arXiv:2301.09810v1 [cs.DM] 24 Jan 2023

异构箱中的均衡分配：记忆的力量*

Dimitrios Los[1], Thomas Sauerwald[1], 和 John Sylvester†[2]

[1]剑桥大学计算机科学与技术系，英国，firstname.lastname@cl.cam.ac.uk
[2]利物浦大学计算机科学系，英国，john.sylvester@liverpool.ac.uk

2023年1月25日

摘要

我们考虑将 m 个球（作业）分配到 n 个箱（服务器）中。在标准的双重选择过程中，每个步骤 $t = 1, 2, \ldots, m$ 我们先均匀地随机抽取两个 箱，并将一个球放入负载最小的箱中。众所周知，对于任意 $m \geqslant n$，这会产生 $\log_2 \log n + \Theta(1)$ 的差距（即最大负载与平均负载之间的差值），且该结果以高概率成立。本文中，我们研究 Memory 过程 [29]：该过程中每个步骤不再进行两次选择，而只抽取一个箱，但我们可以访问一个能够存储某个箱位置的缓存。Mitzenmacher、Prabhakar 和 Shah [25]证明，在轻负载情形下（$m = n$），Memory 过程可实现 $\mathcal{O}(\log \log n)$ 的差距。

我们将 Mitzenmacher 等人的设定从两个方面加以扩展。首先，允许球的数量 m 任意取值，这也涵盖了 $m \geqslant n$ 的重负载情形，而这是一种颇具挑战性的情况。其次，我们采用 Wieder 的异构箱模型 [32], ，其中箱子的抽样分布可以偏置至任意 乘法常数。稍令人意外的是，我们证明，即使在这种设定下，Memory 过程仍能达到$\mathcal{O}(\log \log n)$ 的差距界。这与双重选择（或任何满足 $d = \mathcal{O}(1)$ 的 d-选择）过程形成鲜明对比；已知在这些过程中，差距会随 $m \to \infty$ [32] 发散。此外，我们还证明，对于任何独立于 m （但可能依赖于 n）的抽样分布，Memory 过程的差距都可以用一个与 m 无关的界来限制。最后，我们证明，在另一种放宽的设定下，Memory 的紧确差距界为 $\mathcal{O}(\log n)$：该设定包含异构（加权）球，且缓存只能维持两个步骤。

1 引言

在这项工作中，我们研究球箱过程，其目标是将 m 个球（作业或任务）依次分配到 n 个箱子（处理器或服务器）中。球箱框架（又称均衡分配） [3] 是许多资源分配和存储问题的一种常用抽象，例如负载均衡、调度或哈希（参见综述 [26, 33]）。为了高效且去中心化地分配球，通常采用随机策略：为每个球抽取若干个箱子，然后将该球分配到其中一个箱子。该框架对理论和实践产生的深远影响，得到了“*ACM Paris Kanellakis* 理论与实践奖” 2020 [2]的认可。

*发表于 SODA 的论文完整版 2023 [22]

†J.S. 获得了英国工程与自然科学研究委员会项目 EP/T004878/1：利用图结构的多层算法学的资助。

众所周知，若将每个球独立且均匀地放入一个随机选定的箱子（称为单选），那么当 $m = n$ 时，最大负载以高概率为 $\Theta(\log n / \log \log n)$[1]；当 $m \geqslant n \log n$ 时，则以高概率为 $m/n + \Theta(\sqrt{(m/n) \log n})$。Azar、Broder、Karlin 和 Upfal [3]（以及 Karp、Luby 和 Meyer auf der Heide [16]的隐含结果）证明了一个卓越的结论：若将每个球放入 $d \geqslant 2$ 个随机选定的箱子中负载较小的那个箱子，那么当 $m = n$ 时，最大负载以高概率降至 $\log_d \log n + \mathcal{O}(1)$。从 $d = 1$（单选）到 $d = 2$（双重选择）的这一显著改进被称为“双重选择幂”。

后来，Berenbrink、Czumaj、Steger 和 Vöcking [6] 将 d-Choice 的分析扩展到了所谓的“重负载情形”，其中 $m \geqslant n$ 可以任意大。特别是，对于 Two-Choice，他们证明以高概率有一个上界：差距（最大负载与平均负载之差）为 $\log_2 \log n + \mathcal{O}(1)$。与轻负载情形相比，重负载情形更具挑战性，因为当 m足够大时，可能会遇到任意糟糕的配置[2]。 [27] 采用一种基于双曲余弦势函数的通用技术，分析了重负载情形下的一大类过程。Talwar 和 Wieder [30] 将这一技术加以扩展，从而为重负载情形下的 d-Choice 得到了一个稍弱的上界。

记忆设置。 本文重新考察 Shah 和 Prabhakar [29]引入的 Memory 过程。在 (d, D)-Memory 中，我们可以将 D 个箱子存入缓存。此外，在每个步骤中，会抽取 d个选择，将一个球分配到这些箱子中负载最小的那个箱子，然后更新缓存，使其包含 $d + D$ 个箱子中负载最小的 D 个。Mitzenmacher、Prabhakar 和 Shah [25] 证明，在轻负载情形下，对于 $f(d) \in (2d, 2d + 1)$，$(d, 1)$-Memory 的差距为 $\log_{f(d)} \log n + \mathcal{O}(1)$，这表明使用一个缓存大致相当于将选择次数加倍。Luczak 和 Norris [23] 在类似的排队场景中（称为“带记忆的超市模型”）证明了相同的上界。本文聚焦于 $(1, 1)$-Memory 的情形（下文简称 Memory），因为事实证明，这一版本本身就能带来有趣的见解。

一种相关的经典启发式方法是粘性路由，即只要节点之间的路由未发生严重拥塞，就会重复使用该路由。这种方法已在电话网络 [12] 和交换机调度算法 [11]中得到分析和应用。与均衡分配类似，Kirsch、Mitzenmacher 和 Wieder [17]证明了在布谷鸟哈希中使用容量为 1 的缓存所带来的性能提升。

异构箱。 在点对点系统、分布式哈希表和去中心化存储中心等许多应用中，箱子可能并不总是相同的 [8, 14, 32]。Byers、Considine 和 Mitzenmacher [8] 引入了一种模型，其中采样分布与均匀分布的偏差不超过一个常数因子（事实上，最大的采样概率可能比均匀分布高出一个对数因子）。他们证明，只要满足 $m = n$，仍可得到与均匀情形相同的渐近差距界。之后，[32] 将重负载情形推广到如下设定：箱子按照 (a, b) 偏置采样分布进行采样；这意味着对于实数 $a \geqslant 1$，$b \geqslant 1$，每个箱子 $i \in [n]$ 的采样概率 s_i 满足 $\frac{1}{an} \leqslant s_i \leqslant \frac{b}{n}$。他们证明了一个紧确的二分结果：对于任意常数 $a, b \geqslant 1$，存在一个$d' = d'(a, b)$，使得 d-Choice 配合 $d \geqslant (1 + \varepsilon) d'$ 可达到 $\mathcal{O}(\log \log n)$ 的差距；而对于d-Choice 配合 $d \leqslant (1 - \varepsilon) d'$，则存在一些采样分布，使差距随 m发散。

1.1 我们的结果

在本文中，我们将 Memory 过程的分析扩展到重负载情形，并证明即使采用 (a, b)-偏置采样分布，该过程仍能以高概率保持较小的差距。我们

[1]一般而言，“高概率”指概率至少为 $1 - n^{-c}$，其中 $c > 0$ 为某个常数。[2]严格来说，对于任意 $f_1 := f_1(n) \ldots, f_{n-1} := f_{n-1}(n) \in \mathbb{Z}$(以及 $f_n = -f_1 - \ldots - f_{n-1}$)，都存在 $g := g(n, f_1, \ldots, f_n)$，使得以高概率存在一个步骤 $0 \leqslant t \leqslant g$，满足 $x^t = \left(\frac{t}{n} + f_1, \ldots, \frac{t}{n} + f_{n-1}, \frac{t}{n} + f_n\right)$。

我们用 Gap(m) 表示分配 m 个球后最大负载与平均负载之间的差值。

定理 1.1。 考虑具有(a,b)-偏置采样分布的MEMORY 过程，其中常数 $a,b \geqslant 1$ 任意。则存在常数 $\kappa := \kappa(a,b) > 0$，使得对每个步骤$m \geqslant 1$,

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \leqslant \kappa \cdot \log\log n\,\right] \geqslant 1 - n^{-3}.$$

据我们所知，即使在均匀抽样分布的情形下（也就是说，$a = b = 1$），在重负载情形 $m \geqslant n$ 下， $\mathcal{O}$(loglog n) 的上界此前也尚未知晓（最佳结果是 [21] 给出的 $\mathcal{O}$(log n) 差距界）。

对于任意 常数$a, b > 1$，Memory 仍能达到双对数差距，这表明它相较于双重选择（或常数为 d 时的 d-选择）具有显著优势。Wieder [32] 指出，只有对于某些 （足够小的）常数 (a,b)，最大负载才是 n 的双对数；否则最大负载可能无界。也就是说，如果常数（a,b）较大，那么最大负载会随着 $m \to \infty$ 发散。

我们还通过给出一个下界补充了上界，从而证明该结果在乘法常数意义下是紧的:

定理 1.2。 考虑采用均匀采样分布的 MEMORY 过程。那么存在一个常数 $\kappa > 0$，使得对于每个步骤 $m \geqslant n$,

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \geqslant \kappa \cdot \log\log n\,\right] \geqslant 1 - n^{-1}.$$

定理 1.1 和定理 1.2 确定了当 (a,b)-偏置采样分布中的 a,b 为常数时，Memory 的渐近行为；而我们的下一个结果探讨了更广泛的采样分布变化：只需 $s_{\min}$ 独立于 m （但可能依赖于 n）有一个下界即可[3]。当然，这种一般性的代价是界不再紧密。不过，这足以表明它与双重选择又有一个显著差异：对于 (10,10)-偏置采样分布，双重选择的差距就已经在 $m \to \infty$ 中发散（见 [32]）。

定理 1.3。 考虑采用任意满足 $s_{\min} := \min_{i\in[n]} s_i > 0$ 的采样分布 $s = (s_i)^{i\in[n]}$ 的 MEMORY 过程。那么对于每个步骤 $m \geqslant 1$，都有

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \leqslant \frac{n^8}{s_{\min}^{10}}\,\right] \geqslant 1 - n^{-1}.$$

定理 1.1 证明的第一部分基于对 Memory 的一种“泄漏”版本进行分析，我们称之为 d-Weak-Memory。在这个版本中，每经过 d 步就会重置缓存，并且只有在这些步骤中才会更新负载信息（更正式的描述见第2节 2.2）。对于$d = 2$，这可以解释为 $(1+\beta)$-过程[4] 的一种节省采样次数的变体，其中 $\beta = 1/2$：第一个球采用单选分配，第二个球则放入最近两次选择中负载最小的那个。该过程每个球恰好采样一次，而原始 $(1+\beta)$-过程每步平均会进行 $1+\beta$ 次采样。

定理 1.4 (定理 4.1 的推论**)**。 对于任意 $a,b \geqslant 1$，存在常数 $d := d(a,b) \geqslant 2$, $\kappa := \kappa(a,b) > 0$，使得对于具有 (a,b)-偏置采样分布的 d-弱记忆 过程，以及每个步骤 $m \geqslant 1$,

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \leqslant \kappa \cdot \log n\,\right] \geqslant 1 - n^{-2}.$$

[3]排队场景中的一项相关结果可见于 [29, 定理 1], ，但其中并未量化差距。[4]回顾一下，$(1+\beta)$-过程 [27] 在每一步以概率 β执行双重选择，否则执行单选。

请注意，选择足够大的 d （取决于 a 和 b）非常重要。例如，否则可以选择 $a > d$ ，而差距会在 $m \to \infty$[5] 中发散。

通过将 d-Weak-Memory 与 $(1+\beta)$-process [27] 联系起来，很容易就能得到 $\Omega(\log n)$ 的紧下界（引理 7.3）。这两个结果表明，只要常数 d 足够大，差距就可能达到 $\Theta(\log n)$。

最后，事实证明，这种宽松设定的分析甚至可以推广到球带有权重的情形，只要权重是从矩母函数有限的分布中独立抽取的（ [27] 也针对不同的过程研究过这种情形）。

定理 1.5。 考虑具有均匀采样分布的 2-Weak-Memory 过程。此外，假设每个球的权重都从满足$\mathbf{E}[W] = 1$ 且 $\mathbf{E}[e^{\lambda W}] < \infty$ 的分布中独立抽取，其中 $\lambda > 0$ 为常数。那么存在一个常数 $\kappa := \kappa(W) > 0$，使得对每个步骤 $m \geqslant 1$,

$$\mathbf{Pr}\left[\operatorname{Gap}(m) \leqslant \kappa \cdot \log n\right] \geqslant 1 - n^{-2}.$$

我们认为，该结果可以推广到任意常数 $d \geqslant 2$ 对应的 d-Weak-Memory 情形，但为简洁起见，这里我们聚焦于这一特殊情形 $d = 2$ 。

1.2 挑战与技术

分析 Memory 过程的一大挑战，在于缓存引入了强烈的长期依赖性。这使得 Memory 与双重选择大不相同，尽管两者在差距界方面表面上颇为相似。事实上，Memory 能够存储并填满一个轻箱，这对于在有偏采样分布下实现 $\mathcal{O}(\log \log n)$ 的差距至关重要。相比之下，双重选择在不同步骤之间的决策本质上相互独立（如果不考虑负载向量的变化）。形式化地说，双重选择可以用一个时不变分布向量 p 表示，该向量规定了每一步将一个球分配到第 i 重的箱的概率 p_i 。这种分析方法结合双侧指数势函数（又称双曲余弦势函数），最早由 [27] 提出（此后也用于例如 [1, 21, 30]） 。该技术足以证明一个与 m 无关的 $\mathcal{O}(\log n)$ 差距界，并为推导更紧的差距界提供了有用的起点，正如 [19, 30] 所展示的那样。

对于 Memory，确立这一起始点，即基础情形，本身就颇具挑战。为减少依赖关系，我们将 Memory 化简为 d-Weak-Memory，这是 Memory 的一个版本，每经过 $d = \mathcal{O}(1)$步就重置缓存，并使用过时的负载信息。我们通过将 Memory 与 d-Weak-Memory 进行耦合来形式化这一化简；该耦合将前者的双曲势能界定为不超过后者的双曲势能。

接着，我们证明在 d-Weak-Memory 中，任意步骤的双曲势能在期望意义下为 $\mathcal{O}(n)$。尽管这一结果本身足以得到一个 $\mathcal{O}(\log n)$ 差距界，却无法证明负载分布以高概率呈现指数集中。为了得到这一更强的结论，我们研究了采用两个平滑参数的两个版本的双曲势能之间的相互作用。除对 Memory 的分析外，这一关于 d-Weak-Memory 的结果或许也具有独立意义，因为它揭示了缓存周期性重置的 Memory 版本的某些行为。最后，我们指出，这一集中性结果也有助于证明 Memory 的下界 $\Omega(\log \log n)$，且该下界在所有步骤中均以高概率成立 $m \geqslant n$。

在建立基础情形后，我们开始着手研究具有所有长期相关性的实际 Memory 过程。我们采用一种巧妙的划分方式，将步骤按轮次分隔开

[5]考虑某个具有 $s_i = 1/(2d^2n)$ 的抽样分布 s，其中 $i \in [n]$。在运行 d 步的过程中，该箱 i 被选中的概率至多为 $d \cdot 1/(2d^2n)$，因此，在任意运行 d 步的过程中，分配给 i 的球的期望数量小于 $d \cdot d \cdot 1/(2d^2n) = 1/(2n)$。这意味着，箱 i 的负载与平均值之差会随着 $m \to \infty$而发散；因此，差距也必然发散。

（见图1），并通过一种“折叠”程序归纳地定义轮次；只有在一轮开始时或连续足够多步都没有抽中轻箱，才能开始新一轮。具体而言，在折叠过程中，每轮开始时我们都会使用一个全新的缓存，因此分配采用单选。要证明 $\mathcal{O}(\log n)$ 的差距，只需每两步重置一次缓存（这一过程称为 2-Reset-Memory）。然而，要证明 $o(\log n)$ 的差距，我们需要考察长度为 $\omega(1)$ 的区间，以抵消这种可能不利的分配。本质上，我们证明两轮之间的时间足够长，且这一结论以足够高的概率成立；这意味着系统会强烈地远离重箱，从而降低它们对势函数的贡献。按照分层归纳的思路，我们分析一系列越来越陡峭、超指数增长的双曲势函数，并证明它们各自以高概率呈线性增长。完成 $\mathcal{O}(\log\log n)$ 步归纳后，我们最终得到所需的 $\mathcal{O}(\log\log n)$ 差距界。

与 [18, 20, 21, 27] 中的分析相比，主要区别在于，我们需要考虑势函数在越来越长、步数不固定的区间内的变化，这也要求采用形式略有不同的超指数势函数。另一项额外的挑战是，在基础情形中处理（a， b）偏置采样分布时需要格外谨慎，尤其是在计算将 d-Weak-Memory 与 Memory 耦合时的分配概率。

1.3 进一步相关工作

在 [13], 中，Godfrey 引入了一种具有相关选择的 d-选择模型，可将其视为超图上的平衡分配。Godfrey 的结果后来在 [4, 15], 中得到改进，而 [5]则研究了一种具有箱容量的相关模型。

从另一个角度来看，Vöcking [31]表明，通过精心设计的非对称协议，可以提升 d-选择的性能。在他的协议中， d 个样本均匀抽取，每个样本分别来自 d 个互不相交的箱子组之一，并以非对称方式打破平局。组合学中也使用了多个势函数来分析各种随机贪心过程。具体而言，[7] 使用了一个层次化鞅集，其中每个鞅都用于控制集合中位于其下方的鞅。

1.4 内容概览

第2节介绍必要的记号，并对所有过程作出更正式的定义。第3节简要概述了我们关于差距上界的证明，该证明进一步分为基础情形（第 3.1 节）和分层归纳步骤（第3.2节）。相应地，基础情形的完整证明推迟至第4节，分层归纳的完整证明则推迟至第 5 节。第6节证明了任意偏置分布和权重下的界。下界在第7节中证明。最后，我们在第8节总结主要结果，并指出一些开放问题。

2 记号

2.1 基本定义

我们考虑将 m 个球分配到 n 个箱子中，这些箱子标记为 $[n] := \{1, 2, \dots, n\}$。暂且假设这些 m 球没有权重（或者等价地说，所有球的权重均为 1）。对于任意步骤$t \geqslant 0$， x^t 是一个 n 维的**负载向量**，其中 x_i^t 表示在前 t 次分配中分配到箱子 i 中的球数。特别地，对每个 $i \in [n]$，都有 $x_i^0 = 0$ 。步骤 t 处的**差距** 定义为

$$\mathrm{Gap}(t) = \max_{i\in[n]} x_i^t - \frac{t}{n}.$$

将负载向量 x 按顺序排列也很方便。为此，重新标记 n 个箱子，使得 y^t 是 x^t 的一个排列（$\binom{t}{n}$、……、$\binom{t}{n}$），并且 $y_1^t \geqslant y_2^t \geqslant \cdots \geqslant y_n^t$。注意，$\sum_{i\in[n]} y_i^t = 0$ 且 $\mathsf{Gap}(t) = y_1^t$。此外，若对于所有 $1 \leqslant k \leqslant n$，前缀和都满足：$\sum_{i=1}^{k} v_i \geqslant \sum_{i=1}^{k} u_i$，则称向量 $v =$（v_1，v_2，……，v_n）主序 $u =$（u_1，u_2，……，u_n）。

以下 [27], 许多分配过程都可以用一个随时间不变的概率分配向量 $p = (p_i)^{i\in[n]}$来描述，使得在每个步骤 $t \geqslant 1$, p_i 表示将一个球分配到负载最高的第 i个箱子的概率。

我们用 $\mathfrak{F}^t$ 表示该过程截至步骤 t的滤过，特别是它揭示了前 t 个球的抽样与分配情况，包括负载向量 x^t。对于随机变量Y，Z，若 Y随机小于 Z（等价地，若 Y被 Z随机支配），则记作 $Y \preceq Z$；如果对所有实数 x都有$\mathbf{Pr}\left[Y \geqslant x\right] \leqslant \mathbf{Pr}\left[Z \geqslant x\right]$。

2.2 过程

本文研究平衡分配过程，其中抽样分布（即过程请求一个“随机”箱时提供的箱子索引的分布）不一定是均匀分布。令 $\mathcal{S}$ 为 n 个箱子上的一个抽样分布，它是一个向量$(s_1, \ldots, s_n)$，并用它来定义我们的过程。我们强调，尽管在分析中，我们通常会在每个步骤通过重新标记箱子，将负载向量按降序排列，但抽样分布为每个箱子分配一个固定概率，不会因重新标记而改变。特别地，$(s_1, \ldots, s_n)$不应与$(p_1, \ldots, p_n)$混淆。

我们首先对任意整数 $d \geqslant 1$ 的 d-选择过程给出形式化描述。

> d-选择（$\mathcal{S}$）过程：
> 迭代：对于每个步骤 $t \geqslant 0$，根据 $\mathcal{S}$ 独立抽取 d 个箱子 $i_1, \ldots, i_d$。令$i \in \{i_1, \ldots, i_d\}$ 为满足 $x_i^t = \min\{x_{i_1}^t, \ldots, x_{i_d}^t\}$ 的一个箱子；若出现并列，则随机打破平局。然后更新：
> $$x_i^{t+1} = x_i^t + 1.$$

显然，对于均匀采样分布，双重选择的概率分配向量为

$$p_i = \frac{2i-1}{n^2}, \qquad \text{for all } i \in [n].$$

我们的下一个过程是将 Memory 过程推广到非均匀箱子的一般化过程。该过程最初由 Prabhakar 和 Shah [29] 针对均匀采样分布提出。

> Memory（$\mathcal{S}$）过程：
> 迭代：将缓存箱 $b = \emptyset$ 设定为初始值，以初始化过程。对于每个步骤 $t \geqslant 0$，根据 $\mathcal{S}$ 抽取一个箱 i，并更新：
> $$\begin{cases} x_i^{t+1} = x_i^t + 1 & \text{if } x_i^t < x_b^t \text{ or } b = \emptyset \qquad (\text{also update cache } b = i), \\ x_i^{t+1} = x_i^t + 1 & \text{if } x_i^t = x_b^t, \\ x_b^{t+1} = x_b^t + 1 & \text{if } x_i^t > x_b^t. \end{cases}$$

接下来，我们介绍自己对 Memory 的一种变体：每经过 d 步就重置缓存；对于两次重置之间各步骤的分配，过程只能根据上次重置时按负载对箱排序的结果来作出分配决策。

d-弱记忆($\mathcal{S}$) 过程：

迭代：对于每个满足 $k \in \mathbb{N}$ 的步骤 $t = d \cdot k$，固定任意一种排序 $\sigma : [n] \to [n]$，使得对所有 $i, j \in [n]$, $\sigma(i) \leqslant \sigma(j)$ 当且仅当 $x^t_{\sigma(i)} \geqslant x^t_{\sigma(j)}$。然后根据 $\mathcal{S}$ 对箱子 i 进行采样，并更新：

$$x_i^{t+1} = x_i^t + 1, \qquad (\text{also update cache } b = i).$$

对于每个步骤 $t = d \cdot k + j$，其中 $k \in \mathbb{N}$ 和 $1 \leqslant j < d$，按照 $\mathcal{S}$抽样一个箱 i，并进行更新：

$$\begin{cases} x_i^{t+1} = x_i^t + 1 & \text{if } \sigma(i) > \sigma(b) \qquad (\text{also update cache } b = i), \\ x_b^{t+1} = x_b^t + 1 & \text{if } \sigma(i) \leqslant \sigma(b). \end{cases}$$

我们将步骤 $d \cdot k, \ldots, d \cdot (k+1) - 1$的序列称为一次运行，其中任意 $k \geqslant 0$均成立。对于 $d = 2$，这一过程可视为针对 $\beta = 1/2$ 的$(1+\beta)$过程的样本高效变体（采用均匀采样分布 $\mathcal{U}$时）。

2-弱记忆($\mathcal{U}$) 过程：

迭代：对于每个步骤 $2t \geqslant 0$，独立且均匀地抽样两个箱子 $i_1, i_2 \in [n]$，并进行更新：

$$\begin{cases} x_{i_1}^{2t+1} & = x_{i_1}^{2t} + 1, \\ x_i^{2t+2} & = x_i^{2t+1} + 1, \end{cases}$$

其中 $i \in \{i_1, i_2\}$满足 $x_i^{2t} = \min \{x_{i_1}^{2t}, x_{i_2}^{2t}\}$。

我们强调，对于 $d > 2$，d-Weak-Memory($\mathcal{S}$) 过程在某个步骤 t 中可能会出现 $x_i^t = x_b^t$, 以及$\sigma(i) > \sigma(b)$，这会使箱 i 的负载增加，同时也会导致缓存中的箱变为 i。这与 Memory 不同，后者会更“聪明”地增加i 的负载，而不更新缓存。

另一个值得进一步研究的相关过程，是d-Weak-Memory($\mathcal{S}$)的以下变体：缓存每隔 d 步也会重置，但负载信息每一步都会更新。从这个意义上说，该过程介于 d-Weak-Memory($\mathcal{S}$) 和 Memory($\mathcal{S}$)之间。

d-Reset-Memory($\mathcal{S}$) 过程：

迭代：对于满足 $k \in \mathbb{N}$ 的 $t = d \cdot k$，根据 $\mathcal{S}$ 采样一个箱 i 并更新：

$$x_i^{t+1} = x_i^t + 1 \qquad \text{and update cache } b = i.$$

对于 $t = d \cdot k + j$，其中 $k \in \mathbb{N}$ 且 $1 \leqslant j < d$，根据 $\mathcal{S}$ 采样一个箱 i，并更新：

$$\begin{cases} x_i^{t+1} = x_i^t + 1 & \text{if } x_i^t < x_b^t \qquad (\text{also update cache } b = i), \\ x_i^{t+1} = x_i^t + 1 & \text{if } x_i^t = x_b^t, \\ x_b^{t+1} = x_b^t + 1 & \text{if } x_i^t > x_b^t. \end{cases}$$

