(X, \mu) \leq \varepsilon$, 其中, μ 是所有合法 q -染色的均匀分布, $O(\cdot)$ 记号隐藏了一个只与 δ 相关的常数因子.

定理 5.5 研究在最大度数足够大、围长足够长的图上的随机染色采样问题, 其中, 围长是指图中最小环的长度. 在这类特殊的图上, 对颜色数的要求可以改进 $q \geq (\alpha^* + \delta)\Delta$. 特殊图上的图染色问题是经典采样理论的重要课题之一. 条件 $q \geq (\alpha^* + \delta)\Delta$ 在研究空间混合性质^[62,63]和马尔可夫链的混合时间^[64-70]中都有重要的应用.

图染色问题是经典采样理论的重要模型, $q \geq (2 + \delta)\Delta$ 和 $q \geq (\alpha^* + \delta)\Delta$ 是经典串行采样算法的两个重要收敛条件. 定理 5.4 和定理 5.5 说明分布式算法也能在这两个条件下收敛, 且根据后文第 6.1 节的结论, $O(\log n)$ 是分布式采样图染色的最优收敛时间. 这说明, 局部梅特罗波利斯算法是一个高度并行的算法, 且在重要的模型中能够表现出与经典串行算法相当的性能.

5.3 梅特罗波利斯算法的分布式模拟

文献[24]给出了一个在分布式计算模型上模拟串行梅特罗波利斯算法的策略. 考虑一个定义在图 $G=(V, E)$ 上的自旋系统, 自旋系统的梅特罗波利斯算法可以抽象成算法 5 的形式.

算法 5. 自旋系统的抽象梅特罗波利斯算法.

输入: 一个定义在图 $G=(V, E)$ 上, 样本空间为 Σ^V 的自旋系统;

- 1 令 $X_0 \in \Sigma^V$ 为一个初始配置;
- 2 **for** $t=1$ to T **do**
- 3 随机、均匀地选择一个点 $v \in V$, 并令 $c = X_{t-1}(v)$;
- 4 从分布 ν_v 中采样一个候选值 $c' \in Q$, 构造新配置 $X' \in Q^V$, 满足 $X'(v) = c'$, $X'(V \setminus \{v\}) = X_{t-1}(V \setminus \{v\})$;
- 5 以概率 $f_{c,c'}^v(X_{t-1}(\Gamma_v))$ 令 $X_t \leftarrow X'$; 以概率 $1 - f_{c,c'}^v(X_{t-1}(\Gamma_v))$, 令 $X_t \leftarrow X_{t-1}$;
- 6 **return** X_N ;

在算法 5 中, 每个点 $v \in V$ 都有一个候选值分布 ν_v . 对于任意 $v \in V$ 、任意 $c, c' \in \Sigma$, 接受率函数:

$$f_{c,c'}^v : \Sigma^{\Gamma_v} \rightarrow [0, 1].$$

在点 v 当前值为 c 、候选值为 c' 时, 把点 v 当前邻居的配置映射成一个接受概率. 算法 5 每次更新都随机、等概率地选择一个点 $v \in V$, 按照分布 ν_v 采样一个候选值, 再利用接受率函数计算接受候选值的概率. 算法 5 重复 N 次更新操作之后, 输出当前的随机样本.

很多特殊的梅特罗波利斯算法都能解释成算法 5 的特例.

例 1: 算法 3. 在这种梅特罗波利斯算法中, 对于任意 $v \in V$, 候选值分布 $\nu_v = \tilde{f}_v$. 接受概率满足:

$$\forall v \in V, c, c' \in \Sigma, \forall \sigma \in \Sigma^{F_v}, f_{c,c'}^v(\sigma) = \prod_{\{u,v\} \in E} \tilde{f}_{uv}(c', \sigma_u).$$

例 2: 图 q -染色问题的梅特罗波利斯算法. 在这种梅特罗波利斯算法中, 对于任意 $v \in V$, 候选值分布 ν_v 是 $[q]$ 上的均匀分布. 接受概率满足:

$$\forall v \in V, c, c' \in \Sigma, \forall \sigma \in \Sigma^{F_v}, f_{c,c'}^v(\sigma) = \prod_{\{u,v\} \in E} \mathbb{1}[c' = \sigma_u].$$

上述梅特罗波利斯算法 $(X_t)_{t \geq 0}$ 是一个本质串行的算法. 另一方面来看, $(X_t)_{t \geq 0}$ 是一个连续时间马尔可夫链 $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$ 的离散化版本. 连续时间马尔可夫链是一个天然并行的过程. 事实上, 早在计算机科学家研究离散马尔可夫链之前, 物理学家已经开始用连续时间马尔可夫对物理系统进行建模^[71].

在定义连续时间马尔可夫链之前, 先定义速率 k 泊松时钟(Poisson clock). 令 $\{x_i\}_{i=1}^\infty$ 为一串独立同分布的服从指数分布的随机变量且 $\mathbb{E}[x_i] = 1/k$, 泊松时钟从 0 时刻开始运行, 在 t_1, t_2, t_3, \dots 时刻敲响且 $t_i = \sum_{j \leq i} x_j$. 算法 6 给出了连续时间梅特罗波利斯算法的定义.

算法 6. 自旋系统的连续时间梅特罗波利斯算法.

输入: 一个定义在图 $G=(V,E)$ 上, 样本空间为 Σ^V 的自旋系统;

- 1 令 $Y \in \Sigma^V$ 为一个初始配置;
- 2 每个点 $v \in V$ 放置一个独立的速率 1 泊松时钟;
- 3 **for** $t=0$ 到 $T \in \mathbb{R}$ **do**
- 4 \perp 当点 v 上的泊松时钟敲响, $Y(v)$ 按照算法 5 第 4 行、第 5 行的规则更新;
- 5 **return** Y ;

命题 5.1(连续时间和离散时间马尔可夫链的关系^[32]). 令 $(X_t)_{t \in \mathbb{N}}$ 为离散时间梅特罗波利斯算法, 令 $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$ 为连续时间梅特罗波利斯算法, 如果 $X_0=Y_0$, 则对于任意 $T \in \mathbb{R}_{\geq 0}$, Y_T 和 $X_{N(T)}$ 同分布, 其中, 随机变量 $N(T)$ 服从均值为 nT 的泊松分布且 $n=|V|$.

令 $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$ 为一个抽象的连续时间马尔可夫链. 考虑在 LOCAL 模型下模拟 $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$. 每个点 $v \in V$ 的输入为值域 Σ 、分布 ν_v 、接受率函数 $(f_{c,c'}^v)_{c,c' \in \Sigma}$ 、初始值 $Y_0(v)$ 以及一个时间 $T \in \mathbb{R}_{\geq 0}$. 模拟算法结束时, 每个点输出随机值 $Y_T(v)$, 满足 $Y_T=(Y_T(v))_{v \in V}$, 正好是连续时间梅特罗波利斯算法 T 时刻的随机状态. 文献[24]证明: 如果接受率函数满足如下利普希茨条件, 则存在高效模拟算法.

条件 5.1(利普希茨条件). 对于一个常数 $C \geq 0$ 使得对于任意 $\{u,v\} \in E$, 任意 $a,b,c \in \Sigma$:

$$\mathbb{E}_{c' \sim \nu_v} [\delta_{u,a,b} f_{c,c'}^v] \leq \frac{C}{\Delta},$$

其中, Δ 是图 $G=(V,E)$ 的最大度数, 算子 $\delta_{u,a,b}$ 定义为

$$\delta_{u,a,b} f_{c,c'}^v \triangleq \max_{\sigma, \tau} |f_{c,c'}^v(\sigma) - f_{c,c'}^v(\tau)|.$$

上式中的最大值枚举所有的 $\sigma, \tau \in \Sigma^{F_v}$, 满足 $\sigma_u=a, \tau_u=b$; 且对所有 $w \neq u, \sigma_w=\tau_w$.

令 n 为通信网络 G 的点数, 如果一个事件发生的概率至少为 $1-O(1/n)$, 则我们称这个事件高概率发生.

定理 5.6(梅特罗波利斯算法的分布式模拟^[24]). 存在一个分布式 LOCAL 算法, 它可以正确地模拟抽象连续时间梅特罗波利斯算法 $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$ 到时刻 T , 算法的时间复杂度高概率为 $O(\Delta T + \log n)$ 轮. 进一步地, 如果梅特罗波利斯算法的接受率函数满足条件 5.1, 则模拟算法的时间复杂度高概率为 $O(T + \log n)$ 轮, 其中, $O(\cdot)$ 记号隐藏了一个只与 C 相关的常数因子.

评注 5.1. 连续时间马尔可夫链的 Y_T 和离散时间马尔可夫链的 $X_{N(T)}$ 同分布, 其中, $N(T)$ 服从均值为 nT 的泊松分布. 根据泊松分布的概率集中性质, 定理 5.6 证明, 存在一个分布式 LOCAL 算法, 它可以用 $O(\Delta N/n + \log n)$ 轮的时间高概率地模拟 N 步离散时间梅特罗波利斯算法. 进一步地, 如果梅特罗波利斯算法的接受率函

数满足条件 5.1, 则它可以用 $O(N/n+\log n)$ 轮的时间高概率地模拟 N 步离散时间梅特罗波利斯算法.

在一些具体模型上, 条件 5.1 可以简化为如下条件.