3 定理 1.1 证明概要

本节将从宏观层面概述定理 1.1 的证明，因为这是本文迄今为止最重要的结果，也占据了分析工作的绝大部分。

粗略地说，我们通过分层归纳，为一系列越来越陡峭的势函数证明一系列界。第 j 个这样的势函数（其中 $j \leqslant j_{\max}$）大致具有

形式 $\sum_{i\in[n]} e^{f_j\cdot(y_i^m-g_j)}$，适用于函数 $f_j := f_j(n)$ 和 $g_j := g_j(n)$。随着 j 增大，函数 $f_j(n)$ 和 $g_j(n)$ 随 n 增长而趋于更快地 ∞。对于最后一“层” $j = j_{\max}$，我们有 $f_j(n) = \Theta(\log n)$，以及 $g_j(n) = \Theta(\log\log n)$，因此，在第 $j_{\max}$ 层，势函数的界为 $\mathcal{O}(n)$，这意味着 $\mathrm{Gap}(m) = \max_{i\in[n]} y_i^m = \mathcal{O}(\log\log n)$。采用分层归纳的原因是：在每一层 j，第 j 层势函数的界都能给出负载至少达到给定值的箱子数量的尾界；正是利用这个界，才能证明第 $j+1$ 层势函数的界。

为启动这一归纳法，我们使用一个关于双曲余弦势函数的界。该界本身就足以证明该过程的 $\mathcal{O}(\log n)$ 差距界（回顾一下，我们最终的目标是 $\mathcal{O}(\log\log n)$）。要使这一论证严谨，还面临许多挑战。我们首先更详细地概述基础情形，随后介绍分层归纳。

3.1 基础情形

对于步骤 t 中的归一化负载向量 $(y_i^t)_{i\in[n]}$，带有平滑参数 $\alpha > 0$ 的 双曲余弦势函数 [27] 定义为

$$\Gamma^t := \sum_{i=1}^{n} e^{\alpha y_i^t} + \sum_{i=1}^{n} e^{-\alpha y_i^t}.$$

该势函数会根据箱子的归一化负载，以指数成本惩罚负载过高或欠载的箱子。对于分层归纳的基础情形，我们将使用以下事实：对于某个充分小的常数 $\alpha = \Theta(1)$，以高概率，任意步骤处的 $\Gamma^t = \mathcal{O}(n)$。由此我们可以证明，归一化负载至少为 v 的箱子数量至多为 $\mathcal{O}(n\cdot e^{-\alpha v})$，这对于分层归纳（第3.2节）至关重要。

定理 4.3 (重述)。考虑采用任意MEMORY 过程，其采样分布具有(a,b)偏差，且常数为 $a,b\geqslant 1$。那么，存在常数 $c := c(a,b) \geqslant 1$和 $0 < \alpha :=\alpha(a,b) < 1$，使得对于势函数 $\Gamma := \Gamma(\alpha)$以及任意步骤 $t\geqslant 1$，我们有

$$\mathbf{Pr}\left[\bigcap_{u\in[t,t+n\log^8 n]} \{\Gamma^u \leqslant 6cn\}\right] \geqslant 1 - n^{-4}.$$

为了证明这一结果，我们首先需要证明，经过足够多步之后，势函数的期望会呈下降趋势。

定理 *4.1*（重述）。考虑采用任意 *(a,b)*-偏差采样分布的 *Memory* 过程，其中常数为 $a,b\geqslant 1$。那么，存在常数 $\alpha' := \alpha'(a,b) > 0$, $c := c(a,b) \geqslant 1$和 $d := d(a,b) \geqslant 2$，使得对于势函数 $\Gamma := \Gamma(\alpha)$，任意 $\alpha \leqslant \alpha'$ 以及任意步骤$t\geqslant 1$，

$$\mathbf{E}\left[\Gamma^{t+d} \;\middle|\; \mathfrak{F}^t\right] \leqslant \Gamma^t\cdot\left(1-\frac{\alpha}{c\cdot n}\right) + c\cdot\alpha.$$

同样的界也适用于 d-WEAK-MEMORY 过程。

直接证明定理 4.1 的难点在于，Memory 过程中缓存的内容会使每个步骤所做的选择之间产生极长距离的依赖关系。为绕过这一问题，我们让 Memory 过程每隔 d 步就“忘掉”缓存中的内容，然后执行一次单选分配，以获得一个新的缓存箱。我们将这一过程称为d-Reset-Memory。

尽管 d-Reset-Memory 的分配之间只有有界范围的依赖，但这一过程仍有些难以理解，例如，很难写出概率分配向量——也就是说，在给定步骤中，它将一个球分配给第 i 重的箱的概率。因此，我们进一步限制 d-Reset-Memory，使其所有比较

都使用过时信息（即自上次清空缓存以来的信息）。我们将这一过程称为d-Weak-Memory。对于一类我们称为（a, b）阶梯分布的抽样分布，它的概率分配向量更容易处理。幸运的是，我们可以证明，在所有（a, b）偏置分布中，这些是 d-Weak-Memory 的最坏情况采样分布。有了这一结果，我们可以证明， d-Weak-Memory 的概率分配向量会偏向低负载箱；这足以证明定理 4.1 对 d-Weak-Memory 过程成立。接着，通过耦合，我们可以将这一结果与 Memory 过程的负载下降联系起来。

在定理 4.1 中建立期望上界和漂移不等式后，我们能够证明，对于足够小的常数平滑参数，以高概率，双曲余弦势函数为 $\mathcal{O}(n)$。为此，我们证明了一个关于双曲余弦势函数的一般集中性结果——定理 4.2；该结果与 [19, Section 4] 中的结果类似。证明依赖于两个具有不同平滑参数 α_1 和 α_2 的双曲余弦势函数实例 Γ_1 和 Γ_2 之间的相互作用，其中 α_1 比 α_2大一个常数倍；并且对于任意满足 $\Gamma_1^t = \text{poly}(n)$ 的步骤t，也有 $|\Gamma_2^{t+1} - \Gamma_2^t| \leqslant n^{1/3}$。通过对这样的坏事件进行条件化：在足够长的区间内， Γ_1 =poly(n)（由马尔可夫不等式和并集界可知，该情况以高概率发生），我们能够利用一种与独立有界差分法思路相近的不等式，证明 Γ_2 稳定在 $\mathcal{O}(n)$。

基础情形的完整自足证明见第4节。

3.2 分层归纳

（完整）势函数。我们将对超指数势函数进行分层归纳，这些势函数与 [19, 第 6] 节和 [20, 第 6], 节中使用的类似，但也有一些不同之处（见第33页的讨论）。现在我们将定义这些势函数（为避免过多技术细节，常数的精确定义以及这些势函数的其他版本留待第 5 节介绍）。对于 $1 \leqslant j < j_{\max} = \Theta(\log\log n)$，超指数势函数定义如下：

$$\Phi_j^t := \sum_{i=1}^{n} \Phi_{j,i}^t := \sum_{i=1}^{n} e^{\alpha_2 \cdot v^j \cdot (y_i^t - z_j)^+},$$

其中， $\alpha_2 > 0$ 是一个常数， $x^+ = \max\{x, 0\}$, $z_j := \frac{5v}{\alpha_2} \cdot j$，且 v 是一个足够大的常数。我们的目标是证明 $\Phi_{j_{\max}-1}^t = \mathcal{O}(n)$，这将推出$\max_{i\in[n]} y_i^t \leqslant z_{j_{\max}-1} + \frac{5v^2}{\alpha_2^2} = \mathcal{O}(\log\log n)$。为了证明集中性，我们还使用该势函数的第二个版本，记为 Ψ_j^t，其定义方式与 Φ_j^t相同，但采用了更大的平滑参数 $\alpha_1 > \alpha_2$。这种相互配合的思路与基础情形中的 Γ_1 和 Γ_2 类似。

在进一步概述 $\mathcal{O}(\log\log n)$ 界的证明之前，我们先定义折叠过程（图 1 展示了示意图）。从宏观上看，折叠过程将步骤划分为若干运行段，并以轮次将它们隔开。粗略地说，只要我们最近抽到的是一个轻箱，每个运行段就会继续。只有在这种情况不再成立时，才会开始新一轮。出于技术原因，只要球不会进入负载更高的箱，我们也允许在分配球时有一定的灵活性。

折叠过程。 在分层归纳的第 j 层（对于 $1 \leqslant j \leqslant j_{\max} - 1$），我们将分析以下折叠过程；Memory 是该过程的一个实例，我们将在引理 5.1 中很快验证这一点。为此，我们将步骤分组成连续的轮次（长度各不相同），并将一轮中的第 s 个步骤称为子步骤 s。此外，令 $y^{r,\,i\,\,s}$ 表示第 r 轮第 s 个子步骤之后箱 i 的归一化负载。然后，我们将折叠过程 定义如下：

- 对于每轮 $r \geqslant 0$，按照抽样分布 $\mathcal{S}$ 抽取箱 $i := i(r) \in [n]$：
  - **情况 A：** 如果 $y^{r,\,i\,\,0} \geqslant z_{j-1} + \frac{2v}{\alpha_2}$，则将一个球分配给任意一个箱 ℓ，满足$y^{r,\,\ell\,\,0} \leqslant y^{r,\,i\,\,0}$，然后进入下一轮次。

– **情况 B：** 否则，在当前轮次 r 中开始一系列连续阶段，每个阶段包含 $\frac{v}{\alpha_2}$ 个子步骤（也就是说，每个阶段 $k \geqslant 1$ 由子步骤 $s \in [(k-1)\cdot\frac{v}{\alpha_2},\ k\cdot\frac{v}{\alpha_2}]$ 组成。）。在每个子步骤 s 中，我们根据 $\mathcal{S}$ 抽取一个箱 $i = i(r,\ s)$，并将一个球分配给任意一个箱 ℓ，满足 $y^{r,}_{\ell}{}^{s} \leqslant z_{j-1} + \frac{4v}{\alpha_2}$。在每个阶段结束时，如果以下两个条件之一成立，我们也会结束该轮次：

* **条件 1：** 在当前阶段的任何子步骤中，我们都没有在对应的子步骤 s采样到负载为 $y^{r,}_{\ell}{}^{s} < z_{j-1} + \frac{2v}{\alpha_2}$ 的箱子 ℓ。
* **条件 2：** 我们已完成 $k_j := e^{v^{j+1}} \cdot \log^3 n \leqslant n^{1/7}$ 个阶段。

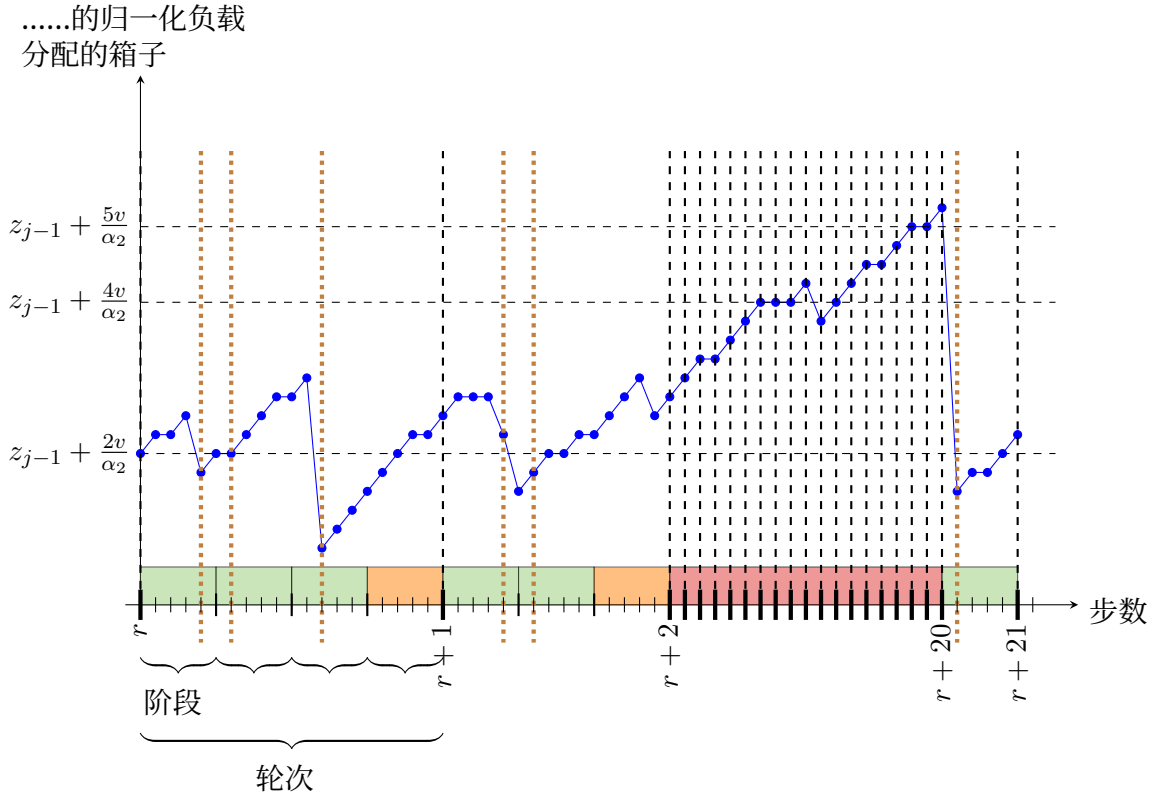

图1： 折叠过程中各阶段和轮次的示意图。棕色线条表示某一阶段内首次采样到轻箱的子步骤（如轮次 $r+1$ 的第二阶段所示，这并不一定意味着该箱将用于分配或缓存）。如图所示，只有在经历一长串红色轮次之后，才可能向归一化负载高于 $z_{j-1} + \frac{5v}{\alpha_2}$ 的箱进行分配。

—

有了轮次的定义之后，我们接下来分析势函数 Φ_j和 Ψ_j。我们的首要任务是为这些势函数推导下降不等式。与以往分析相比，一个挑战是我们需要考虑越来越长的时间区间（随着j 增加），这些区间就是通过折叠过程正式定义的轮次。

恢复。 通过不同势函数之间的相互作用——其中一些针对轮次定义，另一些针对步骤定义——我们得到了所需的下降不等式，并由此得出：在条件化于 Φ_{j-1} 势函数较小的情况下，势函数 Φ_j 也必定会在不太遥远的未来某个步骤中较小。

稳定化。 为了证明 Φ_j 保持较小，我们利用选择不同平滑参数所带来的灵活性，并证明对另一个势函数的弱界

Ψ_j 意味着 Φ_j （其平滑参数较小）最多只会发生较小的次线性变化。这使我们能够按照有界独立差分法的思路，应用带有坏事件的鞅集中不等式（定理 A.10）。

分层归纳。 在第 3.1 节确立基础情形后，我们便可以通过执行归纳步骤（$j \to j+1$, $j < j_{\max} = \mathcal{O}(\log\log\ n)$）将所有部分整合起来。对于每次迭代 j，我们考虑越来越陡峭的势函数 Φ_j。此外，在归纳的每一层，我们都将时间区间稍稍向前移动，并以一个加法常数温和地增加势函数的偏移量。由此得到以下关键引理，正式表述如下：

定理 5.13 (重述)。 考虑具有任意(a,b)-偏置采样分布的MEMORY 过程，其中 $a,b \geqslant 1$ 为常数。那么，对于任意步骤 $t \geqslant 1$ 和 $1 \leqslant j \leqslant j_{\max}-1$，定义$\beta_j := t + 2jn\log^6 n$，并令 $C \geqslant 6$ 如(5.1) 中所定义。那么，假设以下条件成立：

$$\mathbf{Pr}\left[\bigcap_{s\in[\beta_{j-1},t+n\log^8 n]} \{\Phi_{j-1}^s \leqslant 2Cn\}\right] \geqslant 1 - \frac{(\log n)^{11(j-1)}}{n^4},$$

那么以下结论也成立，

$$\mathbf{Pr}\left[\bigcap_{s\in[\beta_j,t+n\log^8 n]} \{\Phi_j^s \leqslant 2Cn\}\right] \geqslant 1 - \frac{(\log n)^{11j}}{n^4}.$$

一旦证明了这个引理，从基础情形（定理 4.3）出发，并对迭代 $1 \leqslant j \leqslant j_{\max}$ 使用并集界，即可得到所需的差距，从而得出 $\Phi_{j_{\max}}^m = \mathcal{O}(n)$；这立即给出差距界 $\mathcal{O}(\log\log\ n)$，且以高概率成立。至此，证明纲要结束；完整且自洽的证明见第 5 节。

4 基础情形的证明

为便于阅读并使证明自成一体，我们重复介绍了第 3.1 节中的若干定义和说明。

在 [21], 中，作者证明了对于采用均匀采样分布的 Memory 过程，在任意步骤 $m \geqslant 1$, 均有界 Gap(m) $= \mathcal{O}(\log n)$，且该结论以高概率成立。这里，我们将采用不同的分析方法显著加强这一结果，证明对于任意步骤 $m \geqslant 1$，在采用 (a, b)-偏置采样分布的 Memory 过程中，双曲余弦势函数至少在 $n\log^8 n$ 步内为 $\mathcal{O}(n)$，且该结论以高概率成立。这将构成第 5 节中更精细分析的基础情形，同时也意味着更强的保证：欠载以高概率满足 $-\min_{i\in[n]} y_i^t = \mathcal{O}(\log\ n)$。

分析将使用 d-弱 Memory（$\mathcal{S}$）进行，该过程每 d 步重置一次缓存。因此，在步骤 t，球通过单选进行分配，并将该箱加入缓存。随后在步骤 $t+1$、.....、$t+d$，球从箱中采样，并分配给样本和缓存中负载最小的箱。该过程还会基于过时信息进行比较，我们可以证明，从某种将形式化的意义上说，该过程的分配效果劣于 Memory（引理 4.5）。通过共同分析这 d 步，我们仍然能够证明势函数在 d 步内下降，并推导出其以高概率为 $\mathcal{O}$（n）。对于 $d=2$, 该过程可以解释为（$1+\beta$）的一种变体，其中 $\beta = 1/2$, 箱的选择彼此相关。也就是说，第 $2t$ 步的样本会在第 $2t+1$ 步被重复使用。因此，实际上每次分配只需要一个样本；尽管如此，该过程仍能使差距以高概率为 $\mathcal{O}$（$\log\ n$）。

4.1 预备知识

我们将带平滑参数 $\alpha > 0$ 的双曲余弦势函数 Γ 定义为

$$\Gamma^t := \Gamma^t(\alpha) := \Phi^t(\alpha) + \Psi^t(\alpha),$$

其中 Φ^t 是超载指数势函数

$$\Phi^t := \Phi^t(\alpha) := \sum_{i=1}^{n} \Phi_i^t := \sum_{i=1}^{n} e^{\alpha y_i^t},$$

且 Ψ^t 是欠载指数势函数

$$\Psi^t := \Psi^t(\alpha) := \sum_{i=1}^{n} \Psi_i^t := \sum_{i=1}^{n} e^{-\alpha y_i^t}.$$

根据 [32],，回顾一下：若对所有$i \in [n]$都有 $\frac{1}{an} \leqslant s_i \leqslant \frac{b}{n}$，则称抽样分布（$s_1, \ldots, s_n$）是（$a, b$）-偏置的，其中 s_i是箱 i 被抽样的概率，且 $a, b > 1$。注意，由于（$s_1, \ldots, s_n$）是一个概率分布，我们总可以假设 $b \leqslant n$。对于满足 M的实数 $a, b \geqslant 1$，若:= ^{n}（$a-1$）$^{ab-1}$是一个整数，则我们将（a, b）阶梯分布定义为向量 $\left(\frac{b}{n}, \ldots, \frac{b}{n}, \frac{1}{an}, \ldots, \frac{1}{an}\right)$，其中前 ^{n}（$a-1$）$^{ab-1}$个指标的值为 $\frac{b}{n}$，其余指标的值为 $\frac{1}{an}$。

在第4.2节中，我们将证明，对于任意(a, b)-偏置抽样分布，Memory 过程中的双曲余弦势函数在期望意义下为 $\mathcal{O}(n)$（其中 $a, b \geqslant 1$ 为任意常数）：

定理 *4.1*。考虑具有任意(a, b)-偏置抽样分布的 *Memory* 过程，其中常数为 $a, b \geqslant 1$。那么，存在常数 $\alpha' := \alpha'(a, b) > 0$, $c := c(a, b) \geqslant 1$ 以及 $d := d(a, b) \geqslant 2$，使得对于势函数 $\Gamma := \Gamma(\alpha)$，对任意 $\alpha \leqslant \alpha'$ 以及任意步骤 $t \geqslant 1$,

$$\mathbf{E}\left[\, \Gamma^{t+d} \;\middle|\; \mathfrak{F}^t \,\right] \leqslant \Gamma^t \cdot \left(1 - \frac{\alpha}{c \cdot n}\right) + c \cdot \alpha.$$

同一界对 d-Weak-Memory 过程也成立。

在第4.3节中，我们将证明以下关于双曲势能的一般集中不等式；该不等式的一种形式已在 [19, 第 4]节中证明。

定理 4.2（双曲余弦势函数集中性）。考虑任意一个过程 $\mathcal{P}$，其中每个步骤至多分配 $d \in \mathbb{N}_+$ 个球，并考虑任意常数 $\kappa \geqslant 6$。此外，假设对于此过程，双曲势函数 $\Gamma_1 := \Gamma_1(\alpha_1)$ 和 $\Gamma_2 := \Gamma_2(\alpha_2)$ 分别采用平滑参数 $0 < \alpha_1 < 1/(2d)$ 和 $\alpha_2 \leqslant \frac{\alpha_1}{12\kappa}$，存在一个 $\varepsilon > 0$（其中 $\alpha_2 \varepsilon \geqslant n^{-1/6}$）以及常数 c_1、$c_2 > 0$（其中 $c_1 \leqslant c_2$），使得对于任意步骤 $t \geqslant 1$,

$$\mathbf{E}\left[\, \Gamma_1^{t+1} \;\middle|\; \mathfrak{F}^t \,\right] \leqslant \Gamma_1^t \cdot \left(1 - \frac{c_1 \alpha_1 \varepsilon}{n}\right) + c_2 \alpha_1 \varepsilon,$$

and

$$\mathbf{E}\left[\, \Gamma_2^{t+1} \;\middle|\; \mathfrak{F}^t \,\right] \leqslant \Gamma_2^t \cdot \left(1 - \frac{c_1 \alpha_2 \varepsilon}{n}\right) + c_2 \alpha_2 \varepsilon.$$

那么，对于 $c := 2 - \frac{c_2}{c_1} \geqslant 2$，对于任意步骤 $t \geqslant 1$,

$$\mathbf{Pr}\left[\, \Gamma_2^t \leqslant 3cn \,\right] \geqslant 1 - n^{-\kappa}.$$

接下来，利用定理 4.2，我们将推导出，对于足够小的平滑参数$\alpha > 0$，以高概率，双曲余弦势函数 Γ 是 $\mathcal{O}(n)$:

定理 4.3。 考虑具有任意(a,b)偏置采样分布的MEMORY 过程，其中常数为 $a,b \geqslant 1$。那么，存在常数 $c := c(a,b) \geqslant 1$和 $0 < \alpha := \alpha(a,b) < 1$，使得对于势函数 $\Gamma := \Gamma(\alpha)$以及任意步骤 $t \geqslant 1$，我们有

$$\mathbf{Pr}\left[\bigcap_{u\in[t,t+n\log^8 n]} \{\Gamma^u \leqslant 6cn\}\right] \geqslant 1 - n^{-4}.$$

4.2 双曲余弦势函数上的期望界

本节旨在证明定理 4.1，该定理给出了双曲余弦势函数 Γ 在其至少为某个适当大的常数乘以 n 时的期望下降。为此，我们将利用 [18, 定理 3.1], ，这是对 [27] 中主定理的推广。这将使我们能够证明 Γ 的下降不等式，并在 $\alpha > 0$ 是充分小的常数时界定其期望。在陈述该定理之前，我们先回顾概率向量满足的以下条件：

- **条件** $\mathcal{C}_1$：存在一个常数 $\delta \in (0,1)$ 和一个（不一定是常数的） $\varepsilon \in (0,1)$，使得对于任意 $1 \leqslant k \leqslant \delta \cdot n$，

$$\sum_{i=1}^{k} p_i \leqslant (1-\varepsilon)\cdot\frac{k}{n},$$

对于任意 $\delta \cdot n + 1 \leqslant k \leqslant n$，也类似地有，

$$\sum_{i=k}^{n} p_i \geqslant \left(1+\varepsilon\cdot\frac{\delta}{1-\delta}\right)\cdot\frac{n-k+1}{n}.$$

此类向量的一个例子是双选概率分配向量 p，其中$p_i = \frac{2i-1}{n^2}$，它满足条件 $\mathcal{C}_1$，其中 $\delta = \frac{1}{4}$和 $\varepsilon = \frac{1}{2}$（例如， [18, 命题 2.3]）。

定理 4.4（定理 3.1 见 [18]）。 考虑任意满足条件 $\mathcal{C}_1$的概率向量 p，其中常数 $\delta \in (0,1)$ 且 $\varepsilon > 0$，以及满足以下条件的任意负载向量 x： $\Phi := \Phi(\alpha, x)$， $\Psi := \Psi(\alpha, x)$ 且$\Gamma := \Gamma(\alpha, x)$。进一步，对某些 K，定义 $\kappa > 0$，

$$\Delta\overline{\Phi} := \sum_{i=1}^{n} \Phi_i \cdot \left(\left(p_i - \frac{1}{n}\right)\cdot\kappa\cdot\alpha + K\cdot\kappa\cdot\frac{\alpha^2}{n}\right), \tag{4.1}$$

and

$$\Delta\overline{\Psi} := \sum_{i=1}^{n} \Psi_i \cdot \left(\left(\frac{1}{n} - p_i\right)\cdot\kappa\cdot\alpha + K\cdot\kappa\cdot\frac{\alpha^2}{n}\right). \tag{4.2}$$

那么，存在一个常数 $c := c(\delta) > 0$，使得对于任意 $0 < \alpha <$ 最小值$\{1, \frac{\varepsilon\delta}{8K}\}$，—

$$\Delta\overline{\Gamma} := \Delta\overline{\Phi} + \Delta\overline{\Psi} \leqslant -\frac{\varepsilon\delta}{8}\cdot\kappa\cdot\frac{\alpha}{n}\cdot\Gamma + c\cdot\kappa\cdot\varepsilon\cdot\alpha.$$