1. 图染色

硬核模型 $(V, E, [q])$ 定义见第 2.2 节. 自然的梅特罗波利斯算法定义为: 对每个点 $v \in V$, ν_v 定义为 $[q]$ 上的均匀分布, 接受率函数定义为: 对于任意点 $v \in V$,

$$\forall c, c' \in [q], \tau \in [q]^{\Gamma_v}, f_{c,c'}^v(\tau) = \prod_{u \in \Gamma_v} \mathbb{1}[\tau_u \neq c']$$

条件 5.1 变为

$$\exists \text{常数 } C > 0, q \geq C\Delta,$$

而图染色的唯一性条件为 $q \geq \Delta + 1$ [72,73].

2. 硬核模型

硬核模型 (V, E, λ) 的定义见第 2.2 节. 自然的梅特罗波利斯算法定义为: 对每个点 $v \in V$, ν_v 定义在 $\{0, 1\}$ 上满足 $\nu_v(0) = \frac{1}{1+\lambda}$ 且 $\nu_v(1) = \frac{\lambda}{1+\lambda}$, 接受率函数定义为: 对于任意点 $v \in V$,

$$\forall c, c' \in \{0, 1\}, \tau \in \{0, 1\}^{\Gamma_v}, f_{c,c'}^v(\tau) = \prod_{u \in \Gamma_v} \mathbb{1}[\tau_u + c' \leq 1]$$

条件 5.1 变为

$$\exists \text{常数 } C > 0, \lambda \leq \frac{C}{\Delta},$$

而硬核模型的唯一性条件为 $\lambda \leq \frac{(\Delta-1)^{(\Delta-1)}}{(\Delta-2)^\Delta}$ [3,4].

3. 伊辛模型

伊辛模型 (V, E, β) 的定义见第 2.2 节. 自然的梅特罗波利斯算法定义为: 对每个点 $v \in V$, ν_v 是 $\{-1, +1\}$ 上的均匀分布, 接受率函数定义为: 对于任意点 $v \in V$,

$$\forall c, c' \in \{-1, +1\}, \tau \in \{-1, +1\}^{\Gamma_v}, f_{c,c'}^v(\tau) = \prod_{u \in \Gamma_v} \exp(\min(0, \beta(c' - c) \sum_{u \in \Gamma_v} \tau_u)).$$

条件 5.1 变为

$$\exists \text{常数 } C > 0, 1 - \exp(-2|\beta|) \leq \frac{C}{\Delta},$$

而伊辛模型的唯一性条件为 $1 - \exp(2|\beta|) \leq \frac{2}{\Delta}$ [28,74].

对于很多自旋系统, 唯一性条件(uniqueness condition)是吉布斯分布可被高效采样的必要条件. 这说明, 即使梅特罗波利斯算法收敛慢, 模拟算法依然可以高效地模拟它. 定理 5.6 不会与已知的采样下界结果产生矛盾, 这是因为, 算法的目标只是正确地模拟梅特罗波利斯算法, 算法并不要求梅特罗波利斯算法快速收敛.

模拟算法的关键技术是提前解决更新(resolve update in advance). 假设算法要把连续时间马尔可夫链 $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$ 模拟到时刻 T . 算法的框架描述见算法 7.

算法 7. 梅特罗波利斯算法的分布式模拟.

点 $v \in V$ 处的算法

阶段 I:

- 本地模拟泊松时钟生成时刻 T 之前所有更新时间 $0 < t_1^v < t_2^v < \dots < t_{m_v}^v < T$, 本地从概率分布 ν_v 中采样所有更新使用的候选随机值 $c_1^v, c_2^v, \dots, c_{m_v}^v \in \mathcal{S}$;
- 把这些信息以及 $Y_0(v)$ 发送给所有的邻居. 接受邻居阶段 I 发送的消息并进入阶段 II.

阶段 II:

For $i=1$ 到 m_v **do**

- 接收来自邻居的消息;
- (*)只要 v 得到足够多的信息, 则解决第 i 个更新.

解决点 v 的第 i 个更新, 并把更新处理的结果(接受或拒绝)发送给所有的邻居.

模拟算法有两个阶段. 在阶段 I 中, 每个点 $v \in V$ 本地生成所有要用到的随机比特, 并且把这些随机比特发送给所有的邻居. 所以在阶段 II 的开始, 点 v 知道每个邻居的初始值、时刻 T 之前所有更新时间以及更新的候选随机值. 在阶段 II, 点 v 处的算法在(*)行解决更新. 一个最直接的策略去实现(*)行是点 v 解决时刻 t 的更新当且仅当点 v 知道邻居时刻 t 之前所有更新的处理结果. 这时可以 v 知道 $Y_t(\Gamma_v)$, 这样 v 就能解决时刻 t 的更新. 因为这个策略不允许相邻的节点同时更新, 所以模拟的速率会损失 Δ 倍, 只能得到 $O(\Delta T + \log n)$ 的时间复杂度.

文献[24]给出了一个提前解决更新的策略. 考虑图染色模型, 当点 v 一次更新生成一个随机颜色 c' 后, 如果下列两个事件之一发生, 则此更新可以被处理.

- $\forall u \in \Gamma_v, Y_t(u) \neq c'$, 此时新颜色会被接受;
- $\exists u \in \Gamma_v, Y_t(u) = c'$, 此时新颜色会被拒绝.

该技术最关键的观察是: 即使状态 $Y_t(\Gamma_v)$ 没有被完全确定, 更新依然有可能被提前解决. 这是因为, 对于一个邻居 u, v 即使只知道 u 在某个时刻 $t_u < t$ 的准确颜色, v 也可以推断出 u 在时刻 t 所有可能颜色 $Y_t(u)$ 的集合. 这是因为, $Y_t(u)$ 要么是 $Y_{t_u}(u)$, 要么是时刻 t_u 和 t 之间某个更新生成的候选颜色. 所以, 如果 c' 和所有邻居的所有可能颜色都不冲突, 那么这个更新就可以提前被接受. 如果存在一个邻居, 它在时刻 t 的颜色只能是 c' , 那么这个更新就可以提前被拒绝. 所以, 在 $Y_t(\Gamma_v)$ 被完全确定之前, 算法依然有可能解决当前更新.

对于一般的连续时间马尔可夫链 $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$, 假设时刻 t 有一个候选值为 $c' \in \Sigma$ 的更新. 为了解决这个更新, 点 v 需要抛一枚正面向上概率为 $f_{c,c'}^v(Y_t(\Gamma_v))$ 的硬币, 其中, $c = Y_{t-}(v)$ 是点 v 在更新之前的取值. 与之前一样, 点 v 可以推出邻居 u 在时刻 t 所有可能取值 $Y_t(u)$ 的集合. 对所有邻居所有可能的取值做笛卡尔积, 点 v 可以得出所有可能配置 $Y_t(\Gamma_v)$ 的集合. 给定当前可能配置的集合, 点 v 可以算出 $f_{c,c'}^v(Y_t(\Gamma_v))$ 的上界和下界, 算法就可以利用上下界尝试提前解决更新. 算法从区间 $[0,1]$ 上采样一个均匀的实数 β (对于每个更新, β 只采样 1 次). 如果 β 比当前 $f_{c,c'}^v(Y_t(\Gamma_v))$ 的下界小, 则接受更新; 如果 β 比当前 $f_{c,c'}^v(Y_t(\Gamma_v))$ 的上界大, 则拒绝更新; 否则, 等待接收更多来自邻居的信息, 进一步缩小可能的配置集合. 随着点 v 收到越来越多的信息, 这个可能配置的集合就会越来越小, 直到最后, 这个集合只剩下唯一确定的配置 $Y_t(\Gamma_v)$. 所以, 更新最终一定可以被解决.

模拟算法的具体实现细节可参考文献[24].

5.4 局部拒绝采样算法

最后, 介绍一类分布式精确采样算法——局部拒绝采样算法^[23,26]. 这类算法起初是为了解决洛瓦兹局部引理的采样问题^[26], 之后又被应用到分布式采样^[23,26]、精确采样^[25,36,38,39]和动态采样问题^[23,25]上, 一个著名的应用是解决了重要的公开问题——network reliability^[35,36].

考虑一个分布式计算网络 $G=(V,E)$, 令 (V,Σ,\mathcal{F}) 是一个图 $G=(V,E)$ 上的一个局部吉布斯分布, \mathcal{F} 是一系列约束 $(S,f) \in \mathcal{F}$, 其中, $S \subseteq V$ 是约束的作用域, $f: \Sigma^S \rightarrow \mathbb{R}_{\geq 0}$ 是约束函数. 不失一般性地, 我们假设所有约束的作用域互不相同, 如果存在多个约束作用域一样, 则可以把这些约束乘起来得到一个新的约束. 我们可以用一个超图 $h=(V,E_H)$ 来对吉布斯分布 (V,Σ,\mathcal{F}) 进行建模, 其中, $E_H = \{S | (S,f) \in \mathcal{F} \wedge |S| \geq 2\}$ 是一系列超边的集合. 对于每个点 $v \in V$, 如果存在约束 $(\{v\}, f) \in \mathcal{F}$, 定义约束 $\tilde{f}_v: \Sigma \rightarrow \mathbb{R}_{\geq 0}$ 为

$$\forall c \in \Sigma, \tilde{f}_v(c) = \frac{f_{\{v\}}(c)}{\sum_{c' \in \Sigma} f_{\{v\}}(c')};$$

如果不存在约束 $(\{v\}, f) \in \mathcal{F}$, 定义约束 $\tilde{f}_v: \Sigma \rightarrow \mathbb{R}_{\geq 0}$ 为

$$\forall c \in \Sigma, \tilde{f}_v(c) = \frac{1}{|\Sigma|};$$

重申 $E_H = \{S|(S,f) \in \mathcal{F} \wedge |S| \geq 2\}$. 对于任意超边 $h \in E_H$, 假设它对应的约束是 (S,f) , 则一定有 $h=S$. 定义约束函数 $\tilde{f}_h: \Sigma^h \rightarrow \mathbb{R}_{\geq 0}$ 为

$$\forall \sigma \in \Sigma^h, \tilde{f}_h(\sigma) = \frac{f(\sigma)}{\max_{\tau \in \Sigma^h} f(\tau)}.$$

令 $\tilde{\mathcal{F}} = \{(\{v\}, \tilde{f}_v) | v \in V\} \cup \{(h, \tilde{f}_h) | h \in E_H\}$. 容易验证, $(V, \Sigma, \tilde{\mathcal{F}})$ 和 (V, Σ, \mathcal{F}) 表示同一个吉布斯分布. 超图 $h=(V, E_H)$ 的点和超边的集合对应了 $\tilde{\mathcal{F}}$ 中的约束. 每个点上的约束 \tilde{f}_v 是一个概率分布, 每个超边上的约束 \tilde{f}_h 的值域是 $[0,1]$. 因为 (V, Σ, \mathcal{F}) 是一个通信网络 $G=(V,E)$ 上的局部吉布斯分布. 所以 $h=(V, E_H)$ 上的任意超边 h 在 G 上的直径都是一个常数.

为了描述算法, 先引入一些记号. 令 $A \subseteq V$ 是一个变量的集合, 定义 A 的内部超边为

$$E(A) \triangleq \{h \in E_H | h \subseteq A\};$$

定义 A 的边界超边为

$$\delta(A) \triangleq \{h \in E_H | h \not\subseteq A \wedge h \cap A \neq \emptyset\}.$$

定义与 A 相关的所有超边为

$$E^+(A) = E(A) \cup \delta(A).$$

局部拒绝采样算法的设计思想来源于算法 8 中的原始的拒绝采样(rejection sampling)算法.

原始拒绝采样算法(算法 8)每个点 $v \in V$ 从分布 \tilde{f}_v 中独立采样 $X_v \in \Sigma$, 之后, 每个超边 $h \in E_H$ 独立地抛一枚硬币, 硬币正面向上($F_h=0$, $F_h \in \{0,1\}$ 是一个指示抛硬币结果的变量)的概率为 $\tilde{f}_h(X_h)$. 如果所有硬币都是正面向上, 则算法输出 X ; 否则, 重新执行上述过程. 显然, 拒绝采样算法是一个精确采样算法, 但是它有两个明显的缺陷.

- 拒绝采样算法每次尝试成功要求所有硬币都是正面向上. 对于很多吉布斯分布, 算法一次尝试只有指数级小的成功概率, 所以算法的期望运行时间不是多项式;
- 拒绝采样算法不能用于分布式采样. 因为一次尝试失败后需要所有点重新采样, 分布式计算不能执行全局性的操作.

算法 8. 原始拒绝采样算法.

输入: 一个由 $(V, \Sigma, \tilde{\mathcal{F}})$ 定义的吉布斯分布 $\mu = \mu_{(V, \Sigma, \tilde{\mathcal{F}})}$, 建模 $(V, \Sigma, \tilde{\mathcal{F}})$ 的超图 $H=(V, E_H)$;

- 1 **repeat**
- 2 每个点 $v \in V$ 独立地从分布 \tilde{f}_v 中采样 $X_v \in \Sigma$, 令 $X=(X_v)_{v \in V}$;
- 3 每个超边 $h \in E_H$ 独立地采样 $F_h \in \{0,1\}$, 使得 $\Pr[F_h=0] = \tilde{f}_h(X_h)$;
- 4 **until** 所有超边 $h \in E_H$ 都有 $F_h=0$;
- 5 **Return** X .

局部拒绝采样算法是对原始的拒绝采样算法的一种改进. 当一次尝试有超边失败($F_h=1$)时, 算法不需要全局重新采样, 而是只重新采样在失败超边一个局部范围内的点, 然后再构造新的失败超边. 重复这个过程, 直到所有超边都成功. 这样就得到了一个既高效又支持分布式计算的算法. 文献[23]给出了如下局部拒绝采样算法.

局部拒绝采样算法(算法 9)第 1 行、第 2 行是原始的拒绝采样算法. 但是拒绝采样失败后, 没有全局重新采样, 而是在第 3 行构造失败超边内部点的集合 \mathcal{R} . 每次 while-循环中, 算法只对集合 \mathcal{R} 内部的点进行重采样, 并动态维护 \mathcal{R} 集合. 每次重采样之前, 与 \mathcal{R} 关联的所有超边 $h \in E^+(\mathcal{R})$ 先计算一个纠正系数 κ_h , 计算 κ_h 时, 算法需要枚举所有的 $x \in \Sigma^h$ 满足 $x_{h \cap \mathcal{R}} = X_{h \cap \mathcal{R}}$, 即 x 在 $h \cap \mathcal{R}$ 上的取值要与当前样本 X 在 $h \cap \mathcal{R}$ 上的取值相同; 之后, \mathcal{R} 内部的点重采样; 最后, 与 \mathcal{R} 关联的所有超边以概率 $\kappa_h \cdot \tilde{f}_h(X_h)$ 通过测试. 相对于第 2 行, 第 7 行的概率多乘了一个纠正系数 κ_h , 这是保证算法正确的关键. 当 \mathcal{R} 为空集时, 局部拒绝采样算法结束并输出当前的 X .

算法 9. 局部拒绝采样算法.

输入: 每个点 $v \in V$ 收到值域 Σ 、函数 \tilde{f}_v 以及满足 $v \in V$ 的所有函数 \tilde{f}_h ;

- 1 每个点 $v \in V$ 并行独立地从分布 \tilde{f}_v 中采样 $X_v \in \Sigma$, 令 $X = (X_v)_{v \in V}$;
- 2 每个超边 $h \in E_H$ 并行独立地采样 $F_h \in \{0, 1\}$, 使得 $\Pr[F_h = 0] = \tilde{f}_h(X_h)$;
- 3 $\mathcal{R} \leftarrow \cup_{h \in E_H: F_h=1} h$;
- 4 **while** $\mathcal{R} \neq \emptyset$ **do**
 - 5 每个超边 $h \in E^+(\mathcal{R})$ 并行地计算 $\kappa_h \triangleq \frac{1}{\tilde{f}_h(X_h)} \min_{x \in \Sigma^h: x_{h \cap \mathcal{R}} = X_{h \cap \mathcal{R}}} \tilde{f}_h(x)$ (规定 $\frac{0}{0} = 1$);
 - 6 每个点 $v \in \mathcal{R}$ 并行独立地从分布 \tilde{f}_v 中重新采样 $X_v \in \Sigma$;
 - 7 每个超边 $h \in E^+(\mathcal{R})$ 并行独立地采样 $F_h \in \{0, 1\}$, 使得 $\Pr[F_h = 0] = \kappa_h \cdot \tilde{f}_h(X_h)$;
 - 8 $\mathcal{R} \leftarrow \cup_{h \in E_H: F_h=1} h$
- 9 **return** X .

评注 5.2(局部拒绝采样算法的分布式实现). 算法 9 从全局的视角描述局部拒绝采样算法. 因为 $(V, \Sigma, \tilde{\mathcal{F}})$ 局部吉布斯分布, 所以, 所有超边 $h \in E_H$ 在通信网络 G 上的直径是一个常数. 因此, 算法 9 的第 1-3 行可以用 $O(1)$ 轮实现, 且每次 while-循环也可以用 $O(1)$ 轮实现. 第 3 行和第 8 行的集合 $\mathcal{R} \subseteq V$ 可以局部性地构造, 即图上每个点知道自己是否在当前的集合 \mathcal{R} 中. 当分布式网络内所有点都停止计算时, 算法结束. 如果吉布斯分布由自旋系统定义, 则通信消息的大小也有较小的上界.

利用条件吉布斯(conditional Gibbs)^[23]性质, 可以证明算法 9 的正确性.

定理 5.7(局部拒绝采样算法的正确性^[23,26]). 如果算法 9 停止, 则输出 X 精确服从 $\mu_{(V, \Sigma, \tilde{\mathcal{F}})} = \mu_{(V, \Sigma, \mathcal{F})}$.

文献[23]利用势函数分析法, 得到如下收敛性结论. 给定一个吉布斯分布 (V, Σ, \mathcal{F}) , 令 $H = (V, E_H)$ 为上述定义中对 (V, Σ, \mathcal{F}) 进行建模的超图. 定义:

$$d = d_{(V, \Sigma, \mathcal{F})} \triangleq \max_{h \in E_H} \{|\{h' \in E_H \mid h \cap h' \neq \emptyset \wedge h \neq h'\}|\}$$

即, 超图 h 中一个超边最多与 d 个其他的超边有交集.