证明定理 4.1 的主要难点在于 Memory 过程中存在长程相关性，此外，非均匀采样分布也让情况更加棘手。我们转而分析较弱的 d-Weak-Memory 过程，它更适合应用定理 4.4；然而，非均匀采样分布仍然是个难题。第一步（引理 4.5）是证明，从势函数的角度来看，在 (a,b)-偏置采样向量下， d-Weak-Memory 过程确实比 Memory 过程弱（误差至多为一个较小的加性项），因此，适用于 d-Weak- Memory 的上界也适用于 Memory。随后我们证明，在所有 (a,b)-偏置分布中，

就 d-Weak-Memory 过程的势函数变化而言，(a,b)-阶梯分布是最坏情况（引理 4.6）。由于 (a,b)-阶梯分布形式简单，我们可以计算在运行 d 次分配的过程中，向按固定顺序排列的第 i 个负载最高的箱分配 j 个球的概率（引理 4.8）。利用这些分配概率，我们便能界定 d-Weak-Memory 运行过程中的势下降，并将其与 Memory 的势函数联系起来，从而证明定理 4.1。

记号 在下文中，令 $\mathcal{S}$ 为一个抽样分布，$d \geqslant 1$、$t_0 \geqslant 0$ 为满足 $d \mid t_0$ 的整数。回顾一下，在一次运行开始时，d-Weak-Memory 过程会按负载为箱子固定一个排序 σ，随后在整个运行期间的所有分配决策中都使用这一静态排序 σ。如果运行始于某个时间 t_0，我们可以假设排序 σ 是 $\mathfrak{F}^{t_0}$ 可测的。p 令 $\widehat{p}_i$、$_j := \widehat{p}_i$、$_j(\mathcal{S}, t_0, d, \sigma)$ 表示在始于 t_0 的一次 d-Weak-Memory($\mathcal{S}$) 运行中，将 j 个球分配到箱子 $\sigma^{-1}(i)$ 的概率，其中 $\sigma^{-1}(i)$ 是相对于 σ（即在时间 t_0 固定的排序）而言第 i 重的箱子。类似地，令 p_i、$_j := p_i$、$_j(\mathcal{S}, t_0, d, \sigma)$ 表示 Memory($\mathcal{S}$) 在步骤 t_0，……，t_0+d-1 期间将 j 个球分配到箱子 σ^{-1}（i）的概率。令 $\widehat{\eta}_i^t := \widehat{\eta}_i^t(\mathcal{S}, t_0, d, \sigma)$ 表示在 d-Weak-Memory($\mathcal{S}$) 的前 $t \leqslant d$ 步中添加到箱子 $\sigma^{-1}(i)$ 的球数；令 $\eta_i^t := \eta_i^t(\mathcal{S}, t_0, d, \sigma)$ 表示 Memory($\mathcal{S}$) 中的相应数量。若 $t = d$，则省略上标，记为 $\widehat{\eta}_i = \widehat{\eta}_i^d$ 和 $\eta_i = \eta_i^d$。

注意，$\widehat{p}_{i,\ j}$和 $\widehat{\eta}_i^t$只取决于排序 σ，而 $p_{i,\ j}$和 η_i^t还取决于负载配置 x^t。鉴于 Memory 过程并不遵循 σ，量 $p_{i,\ j}$ 和 η_i^t或许看起来不太自然；然而，在联系两个过程的势函数时，考察这些量会很有用。上下文清楚时，我们通常省略对 $\mathcal{S}$，t_0，d 和 σ的符号依赖关系，尤其是在对 $\mathfrak{F}^{t_0}$进行条件化时。

注意，$\mathbf{E}\left[\widehat{\eta}_i \mid \mathfrak{F}^{t_0}\right] = \sum_{j=0}^{d} \widehat{p}_{i,\ j} \cdot j$ 表示在时刻 t_0开始运行 d-Weak-Memory($\mathcal{S}$) 期间，关于 σ 而言第 i 重的箱所接收的球的期望数量。由于一次运行中会分配 d 个球，因此由 $\widehat{p}_i = \frac{1}{d}\sum_{j=0}^{d} \widehat{p}_{i,\ j} \cdot j$给出的向量 $(\widehat{p}_i)_{i\in[n]}$是一个概率分布。将其称为$d$-Weak-Memory 的分配向量似乎很自然，但严格来说并不准确，因为那样一来就允许分配分数球。因此，我们称之为 d-Weak-Memory 的代理分配向量 。

4.2.1 过程间的耦合

由于缓存引入的分配之间存在长程依赖，Memory 过程相当难以处理。d-Weak-Memory 过程也会每隔 d步“忘记”缓存，并使用负载向量 x^t为步骤 $t+1$、……、$t+d$做出决策。对于正整数 d，我们可以通过对一次包含 d 次分配的运行进行如下耦合，将较弱的 d-Weak-Memory 过程的结果与 Memory 过程联系起来。

引理 *4.5*。考虑任意固定的 $d \geqslant 1$，以及任意$0 < \alpha < 1/d$对应的势函数 $\Phi := \Phi(\alpha)$ 和 $\Psi := \Psi(\alpha)$。固定步骤 t 时的负载向量 x^t，并按负载递减顺序排列箱子。令 μ为任意 (a,b)-偏置采样向量，并固定 $p_{i,j} := p_{i,j}(\mu, d)$ 和 $\widehat{p}_{i,j} := \widehat{p}_{i,j}(\mu, d)$。那么，

$$\alpha \cdot \sum_{i=1}^{n} \Phi_i^t \cdot \sum_{j=0}^{d} p_{i,j} \cdot j \leqslant \alpha \cdot \sum_{i=1}^{n} \Phi_i^t \cdot \sum_{j=0}^{d} \widehat{p}_{i,j} \cdot j + \Phi^t \cdot \frac{\alpha^2}{n} \cdot (2d^3 b),$$

and

$$-\alpha \cdot \sum_{i=1}^{n} \Psi_i^t \cdot \sum_{j=0}^{d} p_{i,j} \cdot j \leqslant -\alpha \cdot \sum_{i=1}^{n} \Psi_i^t \cdot \sum_{j=0}^{d} \widehat{p}_{i,j} \cdot j + \Psi^t \cdot \frac{\alpha^2}{n} \cdot (2d^3 b).$$

证明。 对于第一个结论，我们的目标是给出上界

$$\alpha \cdot \sum_{i=1}^{n} \Phi_i^t \cdot \sum_{j=0}^{d} p_{i,j} \cdot j = \alpha \cdot \mathbf{E}\left[\sum_{i=1}^{n} \Phi_i^t \cdot \eta_i \,\middle|\, \mathfrak{F}^t\right],$$

其中 $\eta_i := \eta_i(\mu, t, d, \sigma)$，且对每个 $i \in [n]$ 以及任意按负载对箱子进行的排序 σ，都有 $\widehat{\eta}_i := \widehat{\eta}_i(\mu, t, d, \sigma)$。我们通过在步骤 $t+1, \ldots, t+d$ 中对两个过程 Memory 和 d-Weak-Memory 采样相同的箱子，构造二者之间的耦合，并从归一化负载向量y^t开始。令 c 为步骤 t 时 Memory 过程的缓存，并令 $\mathcal{D}$ 表示在步骤$t+1, \ldots, t+d$中，没有任何箱子被重复采样，且箱子 c 也未被采样这一事件。

情况 1 [$\mathcal{D}$ 成立]: 在这种情况下，所有采样箱（以及初始缓存 c）都互不相同。因此，Memory 始终存储的箱的负载低于 d-Weak-Memory 的缓存，所以它分配到的箱的负载也低于 d-Weak-Memory 分配到的箱。因此，

$$\alpha \cdot \mathbf{E}\left[\sum_{i=1}^{n} \eta_i \cdot \Phi_i^t - \sum_{i=1}^{n} \widehat{\eta}_i \cdot \Phi_i^t \,\middle|\, \mathfrak{F}^t,\, \mathcal{D} \right] \leqslant 0. \tag{4.3}$$

情况 2 [$\mathcal{D}$ 不成立]: 另一方面，如果存在一个步骤 $s \in [t+1, t+d]$，使得过程确实会再次采样某个箱，或采样箱 c，那么对于此后的任意步骤$r \in [s, t+d]$，Memory 都可能分配到一个箱 i，而 d-Weak-Memory 分配到一个箱 i'，使得$y_i^t \leqslant y_{i'}^t + d$；这是因为在 d 次分配中，一个箱的负载最多变化 d，因此 Memory 的缓存最多比 d-Weak-Memory 的缓存多 d 个球。

为了给事件 $\neg\mathcal{D}$ 发生的概率求上界，我们定义 $\neg\mathcal{D}^r$ 为 $1 \leqslant r \leqslant d$，即在步骤 $t+r$中，我们第二次抽到某个箱子或缓存的事件

$$\begin{aligned}
\mathbf{Pr}\left[\, \neg\mathcal{D} \mid \mathfrak{F}^t \,\right] &\leqslant \mathbf{Pr}\left[\, \neg\mathcal{D}^1 \mid \mathfrak{F}^t \,\right] + \ldots + \mathbf{Pr}\left[\, \neg\mathcal{D}^d \mid \mathfrak{F}^t \,\right] \\
&\leqslant \frac{b}{n} + 2 \cdot \frac{b}{n} + 3 \cdot \frac{b}{n} + \ldots + d \cdot \frac{b}{n} \\
&\leqslant \frac{d^2 b}{n}.
\end{aligned} \tag{4.4}$$

由于 $y_i^t \leqslant y_{i'}^t + d$，项 Φ_i^t 可由项 $\Phi_{i'}^t$ 按如下方式给出上界

$$\Phi_i^t = e^{\alpha y_i^t} \leqslant e^{\alpha \cdot (y_{i'}^t + d)} \leqslant \Phi_{i'}^t \cdot e^{\alpha d} \leqslant \Phi_{i'}^t \cdot (1 + 2\alpha d) = \Phi_{i'}^t + \Phi_{i'}^t \cdot 2\alpha d.$$

利用 $\alpha \leqslant \frac{1}{d}$ 以及对于任意 $|z| \leqslant 1$都有 $e^z \leqslant 1 + 2z$。由于在步骤 $t+1, \ldots, t+d$中分配的不同箱子至多有 d 个$j_1, \ldots, j_d$，因此我们有

$$\alpha \cdot \mathbf{E}\left[\sum_{i=1}^{n} \eta_i \cdot \Phi_i^t - \sum_{i=1}^{n} \widehat{\eta}_i \cdot \Phi_i^t \,\middle|\, \mathfrak{F}^t,\, \neg\mathcal{D} \right] \leqslant \sum_{k=1}^{d} \Phi_{j_k}^t \cdot 2\alpha d \leqslant \Phi^t \cdot 2\alpha d. \tag{4.5}$$

因此，结合 (4.3) 和 (4.5)，我们可以给出期望的上界

$$\begin{aligned}
\alpha \cdot \mathbf{E}\left[\sum_{i=1}^{n} \eta_i \cdot \Phi_i^t - \sum_{i=1}^{n} \widehat{\eta}_i \cdot \Phi_i^t \,\middle|\, \mathfrak{F}^t \right] &= \alpha \cdot \mathbf{E}\left[\sum_{i=1}^{n} \eta_i \cdot \Phi_i^t - \sum_{i=1}^{n} \widehat{\eta}_i \cdot \Phi_i^t \,\middle|\, \mathfrak{F}^t, \mathcal{D} \right] \cdot \mathbf{Pr}\left[\, \mathcal{D} \mid \mathfrak{F}^t \,\right] \\
&\quad + \alpha \cdot \mathbf{E}\left[\sum_{i=1}^{n} \eta_i \cdot \Phi_i^t - \sum_{i=1}^{n} \widehat{\eta}_i \cdot \Phi_i^t \,\middle|\, \mathfrak{F}^t,\, \neg\mathcal{D} \right] \cdot \mathbf{Pr}\left[\, \neg\mathcal{D} \mid \mathfrak{F}^t \,\right] \\
&\leqslant 0 + \Phi^t \cdot 2\alpha^2 d \cdot \frac{d^2 b}{n} = \Phi^t \cdot \frac{2\alpha^2 d^3 b}{n}.
\end{aligned}$$

现在我们对 Ψ 采用类似的方法，用 $-\Psi_{i'}^t$ 给项 $-\Psi_i^t$ 求上界

$$-\Psi_i^t = -e^{-\alpha y_i^t} \leqslant -e^{-\alpha \cdot (y_{i'}^t + d)} \leqslant -\Psi_{i'}^t \cdot e^{-\alpha d} \leqslant -\Psi_{i'}^t + \Psi_{i'}^t \cdot 2\alpha d,$$

利用 $\alpha \leqslant \frac{1}{d}$，并且对于任意 $|z| \leqslant 1$，$e^z \leqslant 1 + 2z$。因此，

$$-\alpha \cdot \mathbf{E}\left[\sum_{i=1}^{n} \eta_i \cdot \Psi_i^t - \sum_{i=1}^{n} \widehat{\eta}_i \cdot \Psi_i^t \,\middle|\, \mathfrak{F}^t \right] \leqslant \Psi^t \cdot \frac{2\alpha^2 d^3 b}{n}. \qquad \square$$

4.2.2 d-弱 Memory 中分配概率的性质

我们对 Memory 过程（定理 4.1）的期望势下降所给出的界，源于对 d-Weak-Memory 过程的期望势下降所给出的界。证明这一点需要应用定理 4.4，该定理要求证明过程的概率分配向量（或其代理）会产生一种偏向，使其避开高负载箱子。

本小节汇集了若干关于 d-弱 Memory 代理分配向量的结果，这些结果将使我们能够应用定理 4.4，证明一次运行中期望势函数会下降。前两个结果使我们能够在 $(a,\ b)$-阶梯分布下分析 d-弱 Memory 的代理分配向量，而不必考虑任意的 $(a,\ b)$-偏置分布。

回忆一下，若对于所有 $1 \leqslant k \leqslant n$，前缀和都满足：$\sum_{i=1}^{k} v_i \geqslant \sum_{i=1}^{k} u_i$，我们就称向量 $v = (v_1,\ v_2,\ \ldots,\ v_n)$ 主序于 $u = (u_1,\ u_2,\ \ldots,\ u_n)$。

引理 4.6。设 $d \geqslant 1$, $t_0 \geqslant 0$ 和 $a, b > 1$ 满足 $M = n \cdot \frac{a-1}{ab-1}$ 是整数。设 ν 为 ($a\ b$ μ $a\ b$ $\mu \preceq$,)-阶梯分布，且 $\sum_{k=1}^{i} \widehat{\eta}_k\ \nu$ $i \in [n]$ $\left(\ \left[\widehat{\eta}_j\ \nu\ \middle|\ \mathfrak{F}^{t_0}\right]\right)_{j\in[n]}$ $\left(\ \left[\widehat{\eta}_j\ \mu\ \middle|\ \mathfrak{F}^{t_0}\right]\right)_{j\in[n]}$ 为任意 (,)-偏置分布。那么，对于任意，$\sum_{k=1}^{i} \widehat{\eta}_k() \,() \,$ 成立。因此，$\mathbf{E}\,() \,$ 主序于 $\mathbf{E}() \,$。

证明。我们将 $\mathcal{N}$（即在 (a,b) 阶梯分布 $\nu = (\nu_1, \ldots, \nu_n)$ 上运行的 d-Weak-Memory 实例）与 $\mathcal{M}$（即在 (a,b) 偏置分布 $\mu = (\mu_1, \cdots, \mu_n)$ 上运行的 d-Weak-Memory 实例）进行耦合。两个过程都从时间 t_0 的相同初始负载配置 x^{t_0} 和排序 σ 开始。

我们给出采样分布 $s_\nu \sim \nu$ 和 $s_\mu \sim \mu$ 的一个耦合 (s_ν, s_μ)，如下所示 s:

对于每个 $k \in [n]$，令 $q_k = \min\{\nu_k, \mu_k\}$ 和 $q = \sum_{k=1}^{n} q_k$。令 $X \sim \mathrm{Ber}(q)$。

- 若 $X =$ 为正面，则以概率 q_k/q 抽取一个箱 $k \in [n]$，并令 (s_ν, s_μ) 为 $= (k, k\ )$.
- 若 $X =$ 为反面，则从 $[0,1]$ 中均匀地抽取 U。令 $s_\nu \in [n]$ 和 $s_\mu \in [n]$ 分别为满足 $\sum_{k=1}^{s_\nu} (\nu_k - q_k)/(1-q) \geqslant U$ 和 $\sum_{k=1}^{s_\mu} (\mu_k - q_k)/(1-q) \geqslant U$ 的最小整数。

要证明这是 ν 和 μ 的一个有效耦合，只需注意，对于任意 $j \in [n]$,

$$\mathbf{Pr}\left[\, s_\nu = j \,\right] = \mathbf{Pr}\left[\, X = \mathsf{heads} \,\right] \cdot \frac{q_j}{q} + \mathbf{Pr}\left[\, X = \mathsf{tails} \,\right] \cdot \frac{\nu_j - q_j}{1-q} = \nu_j,$$

类似地，$\mathbf{Pr}\left[\, s_\mu = j \,\right] = \mu_j$。接着，我们逐个球地耦合 $\mathcal{N}$ 和 $\mathcal{M}$：在每个步骤中，按照该耦合对 (s_ν, s_μ) 进行采样，并将箱样本 s_ν 分配给 $\mathcal{N}$，将 s_μ 分配给 $\mathcal{M}$。我们声称，在上述耦合下，运行的每个步骤 $t \leqslant d$ 都满足以下不变式：

($\mathcal{J}$)：$\mathcal{N}$ 采样到的箱在排序中高于 $\mathcal{M}$ 采样到的箱，即 $s_\nu \leqslant s_\mu$。

显然，在 $X =$ 为正面的每个步骤，不变式 $\mathcal{J}$ 都成立，因为此时 $s_\nu = s_\mu$，也就是说，在该耦合下，两个过程获得相同的箱样本。

否则，如果 $X =$ 为反面，那么对于任意 $j \leqslant M$,

$$\sum_{k=1}^{j} \frac{\nu_k - q_k}{1-q} = \sum_{k=1}^{j} \frac{b/n - q_k}{1-q} \geqslant \sum_{k=1}^{j} \frac{\mu_k - q_k}{1-q}. \tag{4.6}$$

同样，对于任意 $j > M$，由于此时 $q_j = \min\{\nu_j, \mu_j\} = 1/(an)$，我们有

$$\sum_{k=j}^{n} \frac{\nu_k - q_k}{1-q} = \sum_{k=j}^{n} \frac{1/(an) - q_k}{1-q} = 0 \leqslant \sum_{k=j}^{n} \frac{\mu_k - q_k}{1-q},$$

因此，由于 $\sum_{k=j}^{n} \frac{\mu_k - q_k}{1-q} = 1$ 和 $\sum_{k=j}^{n} \frac{\mu_k - q_k}{1-q} = 1$，对于任意 $j > M$，我们得到

$$\sum_{k=1}^{j} \frac{\nu_k - q_k}{1-q} = 1 - \sum_{k=j+1}^{n} \frac{\nu_k - q_k}{1-q} \geqslant 1 - \sum_{k=j+1}^{n} \frac{\mu_k - q_k}{1-q} = \sum_{k=1}^{j} \frac{\mu_k - q_k}{1-q}, \tag{4.7}$$

whe这里我们将空和视为零。因此，由 (4.6) 和 (4.7)，若 $X =$ 为反面，则 $s_\nu \leqslant s$ μ.

注意，对于任意 $i \in [n]$，如果任一过程在某个时刻获得一个箱样本 $j \geqslant i$，那么此后就不会再有球被分配到箱子 $k < i$，因为缓存中会有一个在初始排序中更轻的箱。因此，对于任意 $i \in [n]$，$\sum_{k=1}^{i} \widehat{\eta}_k(\mu) \preceq \sum_{k=1}^{i} \widehat{\eta}_k(\nu)$ 均成立，因为 $\mathcal{J}$ 在运行的每一步都成立。于是，陈述中的第二个结论可由期望的线性性和主序的定义推出。

□

现在我们证明，在阶梯分布下，d-Weak-Memory 的代理分配向量是分段非递减的。

引理 4.7。设 $d \geqslant 1$, $t_0 \geqslant 0$ 和 $a, b > 1$ 满足 $M = n \cdot \frac{a-1}{ab-1}$ 是整数。设 μ 为 (a, b)-阶梯分布，且 $\widehat{\eta}_i^t := \widehat{\eta}_i^t(\nu)$，其中 $t \in [d]$。那么，对于任意满足 $i < i' \leqslant M$ 或 $M < i < i'$ 的 $i, i' \in [n]$，对于任意 $t \in [d]$，均有 $\widehat{\eta}_i^t \preceq \widehat{\eta}_{i'}^t$。

证明。我们将如下方式耦合 $\mathcal{P}$ 和 $\mathcal{P}'$ 这两个具有相同（a，b）步抽样分布、初始负载配置和排序 σ 的 d-弱记忆实例，使其各运行一次：

> 给定一个箱子序列 $(b_j)_{j \in [d]}$，令 $(b'_j)_{j \in [d]}$ 满足：若 $b_j = i$，则 $b'_j = i'$；若 $b_j = i'$，则 $b'_j = i$；否则 $b_j = b'_j$。那么，若 $\mathcal{P}$ 给定为 $(b_j)_{j \in [d]}$，我们通过将序列 $(b'_j)_{j \in [d]}$ 作为输入提供给 $\mathcal{P}'$ 来进行耦合。

注意，由于 $i < i' \leqslant M$ 或 $M < i < i'$，箱 i 和 i' 被抽样的概率相同。因此，将 $(b_j)_{j \in [d]}$ 映射为 $(b'_j)_{j \in [d]}$ 的函数，是从 d-弱记忆中单次运行的输入集合到其自身的保测双射，从而将 $\mathcal{P}$ 耦合到 $\mathcal{P}'$。

回忆一下，d-Weak-Memory 在一次运行中不会更新负载，因此排序 σ是固定的。注意，如果缓存中包含第 ℓ 个箱，那么在同一次运行的后续任何步骤中，都不会有球被分配到任何箱 $b < \ell$。考虑满足 $b_\tau \in \{i, i'\}$的最早时间 τ。如果对任意 $t < \tau$ 都有 $b_t > i'$，那么整个运行期间都不会有球被分配到 i 或 i'，因此陈述中的结论成立。如果 $i < \min_{t<\tau} b_t < i'$，那么在时间 τ，一个球会在 $\mathcal{P}'$ 中被添加到箱 i'，但不会被添加到 $\mathcal{P}$ 中的箱 i（因为缓存中有一个编号更小的箱），所以 $\widehat{\eta}_i^\tau < \widehat{\eta}_{i'}^\tau$；由于之后不会再有球被添加到 i，陈述中的结论成立。否则，在时间 τ，一个球会分别被添加到 $\mathcal{P}$ 中的箱 i 和$\mathcal{P}'$中的箱 i'，因此 $\widehat{\eta}_i^\tau = \widehat{\eta}_{i'}^\tau$，而这两个箱各自占据其对应的缓存。此后，当且仅当选择了一个排序中严格更高的箱 $b \notin \{i, i'\}$ 时，这些箱才会被挤出缓存。由于 $i' > i$，可知每当一个球被分配到 i 时，它也会被分配到 i'，从而证明了结论。

□

本节的最终结果精确确定了阶梯分布下 d-弱 Memory 的分配概率。

引理 4.8. 设 $t_0 \geqslant 0, d \geqslant 1$ 和 $a, b > 1$ 满足 $M = n \cdot \frac{a-1}{ab-1}$ 是整数。设 μ 为(a, b)阶梯分布，σ 为时间 t_0 按负载对箱进行的任意排序，且 $\widehat{p}_{i,j} := \widehat{p}_{i,j}(\nu, t_0, d, \sigma)$。那么，对于任意 $1 \leqslant j \leqslant d$，有

$$\widehat{p}_{i,j} = \begin{cases} \frac{b}{n} \cdot \left(\frac{ib}{n}\right)^{j-1} \cdot \left[1 - \frac{b}{n-b(i-1)} \cdot \left(1 - \left(\frac{b(i-1)}{n}\right)^{d-j}\right)\right] & \text{if } i \leqslant M, \\ \frac{a}{n} \cdot \left(1 - \frac{a(n-i)}{n}\right)^{j-1} \cdot \left[1 - \frac{a}{a(n-i+1)} \cdot \left(1 - \left(1 - \frac{a(n-i+1)}{n}\right)^{d-j}\right)\right] & \text{if } M < i \leqslant n. \end{cases}$$

证明。回想一下，在 σ 下，箱子按负载排序，负载最大的箱位于位置 1，负载最小的箱位于位置 n。d-弱 Memory 过程只能比较时间 t 的负载；具体而言，它所做的所有比较都基于一个在整个运行过程中固定不变的排序 σ。因此，以下所有关于“负载最高”等的表述均以 σ 为准。

注意，如果对排序中负载最高的第 i' 个箱子（其中 $i' > i$）进行采样，那么在当前运行期间，我们永远无法再将球放入负载最高的第 i 个箱子中。还要注意，如果负载最高的第 i 个箱子位于缓存中，那么它将一直保留在缓存中，直到排序中更高的箱子被

采样。设在排序中分别选中等于、严格低于、低于或等于，以及严格高于 i 的元素的概率为 e_i、f_i、 g_i 和 $h_i = 1 - g_i$。那么，我们声称对于 $1 \leqslant j \leqslant d$，有

$$\widehat{p}_{i,j} = \sum_{\ell=0}^{d-j-1} f_i^{\ell} \cdot e_i \cdot g_i^{j-1} \cdot h_i + f_i^{d-j} \cdot e_i \cdot g_i^{j-1}. \tag{4.8}$$

为说明这一结论成立，我们先考虑第二项；这种情况是指只有最后的 j 个球被分配到负载最高的第 i 个箱子，在此之前没有球被分配到该箱子。因此，前 $d-j$ 个球必须分配到负载最高的 $i-1$ 个箱子，这种情况发生的概率为 f_i^{d-j}。接着，我们分配到 i，随后从排序中负载至少与 i 一样高的箱子中抽取 $j-1$ 次样本，这种情况发生的概率为 $e_i \cdot g_i^{j-1}$。再来看求和式；每个求和项对应这样一种情况：分配到第 i个箱子的 j 个球组成的区块，在这次运行中从第 ℓ 个位置开始。这些概率中的前三项与前面类似，而第 h_i 项是在第（$\ell + j + 1$）步中抽到排序中负载高于 i 的箱子的概率。

我们可以利用几何级数简化（4.8）：

$$\widehat{p}_{i,j} = e_i \cdot g_i^{j-1} \cdot \left[h_i \cdot \sum_{\ell=0}^{d-j-1} f_i^{\ell} + f_i^{d-j} \right] = e_i \cdot g_i^{j-1} \cdot \left[h_i \cdot \frac{1 - f_i^{d-j}}{1 - f_i} + f_i^{d-j} \right].$$

现在，注意到 $h_i = 1 - g_i = 1 - f_i - e_i$，我们有

$$\widehat{p}_{i,j} = e_i \cdot g_i^{j-1} \cdot \left[(1 - f_i - e_i) \cdot \frac{1 - f_i^{d-j}}{1 - f_i} + f_i^{d-j} \right] = e_i \cdot g_i^{j-1} \cdot \left[1 - \frac{e_i}{1 - f_i} \cdot (1 - f_i^{d-j}) \right]. \tag{4.9}$$

回顾 $M := {}^{n}\!/_{(a-1)\ ab-1}$ 根据 (a,b)-步骤向量的定义，并注意到

i	e_i	f_i	g_i
$i \leqslant M$	b/n	$b(i-1)/n$	bi/n
$M < i < n$	a/n	$1 - a(n-i+1)/n$	$1 - a(n-i)/n$

(4.10)

对于 $i \leqslant M$，由 (4.10) 和 (4.9) 可得

$$\widehat{p}_{i,j} = \frac{b}{n} \cdot \left(\frac{bi}{n} \right)^{j-1} \cdot \left[1 - \frac{b/n}{1 - b(i-1)/n} \cdot \left(1 - \left(\frac{b(i-1)}{n} \right)^{d-j} \right) \right].$$

因此，化简后即可得到所述结果。对于 $M < i \leqslant n$ 的情形，同理可得，

$$\widehat{p}_{i,j} = \frac{a}{n} \cdot \left(1 - \frac{a(n-i)}{n} \right)^{j-1} \cdot \left[1 - \frac{a}{a(n-i+1)} \cdot \left(1 - \left(1 - \frac{a(n-i+1)}{n} \right)^{d-j} \right) \right],$$

如前所述。 □

4.2.3 期望势能下降不等式的证明

现在我们可以证明本节的第一个定理了。

定理 *4.1*（重述）。考虑具有任意 (a,b)-偏置抽样分布的 *Memory* 过程，其中常数为 $a, b \geqslant 1$。那么，存在常数 $\alpha' := \alpha'(a,b) > 0$, $c := c(a,b) \geqslant 1$和 $d := d(a,b) \geqslant 2$, 使得对于势函数 $\Gamma := \Gamma(\alpha)$、任意 $\alpha \leqslant \alpha'$ 以及任意步骤 $t \geqslant 1$,

$$\mathbf{E}\left[\Gamma^{t+d} \;\middle|\; \mathfrak{F}^t \right] \leqslant \Gamma^t \cdot \left(1 - \frac{\alpha}{c \cdot n} \right) + c \cdot \alpha.$$

d-弱 Memory 过程也满足相同的界。

定理 *4.1* 的证明。设 μ 为给定的 (a,b)-偏置分布。注意，我们可以假设 $a,b>1$，因为唯一的 (1,1)-偏置分布是均匀分布，而对于任意 $a,b>1$，它也是 (a,b)-偏置分布。我们的目标是界定 Memory(μ) 过程运行 d 步期间势函数 Γ 的期望变化，其中 $\Gamma^t=\Phi^t+\Psi^t$。我们先界定 d-Weak-Memory(μ) 过程运行 d 步（即一次运行）期间势函数 Γ 的期望变化。随后，我们将 d-Weak-Memory(μ) 过程中的势函数下降与 Memory(μ) 中的势函数下降联系起来。为进行耦合，我们令两个过程在时间 t 从相同的配置开始，因此，对于所有 $i\in[n]$，都有 $\Phi_i^t=\widehat{\Phi}_i^t$ 和 $\Psi_i^t=\widehat{\Psi}_i^t$。

首先，令 $\widehat{p}_{i,j}:=\widehat{p}_{i,j}(\mu)$。那么，在 d-Weak-Memory(μ) 一次运行过程中，第 i 个负载最高的箱子的过载势 Φ 的期望变化为

$$\mathbf{E}\left[\,\widehat{\Phi}_i^{t+d}\,\middle|\,\mathfrak{F}^t\,\right]=\widehat{\Phi}_i^t\cdot e^{-d\alpha/n}\cdot\sum_{j=0}^{d}\widehat{p}_{i,j}\cdot e^{j\cdot\alpha}=\Phi_i^t\cdot e^{-d\alpha/n}\left(1+\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot(e^{j\cdot\alpha}-1)\right).$$

利用不等式 $e^z\leqslant 1+z+z^2$，对于 $|z|<1$，由于 $\alpha<1/2$，我们有

$$\mathbf{E}\left[\,\widehat{\Phi}_i^{t+d}\,\middle|\,\mathfrak{F}^t\,\right]\leqslant\Phi_i^t\cdot\left(1-\frac{d\alpha}{n}+\frac{(d\alpha)^2}{n^2}\right)\cdot\left(1+\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot\left(j\alpha+(j\alpha)^2\right)\right)\tag{4.11}$$

现在注意到，对于任意 $i\in[n]$ 箱 i，必须选中它才能使其获得一个球，因此，根据伯努利不等式，对于任意 $1\leqslant j\leqslant d$，有 $\widehat{p}_{i,j}\leqslant 1-(1-b/n)^d\leqslant bd/n$。因此，对于任意 $i\in[n]$,

$$\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot j^2\leqslant\frac{bd}{n}\cdot\frac{d(d+1)(2d+1)}{6}\leqslant\frac{bd^4}{n}.\tag{4.12}$$

因此，由于 $\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot j\leqslant\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot j^2\leqslant\frac{bd^4}{n}$，根据 (4.11) 和 (4.12)，我们有

$$\mathbf{E}\left[\,\widehat{\Phi}_i^{t+d}\,\middle|\,\mathfrak{F}^t\,\right]\leqslant\Phi_i^t\cdot\left[1+\alpha d\left(\frac{1}{d}\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot j-\frac{1}{n}\right)+\alpha^2\left(\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot j^2+\mathcal{O}\Big(\frac{bd^5}{n^2}\Big)\right)\right].\tag{4.13}$$

同理，对于欠载势函数 $\mathbf{E}\left[\,\widehat{\Psi}_i^{t+d}\,\middle|\,\mathfrak{F}^t\,\right]=\Psi_i^t\cdot e^{d\alpha/n}\cdot\sum_{j=0}^{d}\widehat{p}_{i,\,j}\cdot e^{-j\cdot\alpha}$，因此

$$\begin{aligned}\mathbf{E}\left[\,\widehat{\Psi}_i^{t+d}\,\middle|\,\mathfrak{F}^t\,\right]&\leqslant\Psi_i^t\cdot\left(1+\frac{d\alpha}{n}+\frac{(d\alpha)^2}{n^2}\right)\cdot\left(1+\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot\left(-j\alpha+(j\alpha)^2\right)\right)\\&=\Psi_i^t\cdot\left[1+\alpha d\left(\frac{1}{n}-\frac{1}{d}\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot j\right)+\alpha^2\left(\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot j^2+\mathcal{O}\Big(\frac{bd^5}{n^2}\Big)\right)\right].\end{aligned}\tag{4.14}$$

我们旨在应用定理 4.4 来界定 $\mathbf{E}[\widehat{\Gamma}^{t+d}\mid\mathfrak{F}^t]$。由 (4.12) 可知，关于过载势函数和欠载势函数的界 (4.13) 和 (4.14) 具有正确的形式，可与定理 4.4 中的项 $\Delta\Phi$ 和 $\Delta\Psi$ 进行比较。然而，我们必须验证由 $\widehat{p}_i=\frac{1}{d}\sum_{j=1}^{d}\widehat{p}_{i,\,j}\cdot j$ 给出的概率分配向量 $(\widehat{p}_i)_{i\in[n]}$ 对于 $i\in[n]$ 满足条件 $\mathcal{C}_1$。我们首先对最重的箱子上的 $\widehat{p}_i$ 进行界定。

我们可以假设 $a,b>1$ 满足 $M=n\cdot\frac{a-1}{ab-1}$ 是整数；否则，我们可以对任意一对满足 $M=n\cdot\frac{a'-1}{a'b'-1}$ 是整数且 $a'\leqslant a$ 和 $b'\geqslant b$ 的常数 $a',b'>1$ 证明该定理。由于 (a',b')-偏置分布类包含 (a,b)-偏置分布，因此结论成立。其次，令 ν 为 (a,b)-阶梯分布。于是，根据引理 4.6，由 d-Weak-Memory(ν) 诱导的向量 $\widehat{p}_i(\nu)$ 主序于相应的向量 $\widehat{p}_i(\mu)$，后者由

任意其他 d-Weak-Memory(μ) 过程诱导，其中 μ 是任意 (a,b)-偏置分布。因此，应用引理 A.2，Φ 和 Ψ 是 Schur 凸函数，所以

$$\mathbf{E}\big[\widehat{\Phi}^{t+d}(\mu)\,\big|\,\mathfrak{F}^t\big]\leqslant\mathbf{E}\big[\widehat{\Phi}^{t+d}(\nu)\,\big|\,\mathfrak{F}^t\big]\quad\text{and}\quad\mathbf{E}\big[\widehat{\Psi}^{t+d}(\mu)\,\big|\,\mathfrak{F}^t\big]\leqslant\mathbf{E}\big[\widehat{\Psi}^{t+d}(\nu)\,\big|\,\mathfrak{F}^t\big].$$

因此，只需对 d-Weak-Memory(ν) 过程中的 $\widehat{p}_i(\nu)$ 进行界定。于是，令 $\widehat{p}_{i,\ j}:=\widehat{p}_{i,\ j}(\nu)$，并假设对于某些 $\kappa:=\kappa(i)$ 和 $\gamma:=\gamma(i)$，有 $\widehat{p}_{i,\ j}\leqslant\kappa\cdot\gamma^{j-1}<1$。那么，根据引理 A.5

$$\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot j\leqslant\kappa\cdot\sum_{j=0}^{d-1}\gamma^j\cdot(j+1)\leqslant\kappa\left(\frac{1}{1-\gamma}+\frac{\gamma}{(1-\gamma)^2}\right).\tag{4.15}$$

根据引理 4.8，对于任意 $i\leqslant M=\underline{{}^{n}{}_{(a-1)\ ab-1}}$ 和 $0\leqslant j\leqslant d$，我们有

$$\widehat{p}_{i,j}=\left(\frac{ib}{n}\right)^{j-1}\cdot\frac{b}{n}\left[1-\frac{b}{n-b(i-1)}\cdot\left(1-\left(\frac{b(i-1)}{n}\right)^{d-j}\right)\right].\tag{4.16}$$

注意到 $b\cdot M=n\cdot\frac{ab-b}{ab-1}=\cancel{n-}\,n\cdot\frac{b-1}{ab-1}\cancel{<n_{\circ}}$ 因此，对于所有 $0\leqslant i\leqslant M$，我们有$\underline{n-b^{\ b}{}_{(i-1)}}\cancel{\geqslant 0}$ 和 $\underline{{}^{b}{}_{(i-1)\ n}}\cancel{<1_{\circ}}$ 由此，对于任意 $i\leqslant M$ 和 $0\leqslant j\leqslant d$，

$$\frac{b}{n-b(i-1)}\cdot\left(1-\left(\frac{b(i-1)}{n}\right)^{d-j}\right)\geqslant 0.$$

因此 $\widehat{p}_{i,\ j}\leqslant\left(\frac{ib}{n}\right)^{j-1}\cdot\frac{b}{n}$ 由 (4.16) 可知，我们可以取 $\kappa(i)\equiv\frac{b}{n}$ 和 $\gamma(i)\equiv\frac{bi}{n}$。于是，由 (4.15)，对于$i\leqslant M$

$$\widehat{p}_i=\frac{1}{d}\cdot\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot j\leqslant\frac{1}{d}\cdot\frac{b}{n}\left(\frac{n}{n-bi}+\frac{n\cdot bi}{(n-bi)^2}\right)=\frac{bn}{d(n-bi)^2}.\tag{4.17}$$

现在我们证明，定理 4.4 的条件由 $(\widehat{p}_i)^{i\in[n]}$ 满足。

断言 4.9. 对于任意 $\varepsilon<1$，存在某个 $d:=d(a,b)\geqslant 2$，使得 d-Weak-Memory(ν) 过程的概率分配向量 $(\widehat{p}_i)^{i\in[n]}$ 满足条件 $\mathcal{C}_1$，其中 $\delta=\frac{a-1}{ab-1}$ 且$\varepsilon<1$。

断言 *4.9* 的证明. 由于 $\widehat{p}_i=\frac{1}{d}\sum_{j=0}^{d}j\cdot\widehat{p}_{i,\ j}$ 与单次运行中分配给 i 的球的期望数量成正比，根据引理 4.7，我们可知对于任意 $i\leqslant M$，$\widehat{p}_i$ 关于 i 单调不减。此外，由 (4.17)，对于 $i\leqslant M=\underline{{}^{n}{}_{(a-1)\ ab-1}}$，有 $\widehat{p}_i\leqslant\frac{bn}{d\ n-bi()^2}$

$$\begin{aligned}\sum_{i=1}^{k}\widehat{p}_i&\leqslant\frac{bn}{d}\int_1^{k+1}\frac{1}{(n-bx)^2}\,\mathrm{d}x\\&=\frac{bn}{d}\left[\frac{1}{b(n-bx)}\right]_1^{k+1}\\&=\frac{n}{d}\left[\frac{1}{n-b(k+1)}-\frac{1}{n-b}\right]\\&=\frac{bnk}{d(n-b)(n-b(k+1))}.\end{aligned}$$

因此，对于任意 $k\leqslant M-1=\underline{{}^{n}{}_{(a-1)}}\,{}_{ab-1}-1$，若我们确保 $n\geqslant 2b$，则

$$\sum_{i=1}^{k}\widehat{p}_i\leqslant\frac{nb}{dn}\cdot\frac{k}{(n/2)\cdot\left(1-b\cdot\frac{a-1}{ab-1}\right)}=\frac{2b(ab-1)}{d(b-1)}\cdot\frac{k}{n}.\tag{4.18}$$

注意到

$$\widehat{p}_M = \frac{bn}{d(n-bM)^2} = \frac{b}{dn}\cdot\frac{(ab-1)^2}{(b-1)^2}$$

因此，将 $\widehat{p}_M$ 的值加到截至 $M-1$ 的总和上界中，得到

$$\sum_{i=1}^{M}\widehat{p}_i \leqslant \frac{2b(ab-1)}{d(b-1)}\cdot\frac{M-1}{n} + \frac{b}{dn}\cdot\frac{(ab-1)^2}{(b-1)^2} \leqslant \frac{2b(ab-1)^2}{d(b-1)^2}\cdot\frac{M}{n}. \tag{4.19}$$

因此，若取 $\delta = M/n = \frac{a-1}{ab-1}$，则条件 $\mathcal{C}_1$ 的第一部分得到满足，因为对于任意 $\varepsilon<1$，若选取一个整数 $d \geqslant \frac{2b}{(ab-1)^2}\cdot\frac{(b-1)^2(1-\varepsilon)}{2b(ab-1)\cdot(b-1)(1-\varepsilon)}$，则由 (4.18) 和 (4.19) 可得

$$\sum_{i=1}^{k}\widehat{p}_i \leqslant (1-\varepsilon)\cdot\frac{k}{n},$$

对于任意 $k \leqslant M$。因此，一旦固定了一个 $\varepsilon<1$，就有 $\sum_{i=1}^{M}\widehat{p}_i \leqslant (1-\varepsilon)\cdot\delta$。由于这些 $\widehat{p}_i$构成一个概率向量，我们必有

$$\sum_{i=M+1}^{n}\widehat{p}_i = 1-\sum_{i=1}^{M}\widehat{p}_i \geqslant 1-(1-\varepsilon)\cdot\delta.$$

现在，由引理 4.7，$\widehat{p}_i$ 关于 $i>M$ 单调不减，因此对于任意 $M+1\leqslant k\leqslant n$，我们有

$$\sum_{i=k}^{n}\widehat{p}_i \geqslant \frac{n-k+1}{(1-\delta)n}\cdot(1-(1-\varepsilon)\cdot\delta) = \frac{n-k+1}{n}\cdot\left(1+\varepsilon\cdot\frac{\delta}{1-\delta}\right),$$

因此，$\widehat{p}_i$'也满足条件 $\mathcal{C}_1$ 的第二部分。 ◊

现在，根据断言 4.9，对于任意 $a,\ b>1$ 存在一个常数整数 $d := d(a,\ b) \geqslant 2$，使得 $(\widehat{p}_i)_{i\in[n]}$ 满足条件 $\mathcal{C}_1$，适用于 $\delta = \frac{a-1}{ab-1}$ 和 $\varepsilon = 1/2$。由 (4.13)、(4.14) 和 (4.12)，存在某个常数 $K := K(a,\ b,\ d) = bd^4 + \mathcal{O}(bd^5/n) \leqslant 2bd^4$，由于 n 足够大，使得 $\mathbf{E}[\widehat{\Phi}_i^{t+d} \mid \mathfrak{F}^t]$、$\mathbf{E}[\widehat{\Psi}_i^{t+d} \mid \mathfrak{F}^t]$、$(\widehat{p}_i)_{i\in[n]}$ 和 K 满足条件 (4.1) 和 (4.2)。因此，若我们固定

$$\delta = \frac{a-1}{ab-1}, \qquad \varepsilon = \frac{1}{2}, \qquad \kappa = d, \qquad \text{and} \qquad K = 2bd^4, \tag{4.20}$$

然后 b 根据定理 4.4，存在常数 $c_1, c_2 \geqslant 1$，使得对于任意 $0<\alpha\leqslant\min\{\frac{1}{d}, \frac{\varepsilon\delta}{d\cdot 8K}$ – ——$\}$，

$$\mathbf{E}\left[\widehat{\Gamma}^{t+d}\,\middle|\,\mathfrak{F}^t\right] = \mathbf{E}\left[\widehat{\Phi}^{t+d}\,\middle|\,\mathfrak{F}^t\right] + \mathbf{E}\left[\widehat{\Psi}^{t+d}\,\middle|\,\mathfrak{F}^t\right] \leqslant \Gamma^t\cdot\left(1-\frac{\alpha}{c_1\cdot n}\right) + c_2\cdot\alpha. \tag{4.21}$$

在证明 d-弱 Memory(μ) 过程中的势函数下降后，仍需将其与 Memory(μ) 过程的势函数联系起来，这正是我们的初始目标。我们将比较 Memory 过程 Φ 的过载势变化与 d-弱 Memory(μ) ⌒过程 Φ的过载势变化。设 $p_{i,j}$ 为在 Memory(μ) 过程中，箱 i 收到 j 个球的概率。若以缓存的值为条件（$\mathfrak{F}^t$ 已知），这些概率便是固定的。注意，采用相同步骤，可以推导出与 (4.13) 类似的界，但适用于 Memory(μ) 过程。因此，对引理 4.5 中表达式的部分应用该界，可得

$$\begin{aligned}\mathbf{E}\left[\Phi^{t+d}\,\middle|\,\mathfrak{F}^t\right] &\leqslant \sum_{i=1}^{n}\Phi_i^t\cdot\left[1+\alpha d\left(\frac{1}{d}\sum_{j=1}^{d}p_{i,j}\cdot j-\frac{1}{n}\right)+\alpha^2\left(d\cdot\sum_{j=1}^{d}p_{i,j}\cdot j+\mathcal{O}\Big(\frac{bd^5}{n^2}\Big)\right)\right]\\ &\leqslant \sum_{i=1}^{n}\Phi_i^t\cdot\left[1+\alpha d\left(\frac{1}{d}\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot j-\frac{1}{n}\right)+\frac{2d^3\alpha^2}{n}+\alpha^2\left(d\cdot\sum_{j=1}^{d}\widehat{p}_{i,j}\cdot j+\mathcal{O}\Big(\frac{bd^4}{n}\Big)\right)\right]\end{aligned}$$

$$\leqslant \mathbf{E}\left[\, \widehat{\Phi}^{t+d} \,\middle|\, \mathfrak{F}^t \,\right] + \Phi^t \cdot \frac{C\alpha^2}{n},$$

对于某个 $C := C(a,b,d)$，其中在最后一个不等式中，我们根据 (4.12) 使用了 $\sum_{j=1}^{d} \widehat{p}_{i,j} \cdot j \leqslant \sum_{j=1}^{d} \widehat{p}_{i,j} \cdot j^2 \leqslant bd^4/n$。类似地，我们有 $\mathbf{E}\big[\Psi^{t+d} \,\big|\, \mathfrak{F}^t\big] \leqslant \mathbf{E}\big[\widehat{\Psi}^{t+d} \,\big|\, \mathfrak{F}^t\big] + \Psi^t \cdot \frac{C\alpha^2}{n}$，因此，

$$\mathbf{E}\left[\, \Gamma^{t+d} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \mathbf{E}\left[\, \widehat{\Gamma}^{t+d} \,\middle|\, \mathfrak{F}^t \,\right] + \Gamma^t \cdot \frac{2C\alpha^2}{n}. \tag{4.22}$$

因此，由 (4.22) 和 (4.21)，存在一个常数 $0 < \alpha := \alpha(a,\ b)$，使得

$$\mathbf{E}\left[\, \Gamma^{t+d} \,\middle|\, x^t \,\right] \leqslant \Gamma^t \cdot \Big(1 - \frac{\alpha}{c_1 \cdot n} + \frac{\alpha^2 \cdot 2C}{n}\Big) + c_2 \cdot \alpha \leqslant \Gamma^t \cdot \Big(1 - \frac{\alpha}{c \cdot n}\Big) + c \cdot \alpha,$$

对某个 $c := c(a,\ b) \geqslant 1$，从而得到所述结论。 □

4.3 双曲余弦势集中

本节将证明一个关于双曲余弦势函数集中性的相当一般的定理。该定理的一个版本曾出现在 [19] 中。

定理 4.2（重述）。考虑任意一个过程 $\mathcal{P}$，其中每个步骤至多分配 $d \in \mathbb{N}_+$ 个球，并考虑任意常数 $\kappa \geqslant 6$。此外，假设对于该过程，关于双曲势函数 $\Gamma_1 :=\Gamma_1(\alpha_1)$ 和 $\Gamma_2 :=\Gamma_2(\alpha_2)$，分别使用平滑参数 $0 < \alpha_1 < 1/(2d)$ 和 $\alpha_2 \leqslant \frac{\alpha_1}{12\kappa}$，存在一个 $\varepsilon > 0$（满足 $\alpha_2\varepsilon \geqslant n^{-1/6}$）以及常数 c_1、$c_2 > 0$（满足 $c_1 \leqslant c_2$），使得对于任意步骤 $t \geqslant 1$，

$$\mathbf{E}\left[\, \Gamma_1^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma_1^t \cdot \Big(1 - \frac{c_1\alpha_1\varepsilon}{n}\Big) + c_2\alpha_1\varepsilon,$$

and

$$\mathbf{E}\left[\, \Gamma_2^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma_2^t \cdot \Big(1 - \frac{c_1\alpha_2\varepsilon}{n}\Big) + c_2\alpha_2\varepsilon.$$

那么，对于 $c := 2 \cdot \frac{c_2}{c_1} \geqslant 2$，任意步骤 $t \geqslant 1$，

$$\mathbf{Pr}\left[\, \Gamma_2^t \leqslant 3cn \,\right] \geqslant 1 - n^{-\kappa}.$$

我们首先说明如何将定理 4.3 作为该定理的一个推论得出。

定理 4.3（重述）。考虑具有任意 (a,b)-偏置采样分布的 Memory 过程，其中常数为 $a, b \geqslant 1$。那么，存在常数 $c := c(a,b) \geqslant 1$ 和 $0 < \alpha := \alpha(a,b) < 1$，使得对于势函数 $\Gamma := \Gamma(\alpha)$ 以及任意步骤 $t \geqslant 1$，我们有

$$\mathbf{Pr}\left[\, \bigcap_{u \in [t,\, t + n\log^8 n]} \{\Gamma^u \leqslant 6cn\} \,\right] \geqslant 1 - n^{-4}.$$

证明。令 $\alpha' := \alpha'(a,b)$、$c := c(a,b)$ 和 $d := d(a,b)$ 为定理 4.1 中的常数。令 $\alpha_1 := \alpha'$，并令 $\alpha_2 := \frac{\alpha'}{12\cdot 6}$，则根据定理 4.1，我们有势函数 $\Gamma_1 := \Gamma_1(\alpha_1)$ $\Gamma_2 :=\Gamma_2(\alpha_2)$ 对任意步骤 $t \geqslant 1$ 均满足下降不等式

$$\mathbf{E}\left[\, \Gamma_1^{t+d} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma_1^t \cdot \Big(1 - \frac{\alpha_1}{c \cdot n}\Big) + c \cdot \alpha_1,$$

and

$$\mathbf{E}\left[\, \Gamma_2^{t+d} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma_2^t \cdot \Big(1 - \frac{\alpha_2}{c \cdot n}\Big) + c \cdot \alpha_2.$$

因此，根据定理 4.2，取 $\kappa := 6$、$c_1 := 1/c$ 和 $c_2 := c$，可得对于 $c := 2 \cdot \frac{c_2}{c_1} \geqslant 2$ 以及任意 $k \in \mathbb{N}$，

$$\mathbf{Pr}\left[\Gamma_2^{k\cdot d} \leqslant 3cn\right] \geqslant 1 - n^{-6}.$$

令 $k_0 := \lfloor t/d \rfloor$ 和 $k_1 := \lceil (t + n\log^8 n)/d \rceil$，并注意到 $k_1 - k_0 + 1 \leqslant 2n\log^8 n$。因此，对 $k_1 - k_0 + 1$ 步应用并集界可得

$$\mathbf{Pr}\left[\bigcap_{k\in[k_0,k_1]} \left\{\Gamma_2^{k\cdot d} \leqslant 3cn\right\}\right] \geqslant 1 - n^{-6}\cdot(2n\log^8 n) \geqslant 1 - n^{-4}.$$