定理 5.8(局部拒绝采样算法收敛性结论^[23]). 令 $\delta > 0$ 为一个常数. 对于任意吉布斯分布 (V, Σ, \mathcal{F}) , 如果对于任意满足 $|\mathcal{S}| \geq 2$ 的约束 $(\mathcal{S}, f) \in \mathcal{F}$ 都有 $f: \Sigma^{\mathcal{S}} \rightarrow [B_{(\mathcal{S}, f)}, 1]$ 且:

$$B_{(\mathcal{S}, f)} \geq \left(1 - \frac{1 - \delta}{d + 1}\right)^{1/2},$$

则局部拒绝采样算法(算法 9)的期望运行轮数为 $O(\log n)$, 且以至少 $1 - \varepsilon$ 的概率, 算法的轮数为 $O\left(\log \frac{n}{\varepsilon}\right)$, 其中, $n = |V|$ 为总点数, 两个 $O(\cdot)$ 记号隐藏了只与 δ 有关的常数因子.

把上述定义应用到伊辛模型 (V, E, β) (定义见第 2.2) 上, 可以得到 $\exp(-2|\beta|) \geq 1 - \frac{1}{4\Delta} + o\left(\frac{1}{\Delta}\right)$ 的收敛条件. 进一步地, 对伊辛模型可以得到如下更紧的条件, 令 $\alpha \approx 2.221\dots$ 是下列方程的根:

$$\alpha = 1 + \frac{2}{1 + \exp(-1/\alpha)}.$$

定理 5.9(算法 9 在伊辛模型上的应用^[23]). 令 $\delta > 0$ 为一个常数. 对于任意伊辛模型 (V, E, β) , 如果:

$$\exp(-2|\beta|) \geq 1 - \frac{1 - \delta}{\alpha\Delta + 1},$$

其中, Δ 是图 $G = (V, E)$ 的最大度数, 则局部拒绝采样算法(算法 9)的期望运行轮数为 $O(\log n)$, 且以至少 $1 - \varepsilon$ 的概率, 算法的轮数为 $O\left(\log \frac{n}{\varepsilon}\right)$, 其中, $n = |V|$ 为总点数, 两个 $O(\cdot)$ 记号隐藏了只与 δ 有关的常数因子.

注意到, 定理 5.8 的条件要求所有的下界 $B_{(\mathcal{S}, f)} > 0$, 即 (V, Σ, \mathcal{F}) 是由柔性约束定义的吉布斯分布. 局部拒绝采

样算法也能用于由硬性约束定义的吉布斯分布. 考虑第 2.2 节定义的硬核模型 (V, E, λ) . 事实上, 算法 9 中描述的局部拒绝采样算法有很多推广的版本^[23,25,26], 算法 10 给出了一个硬核模型的局部拒绝采样算法.

容易验证, 算法 10 也可以在分布式计算模型上实现. 这里强调, 算法 10 不是算法 9 在硬核模型上的应用. 算法 10 相当于在算法 9 的 while-循环开头插入一行:

$$\mathcal{R} \leftarrow \mathcal{R} \cup \delta\mathcal{R}.$$

然后再用扩张之后的 \mathcal{R} 运行算法 10 的 while-循环. 利用硬核模型性质, 扩张之后的 \mathcal{R} 内边界上所有的点 v 一定满足 $X_v=0$. 因此, 扩张之后的 \mathcal{R} 满足: 对于任意 $h \in E^+(\mathcal{R})$, 一定有 $\kappa_h=1$. 由此可以得出算法 9.

文献[26]首先分析了算法 10 的正确性和运行时间; 之后, 文献[23]用更一般条件吉布斯技术重新证明了算法的正确性, 并且改进了算法的收敛条件. 如下定理说明了算法 10 的正确性和收敛时间, 定理使用了文献[23]中的改进版收敛条件.

定理 5.10(硬核模型的局部拒绝采样算法^[23,26]). 对于任意硬核模型 (V, E, λ) , 当算法 10 停止时, 输出 X 精确服从吉布斯分布 $\mu_{(V, E, \lambda)}$. 进一步地, 如果:

$$\lambda \leq \frac{1}{(1+\delta)\sqrt{2\Delta-1}},$$

其中, $\delta > 0$ 是一个常数, Δ 是图 G 的最大度数, 则硬核模型的局部拒绝采样算法(算法 10)的期望运行轮数为 $O(\log n)$, 且以至少 $1-\epsilon$ 的概率, 算法的轮数为 $O\left(\log \frac{n}{\epsilon}\right)$, 其中, $n=|V|$ 为总点数, 两个 $O(\cdot)$ 记号隐藏了只与 δ 有关的常数因子.

算法 10. 硬核模型的局部拒绝采样算法.

输入: 每个点 $v \in V$ 收到值域 Σ 和参数 $\lambda \in \mathbb{R}_{>0}$;

- 1 每个点 $v \in V$ 并行独立地采样 $X_v \in \{0, 1\}$, 概率满足 $\Pr[X_v] = \frac{\lambda}{1+\lambda}$, 令 $X = (X_v)_{v \in V}$;
- 2 **while** 存在连通块 $\mathcal{R} \subseteq V$ 满足 $|\mathcal{R}| \geq 2$ 且 $\forall v \in \mathcal{R}, X_v = 1$ **do**
- 3 **foreach** $v \in \mathcal{R} \cup \delta\mathcal{R}$, 其中, $\delta\mathcal{R} = \{u \notin \mathcal{R} \mid \exists v \in \mathcal{R}, \text{使得 } \{u, v\} \in E\}$ 是 \mathcal{R} 的外边界 **do**
- 4 点 $v \in \mathcal{R}$ 独立地重新采样 $X_v \in \{0, 1\}$, 概率满足 $\Pr[X_v] = \frac{\lambda}{1+\lambda}$;
- 5 **return** X .

与之前所有的分布式采样算法相比, 局部拒绝采样算法是一个精确采样算法, 即算法停止时输出精确服从目标吉布斯分布. 我们也可以给局部拒绝采样算法预先设定一个 $O\left(\log \frac{n}{\epsilon}\right)$ 的运行时间上界, 如果超过这个上界点 v 处的算法没有结束, 则 v 输出 Σ 中任意一个值. 根据定理 5.8-定理 5.10, 这也可以得到一个误差为 ϵ 的近似采样算法. 但是局部拒绝采样算法目前得到的收敛条件比较苛刻, 适用的模型比较受限. 作为一个最近提出的新算法, 局部拒绝采样算法的性能还有待进一步研究.

6 分布式采样复杂性

6.1 分布式采样下界

在本节中, 总结分布式采样问题的下界结论. 令 $G=(V, E)$ 为一个分布式计算网络, 目标分布 μ 为样本空间 Σ^V 上的一个联合分布. 本节的下界结论只使用分布式算法的如下性质.

性质 6.1(分布式算法产生的概率分布的远距离独立性质). 任意一个 t 轮的 LOCAL 算法产生的输出 $X=(X_v)_{v \in V}$ 一定满足:

$$\forall u, v \in V, \text{dist}_G(u, v) > 2t \Rightarrow X_u \text{ 和 } X_v \text{ 相互独立.}$$

性质 6.1 成立, 是因为 LOCAL 模型每个点本地生成的随机比特相互独立, 而一个 t 轮的算法每个点 $v \in V$

只能收集以 v 为中心、距离不超过 t 的所有点上的信息(包括其所产生的随机比特), 所以距离超过 $2t$ 的点收集到的信息无交集, 输出一定相互独立.

下界的证明只用到了性质 6.1. 即使每个点知道整个网络 G (包括每个点的 UID)以及整个联合分布 μ 的定义, 本节证明的下界依然成立. 注意到: 如果每个点知道整个网络 G , 那么很多构造问题, 例如 $(\Delta+1)$ -染色、极大独立集等, 分布式算法可在 $O(1)$ 轮解决. 但是对于采样问题, 依然存在下界. 这说明, 分布式采样问题比构造问题困难, 且分布式采样问题的难度源于不同点上的随机比特相互独立.

很多自旋系统满足如下性质. 考虑链状图(下文简称为链)上的自旋系统, 设其吉布斯分布为 μ . 存在常数 $\delta, \eta > 0$, 使得对于任意长度为 n 的链 P , 在链上的任意两个点 u, v , 存在两个点 u 上的取值 $\sigma_u, \sigma'_u \in \Sigma$, 满足 $\mu_u(\sigma_u) \geq \delta, \mu_u(\sigma'_u) \geq \delta$, 其中, μ_u 是 μ 投影到 u 上的边缘分布, 而且以下性质成立:

$$d_{TV}(\mu_v^{\sigma_u}, \mu_v^{\sigma'_u}) \geq \eta^{dist_P(u,v)},$$

其中, $dist_P(u,v)$ 表示 u, v 在 P 上的距离. 满足此性质的自旋系统包含图的 q 染色模型, 其中, $q=O(1)$. 如果自旋系统满足上述性质, 则对于任意 $\varepsilon \geq \exp(-o(n))$, 对于任意一对满足 $dist_P(u,v)=\Omega(\log(1/\varepsilon))$ 的点对 (u,v) , 由于相关性的存在, 吉布斯分布产生的 (σ_u, σ_v) , 与完全独立的 (X_u, X_v) 之间至少有 ε 的全变差. 对于此类系统上 ε -全变差近似的分布式采样问题, 就可以直接给出一个 $\Omega(\log(1/\varepsilon))$ 下界.