在任意 d 步内，单个箱 $i \in [n]$ 对 Γ_2 的贡献至多变化 $e^{2\alpha d}$倍，即对于任意步骤 s，$\Gamma_{2,\,i}^{s+d} \leqslant e^{2\alpha d}\cdot\Gamma_{2,\,i}^{s}$，并通过聚合 $\Gamma_2^{s+d} \leqslant e^{2\alpha d}\cdot\Gamma_2^s$。因此，对于任意 $k \in \mathbb{N}$和 $0 \leqslant j < d$，对于中间步骤 $k\cdot d + j$，我们有 $\Gamma_2^{k\cdot d+j} \leqslant e^{2\alpha_2 d}\cdot\Gamma_2^{k\cdot d} \leqslant 2\Gamma_2^{k\cdot d}$，这里用到了 $\alpha_2 \leqslant 1/(4d)$。因此，我们得出

$$\mathbf{Pr}\left[\bigcap_{u\in[t,t+n\log^8 n]} \{\Gamma_2^u \leqslant 6cn\}\right] \geqslant 1 - n^{-4}. \qquad \square$$

4.3.1 证明纲要

本节将概述定理 4.2 的证明，并对为何需要选择这两个势函数 Γ_1 和 Γ_2 提供一些直观解释。

我们的目标是证明，对于任意给定的 $t \geqslant 0$，以高概率有 $\Gamma_2^t \leqslant 3cn$。为此，我们将分析区间 $[t - T_r, t]$ 内的步骤，其中 $T_r := \lceil 2 \cdot \frac{4/3+2\kappa}{c_1\alpha_2\varepsilon} \cdot n\log n \rceil$。特别是，在这个我们称为恢复区间的区间内，我们将证明，以高概率，$\Gamma_2^r \leqslant cn$ 至少对一个步骤 $r \in [t - T_r, t]$ 成立；然后证明其会稳定下来，即在$[r, t]$中的所有步骤中都保持较小。

现在，我们将进一步介绍证明中的几个步骤（见图 2）。根据期望上界，对于任意步骤 $t \geqslant 0$，都有 $\mathbf{E}\left[\Gamma_1^t\right] \leqslant cn$。因此，根据马尔可夫不等式，以高概率有 $\Gamma_1^s \leqslant cn^{2\kappa+1}$ 对所有 $s \in [t - T_r, t]$ 均成立。

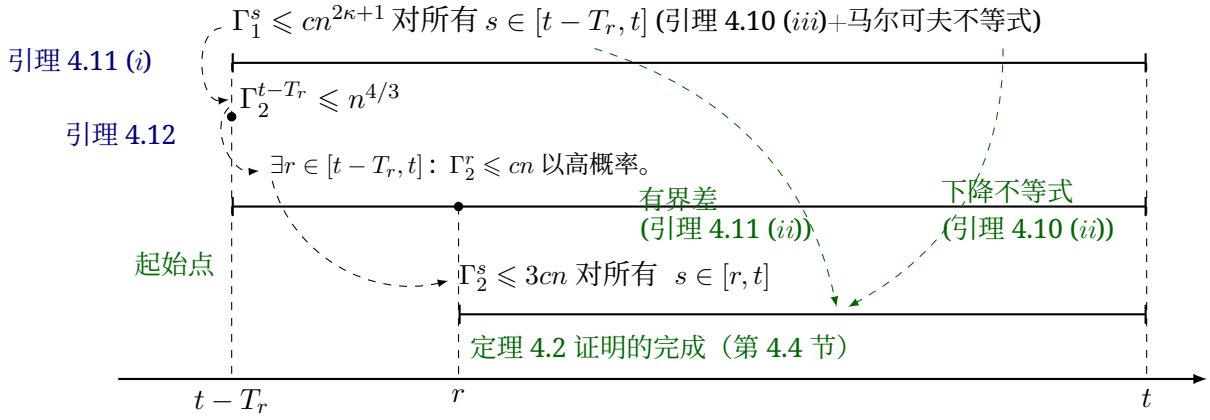

图 2： 定理 4.2 证明的概要。绿色标出的结果用于定理 4.2 中对上鞅应用阿祖马集中不等式（引理 A.8）。

根据 $\alpha_2 \leqslant \frac{\alpha_1}{12\kappa}$ 的选择，我们将证明，当 $\Gamma_1^s \leqslant cn^{2\kappa+1}$ 时，还满足 (*i*)：$\Gamma_2^s \leqslant n^{4/3}$；以及 (*ii*)：$|\Gamma_2^{s+1} - \Gamma_2^s| \leqslant n^{1/3}$（引理 4.11）。第一个条件有助于证明恢复，即 $\Gamma_2^r \leqslant cn$ 对至少一个步骤 $r \in [t - T_r, t]$ 成立（引理 4.12）。然后，从这一步 r 出发并利用第二个条件，我们可以使用集中不等式推导出 Γ_2 趋于稳定，即对所有步骤 $s \in [r, t]$，$\Gamma_2^s \leqslant 3cn$ 均成立（引理 4.13）。

4.3.2 辅助引理

本节将证明一些关于势函数 Γ_1 和 Γ_2 的辅助引理，这些势函数定义见定理 4.2。

引理 4.10。 考虑任意满足定理 *4.2* 前提条件的过程 $\mathcal{P}$ 。那么，对于任意步骤 $t \geqslant 0$,

$$
\begin{aligned}
(i) \quad & \mathbf{E}\left[\, \Gamma_1^{t+1} \,\middle|\, \mathfrak{F}^t, \Gamma_1^t > cn \,\right] \leqslant \Gamma_1^t \cdot \left(1 - \frac{c_1 \alpha_1 \varepsilon}{2n}\right), \\
(ii) \quad & \mathbf{E}\left[\, \Gamma_2^{t+1} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t > cn \,\right] \leqslant \Gamma_2^t \cdot \left(1 - \frac{c_1 \alpha_2 \varepsilon}{2n}\right), \\
(iii) \quad & \mathbf{E}\left[\, \Gamma_1^t \,\right] \leqslant cn.
\end{aligned}
$$

证明。首个陈述。 回忆一下， $c = 2 \cdot \frac{c_2}{c_1} \geqslant 2$。对于第一个陈述，根据假设 s

$$
\begin{aligned}
\mathbf{E}\left[\, \Gamma_1^{t+1} \,\middle|\, \mathfrak{F}^t, \Gamma_1^t > cn \,\right] &\leqslant \Gamma_1^t \cdot \left(1 - \frac{c_1 \alpha_1 \varepsilon}{n}\right) + c_2 \alpha_1 \varepsilon \\
&= \Gamma_1^t \cdot \left(1 - \frac{c_1 \alpha_1 \varepsilon}{2n}\right) - \Gamma_1^t \cdot \frac{c_1 \alpha_1 \varepsilon}{2n} + c_2 \alpha_1 \varepsilon \\
&\leqslant \Gamma_1^t \cdot \left(1 - \frac{c_1 \alpha_1 \varepsilon}{2n}\right) - 2 \cdot \frac{c_2}{c_1} \cdot n \cdot \frac{c_1 \alpha_1 \varepsilon}{2n} + c_2 \alpha_1 \varepsilon = \Gamma_1^t \cdot \left(1 - \frac{c_1 \alpha_1 \varepsilon}{2n}\right).
\end{aligned}
$$

第二个陈述。 同理，我们得到关于 Γ_2 的第二个陈述。

第三个陈述。 由引理 A.7，对于 $a = \frac{c_1 \alpha_1 \varepsilon}{n}$ 和 $b = c_2 \alpha_1 \varepsilon$，由于 $\Gamma_1^0 = 2n \leqslant 2 \cdot \frac{c_2}{c_1} n = cn$, 可得 $\mathbf{E}[\Gamma^t] \leqslant cn$，对任意步骤 $t \geqslant 0$ 均成立。 □

引理 4.11. 考虑任意满足定理 *4.2* 前提条件的过程 $\mathcal{P}$ 。对于任意满足 $\Gamma_1^t \leqslant cn^{2\kappa+1}$ 的步骤 $t \geqslant 0$ ，我们有

$$
\begin{aligned}
(i) \quad & \Gamma_2^t \leqslant n^{4/3}, \\
(ii) \quad & |\Gamma_2^{t+1} - \Gamma_2^t| \leqslant n^{1/3}.
\end{aligned}
$$

证明。 考虑任意一个满足 $\Gamma_1^t \leqslant cn^{2\kappa+1}$ 的步骤 t。我们首先证明，对于任意箱 $i \in [n]$，归一化负载 y_i^t 满足以下界

$$
\Gamma_1^t \leqslant cn^{2\kappa+1} \Rightarrow e^{\alpha_1 \cdot y_i^t} + e^{-\alpha_1 \cdot y_i^t} \leqslant cn^{2\kappa+1} \Rightarrow y_i^t \leqslant \frac{3\kappa}{\alpha_1} \cdot \log n \wedge -y_i^t \leqslant \frac{3\kappa}{\alpha_1} \cdot \log n,
$$

其中，在第二个蕴含中，我们使用了 $\log c + \frac{2\kappa+1}{\alpha_1} \log n \leqslant \frac{3\kappa}{\alpha_1} \log n$；当 n 足够大时，由于 c 是一个常数且 $\kappa \geqslant 6 \geqslant 1$。

第一个结论。 回顾 $\alpha_2 \leqslant \alpha_1 / 12\kappa$。根据 Γ_2^t 的定义以及每个归一化箱负载的界，我们得到

$$
\Gamma_2^t \leqslant 2 \cdot \sum_{i=1}^{n} \exp\left(\alpha_2 \cdot \frac{3\kappa}{\alpha_1} \cdot \log n\right) = 2n \cdot n^{1/4} \leqslant n^{4/3}.
$$

第二个结论。 将 Γ_2^{t+1} 看作 $2n$ 个指数函数值之和，它由 Γ_2^t 得到，只需稍微改变指数 $2n$ 的值。指数的总变化量上界为 $4d$，因为我们将在负载向量 x^t 中增加 d 个条目（每个条目出现两次），并且还将在所有 $2n$ 个指数中将平均负载增加 $\frac{d}{n}$ 。由于 $\exp(\cdot)$ 是凸函数，最大的变化量不超过以下（假设）情形：最大指数增加 $4d$ ，其余指数保持不变,

$$
\left|\Gamma_2^{t+d} - \Gamma_2^t\right| \leqslant \exp\left(\alpha_2 \cdot \left(4d + \max_{i \in [n]} |y_i^t|\right)\right) \leqslant e^{4d\alpha_2} \cdot \exp\left(\alpha_2 \cdot \frac{3\kappa}{\alpha_1} \cdot \log n\right) = e^{4d\alpha_2} \cdot n^{1/4} \leqslant n^{1/3},
$$

利用 $\alpha_2 \leqslant \frac{\alpha_1}{12\kappa}$ 以及 $\alpha_2 \leqslant \alpha_1 \leqslant 1/(2d)$。 □

4.3.3 恢复与稳定化

利用引理 4.10 中的第二和第三个陈述，我们现在将证明定理 4.2 的一个较弱结论，表明对于至少一个步骤 $r \in [t - T_r, t]$，T_r 是恢复区间的长度

$$T_r := \left\lceil 2 \cdot \frac{4/3 + 2\kappa}{c_1 \alpha_2 \varepsilon} \cdot n \log n \right\rceil. \tag{4.23}$$

在此之前，我们先定义一个辅助过程。

辅助过程。设 $\mathcal{P}$ 是满足定理 4.2 前提条件的过程。我们希望对以下条件进行条件化：在长度为 poly(n) 的某个区间内的每个步骤 s，$\mathcal{P}$ 都具有 $\Gamma^s \leqslant cn^{2\kappa+1}$，这样便可推出它满足有界差分条件（引理 4.10），进而应用 Azuma不等式（引理 A.8）。

为此，我们将定义一个辅助过程 $\mathcal{P}_{t_0} := \mathcal{P}_{t_0}(\mathcal{P})$，其中 $t_0 \geqslant 0$ 是任意步骤。令 $\sigma := \inf\left\{s \geqslant t_0 : \Gamma^s > \frac{1}{2} cn^{2\kappa+1}\right\}$。接着，我们定义 $\widetilde{\mathcal{P}}$，使得

- 在 $[0, \sigma$ 个步骤中），它所做的分配与 $\mathcal{P}$ 相同，并且
- 在 $[\sigma, \infty$步中，它分配到当前负载最小的箱，即使用概率分配向量 $q^s = (0, \ldots\ldots, 0, 1)$。

令 $y^s_{\widetilde{\mathcal{P}}}$ 为 $\mathcal{P}_{t_0}$ 在步骤 $s \geqslant 0$时的归一化负载向量。根据引理 4.10（iii）、马尔可夫不等式和并集界可知，对于任意满足 $m - t_0 \leqslant T_r$ 的区间 $[t_0, m]$，两个过程以高概率一致

$$\mathbf{Pr}\left[\bigcap_{s\in[t_0,m]} \left\{y^s_{\mathcal{P}} = y^s\right\}\right] \geqslant \mathbf{Pr}\left[\bigcap_{s\in[t_0,m]} \left\{\Gamma^s_1 \leqslant cn^{2\kappa+1}\right\}\right] \geqslant 1 - n^{-2\kappa} \cdot T_r. \tag{4.24}$$

过程 $\mathcal{P}_{t_0}$ 的定义满足以下性质：

- （**性质 1**）$\mathcal{P}_{t_0}$ 过程对于势函数 $\Gamma_{1,\,\widetilde{\mathcal{P}}}$ 和 $\Gamma_{2,\,\widetilde{\mathcal{P}}}$，在任意步骤 $s \geqslant 0$都满足下降不等式（第一和第二个前提条件）。这是因为对于任意步骤 $s < \sigma$，该过程都遵循 $\mathcal{P}$。对于任意步骤 $s \geqslant \sigma$，该过程都会分配到当前负载最小的箱，因此在给定任意 $\mathfrak{F}^s$ 的情况下，使势函数 $\Gamma^{s+1}_{1\,\widetilde{\mathcal{P}}}$ 最小；这意味着 $\Gamma^{s+1}_{1,\,\widetilde{\mathcal{P}}} \leqslant \mathbf{E}\left[\Gamma^{s+1}_1 \mid \mathfrak{F}^s\right]$，因此显然满足任意下降不等式（$\Gamma_{2,\,\widetilde{\mathcal{P}}}$ 的情况也类似）。

进一步，我们将步骤 t_0 时势函数 Γ_1 较小这一事件定义为：

$$\mathcal{Z}^{t_0} := \left\{\Gamma^{t_0}_{1,\widetilde{\mathcal{P}}} \leqslant \frac{1}{2} cn^{2\kappa+1}\right\}, \tag{4.25}$$

其中，$c \geqslant 1$ 是定理 4.2 中定义的常数。当事件 $\mathcal{Z}^{t_0}$ 成立时，过程 $\widetilde{\mathcal{P}}$ 也满足以下性质（它“实现”了条件化 $\Gamma^s_{1,\,\widetilde{\mathcal{P}}} \leqslant cn^{2\kappa+1}$）：

- （**性质 2**）对于任意步骤 $s \geqslant t_0$，都有

  $$\Gamma^{t_0}_{1,\widetilde{\mathcal{P}}} \leqslant cn^{2\kappa+1},$$

  在任意步骤 $s \in [t_0, \sigma)$，根据 σ 的定义，这一点都成立。对于任意步骤 $s \geqslant \sigma$，球绝不会被分配到负载为 $y^s_i > 0$ 的箱子中，并且每经过 n 步，负载

等于最小负载的至多 个箱子（在步骤 s）都会至少各接收一个球。因此，在任意 n 步内，最大绝对归一化负载都不会增加；而在中间的步骤中，它最多会增加 1，因此，

$$\Gamma_{1,\widetilde{\mathcal{P}}}^{t_0} \leqslant e^{\alpha_1} \cdot \Gamma_{1,\widetilde{\mathcal{P}}}^{\sigma} \leqslant e^{\alpha_1} \cdot \frac{1}{2} c n^{2\kappa+1} \leqslant c n^{2\kappa+1}.$$

引理 4.12 (恢复)。 考虑任意步骤 $t \geqslant 0$ 以及辅助过程 $\mathcal{P}_{t-T_r} := \mathcal{P}_{t-T_r}(\mathcal{P})$，对于任何 $\mathcal{P}$ 满足定理 *4.2* 的前提条件且 $\mathcal{Z}^{t-T_r}$ 为(4.25)中定义的事件。对于任意步骤 $t \geqslant 0$，有

$$\mathbf{Pr}\left[\bigcup_{r \in [t-T_r, t]} \{\Gamma_2^r \leqslant cn\} \,\middle|\, \mathfrak{F}^{t-T_r}, \mathcal{Z}^{t-T_r} \right] \geqslant 1 - 2n^{-2\kappa-1}.$$

证明。如果 $t < T_r$，则该陈述显然成立，因为对于 $r = 0$，确定性地有 $\Gamma_2^r = 2n \leqslant cn$。否则，根据条件 $\mathcal{Z}^{t-T_r}$，我们有 $\{\Gamma_1^{t-T_r} \leqslant cn^{2\kappa+1}\}$ 成立。由引理 4.11（i），可知 $\{\Gamma_2^{t-T_r} \leqslant n^{4/3}\}$ 也成立。

由引理 4.10（ii），对于任意步骤 $s \geqslant 0$,

$$\mathbf{E}\left[\Gamma_2^{s+1} \,\middle|\, \mathfrak{F}^s, \Gamma_2^s > cn \right] \leqslant \Gamma_2^s \cdot \left(1 - \frac{c_1 \alpha_2 \varepsilon}{2n}\right). \tag{4.26}$$

接下来，我们定义步骤 $s \geqslant t - T_r$ 时的“终止”势函数为

$$\widehat{\Gamma}_2^s := \Gamma_2^s \cdot \mathbf{1}_{\bigcap_{r \in [t-T_r, s]} \{\Gamma_2^r > cn\}}.$$

注意，当 $\{\Gamma_1^s \leqslant cn\}$ 时，也有 $\left\{\widehat{\Gamma}_1^s = 0\right\}$ 和 $\left\{\widehat{\Gamma}_1^{s+1} = 0\right\}$。因此，$\widehat{\Gamma}$ 势函数无条件满足不等式 (4.26)，也就是说，对任意 $s \geqslant t - T_r$

$$\mathbf{E}\left[\widehat{\Gamma}_2^{s+1} \,\middle|\, \mathfrak{F}^s \right] \leqslant \widehat{\Gamma}_2^s \cdot \left(1 - \frac{c_1 \alpha_2 \varepsilon}{2n}\right).$$

从 $\widehat{\Gamma}_2^{t-T_r} \leqslant \Gamma_2^{t-T_r} \leqslant n^{4/3}$ 开始，归纳地应用此结论 T_r 步，得到

$$\begin{aligned} \mathbf{E}\left[\widehat{\Gamma}_2^t \,\middle|\, \mathfrak{F}^{t-T_r}, \mathcal{Z}^{t-T_r} \right] &\leqslant \mathbf{E}\left[\widehat{\Gamma}_2^t \,\middle|\, \mathfrak{F}^{t-T_r}, \Gamma_2^{t-T_r} \leqslant n^{4/3} \right] \leqslant \left(1 - \frac{c_1 \alpha_2 \varepsilon}{2n}\right)^{T_r} \cdot n^{4/3} \\ &\overset{(a)}{\leqslant} e^{-\frac{1}{2} c_1 \alpha_2 \varepsilon \cdot \frac{T_r}{n}} \cdot n^{4/3} \overset{(b)}{\leqslant} n^{-2\kappa}, \end{aligned}$$

在 (*a*) 中使用 $1 + u \leqslant e^u$（对任意 u），并在 (*b*) 中使用 $T_r = \left\lceil 2 \cdot \frac{4/3 + 2\kappa}{c_1 \alpha_2 \varepsilon} \cdot n \log n \right\rceil$。因此，由马尔可夫不等式,

$$\mathbf{Pr}\left[\widehat{\Gamma}_2^t \leqslant n \,\middle|\, \mathfrak{F}^{t-T_r}, \mathcal{Z}^{t-T_r} \right] \geqslant 1 - n^{-2\kappa-1}.$$

由于在任意步骤 s，我们都可以确定 $\Gamma_2^s \geqslant 2n$，因此可以得出，当 $\left\{\widehat{\Gamma}_2^t \leqslant n\right\}$ 时，也有 $\left\{\widehat{\Gamma}_2^t = 0\right\}$，于是

$$\mathbf{1}_{\bigcap_{r \in [t-T_r, t]} \{\Gamma_2^r > cn\}} = 0, \text{ implying that } \neg \bigcap_{r \in [t-T_r, t]} \{\Gamma_2^r > cn\} \text{ holds}$$

以至少 $1 - 2n^{-2\kappa}$ 的概率成立，从而证明了该结论。 □

我们现在证明，只要 $\Gamma_2^r \in [cn, 2cn]$ 在某个步骤 $r \in [t - T_r, t]$中成立，那么以高概率，它（i）会一直保持较小，直到步骤 t，也就是说，$\Gamma_2^s \leqslant 3cn$ 对所有 $s \in [r, t]$ 均成立；或者（ii）它会一直保持较小，直到某个步骤 $s \leqslant t$，此时它变得非常小，也就是说，$\Gamma_2^s \leqslant cn$。

~引理 4.13 (稳定化)。 考虑任意步骤 $t \geqslant 0$ 以及辅助过程 $\mathcal{P}_{t-T_r}$：
$\widetilde{\mathcal{P}}_{t-T_r}$ $\mathcal{P}$ $\mathcal{P}$ 4 2 $\mathcal{Z}^{t-T_r}$ = () 对于任何满足定理 .
的前提条件、且其中 为(4.25)中定义的事件的情形。则对于(4.23)中定义的任意步骤 $r \in [t-T_r, t]$ （ T_r），我们有

$$\mathbf{Pr}\left[\bigcap_{s\in[r,t]}\{\Gamma_2^s \leqslant 3cn\} \cup \bigcup_{s\in(r,t]}\left(\bigcap_{u\in[r,s]}\{\Gamma_2^u \leqslant 3cn\} \cap \{\Gamma_2^s \leqslant cn\}\right) \;\middle|\; \mathcal{Z}^{t-T_r}, \mathfrak{F}^r, \Gamma_2^r \in [cn, 2cn]\right] \geqslant 1 - n^{-\frac{4}{3}\kappa}.$$

证明。 考虑任意步骤 $r \in [t-T_r, t]$，使得 $\Gamma_2^r \in [cn, 2cn]$。我们定义停止时间

$$\tau := \inf\{\widetilde{r} > r \colon \Gamma_2^{\widetilde{r}} \leqslant cn\},$$

对于任意步骤 $s \in [r, t]$,

$$X_r^s := \Gamma_2^{s\wedge\tau}.$$

这一设定背后的想法是， X_r^s 构成一个上鞅。为说明这一点，注意到根据引理 4.10（ii），对于任意步骤 $s < \tau$，我们有

$$\mathbf{E}\left[\, X_r^{s+1} \,\middle|\, \mathfrak{F}^s, s < \tau \,\right] \leqslant \mathbf{E}\left[\, \Gamma_2^{s+1} \,\middle|\, \mathfrak{F}^s, \Gamma_2^s \geqslant cn \,\right] \leqslant \Gamma_2^s, \tag{4.27}$$

对于任意步骤 $s \geqslant \tau$,

$$\mathbf{E}\left[\, X_r^{s+1} \,\middle|\, \mathfrak{F}^s, s \geqslant \tau \,\right] = X_r^s. \tag{4.28}$$

回想一下，当 $\mathcal{Z}^{t-T_r}$ 成立时，根据性质 2（见第 4.3.3 节），对于每个步骤 $s \geqslant t - T_r$，都有 $\Gamma_1^s \leqslant cn^{2\kappa+1}$。因此，根据引理 4.11（$ii$），也有$|\Gamma_2^{s+1} - \Gamma_2^s| \leqslant n^{1/3}$。于是，对任意 $s \in [r, t]$ 应用引理 A.8 可得

$$\mathbf{Pr}\left[\, X_r^s \geqslant X_r^r + cn \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^r, \Gamma_2^r \in [cn, 2cn] \,\right] \leqslant \exp\left(-\frac{c^2n^2}{10 \cdot T_r \cdot (n^{1/3})^2}\right) \leqslant 2 \cdot T_r \cdot n^{-2\kappa},$$

利用 $T_r = \mathcal{O}(n \cdot n^{1/6} \cdot \log n)$。另请回忆，对于起始点 r， $X_r^r = \Gamma_2^r \leqslant 2cn$。因此，我们可以得出

$$\mathbf{Pr}\left[\, X_r^s > 3cn \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^r, \Gamma_2^r \in [cn, 2cn] \,\right] \leqslant 2 \cdot T_r \cdot n^{-2\kappa}.$$

对所有 $s \in [r, t]$ 应用并集界，可得

$$\mathbf{Pr}\left[\bigcap_{s=r}^{t}\{X_r^s \leqslant 3cn\} \;\middle|\; \mathcal{Z}^{t-T_r}, \mathfrak{F}^r, \Gamma_2^r \in [cn, 2cn]\right] \geqslant 1 - 3 \cdot T_r^2 \cdot n^{-2\kappa} \geqslant 1 - n^{-\kappa},$$

利用 $\kappa \geqslant 6$。现在，假设 $\bigcap_{s=r}^{t}\{X_r^s \leqslant 3cn\}$ 成立，我们根据停止时间 τ 考虑以下几种情况:

- **情况 1 [$\tau > t$]:** 那么，对于所有步骤 $u \in [r, t]$，都有 $\Gamma_2^u = X_r^u \leqslant 3cn$。
- **情况 2 [$\tau \leqslant t$]:** 那么，对于所有步骤 $u \in [r, \tau]$，都有 $\Gamma_2^u = X_r^u \leqslant 3cn$ 和 $\Gamma_2^\tau \leqslant cn$。因此，对于 $s = \tau > r$，以下事件成立，

$$\bigcup_{s\in(r,t]}\left(\bigcap_{u\in[r,s]}\{\Gamma_2^u \leqslant 3cn\} \cap \{\Gamma_2^s \leqslant cn\}\right).$$

至此，论断得证。 □

4.4 完成定理 4.2 的证明

为了完成定理 4.2 的证明，我们将首先证明关于~辅助过程 $\mathcal{P}_{t-T_r}$的等价命题。

~ ~**引理 4.14**。考虑任意步骤 $t \geqslant 0$ 以及辅助过程 $\mathcal{P}_{t-T_r} \colon = \mathcal{P}_{t-T_r}(\mathcal{P})$ 对于任意满足定理 *4.2* 前提条件且 $\mathcal{Z}^{t-T_r}$ 为(4.25)中定义的事件的$\mathcal{P}$。那么，对于 $c \colon = 2 \cdot \frac{c_2}{c_1} \geqslant 2$, 对于任意步骤 $t \geqslant 0$,

$$\mathbf{Pr}\left[\, \Gamma_2^t \leqslant 3cn \,\middle|\, \mathcal{Z}^{t-T_r} \,\right] \geqslant 1 - n^{-\kappa}.$$

证明。证明将关注步骤 $\in [t-T_r, t]$。首先，通过应用引理 4.12 ，

$$\mathbf{Pr}\left[\bigcup_{r_0 \in [t-T_r, t]} \{\Gamma_2^{r_0} \leqslant cn\} \,\middle|\, \mathfrak{F}^{t-T_r}, \mathcal{Z}^{t-T_r} \right] \geqslant 1 - 2n^{-2\kappa-1}. \tag{4.29}$$

现在考虑任意步骤 $r_0 \in [t-T_r, t]$，并假设 $\Gamma_2^{r_0} \leqslant cn$。我们将时间步 $s \in [r_0, t]$ 划分为红色阶段和绿色阶段（见图 3）：

1. **红色阶段：** 如果 $\Gamma_2^s > cn$，则步骤 s 处于红色阶段。2. **绿色阶段：** 否则，过程处于绿色阶段。

注意，根据 r_0 的选择，在时间 r_0 过程中处于绿色阶段。每个绿色阶段之前都可能有一个红色阶段。显然，对于绿色阶段中的每个步骤 s，都有 $\Gamma_2^s \leqslant cn$。当 s 是绿色阶段之后红色阶段的第一步时，可知$\Gamma_2^s \leqslant e^{d\alpha_2} \cdot \Gamma_2^{s-1} \leqslant 2 \cdot \Gamma_2^{s-1} \leqslant 2cn$，因为 $0 < \alpha_2 < 1/(2d)$。

令 $\mathcal{R}^s$ 表示步骤 s 是红色阶段第一步这一事件，并令 $\mathcal{A}^s$ 表示与 s 处于同一阶段的所有步骤 $u \geqslant s$ 都满足 $\Gamma_2^u \leqslant 3cn$ 这一事件。根据引理 4.13，我们有

$$\mathbf{Pr}\left[\, \mathcal{A}^s \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^s, \mathcal{R}^s \,\right] \geqslant 1 - n^{-\frac{4}{3}\kappa}.$$

For 对于任意事件 $\mathcal{E}_1 \neq \emptyset$ 和 $\mathcal{E}_2$，我们有 **概率** $[\,\mathcal{E}_2 \cup \neg\mathcal{E}_1\,] \geqslant 1 -$ **概率** $[\,\neg\mathcal{E}_2 \mid \mathcal{E}_1\,]$，因此 e

$$\mathbf{Pr}\left[\, \mathcal{A}^s \cup \neg\mathcal{R}^s \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^s \,\right] \geqslant 1 - n^{-\frac{4}{3}\kappa}.$$

对 $[r_0, t]$ 中的所有步骤 s 取并集界，可得

$$\mathbf{Pr}\left[\bigcap_{s \in [r_0, t]} (\mathcal{A}^s \cup \neg\mathcal{R}^s) \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^{r_0}, \Gamma_2^{r_0} \leqslant cn \right] \geqslant 1 - n^{-\frac{4}{3}\kappa} \cdot T_r \geqslant 1 - \frac{1}{2}n^{-\kappa}.$$

当 $\cap_{s \in [r_0, t]}(\mathcal{A}^s \cup \neg\mathcal{R}^s)$ 成立时，所有红色阶段中的步骤 u 均满足 $\Gamma_2^u \leqslant 3cn$。因此，由于绿色阶段中的步骤根据定义都是好的，我们可得

$$\mathbf{Pr}\left[\bigcap_{s \in [r_0, t]} \{\Gamma_2^s \leqslant 3cn\} \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^{r_0}, \Gamma_2^{r_0} \leqslant cn \right] \geqslant 1 - n^{-\frac{4}{3}\kappa} \cdot T_r \geqslant 1 - \frac{1}{2}n^{-\kappa}.$$

因此，通过定义停止时间 $\rho := \inf\{r_0 \geqslant t - T_r \colon \Gamma_2^{r_0} \leqslant cn\}$，我们得到

$$\begin{aligned}
\mathbf{Pr}\left[\, \Gamma_2^t \leqslant 3cn \,\right] &\geqslant \sum_{r_0 = t-T_r}^{t} \mathbf{Pr}\left[\, \Gamma_2^t \leqslant 3cn \,\middle|\, \mathcal{Z}^{t-T_r}, \rho = r_0 \,\right] \cdot \mathbf{Pr}\left[\, \rho = r_0 \,\middle|\, \mathcal{Z}^{t-T_r} \,\right] \\
&\geqslant \sum_{r_0 = t-T_r}^{t} \mathbf{Pr}\left[\bigcap_{s \in [r_0, t]} \{\Gamma_2^s \leqslant 3cn\} \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^{r_0}, \Gamma_2^{r_0} \leqslant cn \right] \cdot \mathbf{Pr}\left[\, \rho = r_0 \,\middle|\, \mathcal{Z}^{t-T_r} \,\right]
\end{aligned}$$

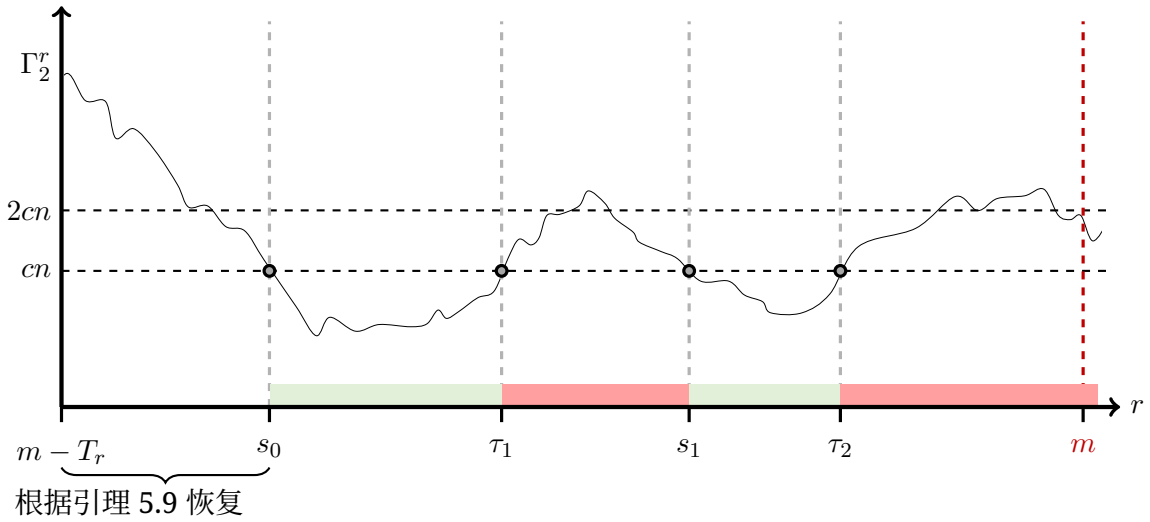

图 3： 绿色阶段表示 Γ_2^r 较小的步骤，红色阶段表示势函数较大且（在期望意义下）下降的步骤。在引理 4.13 中，我们利用集中不等式证明，在红色阶段的每个时刻， $\Gamma_2 \leqslant 3cn$ 都成立。

$$
\begin{aligned}
&\geqslant \left(1 - \frac{1}{2}n^{-\kappa}\right) \cdot \sum_{r_0 = t - T_r}^{t} \mathbf{Pr}\left[\rho = r_0 \mid \mathcal{Z}^{t-T_r}\right] \\
&= \left(1 - \frac{1}{2}n^{-\kappa}\right) \cdot \mathbf{Pr}\left[\rho \leqslant t\right] \\
&\overset{(a)}{\geqslant} \left(1 - \frac{1}{2}n^{-\kappa}\right) \cdot \left(1 - 2n^{-2\kappa}\right) \geqslant 1 - n^{-\kappa},
\end{aligned}
$$

在 (a) 中使用 (4.29)，于是结论得证。 □

现在，我们完成定理 4.2 的证明。

~定理 *4.2* 的证明。 考虑辅助过程 $\mathcal{P}_{t-T_r}$，并令 $\Gamma_{2,\ \tilde{\mathcal{P}}}$ 为其势函数。那么，根据引理 4.14，我们有

$$
\mathbf{Pr}\left[\ \Gamma^t_{2,\widetilde{\mathcal{P}}} \leqslant 3cn \ \middle|\ \mathcal{Z}^{t-T_r}\ \right] \geqslant 1 - n^{-\kappa}. \tag{4.30}
$$

~根据引理 4.10 和马尔可夫不等式，由于 $\mathcal{P}_{t-T_r}$ 和 $\mathcal{P}$ 在每个步骤 $s \leqslant t - T_r$ 都一致，我们有

$$
\mathbf{Pr}\left[\mathcal{Z}^{t-T_r}\right] = \mathbf{Pr}\left[\Gamma_1^{t-T_r} \leqslant \frac{1}{2}cn^{2\kappa+1}\right] \geqslant 1 - 2n^{-2\kappa}. \tag{4.31}
$$

因此，结合 (4.30) 和 (4.31)，我们有

$$
\mathbf{Pr}\left[\Gamma^t_{2,\widetilde{\mathcal{P}}} \leqslant 3cn\right] \geqslant 1 - n^{-\kappa}.
$$

~如（4.24）中所述，以高概率，过程 $\mathcal{P}$ 与 $\mathcal{P}$ 在 $[t - T_r, t]$中一致，因此

$$
\mathbf{Pr}\left[\Gamma_2^t \leqslant 3cn\right] \geqslant \mathbf{Pr}\left[\left\{\Gamma^t_{2,\widetilde{\mathcal{P}}} \leqslant 3cn\right\} \cap \bigcap_{s \in [t-T_r, t]} \left\{y^s = y^s_{\widetilde{\mathcal{P}}}\right\}\right] \geqslant 1 - n^{-2\kappa} - \frac{1}{2}n^{-2\kappa} \cdot T_r \geqslant 1 - n^{-\kappa},
$$

利用 $T_r = \mathcal{O}(n \cdot n^{1/6} \cdot \log n)$。 □

5 分层归纳步骤的证明

为使证明自成一体，我们重复了第3.2节中的一些定义和解释。我们的目标是证明以下定理。

定理 1.1 （重述）。 考虑具有（a, b）偏置抽样分布的MEMORY 过程，其中常数为 $a, b \geqslant 1$。那么存在一个常数 $\kappa := \kappa(a, b) > 0$，使得对于每个步骤 $m \geqslant 1$,

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \leqslant \kappa \cdot \log\log n\,\right] \geqslant 1 - n^{-3}.$$

下文中，所有势函数均针对具有 (a, b)-偏置抽样分布的 Memory 过程定义，其中常数为 a, $b \geqslant 1$。此外，本节中的所有结果均在此假设下成立。

完整势函数： 我们将对超指数势函数使用分层归纳法，类似于 [19, 节 6]和 [20, 节 6], 中采用的方法，但也存在一些差异（见第 33 页的讨论）。现在我们为$1 \leqslant j \leqslant j_{\max} - 1$定义超指数势函数,

$$\Psi_j^t := \sum_{i=1}^{n} \Psi_{j,i}^t := \sum_{i=1}^{n} e^{\alpha_1 \cdot v^j \cdot (y_i^t - z_j)^+},$$

and

$$\Phi_j^t := \sum_{i=1}^{n} \Phi_{j,i}^t := \sum_{i=1}^{n} e^{\alpha_2 \cdot v^j \cdot (y_i^t - z_j)^+},$$

其中 $x^+ = \max\{x, 0\}$，并且在本文余下部分中，我们设

$$z_j := \frac{5v}{\alpha_2} \cdot j, \quad v := \max\{\log(2Cb), 36b\}, \quad C := \max\{6c, 6\}, \quad j_{\max} = \log_v(\frac{\alpha_2}{2v}\log n), \tag{5.1}$$

以及 α_1、α_2、$c > 0$, 其中 $\alpha_1 = 6 \cdot 14 \cdot \alpha_2$ 的定义见第4节中的定理4.2和定理4.3。我们的目标是证明 $\Phi_{j_{\max}-1}^t = \mathcal{O}(n)$，由此可推出

$$\max_{i \in [n]} y_i^t \leqslant z_{j_{\max}-1} + \frac{4v^2}{\alpha_2^2} = \mathcal{O}(\log\log n).$$

折叠过程。在分层归纳的第 j 层（针对 $1 \leqslant j \leqslant j_{\max} - 1$）中，我们将分析以下折叠过程；我们很快会在引理 5.1 中验证，Memory 是其中一个实例。为此，我们将步数划分为连续的轮次（长度各不相同），并将每轮中的第 s 步称为子步骤 s。此外，我们令 $y^{r,\,i\,s}$ 表示第 r 轮第 s 个子步骤之后箱 i 的归一化负载。然后，我们将折叠过程 定义如下:

- 对于每一轮次 $r \geqslant 0$，根据抽样分布 $\mathcal{S}$，抽取箱 $i := i(r) \in [n]$：

  – **情况 A：** 如果 $y^{r,\,i\,0} \geqslant z_{j-1} + \frac{2v}{\alpha_2}$，那么将一个球分配到任意一个满足$y^{r,\,\ell\,0} \leqslant y^{r,\,i\,0}$ 的箱 ℓ，然后进入下一轮次。– **情况 B：** 否则，在当前轮次 r内开始一系列连续的阶段，每个阶段包含 $\frac{v}{\alpha_2}$ 个子步骤（也就是说，每个阶段 $k \geqslant 1$ 由子步骤 $s \in [(k-1) \cdot \frac{v}{\alpha_2}, k \cdot \frac{v}{\alpha_2}]$ 组成）。在每个子步骤 s中，我们根据 $\mathcal{S}$抽取一个箱 $i = i(r, s)$，并将一个球分配到任意一个满足 $y^{r,\,\ell\,s} \leqslant z_{j-1} + \frac{4v}{\alpha_2}$ 的箱 ℓ。每个阶段结束时，如果满足以下两个条件中的任意一个，我们也会结束该轮次:

  * **条件 1：** 在当前阶段的任何子步骤 s 中，我们都没有抽取一个箱 ℓ，其中包含 $y^{r,\,\ell\,s} < z_{j-1} + \frac{2v}{\alpha_2}$ 且在对应的子步骤 s 中。

$*$ **条件 2：**我们已完成 $k_j := e^{v^{j+1}} \cdot \log^3 n \leqslant n^{1/7}$ 个阶段。

折叠过程具有由固定长度的（半）独立区块构成的结构，这有助于在分层归纳中证明上界时进行分析。选择分配到哪个箱的灵活性，使我们能够证明，对于某一系列分配选择，Memory 过程是折叠过程的一个实例。折叠过程的示意图见图 4。

引理 5.1。MEMORY 是上述定义的折叠过程的一个实例。

证明。回想一下，原始 Memory 步骤中的任意步骤 $t \geqslant 1$ 都唯一属于某一轮次 $r \geqslant 0$和子步骤 $s \geqslant 1$。

在情形 A （即 $s = 1$）中，Memory 抽取一个箱 $i = i(r)$，并将其分配到箱 b 或 $i = i(r)$，即 $\ell \in \{i, b\}$，因此按需有 $y^{r,\ 0}_{\ell} \leqslant y^{r,\ 0}_{i}$。

对于情形 B，我们有 $s \geqslant 1$，且子步骤 s 只有在前一（或当前）阶段抽取了一个轻箱时，才可能属于第 r 轮次中的某个阶段；也就是说，存在一个子步骤 $u \in [s - 2 \cdot \frac{v}{\alpha_2}, s]$ 和一次箱抽取 $i = i(r,\ u)$，使得

$$y_i^{r,u} < z_{j-1} + \frac{2v}{\alpha_2}.$$

因此，由于每轮我们都会存储采样箱和缓存箱中负载最小的那个，所以在该阶段的每个子步骤中，我们都能访问一个负载至多为 i （再加上自采样 i 以来放置的球数）的箱。因此，

$$y_\ell^{r,s} \leqslant y_i^{r,s} \leqslant y_i^{r,u} + s - u \leqslant z_{j-1} + \frac{2v}{\alpha_2} + 2 \cdot \frac{v}{\alpha_2} = z_{j-1} + \frac{4v}{\alpha_2}. \qquad \square$$

部分势能：对于恢复阶段，也就是说，为了证明在长度为 $n \cdot \text{polylog}(n)$ 的区间内，某个步骤 s 上以高概率有 $\Phi_j^s \leqslant Cn$ （引理 5.12），我们需要势函数下降得更快，因此将使用以下仅在重箱上定义的“部分”势函数

$$\dot\Psi_j^t := \sum_{i=1}^n \dot\Psi_{j,i}^t := \sum_{i: y_i^t \geqslant z_j} e^{\alpha_1 \cdot v^j \cdot (y_i^t - z_j)}, \quad \text{and} \quad \dot\Phi_j^t := \sum_{i=1}^n \dot\Phi_{j,i}^t := \sum_{i: y_i^t \geqslant z_j} e^{\alpha_2 \cdot v^j \cdot (y_i^t - z_j)},$$

其参数为 $\alpha_1,\ \alpha_2,\ z_j > 0$，定义如上。与 Φ_j （以及 Ψ_j）不同，后者的值始终至少为 n （因为每个箱的贡献至少为 1）， Φ_j （以及 Ψ_j）可能小至 0。此外，$\dot\Phi_j^t \leqslant \Phi_j^t \leqslant \dot\Phi_j^t + n$ （以及 $\dot\Psi_j^t \leqslant \Psi_j^t \leqslant \dot\Psi_j^t + n$）。

轮次上的势能：我们还定义了按轮次 $r \geqslant 0$ 编号的 Φ_j 和 Ψ_j 势函数版本（注意，第一轮的起始步骤可能不等于 0）：

$$\overline{\Psi}_j^r := \sum_{i=1}^n e^{\alpha_1 \cdot v^j \cdot (y_i^{r,0} - z_j)^+}, \quad \text{and} \quad \overline{\Phi}_j^r := \sum_{i=1}^n e^{\alpha_2 \cdot v^j \cdot (y_i^{r,0} - z_j)^+}.$$

类似地，我们定义轮次上的部分势函数

$$\ddot\Psi_j^r := \sum_{i: y_i^t \geqslant z_j} e^{\alpha_1 \cdot v^j \cdot (y_i^{r,0} - z_j)}, \quad \text{and} \quad \ddot\Phi_j^r := \sum_{i: y_i^t \geqslant z_j} e^{\alpha_2 \cdot v^j \cdot (y_i^{r,0} - z_j)}.$$

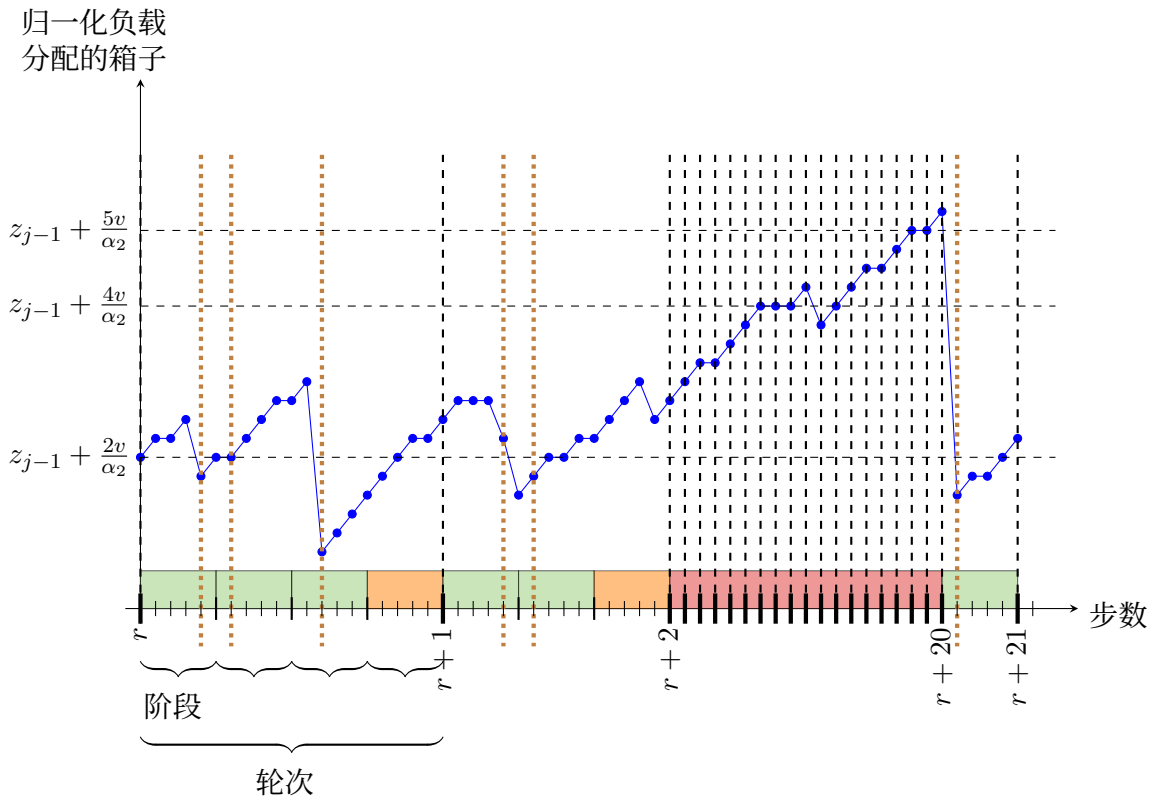

图 4： 折叠过程的阶段和轮次示意图。棕色线条表示在某一阶段中首次采样到轻箱的子步骤（如轮次 $r+1$ 的第二阶段所示，这并不一定意味着该箱将用于分配或缓存）。如图所示，只有经过一长串红色轮次后，才有可能向归一化负载高于 $z_{j-1}+\frac{5v}{\alpha_2}$ 的箱进行分配。

—

与先前应用的差异：这些势函数的形式与 [19, 第 6] 节中用于 $\Theta(\log\log n)$ 分位数的势函数，以及 [20, 第 6] 节中用于 $\Theta(1)$ 加性噪声的势函数相似。然而，分析方式有所不同，因为只有在考虑足够多的步骤（例如，过程的折叠版本）时，势函数才会在期望意义上下降。例如，从缓存负载至少为 z_j+1 的状态开始，势函数 Φ_j 在一步之后的期望值会增加。考虑由多个球组成的轮次会带来若干挑战：

- **问题 1：**每轮我们最多可以分配 k_j 个球，这可能达到 $\Omega(n^{\varepsilon})$。这意味着，从具有 $\Phi_j^{r_0} = \mathcal{O}(e^{\frac{1}{2}\log^3 n})$ 的轮次 r_0 开始，即使具有类似于 [19, 20]，中的下降不等式，例如，

$$\mathbf{E}\left[\overline{\Phi}_j^{r+1} \,\middle|\, \mathfrak{F}^r, \overline{\Phi}_{j-1}^r \leqslant Cn\right] \leqslant \overline{\Phi}_j^r \cdot \left(1-\frac{1}{n}\right)+2,$$

  我们可能需要 $\Omega(n\log^3 n)$ 轮，才能证明势函数的期望值变为 $\mathcal{O}(n)$。在这些轮次中，我们可能分配 $\Omega(n^{1+\varepsilon}\cdot\log^3 n)$ 个球（$k_j\cdot\frac{v}{\alpha}$ 每轮），因此整个分析区间的长度将需要达到 $\omega(n\cdot\mathrm{polylog}(n))$。然而，在分层归纳的每一层（引理 5.13）中，都无法容忍 $\mathrm{poly}(n)$ 的概率下降，因为我们有 $j_{\max}=\Theta(\log\log n)$ 层。

  解决方案：定义势函数 $\ddot{\Phi}_j$，只考虑归一化负载至少为 z_j 的箱子。对于这个势函数，我们可以证明：

$$\mathbf{E}\left[\ddot{\Phi}_j^{r+1} \,\middle|\, \mathfrak{F}^r, \overline{\Phi}_{j-1}^r \leqslant 2Cn\right] \leqslant \ddot{\Phi}_j^r \cdot \left(1-\frac{e^{v^{j+1}}}{n}\right)+e^{-v^j}.$$

  这意味着，从一个满足 $\Phi_j^{r_0}=\mathcal{O}(e^{\frac{1}{2}\log^3 n})$ 的轮次 r_0 开始，我们只需等待 $n\cdot e^{-v^{j+1}}\cdot\log^3 n$ 轮次，即至多 $n\cdot\log^6 n$ 步（至多为 $\mathcal{O}(n\cdot\mathrm{polylog}(n))$），势函数就会变为 $\mathcal{O}(n)$。

- **问题 2：**遗憾的是，对于稳定化，也就是证明在 $n\cdot\mathrm{polylog}(n)$ 步内，以高概率有 $\Phi_j^s=\mathcal{O}(n)$，我们不能只使用部分势函数 Φ_j。原因是，在单个轮次中，势函数可能变化 $\Omega(n)$，因此无法应用有界差分不等式。

  考虑有 $n\cdot e^{-v^j}$ 个箱子（对于 $j=1$）且其负载为 $z_j+\frac{k_j}{n}$ 的情况。此时，在单个轮次中，我们可以只在轻箱中分配 k_j 个球，使得势函数变为 0。因此，对于 $j=1$，势函数会减少 $n\cdot e^{-v^j}\cdot e^{\alpha_2 k_j/n}=\Omega(n)$。这意味着我们无法再应用集中不等式，因为有界差分条件不够强。