进一步地, 对于常数误差 ε , 文献[18,26]证明了 $\Omega(\log n)$ 的下界, 这里以图染色和硬核模型为例给出结论.

定理 6.1(分布式采样的 $\Omega(\log n)$ 下界^[18]). 令 $q \geq 3$ 为一个常数, 并令 $\varepsilon < 1/3$. 对于任意一个以 ε 误差采样一条链上均匀 q -染色的 LOCAL 算法, 其时间复杂度为 $\Omega(\log n)$, 其中, n 是链上的点数.

定理 6.1 不仅对图染色模型成立, 对一大类自然的自旋系统也都成立. 例如, 文献[26]对硬核模型也证明了类似的 $\Omega(\log n)$ 下界结论.

对于有远距离强相关性的自旋系统, 可以证明更强的 $\Omega(diam)$ 下界, 这里的 $diam$ 是通信网络 $G=(V,E)$ 的直径. 注意到, 任何分布式计算问题都可以由 $O(diam)$ 的 LOCAL 算法解决(某个点收集全图信息, 利用不受限的本地计算能力计算答案, 最后全网广播答案). 这说明, 此类分布不存在基于局部信息的采样算法. 考虑第 2.2 节定义的硬核模型 (V,E,λ) . 令 Δ 为图 G 的最大度数. 定义唯一性阈值:

$$\lambda_c(\Delta) \triangleq \frac{(\Delta-1)^{(\Delta-1)}}{(\Delta-2)^\Delta}.$$

在经典的图灵机计算模型中, 如果 $\lambda \leq \lambda_c(\Delta)$, 则存在多项式时间采样算法^[4]; 如果 $\lambda > \lambda_c(\Delta)$, 除非 $NP=RP$, 否则不存在多项式时间采样算法. 在分布式计算模型中, 也有类似的下界结论.

定理 6.2(分布式采样的 $\Omega(diam)$ 下界). 令 $\Delta \geq 3, \lambda > \lambda_c(\Delta)$. 令 $\varepsilon > 0$ 是一个足够小的常数. 对于任意 $N > 0$, 存在一个 $\Theta(N)$ 个点的图 G , 图的最大度数为 Δ 且直径为 $diam=\Omega(N^{1/11})$. 对于图 G 上参数为 λ 的硬核模型, 任何一个能以 ε 全变差误差采样硬核模型吉布斯分布的 LOCAL 算法的时间复杂度为 $\Omega(diam)$.

硬核模型是图上所有独立集的加权分布. 注意到分布式算法构造一个独立集的复杂度为 0, 每个点直接输出不在独立集中即可. 定理 6.2 指出, 在唯一性阈值以外, 采样一个加权独立集的任务不可能被分布式算法完成. 这就说明, 在一些模型上, 分布式采样问题的难度远高于分布式构造问题.

6.2 分布式 Jerrum-Valiant-Varzirani 定理

经典的 Jerrum-Valiant-Varzirani 定理(定理 3.2)指出: 在图灵机计算模型上, 对于有自归约性质的概率分布, 采样问题和计数问题可以在多项式时间内相互归约. 本节在分布式计算模型上研究采样和计数问题之间的归约关系.

重申在分布式计算中, (G, Σ, \mathbf{x}) 表示分布式采样/计数问题实例, 其中, $G=(V,E)$ 是一个 $n=|V|$ 个点 LOCAL 模型的通信网络; Σ 是一个大小为 $q=|\Sigma|=poly(n)$ 的取值范围; $\mathbf{x}=(x_v)_{v \in V}$ 是一个 n 维向量, 每一个 x_v 是一个比特串. 实例 (G, Σ, \mathbf{x}) 唯一编码一个样本空间 Σ^V 上的联合分布 $\mu=\mu_{(G, \Sigma, \mathbf{x})}$. 我们称 μ 为实例 (G, Σ, \mathbf{x}) 的目标分布. 给定实例 (G, Σ, \mathbf{x}) 后, 每个点 $v \in V$ 的输入为 Σ 和 x_v . 相应的分布式采样和推断(计数)问题定义在第 4.1 节中给出.

类似定义 3.1, 首先对分布式采样和计数问题定义有自归约性质的联合分布类.

定义 6.1(分布式计算版本的有自归约性质的联合分布类). 令 \mathfrak{M} 为一个联合分布类. 如果 \mathfrak{M} 满足如下两个性质, 则 \mathfrak{M} 具有分布式计算版本的自归约性质. 对于任意 $\mu_{(G,\Sigma,x)} \in \mathfrak{M}$, 其中, $G=(V,E)$, $\mathbf{x}=(x_v)_{v \in V}$, 对于任意集合 $A \subseteq V$ 上的合法部分配置 $\tau \in \Sigma^A$, 存在一个定义在同一个通信网络上的联合分布 $\mu_{(G,\Sigma,x')} \in \mathfrak{M}$, 满足 $\mu_{(G,\Sigma,x')} = \mu_{(G,\Sigma,x)}^\tau$, 其中, $\mathbf{x}'=(x'_v)_{v \in V}$; 且对于任意 $v \in A$, x'_v 包含 x_v 和 τ_v 的信息, 对于任意 $v \in V \setminus A$, $x'_v = x_v$.

与定义 3.1 相比, 定义 6.1 同样要求 \mathfrak{M} 对条件概率封闭, 不同之处在于: 定义 6.1 要求给定原始分布和条件之后, 相应的条件概率分布信息只要局部性地输入到一些点上.

重申在第 4.1 节的定义中, 分布式算法结束时每个点 $v \in V$ 输出一个值 $F_v \in \{0,1\}$, 表示点 $v \in V$ 局部的算法是否失败, 且算法需要满足 $\sum_{v \in V} \mathbb{E}[F_v] = O(1/n)$, 即算法总是以高概率成功, 其中, $n=|V|$ 表示通信网络的点数. 本节考虑的所有算法都是一个联合分布类 \mathfrak{M} 上的算法, 即: 给定分布类 \mathfrak{M} 上的任意一个具体分布 $\mu \in \mathfrak{M}$, 算法都能解决联合分布 μ 上的采样和推断问题.

6.2.1 一般联合分布上分布式采样和推断归约

首先注意到: 对于分布式推断问题, 点 v 输出的正确性只与输入的问题实例有关, 与其他点的输出无关. 考虑一个有可能失败的分布式推断算法, 因为失败可以本地检测, 利用不受限的本地计算能力, 点 v 可以本地枚举所有可能的随机比特, 然后选择一个让算法成功的随机比特执行. 这样就可以得到一个一定成功的确定分布式近似推断算法.

命题 6.1(随机近似推断算法 \Rightarrow 近似推断算法). 令 $\mathfrak{M}=\{\mu_{(G,\Sigma,x)}\}$ 为任意一类联合分布. 如果存在一个以时间复杂度 $t(n,\delta)$ 作近似推断的随机 LOCAL 算法(全变差误差为任意 $\delta > 0$), 且随机推断算法失败的概率满足 $\sum_{v \in V} \mathbb{E}[F_v] < 1$, 则存在一个以时间复杂度 $t(n,\delta)$ 作近似推断(全变差误差为任意 $\delta > 0$)的确定 LOCAL 算法.

虽然在第 4.1 节的定义中, 我们要求推断算法以高概率成功, 但是命题 6.1 中的归约成立只要求推断算法成功的概率大于 0. 而且命题 6.1 的结论对任意联合分布类 \mathfrak{M} 都成立, 不要求 \mathfrak{M} 满足自归约性质.

对于一般联合分布, 分布式近似推断算法与分布式近似采样算法有如下归约关系.

定理 6.3(近似推断算法 \Rightarrow 近似采样算法^[24]). 令 $\mathfrak{M}=\{\mu_{(G,\Sigma,x)}\}$ 为一个满足定义 6.1 中自归约性质的联合分布类. 如果存在一个以 $t(n,\delta)$ 时间复杂度作近似推断(全变差误差为任意 $\delta > 0$)的 LOCAL 算法, 则存在一个以 $O\left(t\left(n, \frac{\delta}{n}\right) \log^2 n\right)$ 时间复杂度作近似采样(全变差误差为任意 $\delta > 0$)的 LOCAL 算法.

定理 6.4(近似采样算法 \Rightarrow 近似推断算法^[24]). 令 $\mathfrak{M}=\{\mu_{(G,\Sigma,x)}\}$ 为一个满足定义 6.1 中自归约性质的联合分布类. 如果存在一个以 $t(n,\delta)$ 时间复杂度作近似采样(全变差误差为任意 $\delta > 0$)的 LOCAL 算法, 且采样算法失败的概率 $\sum_{v \in V} \mathbb{E}[F_v] < 1$, 则存在一个以 $t(n,\delta)$ 时间复杂度作近似推断(全变差误差为任意 $\delta > 0$)的 LOCAL 算法.

虽然在第 4.1 节的定义中, 我们要求采样算法以高概率成功, 但是定理 6.4 中的归约成立只要求采样算法成功的概率大于 0.

6.2.2 局部吉布斯分布上分布式采样和推断的归约

对于局部吉布斯分布(定义 4.1), 不仅第 6.2.1 节的结论成立, 定理 6.3 的结论可以加强为精确采样.

定理 6.5(近似推断算法 \Rightarrow 精确采样算法^[24]). 令 $\mathfrak{M}=\{\mu_{(G,\Sigma,x)}\}$ 为一类满足定义 6.1 中自归约性质的局部吉布斯分布. 令 $q=|\Sigma|=poly(n)$. 如果存在一个以 $t(n)$ 时间复杂度作近似推断(全变差误差不超过 $1/(5qn^4)$)的 LOCAL 算法, 则存在一个以 $O(t(n)\log^2 n)$ 时间复杂度作精确采样的 LOCAL 算法.

对于局部吉布斯分布, 推断问题的精度也可以加强. 因为有命题 6.1 的存在, 我们可以假设所有的推断算法一定成功. 考虑第 4.1 节定义的分布式近似推断问题. 给定任意问题实例 (G,Σ,\mathbf{x}) 任意误差 $\varepsilon > 0$, 算法结束时, 每个点输出一个 Σ 上的边缘分布 $\tilde{\mu}_v$, 满足 $d_{TV}(\tilde{\mu}_v, \mu_{v,(G,\Sigma,\mathbf{x})}) \leq \varepsilon$, 其中, $\mu_{v,(G,\Sigma,\mathbf{x})}$ 表示目标分布 $\mu_{(G,\Sigma,\mathbf{x})}$ 投影到 v 上的边缘分布. 现在定义一个更强的乘性误差, 令 ν_1 和 ν_2 为 Σ 上的两个概率分布, 定义 $err(\nu_1, \nu_2)$ 为

$$err(\nu_1, \nu_2) = \max_{c \in \Sigma} |\ln \nu_1(c) - \ln \nu_2(c)|.$$

规定, $0/0=1$, $\ln 0 - \ln 0 = \ln \frac{0}{0} = 0$.

• 带乘性误差的分布式近似推断

给定任意问题实例 (G, Σ, \mathbf{x}) 以及任意误差 $\varepsilon > 0$, 算法结束时每个点输出一个 Σ 上的边缘分布 $\tilde{\mu}_v$, 满足 $err(\tilde{\mu}_v, \mu_{v,(G,\Sigma,\mathbf{x})}) \leq \varepsilon$, $\mu_{v,(G,\Sigma,\mathbf{x})}$ 表示目标分布 $\mu_{(G,\Sigma,\mathbf{x})}$ 投影到 v 上的边缘分布.

对于足够小的 $\varepsilon > 0$, 条件 $err(\tilde{\mu}_v, \mu_{v,(G,\Sigma,\mathbf{x})}) \leq \varepsilon$ 可以推出:

$$\forall c \in \Sigma, 1 - \varepsilon \approx \exp(-\varepsilon) = \frac{\tilde{\mu}_v(c)}{\mu_{v,(G,\Sigma,\mathbf{x})}} \leq \exp(\varepsilon) \approx 1 + \varepsilon.$$

这就给出了比全变差误差更强的准确性.

定理 6.6(全变差误差近似推断 \Rightarrow 乘性误差近似推断^[24]). 对于任意一类满足定义 6.1 中自归约性质的局部吉布斯分布 $\mathfrak{M} = \{\mu_{(G,\Sigma,\mathbf{x})}\}$, 如果存在一个以 $t(n, \delta)$ 时间复杂度作近似推断(全变差误差为任意 $\delta > 0$)的 LOCAL 算法, 则存在一个以 $O\left(t\left(n, \frac{\varepsilon}{5qn}\right)\right)$ 时间复杂度作近似推断(乘性误差为 $0 \leq \varepsilon < 1$)的 LOCAL 算法, 其中, $q = |\Sigma| = poly(n)$ 是取值集合的大小.

6.3 分布式采样/计数与强空间混合性质

空间混合性质是吉布斯分布的重要性质, 也是一些吉布斯分布可被高效采样的充分必要条件^[3,4]. 文献 [24] 研究了强空间混合性质与分布式采样的关系, 并证明了: 对于局部吉布斯分布, 一定程度的强空间混合性质是存在高效采样算法的充分必要条件.

令 (V, Σ, \mathcal{F}) 为定义 2.1 中的吉布斯分布, 其中, \mathcal{F} 是一系列约束 (S, f) 的集合. 为了描述强空间混合性质, 我们先引入一些定义.

- 区域吉布斯分布: 令 μ 是 (V, Σ, \mathcal{F}) 定义的吉布斯分布. 任意子集 $R \subseteq V$ 可以诱导出一个 Σ^R 上的吉布斯分布 μ_R , 对于任意 $\sigma \in \Sigma^R$, $\mu_R(\sigma) \propto w_R(\sigma)$, 其中, $w_R(\sigma) = \prod_{(f,S) \in \mathcal{F}: S \subseteq R} f(\sigma_S)$;
- 边缘分布: 给定一个 $A \subseteq V$, 对于任意 $R \subseteq V$ 且满足 $R \supseteq A$, 任意相对于区域吉布斯分布 μ_R 合法的部分配置 $\tau \in \Sigma^A$, 对于任意点 $v \in R$, 区域边缘分布 $\mu_{R,v}^\tau$ 定义为

$$\forall x \in \Sigma, \mu_{R,v}^\tau(x) = \Pr_{Y \sim \mu_R} [Y_v = x | Y_A = \tau] = \frac{\sum_{\sigma \in \Sigma^R: \sigma_A = \tau \wedge \sigma_v = x} w_R(\sigma)}{\sum_{\sigma \in \Sigma^R: \sigma_A = \tau} w_R(\sigma)}.$$

上述分布良定义当且仅当 τ 相对于 μ_R 合法.

令 μ 是一个被 (V, Σ, \mathcal{F}) 定义的吉布斯分布. 令 $H = (V, F)$ 是一个点集为 V , 超边集合为 $F = \{S | (f, S) \in \mathcal{F}\}$ 的超图. 一个子集 $A \subseteq R$ 称为 R 在吉布斯分布 μ 上的一个内边界当且仅当把 A 删除后, 点集 $R \setminus A$ 和 $V \setminus R$ 在超图 $H = (V, F)$ 上不连通. 根据定义和吉布斯分布的条件独立性(命题 2.1), 有如下结论.

命题 6.2. 令 $v \in R \subseteq V$, 且 $A \subseteq R$ 是 R 在 μ 上的一个内边界. 对于任意相对于 μ 的合法配置 $\tau \in \Sigma^A$, 区域吉布斯分布的边缘分布 $\mu_{R,v}^\tau$ 和吉布斯分布的边缘分布 μ_v^τ 满足 $\mu_{R,v}^\tau = \mu_v^\tau$.

令 $G = (V, E)$ 为一个通信网络, 假设 (V, Σ, \mathcal{F}) 是一个 G 上的局部吉布斯分布. 重申 $dist_G(u, v)$ 表示 G 上两点 $u, v \in V$ 的最短路距离. 对任意集合 $S \subseteq V$, 令 $dist_G(u, v) = \min\{dist_G(u, v) | v \in S\}$. 定义如下强空间混合性质.

定义 6.2(强空间混合性质). 令 $\delta_n: \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ 是一个单调不减的函数序列. 令 $\mathfrak{M} = \{\mu_{(G,\Sigma,\mathbf{x})}\}$ 是一类局部吉布斯分布. 如果 \mathfrak{M} 满足如下性质, 则称 \mathfrak{M} 有速率为 δ_n 的强空间混合性质. 对于任意分布 $\mu = \mu_{(G,\Sigma,\mathbf{x})}$, 其中, $G = (V, E)$ 是一个 n 个点图通信网络, 对任意 $R \subseteq V$, 任意 R 在 μ 上的内边界 $A \subseteq R$ 以及两个相对于区域吉布斯分布 μ_R 合法的配置 $\sigma, \tau \in \Sigma^A$, 对任意点 $v \in R$ 都有:

$$d_{TV}(\mu_{R,v}^\sigma, \mu_{R,v}^\tau) \leq \delta_n(dist_G(v, D)),$$

其中, $D = \{u \in A | \sigma_u \neq \tau_u\}$ 是 σ 和 τ 取值不同的集合.

特别地, 对于任意 $0 < \alpha < 1$, 我们称 \mathfrak{M} 有速率为 α 的指数级衰减的强空间混合性质当且仅当函数序列 $\delta(\cdot)$ 满足 $\delta_n(t) \leq poly(n) \cdot \alpha^t$.

分布式推断问题与强空间混合性质有如下等价关系.

定理 6.7(分布式推断与强空间混合性质的联系). 对于任意有自归约性质的局部吉布斯分布类 $\mathfrak{M}=\{\mu_{(G,\Sigma,x)}\}$, 如果存在一个以 $t(n,\delta)$ 时间复杂度作近似推断(全变差误差为任意 $\delta>0$) 的 LOCAL 算法, 则 \mathfrak{M} 有速率为 $\delta_n(t)=2\min\{\delta|t(n,\delta)\leq t-1\}$ 的强空间混合性质; 相反地, 如果 \mathfrak{M} 有速率为 $\delta_n(\cdot)$ 的强空间混合性质, 则存在一个以 $t(n,\delta)=\min\{t|\delta_n(t)\leq \delta\}+O(1)$ 时间复杂度作近似推断(全变差误差为任意 $\delta>0$) 的 LOCAL 算法.

定义 6.2 的强空间混合性质由全变差误差定义, 根据定理 6.6 和定理 6.7, 对于此类分布存在一个乘性误差的分布式推断算法, 所以分布本身一定有乘性误差的强空间混合性质.