  解决方案：在这一部分的分析中，我们使用完整势函数 Φ_j 和停正时间，以确保每次应用集中不等式时分配的球的数量至多为 $n/\log^2 n$。这样，我们便可以应用平滑性论证（断言 5.9），证明在区间的每一步中，势函数均为 $\mathcal{O}(n)$。

5.1 引理 5.13 的证明纲要。

现在，我们将概述引理 5.13 证明中的主要技术步骤（关键步骤的示意图见图 5）。

首先，固定任意 $1\leqslant j\leqslant j_{\max}-1$。然后，归纳假设保证对于所有步数 $u\in[\beta_{j-1}, t+n\log^8 n]$，都有 $\Phi_{j-1}^u=\mathcal{O}(n)$，其中 $\beta_{j-1}:=t+2(j-1)n\log^6 n$。

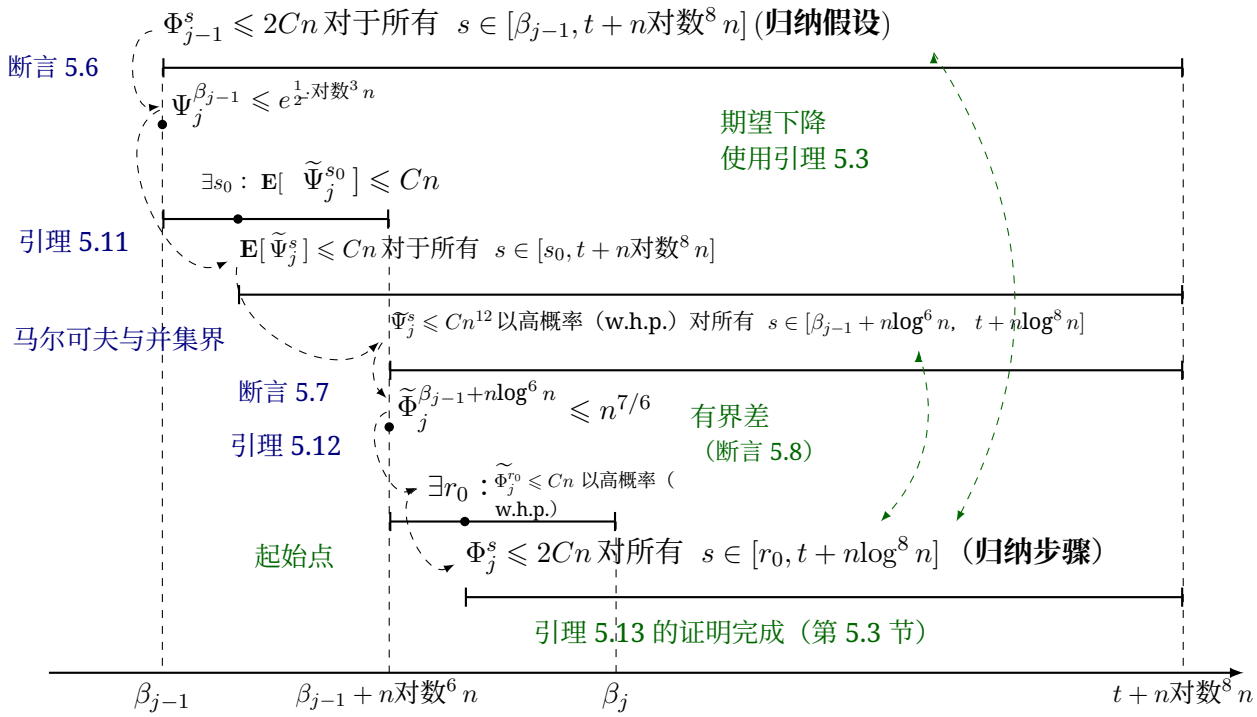

图 5： 引理 5.13 的证明概要。蓝色结果见第 5.2 节；绿色结果则用于在第 5.3 节应用集中不等式（定理 A.10）。

恢复。 当 $\Phi_{j-1}^{\beta_{j-1}} = \mathcal{O}(n)$ 时，通过简单估计可得 $\Psi_j^{\beta_{j-1}} \leqslant e^{\frac{1}{2}\cdot\log^3 n}$（断言 5.6）。利用部分势函数 Ψ_j 的下降不等式（引理 5.3），可得对于任意步骤 $u \in [\beta_{j-1} + n\log^6 n,\ t + n\log^8 n]$，都有 $\mathbf{E}[\Psi_j^u] \leqslant Cn$（引理 5.11）。只需使用马尔可夫不等式和并集界，即可推出对于所有步数 $u \in [\beta_{j-1} + n\log^6 n,\ t + n\log^8 n]$，都有 $\Psi_j^u \leqslant Cn^{12}$。根据两个势函数之间的简单关系，这意味着 $\ddot{\Phi}_j^u \leqslant \Phi_j^u \leqslant n^{7/6}$（断言 5.7）。现在利用部分势函数 $\ddot{\Phi}_j$ 的下降不等式（引理 5.3），可保证以高概率，对于某个单独的步骤 $u \in [\beta_{j-1}, \beta_j]$，有 $\Phi_j^u \leqslant Cn$（引理 5.12）。

稳定化。 为了得到适用于所有步骤 $s \in [\beta_{i-1}, \beta_j]$ 的更强结论，我们将使用集中不等式。关键在于，对于任意满足 $\Psi_j^r \leqslant Cn^{12}$ 的轮次 r，绝对差 $|\Phi_j^{r+1} - \Phi_j^r|$ 至多为 $n^{1/3}$，因为 $14\frac{\alpha_2}{\alpha_1} < 1/3$（断言 5.8）。这一点至关重要，这样从 [9] 到 Φ_j 应用超鞅集中界定理 A.10 后，利用平滑论证（断言 5.9），便可得到对所有步骤 $s \in [\beta_j, t + n\log^8 n]$ 都成立的 $\Phi_j^s = \mathcal{O}(n)$ 保证。

由于证明引理 5.13 需要若干断言和引理，本节余下部分进一步划分如下：

1. 对部分势函数和完整势函数（期望）下降量的分析。（第 5.2.1 节）
2. 涉及一个或两个势函数的（确定性）不等式。（第 5.2.2 节）
3. 基于这些下降结果的辅助（概率）引理。（第 5.2.3 节）

之后，我们将在第 5.3 节完成引理 5.13 的证明。最后，我们将在第 5.4 节完成主定理定理 1.1 的证明。

5.2 引理 5.13 的证明准备

我们定义如下关键事件 $\mathcal{K}_{j-1}^{s}$，适用于任意步骤 s 和 $1 \leqslant j \leqslant j_{\max}-1$。该事件将在分析中频繁使用：

$$\mathcal{K}_{j-1}^{s} := \{\Phi_{j-1}^{s} \leqslant Cn\}.$$

在接下来的论证中，我们将经常使用势函数的“杀死”版本，即假设在所有步骤 $\rho \in [\beta_{j-1}, s]$ 中，$\mathcal{K}_{j-1}^{\rho}$ 均成立：

$$\widetilde{\Phi}_{j}^{s} := \Phi_{j}^{s} \cdot \mathbf{1}_{\cap_{\rho\in[\beta_{j-1},s]}\mathcal{K}_{j-1}^{\rho}} \text{ and } \widetilde{\Psi}_{j}^{s} := \Psi_{j}^{s} \cdot \mathbf{1}_{\cap_{\rho\in[\beta_{j-1},s]}\mathcal{K}_{j-1}^{\rho}}.$$

此外，令 $T_s(r)$ 表示从步骤 s 开始对应于轮次 r 的步骤。

5.2.1 部分势函数和完整势函数下降的分析

现在我们将证明，当 $\Phi_{j-1}^{r} \leqslant Cn$ 成立（由 $\mathcal{K}_{j-1}^{T(r)}$ 推出）时，在 $(a,\ b)$-偏置采样分布下，Memory 过程中的重箱数量有一个上界：

引理 5.2。对于任意 $1 \leqslant j \leqslant j_{\max}-1$ 和任意轮次 $r \geqslant 0$，若 $\Phi_{j-1}^{r} \leqslant Cn$，则满足 $y^{r,}{}_{i}{}^{0} \geqslant z_{j-1} + \frac{2v}{\alpha_2}$ 的箱子 i 的数量至多为

$$\frac{1}{2b} n \cdot e^{-v^j},$$

并且在第 r 轮的任意子步骤 s 中，满足 $y^{r,}{}_{i}{}^{s} \geqslant z_{j-1} + \frac{2v}{\alpha_2}$ 的箱子数量至多为

$$\frac{1}{b} n \cdot e^{-v^j}.$$

证明。回顾一下，$z_j = \frac{5v}{\alpha_2} \cdot j$。任意满足负载 $y^{r,}{}_{i}{}^{0} \geqslant z_{j-1} + \frac{2v}{\alpha_2}$ 的箱 $i \in [n]$ 对 $\Phi_{j-1,\ i}^{r}$ 的贡献上界为

$$\begin{aligned}
\overline{\Phi}_{j-1,i}^{r} &\geqslant e^{\alpha_2 \cdot v^{j-1} \cdot (z_{j-1} + \frac{2v}{\alpha_2} - z_{j-1})^{+}} \\
&= e^{\alpha_2 \cdot v^{j-1} \cdot \frac{2v}{\alpha_2}} \\
&= e^{2v^j}.
\end{aligned}$$

因此，若 $\left\{\Phi_{j-1}^{r} \leqslant Cn\right\}$ 成立，则满足负载 $y^{r,}{}_{i}{}^{0} \geqslant z_{j-1} + \frac{2v}{\alpha_2}$ 的箱子数量至多为

$$Cn \cdot e^{-2v^j} \leqslant \frac{1}{2b} n \cdot e^{-v^j},$$

利用 $j \geqslant 1$ 时 $e^{-v^j} \geqslant e^{-v}$，以及 $v \geqslant \log(2Cb)$。

注意，在单轮中，我们至多向 $k_j \cdot \frac{v}{\alpha_2}$ 个箱子分配。因此，在第 r 轮的任意子步骤 s 中，满足 $y^{r,}{}_{i}{}^{s} \geqslant z_{j-1} + \frac{2v}{\alpha_2}$ 的箱子数量至多可以增长到

$$\frac{1}{2b} n \cdot e^{-v^j} + k_j \cdot \frac{2v}{\alpha_2} \leqslant \frac{1}{b} n \cdot e^{-v^j},$$

对于足够大的 n，利用 $e^{-v^j} \geqslant e^{-v^{j_{\max}}} \geqslant e^{-\frac{\alpha_2}{2v}\log n} \geqslant n^{-1/2}$（因为 $\alpha_2 \leqslant 1$ 和 $v \geqslant 1$），以及 $k_j = e^{v^{j+1}} \cdot \log^3 n \leqslant e^{v^{j_{\max}}} \cdot \log^3 n \leqslant e^{\frac{\alpha_2}{2v}\log n} \cdot \log^3 n \leqslant n^{1/7}$（因为 $\alpha_2 \leqslant 4$ 和 $v \geqslant 36$）。 □

引理 5.2 中关于重箱数量的这些界，使我们能够证明归纳法中跨层的势函数下降。

引理 5.3。 对于任意 $1 \leqslant j \leqslant j_{\max}-1$ 和任意轮次 $r \geqslant 0$，我们有

$$\mathbf{E}\left[\ \ddot{\Phi}_j^{r+1} \ \middle|\ \mathfrak{F}^r, \mathcal{K}_{j-1}^r\right] \leqslant \ddot{\Phi}_j^r \cdot \left(1 - \frac{e^{v^{j+1}}}{n}\right) + e^{-v^j/2},$$

and

$$\mathbf{E}\left[\ \ddot{\Psi}_j^{r+1} \ \middle|\ \mathfrak{F}^r, \mathcal{K}_{j-1}^r\right] \leqslant \ddot{\Psi}_j^r \cdot \left(1 - \frac{e^{v^{j+1}}}{n}\right) + e^{-v^j/2}.$$

证明。 根据引理 5.2，负载为 y^r 的箱子数量 $\ell \in [n]$，$\ell^{\,s} \geqslant z_{j-1} + \frac{2v}{\alpha_2}$ 在轮次 r 的所有子步骤 s 中至多为 $\frac{1}{b}n \cdot e^{-v^j}$。

现在我们将分析箱子 $i \in [n]$ 对势函数 $\ddot{\Phi}_j^{r+1}$ 的期望贡献。

情形 1 [$y_i^{r,0} > z_j$]: 当我们抽取一个满足 $y^{r,\ell^{\,s}} \geqslant z_{j-1} + \frac{2v}{\alpha_2}$ 的重箱子 $\ell \in [\sout{n}]$ 时（发生的概率至多为 $\frac{b}{n} \cdot \frac{n}{b} \cdot e^{-v^j} = \sout{e^{-v^j}}$），平均值会增加 $1/n$。否则，我们抽取一个轻箱子，并开始一个新的阶段。因此，由于箱子 i 是重箱，我们有

$$\begin{aligned}
&\mathbf{E}\left[\ \ddot{\Phi}_{j,i}^{r+1} \ \middle|\ \mathfrak{F}^r, \mathcal{K}_{j-1}^r\right] \\
&\quad\leqslant \ddot{\Phi}_{j,i}^r \cdot \Bigg(\underbrace{\frac{b}{n} \cdot e^{\alpha_2 v^j \cdot (1-1/n)}}_{\text{Bin } i \text{ is chosen}} + \underbrace{\left(e^{-v^j} - \frac{b}{n}\right) \cdot e^{-\alpha_2 v^j/n}}_{\text{Another heavy bin is chosen}} + \underbrace{(1-e^{-v^j}) \cdot \sum_{k=1}^{k_j} A_k}_{\text{A light bin is chosen}}\Bigg),
\end{aligned} \tag{5.2}$$

其中 A_k 是恰好执行 k 个阶段时对势函数的期望贡献。每个后续阶段开始的概率至少为 $1 - e^{-v^j} \cdot v/\alpha_2$。因此，根据支配关系，我们有（因为 $e^{-v^{j+1}} \cdot \alpha_2 \leqslant 1$）

$$p = e^{-v^{j+1}/\alpha_2} \leqslant (\ )$$

$$\sum_{k=1}^{k_j} A_k \leqslant \sum_{k=1}^{k_j} \underbrace{\frac{(1-p)^{k-1} \cdot p}{1-(1-p)^{k_j}}}_{\text{Probability of exactly } k \text{ phases}} \cdot \underbrace{e^{-kv^{j+1}/n}}_{\substack{\text{Change of } \ddot{\Phi}_{j,i}^r \text{ due} \\ \text{to change in average}}},$$

利用经过 k 个阶段后平均负载变化了 $\frac{kv}{n\alpha_2}$，且在一个阶段中我们不会向重箱分配负载，也就是说，我们不会向箱子 i 分配负载（因为它仍然是重箱）。代入 $p = e^{-v^{j+1}}$，得到

$$\sum_{k=1}^{k_j} A_k \leqslant e^{-v^{j+1}/n} \cdot e^{-v^{j+1}} \cdot \frac{1}{1-(1-e^{-v^{j+1}})^{k_j}} \cdot \sum_{k=0}^{k_j-1} \left((1-e^{-v^{j+1}}) \cdot e^{-v^{j+1}/n}\right)^k. \tag{5.3}$$

我们首先对 (5.3) 右侧的和进行上界估计：

$$\begin{aligned}
\sum_{k=0}^{k_j-1} \left((1-e^{-v^{j+1}}) \cdot e^{-v^{j+1}/n}\right)^k &\leqslant \frac{1}{1-(1-e^{-v^{j+1}}) \cdot e^{-v^{j+1}/n}} \\
&\overset{(a)}{\leqslant} \frac{1}{1-(1-e^{-v^{j+1}}) \cdot (1-\frac{1}{2} \cdot \frac{v^{j+1}}{n})} \\
&= \frac{1}{e^{-v^{j+1}}} \cdot \frac{1}{1-\frac{1}{2} \cdot \frac{v^{j+1}}{n} + \frac{1}{2} \cdot \frac{v^{j+1}}{n} \cdot e^{v^{j+1}}} \\
&\overset{(b)}{\leqslant} \frac{1}{e^{-v^{j+1}}} \cdot \frac{1}{1+\frac{1}{4} \cdot \frac{v^{j+1}}{n} \cdot e^{v^{j+1}}} \\
&\overset{(c)}{\leqslant} \frac{1}{e^{-v^{j+1}}} \cdot \left(1 - \frac{1}{8} \cdot \frac{v^{j+1}}{n} \cdot e^{v^{j+1}}\right),
\end{aligned} \tag{5.4}$$

在 (a) 中使用 $e^z \geqslant 1+\frac{1}{2}z$（对于 $|z|<1$），并使用 $v^{j+1} \leqslant v^{j_{\max}} \leqslant \log\, n = o(n)$（因为 $j+1 \leqslant j_{\max} = \log_v(\frac{\alpha_2}{2v}\log\, n)$）；在 (b) 中使用 $e^{v^{j+1}} \geqslant 2$（因为 $v^{j+1} \geqslant v \geqslant 36$）；在 (c) 中使用 $\frac{1}{1+z} \leqslant 1-\frac{z}{2}$（对于$0<z\leqslant 1$）（引理 A.3），以及 $\frac{1}{4}\cdot\frac{v^{j+1}}{n}\cdot e^{v^{j+1}} \leqslant \frac{1}{4}\cdot\log n\cdot n^{-1/2}$。

将 (5.4) 中的界应用于 (5.3)，得到

$$\begin{aligned}
\sum_{k=1}^{k_j} A_k &\leqslant e^{-v^{j+1}/n}\cdot\frac{1}{1-(1-e^{-v^{j+1}})^{k_j}}\cdot\Big(1-\frac{1}{8}\cdot\frac{v^{j+1}}{n}\cdot e^{v^{j+1}}\Big)\\
&\overset{(a)}{\leqslant} e^{-\alpha_2 v^j/n}\cdot\frac{1}{1-(1-e^{-v^{j+1}})^{k_j}}\cdot\Big(1-\frac{1}{8}\cdot\frac{v^{j+1}}{n}\cdot e^{v^{j+1}}\Big)\\
&\leqslant e^{-\alpha_2 v^j/n}\cdot\frac{1}{1-\exp\big(-e^{-v^{j+1}}\cdot k_j\big)}\cdot\Big(1-\frac{1}{8}\cdot\frac{v^{j+1}}{n}\cdot e^{v^{j+1}}\Big)\\
&\overset{(b)}{\leqslant} e^{-\alpha_2 v^j/n}\cdot\Big(1+2\exp\Big(-e^{-v^{j+1}}\cdot k_j\Big)\Big)\cdot\Big(1-\frac{1}{8}\cdot\frac{v^{j+1}}{n}\cdot e^{v^{j+1}}\Big)\\
&= e^{-\alpha_2 v^j/n}\cdot\big(1+2\exp\big(-\log^3 n\big)\big)\cdot\Big(1-\frac{1}{8}\cdot\frac{v^{j+1}}{n}\cdot e^{v^{j+1}}\Big)\\
&\overset{(c)}{\leqslant} e^{-\alpha_2 v^j/n}\cdot\Big(1-\frac{1}{9}\cdot\frac{v^{j+1}}{n}\cdot e^{v^{j+1}}\Big),
\end{aligned}\tag{5.5}$$

其中，在 (a) 中，$\alpha_2 \leqslant 1$ and $v \geqslant 1$；在 (b) 中，由引理 A.4，$\frac{1}{1-\varepsilon} \leqslant 1+2\varepsilon$ for $0<\varepsilon\leqslant 1/2$；在 (c) 中，$2\exp\big(-\log^3 n\big) = o(n^{-2})$，而 $\frac{v^{j+1}}{n}\cdot e^{v^{j+1}} \geqslant \frac{1}{n}$。

回到 (5.2)，我们有

$$\begin{aligned}
\mathbf{E}\Big[\,\ddot{\Phi}_{j,i}^{r+1}\,\Big|\,\mathfrak{F}^r\,\Big] &\leqslant \ddot{\Phi}_{j,i}^{r}\cdot e^{-\alpha_2 v^j/n}\cdot\Big(\frac{b}{n}\cdot e^{\alpha_2 v^j}+\Big(e^{-v^j}-\frac{b}{n}\Big)+(1-e^{-v^j})\cdot e^{\alpha_2 v^j/n}\cdot\sum_{k=1}^{\infty}A_k\Big)\\
&= \ddot{\Phi}_{j,i}^{r}\cdot e^{-\alpha_2 v^j/n}\cdot\Big(1+\frac{b}{n}\cdot(e^{\alpha_2 v^j}-1)+(1-e^{-v^j})\cdot\Big(e^{\alpha_2 v^j/n}\cdot\sum_{k=1}^{\infty}A_k-1\Big)\Big)\\
&\overset{(5.5)}{\leqslant} \ddot{\Phi}_{j,i}^{r}\cdot e^{-\alpha_2 v^j/n}\cdot\Big(1+\frac{b}{n}\cdot e^{\alpha_2 v^j}-(1-e^{-v^j})\cdot\Big(\frac{1}{9}\cdot\frac{v^{j+1}}{n}\cdot e^{v^{j+1}}\Big)\Big)\\
&\overset{(a)}{\leqslant} \ddot{\Phi}_{j,i}^{r}\cdot e^{-\alpha_2 v^j/n}\cdot\Big(1+\frac{b}{n}\cdot e^{\alpha_2 v^j}-\frac{2b}{n}\cdot e^{v^{j+1}}\Big)\\
&\overset{(b)}{\leqslant} \ddot{\Phi}_{j,i}^{r}\cdot\Big(1-\frac{e^{v^{j+1}}}{n}\Big),
\end{aligned}$$

利用 (a) 中的 $1-e^{-v^j} \geqslant 1-e^{-v} \geqslant 1/2$ 和 $v^j \geqslant v \geqslant 36b$，以及 (b) 中的 $\alpha_2 \leqslant 1$ 和 $b \geqslant 1$。**情况 2 [$y_i^{r,0} \in (z_j-1+\frac{1}{n},\ z_j]$]：** 对于这种情况，我们将讨论所有此类箱子 i 的总贡献。在第 r 轮中，我们最多只能向 i 分配一个球，因为 $y^r_{\cdot,\,i}{}^0 > z_{j-1}+\frac{2v}{\alpha_2}$（这意味着我们只能在第一个子步骤中向这个箱子分配球）。在这种情况下，该箱子的势函数 $\Phi^r_{j,\,i}$从 0 增加到至多 $e^{\alpha_2 v^j}$。只有在我们选中一个负载为 $y^r_\ell \geqslant z_{j-1}+\frac{2v}{\alpha_2}$ 的重箱 $\ell\in[n]$ 时才会发生这种情况，因此其概率至多为 e^{-v^j}。因此，总体而言，我们有

$$\begin{aligned}
\sum_{\ell:y_\ell^{r,0}\in(z_j-1+\frac{1}{n},z_j]}\mathbf{E}\Big[\,\ddot{\Phi}_{j,\ell}^{r+1}\,\Big|\,\mathfrak{F}^r\,\Big] &\leqslant e^{\alpha_2 v^j}\cdot e^{-v^j}\\
&\overset{(a)}{\leqslant} e^{-v^j/2}\\
&\overset{(b)}{=} \sum_{\ell:y_\ell^{r,0}\in(z_j-1+\frac{1}{n},z_j]}\ddot{\Phi}_{j,\ell}^{r}\cdot\Big(1-\frac{e^{v^{j+1}}}{n}\Big)+e^{-v^j/2},
\end{aligned}$$

不等式 (a) 成立是因为 $\alpha_2 \leqslant 1/2$ 和 (b)，因为对于每个 ℓ 满足 $y^{r,0}_\ell \in$，有 $\ddot{\Phi}^r_{j,\ell} = 0$（$z_j - 1 + \frac{1}{n}$，$z_j$]。**情况 3(a) [**$y^{r,0}_i \leqslant z_{j-1} + \frac{2v}{\alpha_2}$**]：** 在第 r 轮中，我们可以向箱子 i 分配多个球，但其负载始终至多为 $z_{j-1} + \frac{4v}{\alpha_2} < z_j$，因此 $\ddot{\Phi}^{r+1}_{j,i} = \ddot{\Phi}^r_{j,i} = 0$。

情况 3(b) [$y^{r,0}_i \in (z_{j-1} + \frac{2v}{\alpha_2}, z_j - 1 + \frac{1}{n}$]**：** 在第 t 轮，我们最多只能向 i 分配一个球，因为 $y^{r,0}_i > z_{j-1} + \frac{2v}{\alpha_2}$（这意味着我们只能在第一步向这个箱分配）；因此，与情况 3(a) 一样，$\ddot{\Phi}^{r+1}_{j,i} = \ddot{\Phi}^r_{j,i} = 0$。

因此，将三种情况合并起来，我们就得到了关于 Φ_j 的结论。同样的分析也适用于 Ψ_j，因为我们只用到了 $\alpha_2 \leqslant 1/2$（这对 α_1 也成立）和 $\alpha_1 \geqslant \alpha_2$。 □

这意味着，当势函数较大时，我们会得到乘法意义上的期望下降 p.