推论 6.1(加性与乘性误差的强空间混合性质^[24]). 一个局部吉布斯分布类 \mathfrak{M} 相对于全变差误差有速率为 α 的指数级衰减的强空间混合性质当且仅当 \mathfrak{M} 相对于乘性误差有速率为 α 的指数级衰减的强空间混合性质.

结合定理 6.7 和定理 6.5 可得到如下推论.

推论 6.2(强空间混合性质 \Rightarrow 精确采样算法^[24]). 令 $\mathfrak{M}=\{\mu_{(G,\Sigma,x)}\}$ 为一个有自归约性质的局部吉布斯分布类. 如果 \mathfrak{M} 有速率为 α 的指数级衰减的强空间混合性质, 则存在一个以 $O\left(\frac{1}{1-\alpha}\log^3 n\right)$ 时间复杂度作精确采样的 LOCAL 算法.

结合已有的强空间混合性质的结论, 推论 6.2 可以得到如下 LOCAL 模型上的精确采样算法.

- 因为图匹配的均匀分布有速率为 $1-\Omega(1/\sqrt{\Delta})$ 的指数级衰减的强空间混合性质^[15], 所以此条件下存在一个以 $O(\sqrt{\Delta}\log^3 n)$ 时间复杂度精确采样均匀图匹配的 LOCAL 算法;
- 因为硬核模型 (V,E,λ) (定义见第 2.2 节)在 $\lambda<\lambda_c(\Delta)=(\Delta-1)^{(\Delta-1)}/(\Delta-2)^\Delta$ 时有指数级衰减的强空间混合性质^[4], 所以此条件下存在一个以 $O(\log^3 n)$ 时间复杂度精确采样硬核模型的 LOCAL 算法;
- 因为图的 q 染色问题在 $q\geq \alpha\Delta+\beta(\alpha)$ 且图不含三角形时有强空间混合性质, 其中, $\alpha\geq \alpha^*$ 且 $\alpha^*=1.763\dots$ 满足 $\alpha^*=\exp(1/\alpha^*)$ ^[75], 所以此条件下存在一个以 $O(\log^3 n)$ 时间复杂度精确采样均匀图染色的 LOCAL 算法;
- 因为一般的反铁磁 2-自旋模型在唯一性条件以内有指数级衰减的强空间混合性质, 所以此条件下存在一个以 $O(\log^3 n)$ 时间复杂度精确采样反铁磁 2-自旋模型的 LOCAL 算法^[51];
- 因为加权超图匹配模型在唯一性条件以内有指数级衰减的强空间混合性质, 所以此条件下存在一个以 $O(\log^3 n)$ 时间复杂度精确采样加权超图匹配的 LOCAL 算法^[76].

上述所有模型和条件的具体定义可以在相应的参考文献中找到. 其中有些模型是边上的分布(图匹配、超图匹配), 它们可以转换成相应线图(line graph)上的分布, 这个转化保留了原图上对距离的定义.

对于硬核模型, 当 $\lambda>\lambda_c(\Delta)$ 时, 根据定理 6.2, 分布式采样算法的复杂度为 $\Omega(\text{diam})$; 而当 $\lambda<\lambda_c(\Delta)$ 时, 则存在一个 $\text{polylog}(n)$ 复杂度的分布式采样算法. 硬核模型在阈值 $\lambda_c(\Delta)$ 的计算相变现象是图灵机模型上的经典结论, 而这个计算相变结论在分布式计算上依然成立.

7 公开问题

目前, 关于分布式采样理论的研究才刚刚起步, 该领域依然存在很多公开问题.

- 先前的一些工作为分布式采样设计了专门的算法, 例如分布式马尔可夫链和局部拒绝采样算法. 相对于串行采样算法, 这些算法的收敛条件相对苛刻. 为分布式采样研究专门的算法设计和分析技术, 依然是一个重要的研究方向;
- 在分布式计算模型上精确模拟串行采样算法, 这是一种重要的算法设计方案, 因为由此得出的分布式采样算法有着与串行采样算法一致的性能. 目前的技术只适用于梅特罗波利斯算法, 如何为一般的串行马尔可夫链设计模拟算法, 是一个重要的公开问题;
- 分布式采样复杂性的研究才刚刚起步, 目前只研究了几个基本的问题. 复杂性上还有很多值得进一步探索的问题, 例如为分布式采样定义复杂度类、复杂度层级等;

- 分布式采样复杂度的研究结果和研究技术也能应用到其他领域. 在复杂性的研究中, 目前的一些归约需要使用到分布式模型不受限制的本地计算能力. 如果能够设计一个在计算上更加高效的归约, 那么分布式模型上的结果就能应用到图灵机模型上. 在分布式采样下界的证明中, 证明的关键是利用算法能产生的分布与目标分布之间的矛盾, 这种技术能否用于其他分布式计算问题下界的证明, 也值得深入研究.

References:

- [1] Sinclair A, Jerrum M. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 1989, 82(1): 93–133.
- [2] Jerrum M, Valiant LG, Vazirani VV. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 1986, 43: 169–188.
- [3] Sly A. Computational transition at the uniqueness threshold. In: *Proc. of the 51st Annual IEEE Symp. on Foundations of Computer Science (FOCS)*. 2010. 287–296.
- [4] Weitz D. Counting independent sets up to the tree threshold. In: *Proc. of the 38th Annual ACM Symp. on Theory of Computing (STOC)*. ACM, 2006. 140–149.
- [5] Ahmed A, Aly M, Gonzalez JE, Narayanamurthy S, Smola AJ. Scalable inference in latent variable models. In: *Proc. of the 5th ACM Int'l Conf. on Web Search and Data Mining (WSDM)*. 2012. 123–132.
- [6] Daskalakis C, Dikkala N, Jayanti S. HOGWILD!—Gibbs can be PanAccurate. In: *Proc. of the 31st Advances in Neural Information Processing Systems (NIPS)*. 2018. 32–41.
- [7] De Sa C, Olukotun K, Ré C. Ensuring rapid mixing and low bias for asynchronous Gibbs sampling. In: *Proc. of the 33rd Int'l Conf. on Machine Learning (ICML)*. 2016. 1567–1576.
- [8] De Sa C, Zhang C, Olukotun K, Ré C. Rapidly mixing Gibbs sampling for a class of factor graphs using hierarchy width. In: *Proc. of the 29th Annual Conf. on Neural Information Processing Systems (NIPS)*. 2015. 3097–3105.
- [9] Doshi-Velez F, Mohamed S, Ghahramani Z, Knowles DA. Large scale nonparametric Bayesian inference: Data parallelisation in the Indian buffet process. In: *Proc. of the 23rd Annual Conf. on Neural Information Processing Systems (NIPS)*. 2009. 1294–1302.
- [10] Gonzalez JE, Low Y, Gretton A, Guestrin C. Parallel Gibbs sampling: From colored fields to thin junction trees. In: *Proc. of the 14th Int'l Conf. on Artificial Intelligence and Statistics (AISTATS)*. 2011. 324–332.
- [11] Kandasamy K, Krishnamurthy A, Schneider J, Póczos B. Parallelised Bayesian optimisation via thompson sampling. In: *Proc. of the 21st Int'l Conf. on Artificial Intelligence and Statistics (AISTATS)*. 2018. 133–142.
- [12] Newman D, Asuncion A, Smyth P, Welling M. Distributed inference for latent Dirichlet allocation. In: *Proc. of the 20th Int'l Conf. on Neural Information Processing Systems (NIPS)*. 2007. 1081–1088.
- [13] Smyth P, Welling M, Asuncion AU. Asynchronous distributed learning of topic models. In: *Proc. of the 22nd Annual Conf. on Neural Information Processing Systems (NIPS)*. 2009. 81–88.
- [14] Yan F, Xu N, Qi Y. Parallel inference for latent Dirichlet allocation on graphics processing units. In: *Proc. of the 23rd Annual Conf. on Neural Information Processing Systems (NIPS)*. 2009. 2134–2142.
- [15] Bayati M, Gamarnik D, Katz D, Nair C, Tetali P. Simple deterministic approximation algorithms for counting matchings. In: *Proc. of the 39th Annual ACM Symp. on Theory of Computing (STOC)*. ACM, 2007. 122–127.
- [16] Bhandari S, Chakraborty S. Improved bounds for perfect sampling of k -colorings in graphs. In: *Proc. of the 52nd Annual ACM Symp. on Theory of Computing (STOC)*. 2020. 631–642.
- [17] Chen S, Delcourt M, Moitra A, Perarnau G, Postle L. Improved bounds for randomly sampling colorings via linear programming. In: *Proc. of the 30th Annual ACM-SIAM Symp. on Discrete (SODA)*. 2019. 2216–2234.
- [18] Feng W, Sun Y, Yin Y. What can be sampled locally? In: *Proc. of the 36th ACM Symp. on Principles of Distributed Computing (PODC)*. 2017. 121–130.
- [19] Linial N. Distributive graph algorithms Global solutions from local data. In: *Proc. of the 28th IEEE Annual Symp. on Foundations of Computer Science (FOCS)*. 1987. 331–335.
- [20] Naor M, Stockmeyer L. What can be computed locally? *SIAM Journal on Computing (SIAMP)*, 1995, 24(6): 1259–1277.

- [21] Feng W, Hayes TP, Yin Y. Distributed symmetry breaking in sampling (optimal distributed randomly coloring with fewer colors). arXiv preprint arXiv: 1802.06953, 2018.
- [22] Fischer M, Ghaffari M. A simple parallel and distributed sampling technique: Local glauber dynamics. In: Proc. of the 32nd Int'l Symp. on Distributed Computing (DISC). 2018. 26:1–26:11.
- [23] Feng W, Vishnoi NK, Yin Y. Dynamic sampling from graphical models. In: Proc. of the 51st Annual ACM Symp. on Theory of Computing (STOC). ACM, 2019. 1070–1081.
- [24] Feng W, Yin Y. On local distributed sampling and counting. In: Proc. of the 37th ACM Symp. on Principles of Distributed Computing (PODC). 2018. 189–198.
- [25] Feng W, Guo H, Yin Y. Perfect sampling from spatial mixing. arXiv preprint arXiv: 1907.06033, 2019.
- [26] Guo H, Jerrum M, Liu J. Uniform sampling through the Lovász local lemma. *Journal of the ACM (JACM)*, 2019, 66(3): 18:1–18:31.
- [27] Hayes TP, Sinclair A. A general lower bound for mixing of single-site dynamics on graphs. *The Annals of Applied Probability*, 2007, 17(3): 931–952.
- [28] Sly A, Sun N. Counting in two-spin models on d -regular graphs. *The Annals of Probability*, 2014, 42(6): 2383–2416.
- [29] Mezard M, Montanari A. *Information, Physics, and Computation*. Oxford University Press, 2009.
- [30] Bishop CM. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [31] Koller D, Friedman N, Bach F. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.
- [32] Levin DA, Peres Y. *Markov Chains and Mixing Times*. American Mathematical Soc., 2017.
- [33] Fill JA. An interruptible algorithm for perfect sampling via Markov chains. In: Proc. of the 29th Annual ACM Symp. on Theory of Computing (STOC). 1997. 688–695.
- [34] Fill JA, Huber M. The randomness recycler: A new technique for perfect sampling. In: Proc. of the 41st IEEE Annual Symp. on Foundations of Computer Science (FOCS). IEEE, 2000. 503–511.
- [35] Guo H, He K. Tight bounds for popping algorithms. *Random Structures & Algorithms*, 2020, 57(2): 371–392.
- [36] Guo H, Jerrum M. A polynomial-time approximation algorithm for all-terminal network reliability. *SIAM Journal on Computing (SICOMP)*, 2019, 48(3): 964–978.
- [37] Propp JG, Wilson DB. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures & Algorithms*, 1996, 9(1–2): 223–252.
- [38] Huber M. Perfect sampling using bounding chains. *The Annals of Applied Probability*, 2004, 14(2): 734–753.
- [39] Huber M. *Perfect Simulation*. Chapman and Hall/CRC, 2016.
- [40] Valiant LG. The complexity of computing the permanent. *Theoretical Computer Science*, 1979, 8(2): 189–201.
- [41] Jerrum M. *Counting, Sampling and Integrating: Algorithms and Complexity*. Springer Science & Business Media, 2003.
- [42] Jerrum M, Sinclair A, Vigoda E. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM (JACM)*, 2004, 51(4): 671–697.
- [43] Jerrum M. A very simple algorithm for estimating the number of k -colorings of a low-degree graph. *Random Structures & Algorithms*, 1995, 7(2): 157–165.
- [44] Jerrum M, Sinclair A. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing (SICOMP)*, 1993, 22(5): 1087–1116.
- [45] Bezáková I, Štefankovič D, Vazirani VV, Vigoda E. Accelerating simulated annealing for the permanent and combinatorial counting problems. *SIAM Journal on Computing (SICOMP)*, 2008, 37(5): 1429–1454.
- [46] Huber M. Approximation algorithms for the normalizing constant of Gibbs distributions. *The Annals of Applied Probability*, 2015, 25(2): 974–985.
- [47] Kolmogorov V. A faster approximation algorithm for the Gibbs partition function. In: Proc. of the 31st Annual Conf. on Learning Theory (COLT). PMLR, 2018. 228–249.
- [48] Štefankovič D, Vempala S, Vigoda E. Adaptive simulated annealing: A near-optimal connection between sampling and counting. *Journal of the ACM (JACM)*, 2009, 56(3): 18:1–18:36.

- [49] Qiu GL, Zhang CH. Approximability of partition functions of ferromagnetic two-state spin systems. *Computer Science*, 2020, 47(5): 22–26 (in Chinese with English abstract).
- [50] Zhang CH. Approximation algorithms for counting problems [Ph.D. Thesis]. Shanghai: Shanghai Jiao Tong University, 2016 (in Chinese with English abstract).
- [51] Li L, Lu P, Yin Y. Correlation decay up to uniqueness in spin systems. In: Proc. of the 24th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA). SIAM, 2013. 67–84.
- [52] Barvinok A. *Combinatorics and Complexity of Partition Functions*. Springer, 2016.
- [53] Patel V, Regts G. Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials. *SIAM Journal on Computing (SICOMP)*, 2017, 46(6): 1893–1919.
- [54] Moitra A. Approximate counting, the Lovász local lemma, and inference in graphical models. *Journal of the ACM (JACM)*, 2019, 66(2): 10:1–10:25.
- [55] Dyer ME, Frieze AM, Kannan R. A random polynomial time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)*, 1991, 38(1): 1–17.
- [56] Jerrum M. A very simple algorithm for estimating the number of k -colorings of a low-degree graph. *Random Structures & Algorithms*, 1995, 7(2): 157–165.
- [57] Vigoda E. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 2000, 41(3): 1555–1569.
- [58] Anari N, Liu K, Gharan SO. Spectral independence in high-dimensional expanders and applications to the hardcore model. In: Proc. of the 61st Annual Symp. on Foundations of Computer Science (FOCS). 2020. 1319–1330.
- [59] Dobrushin RL. Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, 1970, 15(3): 458–486.
- [60] Hayes TP. A simple condition implying rapid mixing of single-site dynamics on spin systems. In: Proc. of the 47th Annual IEEE Symp. on Foundations of Computer Science (FOCS). 2006. 39–46.
- [61] Dyer M, Goldberg LA, Jerrum M. Dobrushin conditions and systematic scan. In: Proc. of the 10th Int'l Workshop on Randomization and Computation (RANDOM). Springer, 2006. 327–338.
- [62] Gamarnik D, Katz D, Misra S. Strong spatial mixing of list coloring of graphs. *Random Structures & Algorithms*, 2015, 46(4): 599–613.
- [63] Goldberg LA, Martin R, Paterson M. Strong spatial mixing with fewer colors for lattice graphs. *SIAM Journal on Computing (SICOMP)*, 2005, 35(2): 486–517.
- [64] Chen Z, Galanis A, Štefankovič D, Vigoda E. Rapid mixing for colorings via spectral independence. In: Proc. of the 32nd ACM-SIAM Symp. on Discrete Algorithms (SODA). 2021. 1548–1557.
- [65] Chen Z, Liu K, Vigoda E. Optimal mixing of glauber dynamics: Entropy factorization via high-dimensional expansion. arXiv preprint arXiv: 2011.02075, 2020.
- [66] Dyer M, Frieze A, Hayes TP, Vigoda E. Randomly coloring constant degree graphs. *Random Structures & Algorithms*, 2013, 43(2): 181–200.
- [67] Dyer ME, Frieze AM. Randomly coloring graphs with lower bounds on girth and maximum degree. *Random Structures & Algorithms*, 2003, 23(2): 167–179.
- [68] Feng W, Guo H, Yin Y, Zhang C. Rapid mixing from spectral independence beyond the Boolean domain. In: Proc. of the 32nd ACM-SIAM Symp. on Discrete Algorithms (SODA). SIAM, 2021. 1558–1577.
- [69] Hayes TP. Local uniformity properties for Glauber dynamics on graph colorings. *Random Structures & Algorithms*, 2013, 43(2): 139–180.
- [70] Hayes TP, Vigoda E. Coupling with the stationary distribution and improved sampling for colorings and independent sets. *The Annals of Applied Probability*. 2006, 16(3): 1297–1318.
- [71] Glauber RJ. Time-dependent statistics of the Ising model. *Journal of Mathematical Physics*, 1963, 4(2): 294–307.
- [72] Galanis A, Štefankovič D, Vigoda E. Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. *Journal of the ACM (JACM)*, 2015, 62(6): 50.
- [73] Jonasson J. Uniqueness of uniform random colorings of regular trees. *Statistics & Probability Letters*, 2002, 57(3): 243–248.

- [74] Sinclair A, Srivastava P, Thurley M. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *Journal of Statistical Physics*, 2014, 155(4): 666–686.
- [75] Gamarnik D, Katz D. Correlation decay and deterministic FPTAS for counting list-colorings of a graph. In: *Proc. of the 18th Annual ACM-SIAM Symp. on Discrete (SODA)*. SIAM, 2007. 1245–1254.
- [76] Song R, Yin Y, Zhao J. Counting hypergraph matchings up to uniqueness threshold. *Information and Computation*, 2019, 266: 75–96.

附中文参考文献:

- [49] 邱国良, 张驰豪. 铁磁性双态自旋系统配分函数的可近似性. *计算机科学*, 2020, 47(5): 22–26.
- [50] 张驰豪. 计数问题的近似算法 [博士学位论文]. 上海: 上海交通大学, 2016.



凤维明(1994—), 男, 博士生, 主要研究领域为理论计算机科学.



尹一通(1981—), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为理论计算机科学.