推论 5.4. 对于任意 $1 \leqslant j \leqslant j_{\max} - 1$ 和任意第 $r \geqslant 0$ 轮，我们有

$$\mathbf{E}\left[\, \ddot{\Psi}^{r+1}_j \,\middle|\, \mathfrak{F}^r, \ddot{\Psi}^r_j \geqslant 2n \,\right] \leqslant \ddot{\Psi}^r_j \cdot \Big(1 - \frac{e^{v^{j+1}}}{2n}\Big),$$

and

$$\mathbf{E}\left[\, \ddot{\Phi}^{r+1}_j \,\middle|\, \mathfrak{F}^r, \ddot{\Phi}^r_j \geqslant 2n \,\right] \leqslant \ddot{\Phi}^r_j \cdot \Big(1 - \frac{e^{v^{j+1}}}{2n}\Big).$$

证明。 利用引理 5.3,

$$\begin{aligned}
\mathbf{E}\left[\, \ddot{\Psi}^{r+1}_j \,\middle|\, \mathfrak{F}^r, \ddot{\Psi}^r_j \geqslant 2n \,\right] &\leqslant \ddot{\Psi}^r_j \cdot \Big(1 - \frac{e^{v^{j+1}}}{n}\Big) + e^{-v^j/2} \\
&= \ddot{\Psi}^r_j \cdot \Big(1 - \frac{e^{v^{j+1}}}{2n}\Big) - \ddot{\Psi}^r_j \cdot \frac{e^{v^{j+1}}}{2n} + e^{-v^j/2} \\
&\leqslant \ddot{\Psi}^r_j \cdot \Big(1 - \frac{e^{v^{j+1}}}{2n}\Big) - e^{v^{j+1}} + e^{-v^j/2} \\
&\leqslant \ddot{\Psi}^r_j \cdot \Big(1 - \frac{e^{v^{j+1}}}{2n}\Big).
\end{aligned}$$

以完全相同的方式处理 Φ_j，即可得到第二个结论。 □

对于分层归纳步骤（引理 5.13）中的稳定阶段，我们使用以下关于完整 Φ_j 势函数的下降不等式:

引理 5.5。 对于任意 $1 \leqslant j \leqslant j_{\max} - 1$ 和任意轮次 $r \geqslant 0$，有

$$\mathbf{E}\left[\, \overline{\Phi}^{r+1}_j \,\middle|\, \mathfrak{F}^r, \mathcal{K}^r_{j-1} \,\right] \leqslant \overline{\Phi}^r_j \cdot \Big(1 - \frac{1}{n}\Big) + 2,$$

and

$$\mathbf{E}\left[\, \overline{\Phi}^{r+1}_j \,\middle|\, \mathfrak{F}^r, \mathcal{K}^r_{j-1}, \overline{\Phi}^r_j \geqslant 4n \,\right] \leqslant \overline{\Phi}^r_j \cdot \Big(1 - \frac{1}{2n}\Big).$$

证明。 对于第一个陈述，分析过程与引理 5.3 类似，其中情况 1 保持不变:

情况 2 [$y^{r,0}_i \in (z_j - 1 + \frac{1}{n}, z_j$]**：** 对于这种情况，我们将再次讨论所有此类箱子 i 的总贡献。在第 r 轮中，我们最多只能向 i 分配一个球，因为 $y^{r,0}_i > z_{j-1} + \frac{2v}{\alpha_2}$（这意味着我们只能在第一个步骤中向这个箱子分配球）。在这种情况下，该箱子的势函数 $\ddot{\Phi}^r_{j,i}$ 从 0 增加到至多 $e^{\alpha_2 v^j}$。只有当我们选中一个负载为 $y^r_\ell \geqslant z_{j-1} + \frac{2v}{\alpha_2}$ 的箱子 $\ell \in [n]$ 时才会发生这种情况，因此其概率至多为 e^{-v^j}。因此我们有

$$\sum_{i : y^{r,0}_i \in (z_j - 1 + \frac{1}{n}, z_j]} \mathbf{E}\left[\, \overline{\Phi}^{r+1}_{j,i} \,\middle|\, \mathfrak{F}^r \,\right] \leqslant \sum_{i : y^{r,0}_i \in (z_j - 1 + \frac{1}{n}, z_j]} \overline{\Phi}^r_{j,i} + e^{\alpha_2 v^j} \cdot e^{-v^j}$$

$$\overset{(a)}{\leqslant} \sum_{i:y_i^{r,0}\in(z_j-1+\frac{1}{n},z_j]} \overline{\Phi}_{j,i}^r + 1$$
$$\overset{(b)}{=} \sum_{i:y_i^{r,0}\in(z_j-1+\frac{1}{n},z_j]} \left(\overline{\Phi}_{j,i}^r \cdot \left(1-\frac{1}{n}\right)+\frac{1}{n}\right)+1,$$

其中，不等式 (a) 成立是因为 $\alpha_2 \leqslant 1/2$，而 (b) 成立是因为对于每个满足 $y^{r,0}_i \in$ 的 i，$\overline{\Phi}^r_{j,i}=1$ $(z_j-1+\frac{1}{n},\ z_j]$。**情况 3(a)** **[$y_i^{r,0} \leqslant z_{j-1}+\frac{2v}{\alpha_2}$]:** 在第 r 轮中，我们可以向箱子 i 分配多个球，但其负载始终至多为 $z_{j-1}+\frac{4v}{\alpha_2} \leq z_j$，因此 $\Phi_j^{r+1,i} = \Phi_j^{r,i}=1$。因此，

$$\mathbf{E}\left[\ \overline{\Phi}_{j,i}^{r+1}\ \middle|\ \mathfrak{F}^r\ \right] = \overline{\Phi}_{j,i}^r \cdot \left(1-\frac{1}{n}\right)+\frac{1}{n}.$$

情况 3(b) **[$y_i^{r,0} \in (z_{j-1}+\frac{2v}{\alpha_2},\ z_j-1+\frac{1}{n}]$]:** 在第 r 轮中，我们最多只能向 i 分配一个球，因为 $y^{r,i,0} > z_{j-1}+\frac{2v}{\alpha_2}$（这意味着我们只能在第一步向这个箱分配球）；因此，与情况 3(a) 一样，$\Phi_j^{r+1,i}=\Phi_j^{r,i}=1$。因此，

$$\mathbf{E}\left[\ \overline{\Phi}_{j,i}^{r+1}\ \middle|\ \mathfrak{F}^r\ \right] = \overline{\Phi}_{j,i}^r \cdot \left(1-\frac{1}{n}\right)+\frac{1}{n}.$$

因此，对这三种情况进行汇总，我们就得到了关于 Φ_j 的结论。

对于第二个结论，

$$\mathbf{E}\left[\ \overline{\Phi}_j^{r+1}\ \middle|\ \mathfrak{F}^r, \mathcal{K}_{j-1}^r, \overline{\Phi}_j^r \geqslant 4n\ \right] \leqslant \overline{\Phi}_j^r \cdot \left(1-\frac{1}{2n}\right) - \overline{\Phi}_j^r \cdot \frac{1}{2n} + 2 \leqslant \overline{\Phi}_j^r \cdot \left(1-\frac{1}{2n}\right). \qquad \square$$

5.2.2 势函数之间的确定性关系

我们总结了关于势函数 Φ_j^s 和 Ψ_j^s 的若干基本事实，它们与参数由 (5.1) 固定的 $(a,\ b)$-偏置采样分布上的 Memory 过程有关。

断言 5.6。对于任意 $1 \leqslant j \leqslant j_{\max}-1$ 以及任意步骤 $s \geqslant 0$,

$$\Phi_{j-1}^s \leqslant 2Cn \quad \Rightarrow \quad \Psi_j^s \leqslant \exp\left(\frac{1}{2}\cdot \log^3 n\right).$$

证明。假设 $\Phi_{j-1}^s \leqslant 2Cn$，则对于任意箱 $i \in [n]$,

$$\Phi_{j-1,i}^s = \exp\left(\alpha_2 \cdot v^{j-1} \cdot \left(y_i^s - \frac{5v}{\alpha_2}\cdot(j-1)\right)^+\right) \leqslant 2Cn.$$

重新排列后可得

$$y_i^s \leqslant \frac{\log(2Cn)}{\alpha_2}\cdot v^{-(j-1)} + \frac{5v}{\alpha_2}\cdot(j-1) \leqslant \log^{1.5} n,$$

当 n 足够大时，因为 $j \leqslant j_{\max} = \mathcal{O}(\log\log n)$。

因此，我们可以将 Ψ_j^s 的上界设为

$$\Psi_j^s \leqslant n \cdot \exp\left(\alpha_1 \cdot v^j \cdot \log^{1.5} n\right) \leqslant \exp\left(\frac{1}{2}\cdot \log^3 n\right),$$

当 n 足够大时，利用 $v^j \leqslant v^{j_{\max}} \leqslant \log n$。 $\square$

下一个断言界定了 Φ_j，前提是 $\Psi_j = \text{poly}(n)$。

断言 5.7。对于任意 $1 \leqslant j \leqslant j_{\max}-1$ 以及任意步骤 $s \geqslant 0$，如果 $\Psi_j^s \leqslant Cn^{12}$，那么对于任意箱子 $i \in [n]$，我们都有 $\Phi_{j,\,i}^s \leqslant n^{1/6}$ 因此对所有箱子进行汇总可得 $\Phi_j^s \leqslant n^{7/6}$。

证明。我们首先证明，当 $\Psi_j^s \leqslant Cn^{12}$ 时，对任意箱 $i \in [n]$，

$$y_i^s \leqslant z_j + \frac{14}{\alpha_1} \cdot (\log n) \cdot v^{-j}.$$

反设存在某个箱 i，使得 $y_i^s > z_j + \frac{14}{\alpha_1} \cdot (\log n) \cdot v^{-j}$，则可得$\Psi_j^s > \exp(\alpha_1 \cdot \frac{14}{\alpha_1} \cdot \log n) = n^{14}$，这就产生了矛盾。

接下来，我们对任意箱 $i \in [n]$ 的贡献求上界，

$$\begin{aligned}\Phi_{j,i}^s &= \exp\bigl(\alpha_2 \cdot v^j \cdot \bigl(y_i^s - z_j\bigr)^+\bigr) \\ &\leqslant \exp\left(\alpha_2 \cdot v^j \cdot \frac{14}{\alpha_1} \cdot (\log n) \cdot v^{-j}\right) \\ &= \exp\left(\frac{14 \cdot \alpha_2}{\alpha_1} \cdot \log n\right) = n^{1/6}, \end{aligned} \tag{5.6}$$

因为 $\alpha_1 = 6 \cdot 14 \cdot \alpha_2$。因此，对所有箱子求和可得，

$$\Phi_j^s \leqslant n \cdot n^{1/6} = n^{7/6}.$$ □

下面的断言对于应用引理 5.13 中的集中不等式至关重要，因为第三个结论给出了 Φ^r 的最大加性变化上界（假设 $\Psi^r = \mathrm{poly}(n)$）：

断言 5.8。对于任意 $1 \leqslant j \leqslant j_{\max}-1$ 以及任意轮次 $r \geqslant 0$，若 $\Psi_j^r \leqslant Cn^{12}$，则 $|\Phi_j^{r+1} - \Phi_j^r| \leqslant n^{1/3}$。

证明。我们首先推导 Φ_j^{r+1} 的下界和上界。根据断言 5.7，在任意满足 $\Psi_j^r \leqslant Cn^{12}$ 的轮次 r 中，对于每个箱都有 $\Phi_{j,\,i}^r \leqslant n^{1/6}$。现在来看上界，注意在任意轮次 $r+1$ 中，我们最多只能向一个重箱分配。设 $i \in [n]$ 为该重箱，则

$$\overline{\Phi}_j^{r+1} \leqslant \overline{\Phi}_j^r + \overline{\Phi}_{j,i}^r \cdot \exp\bigl(\alpha_2 \cdot v^j\bigr) \leqslant \overline{\Phi}_j^r + n^{1/6} \cdot n^{1/6} = \overline{\Phi}_j^r + n^{1/3},$$

利用 $v^j \leqslant v^{j_{\max}} \leqslant \log n$，根据 (5.1) 并应用 (5.6)。对于下界，我们悲观地假设，在轮次 $r+1$ 的每个步骤中，所有箱负载都减少 $1/n$。因此，由于步骤数至多为 $k_j \cdot \frac{v}{\alpha_2}$

$$\overline{\Phi}_j^{r+1} \geqslant \overline{\Phi}_j^r \cdot \exp\left(-\frac{v^{j+1}}{n} \cdot k_j\right) \geqslant \overline{\Phi}_j^r \cdot \left(1 - \frac{v^{j+1}}{n} \cdot k_j\right) \geqslant \Phi_j^r - \frac{n \cdot n^{1/6}}{n} \cdot (\log n) \cdot n^{1/7} \geqslant \overline{\Phi}_j^r - n^{1/3},$$

利用 $e^x \geqslant 1 + x$（对于任意 x），$v^{j+1} \leqslant v^{j_{\max}} \leqslant \log n$，根据断言 5.7，对于 $s = T(r)$ 有 $\Phi_j^r \leqslant n \cdot n^{1/6}$，且 $k_j \leqslant n^{1/7}$。结合这两个界，即得所述结论。 □

下一个断言是一个简单的“平滑性”论证，说明势函数在 $n/\log^2 n$ 步数内不会快速下降。推导很简单，依据是平均负载在这些步数内的变化不超过 $1/\log^2 n$。

断言 5.9. 对于任意 $1 \leqslant j \leqslant j_{\max}-1$、任意步骤 $s \geqslant 0$ 以及任意步骤 $u \in [s, s + n/\log^2 n]$，均有 $\Phi_j^u \geqslant 0.99 \cdot \Phi_j^s$。

证明。经过 $u - s$ 步数后，归一化负载最多下降 $\frac{u-s}{n} \leqslant \frac{1}{\log^2 n}$。因此，对于任意箱 $i \in [n]$，

$$\begin{aligned}\Phi_{j,i}^u &= e^{\alpha_2 \cdot v^j \cdot (y_i^u - z_j)^+} \geqslant e^{\alpha_2 \cdot v^j \cdot (y_i^s - \frac{u-s}{n} - z_j)^+} \geqslant e^{\alpha_2 \cdot v^j \cdot (y_i^s - z_j)^+ - \alpha_2 \cdot v^j \cdot \frac{1}{\log^2 n}} = \Phi_{j,i}^s \cdot e^{-\frac{\alpha_2 \cdot v^j}{\log^2 n}} \\ &\geqslant \Phi_{j,i}^s \cdot e^{-o(1)} \geqslant 0.99 \cdot \Phi_{j,i}^s,\end{aligned}$$

对于足够大的 n，利用 $v^j \leqslant v^{j_{\max}} \leqslant \log n$ 和 $\alpha_2 \leqslant 1$。对所有箱子求和后，即得结论。

□

下一个结论将第 5 节开头定义的完整势函数与部分势函数联系起来 .

断言 5.10. 对于任何包含一个或多个阶段的轮次 $r \geqslant 0$，对于轮次 r的任意子步骤 $s \geqslant 0$，都有(i) $\Phi^{r}{}_{,j}{}^{s} \leqslant \Phi_{j}^{r}$, (ii) $\overline{\Psi}^{r}{}_{,j}{}^{s} \leqslant \Psi_{j}^{r}$, (iii) $\dot{\Phi}^{T}{}_{j}^{(r+s)} \leqslant \ddot{\Phi}_{j}^{r)}$以及$(iv)$ $\dot{\Psi}^{T}{}_{j}^{(r+s)} \leqslant \ddot{\Psi}_{j}^{r)}$。

证明. 对于一轮中的所有子步骤，我们都分配到一个不会影响势函数的箱中。因此，势函数只能减小（因为平均负载发生了变化）。 □

5.2.3 势函数上的辅助概率引理（恢复）

~第一个引理证明，对于所有轮次，$\Psi_{j}^{s} = \Psi_{j}^{s} \cdot \mathbf{1}_{\cap_{r\in[\beta_{j-1},s]}\mathcal{K}_{j-1}^{r}}$ 的期望值都很小，为 $s \geqslant \beta_{j-1} + n\log^{6} n$。

引理 5.11。对于任意 $1 \leqslant j \leqslant j_{\max} - 1$，对任意步骤 $s \geqslant \beta_{j-1} + n\log^{6} n$，都有 ~$\mathbf{E}[\Psi_{j}^{s}] \leqslant Cn$。

证明。由引理 5.3，对于从步骤 β_{j-1}开始计数的任意轮次 $r \geqslant 0$，

$$\mathbf{E}\left[\ \ddot{\Psi}_{j}^{r+1} \ \middle|\ \mathfrak{F}^{r}, \mathcal{K}_{j-1}^{r}\right] \leqslant \ddot{\Psi}_{j}^{r} \cdot \left(1 - \frac{e^{v^{j+1}}}{n}\right) + e^{-v^{j}/2}. \tag{5.7}$$

现在，我们定义步骤 β_{j-1}之后任意轮次 r 的势函数，

$$\widehat{\Psi}_{j}^{r} := \ddot{\Psi}_{j}^{r} \cdot \mathbf{1}_{\cap_{\rho\in[\beta_{j-1},T_{\beta_{j-1}}(r)]}\mathcal{K}_{j-1}^{\rho}}.$$

注意，尽管这个势函数带有帽子，但它与 d-Weak-Memory Memory 过程无关。接下来观察到，每当 $\mathbf{1}_{\cap_{\rho\in[\beta_{j-1},T_{\beta_{j-1}}(r)]}\mathcal{K}_{j-1}^{\rho}} = 0$ 时，便可确定性地推出$\widehat{\Psi}_{j}^{r+1} = 0$，因此由 (5.7) 可得，

$$\mathbf{E}\left[\ \widehat{\Psi}_{j}^{r+1} \ \middle|\ \mathfrak{F}^{r}\right] \leqslant \widehat{\Psi}_{j}^{r} \cdot \left(1 - \frac{e^{v^{j+1}}}{2n}\right) + e^{-v^{j}/2}. \tag{5.8}$$

要得到$\mathbf{E}\left[\ \widehat{\Psi}_{j}^{r}\right]$ 的上界，只需得到$\mathbf{E}\left[\ \widehat{\Psi}_{j}^{r} \ \middle|\ \mathcal{K}_{j-1}^{\beta_{j-1}}\right]$ 的上界，因为

$$\mathbf{E}\left[\ \widehat{\Psi}_{j}^{r}\right] = \mathbf{E}\left[\ \widehat{\Psi}_{j}^{r} \ \middle|\ \mathcal{K}_{j-1}^{\beta_{j-1}}\right] \cdot \mathbf{Pr}\left[\mathcal{K}_{j-1}^{\beta_{j-1}}\right] + 0 \cdot \mathbf{Pr}\left[\neg\mathcal{K}_{j-1}^{\beta_{j-1}}\right] \leqslant \mathbf{E}\left[\ \widehat{\Psi}_{j}^{r} \ \middle|\ \mathcal{K}_{j-1}^{\beta_{j-1}}\right]. \tag{5.9}$$

当 $\mathcal{K}_{j-1}^{\beta_{j-1}}$ $\Phi_{j-1}^{\beta_{j-1}} \leqslant 2Cn$ 5 6 $\widehat{\Psi}_{j}^{\beta_{j-1}} \leqslant \Psi_{j}^{\beta_{j-1}} \leqslant$成立时，根据定义可 $e^{\frac{1}{2}\cdot}{}^{3}n$ 6 5 8 $b = e^{-v^{j}}$ 知成立，并由命题 . 得到，log。将引理 A. 应用于 (.)（其中 $a = 1 - \frac{1}{2n}\cdot e^{v^{j+1}}$ 且），对于任意满足 $r \geqslant n\cdot e^{-v^{j+1}}\cdot \log^{3} n$ 的轮次，若轮次 $r = 0$ 从步骤 β_{j-1}开始，则有

$$\begin{aligned}
\mathbf{E}\left[\ \widehat{\Psi}_{j}^{r} \ \middle|\ \mathcal{K}_{j-1}^{\beta_{j-1}}\right] &\leqslant \mathbf{E}\left[\ \widehat{\Psi}_{j}^{r} \ \middle|\ \mathfrak{F}^{\beta_{j-1}}, \widetilde{\Psi}_{j}^{\beta_{j-1}} \leqslant e^{\frac{1}{2}\cdot\log^{3} n}\right] \\
&\leqslant \left(1 - \frac{e^{v^{j+1}}}{2n}\right)^{r} \cdot e^{\frac{1}{2}\cdot\log^{3} n} + \frac{2n\cdot e^{-v^{j}}}{e^{v^{j+1}}} \\
&\leqslant \left(1 - \frac{e^{v^{j+1}}}{2n}\right)^{n\cdot e^{-v^{j+1}}\cdot\log^{3} n} \cdot e^{\frac{1}{2}\cdot\log^{3} n} + 2n.
\end{aligned}$$

N现在，利用不等式 $e^{x} \geqslant 1 + x$（对于任意 x）和 $C \geqslant 6$（由 (5.1) 得出），我们得到

$$\mathbf{E}\left[\ \widehat{\Psi}_{j}^{r} \ \middle|\ \mathcal{K}_{j-1}^{\beta_{j-1}}\right] \leqslant e^{-\frac{1}{2}\cdot\log^{3} n} \cdot e^{\frac{1}{2}\cdot\log^{3} n} + 2n \leqslant 1 + 2n \leqslant \frac{C}{2}n.$$

将其与 (5.9) 结合，对于任意轮次 $r \geqslant n\cdot e^{-v^{j+1}}\cdot\log^{3} n$,

$$\mathbf{E}\left[\ \widehat{\Psi}_{j}^{r}\right] \leqslant \frac{C}{2}n,$$

这意味着

$$\mathbf{E}\left[\ddot{\Psi}_j^r\cdot\mathbf{1}_{\cap_{\rho\in[\beta_{j-1},T_{\beta_{j-1}}(r)]}\mathcal{K}_{j-1}^\rho}\right]\leqslant\frac{C}{2}n.$$

因此,

$$\begin{aligned}\mathbf{E}\left[\widetilde{\Psi}_j^{T_{\beta_{j-1}}(r)}\right]&=\mathbf{E}\left[\overline{\Psi}_j^r\cdot\mathbf{1}_{\cap_{\rho\in[\beta_{j-1},T_{\beta_{j-1}}(r)]}\mathcal{K}_{j-1}^\rho}\right]\\&\overset{(a)}{\leqslant}\mathbf{E}\left[\dot{\Psi}_j^r\cdot\mathbf{1}_{\cap_{\rho\in[\beta_{j-1},T_{\beta_{j-1}}(r)]}\mathcal{K}_{j-1}^\rho}\right]+\mathbf{E}\left[n\cdot\mathbf{1}_{\cap_{\rho\in[\beta_{j-1},T_{\beta_{j-1}}(r)]}\mathcal{K}_{j-1}^\rho}\right]\\&\leqslant\frac{C}{2}n+n,\end{aligned}$$

利用 (a) 中的 $\Psi_j^r\leqslant\dot{\Psi}_j^r+n$。由于每轮至多分配 $k_j=e^{v^{j+1}}\cdot\log^3 n$ 个球，因此在 $n\cdot e^{-v^{j+1}}\cdot\log^3 n$ 轮中至多分配 $n\log^6 n$ 个球。因此，根据断言 5.9 和断言 5.10，对于任意步骤 $s\geqslant T_{\beta_{j-1}}(r)\geqslant\beta_{j-1}+n\log^6 n$,

$$\mathbf{E}\left[\widetilde{\Phi}_j^s\right]\leqslant\frac{C/2+1}{0.99}\cdot n\leqslant Cn,$$

利用由 (5.1) 得到的 $C\geqslant 6$。 □

~现在我们转而考虑另一个势函数 $\Phi_j^s=\Phi_j^s\cdot\mathbf{1}_{\cap_{r\in[\beta_{j-1},s]}\mathcal{K}_{j-1}^r}$，并证明以高概率，在 $[\beta_{j-1}+n\log^3 n,\ \beta_j]$ 中至少有一轮满足它是线性的至少一轮。

引理 5.12。对于任意 $1\leqslant j\leqslant j_{\max}-1$，有:

$$\mathbf{Pr}\left[\bigcup_{r\in[\beta_{j-1}+n\log^6 n,\beta_j]}\left\{\widetilde{\Phi}_j^r\leqslant Cn\right\}\right]\geqslant 1-n^{-5}.$$

*Pro*的。令 $t_0:=\beta_{j-1}+n\log^6 n$。利用引理 5.11 和马尔可夫不等式，我们得到

$$\mathbf{Pr}\left[\widetilde{\Psi}_j^{t_0}\leqslant Cn^{12}\right]\geqslant 1-n^{11}.$$

利用断言 5.7 以及 $\{\widetilde{\Phi}_j^{t_0}=0\}$成立当且仅当 $\{\widetilde{\Psi}_j^{t_0}=0\}$成立这一事实，我们得到 $\{\widetilde{\Psi}_j^{t_0}\leqslant Cn^{12}\}$蕴含 $\{\widetilde{\Phi}_j^{t_0}\leqslant n^{7/6}\}$，因此

$$\mathbf{Pr}\left[\widetilde{\Phi}_j^{t_0}\leqslant n^{7/6}\right]\geqslant 1-n^{11}.\tag{5.10}$$

现在假设 $\{\widetilde{\Phi}_j^{t_0}\leqslant n^{7/6}\}$ 成立。对于步骤 t_0之后的任意轮次 $r\geqslant 0$，我们定义

$$\widehat{\Phi}_j^r:=\ddot{\Phi}_j^r\cdot\mathbf{1}_{\cap_{\rho\in[\beta_{j-1},T_{t_0}(r)]}\mathcal{K}_{j-1}^\rho}\cdot\mathbf{1}_{\cap_{\rho\in[0,r]}\ddot{\Phi}_j^\rho>\frac{C}{2}n}.$$

根据推论 5.4，由于 $C\geqslant 6$，对于步骤 t_0之后的任意轮次 r，

$$\mathbf{E}\left[\ddot{\Phi}_j^{r+1}\,\middle|\,\mathfrak{F}^r,\ddot{\Phi}_j^r>Cn\right]\leqslant\ddot{\Phi}_j^r\cdot\left(1-\frac{e^{v^{j+1}}}{2n}\right).$$

由于每当 $\{\ddot{\Phi}_j^r=0\}$ 时，$\{\ddot{\Phi}_j^{r+1}=0\}$ 都确定性地成立，因此对于步骤 t_0 之后的任意轮次 r，我们也有:

$$\mathbf{E}\left[\widehat{\Phi}_j^{r+1}\,\middle|\,\mathfrak{F}^r\right]\leqslant\widehat{\Phi}_j^r\cdot\left(1-\frac{e^{v^{j+1}}}{2n}\right).\tag{5.11}$$

根据归纳从步骤 t_0 开始，对 $\Delta := n \cdot e^{-v^{j+1}} \cdot \log^2 n$ 轮次反复应用 (5.11)，我们有

$$\begin{aligned}\mathbf{E}\left[\widehat{\Phi}_j^{\Delta} \;\middle|\; \mathfrak{F}^{t_0}, \widetilde{\Phi}_j^{t_0} \leqslant n^{7/6}\right] &\leqslant \widetilde{\Phi}_j^{t_0} \cdot \left(1 - \frac{e^{v^{j+1}}}{2n}\right)^{T} \\ &\leqslant \widetilde{\Phi}_j^{t_0} \cdot \left(1 - \frac{e^{v^{j+1}}}{2n}\right)^{n \cdot e^{-v^{j+1}} \cdot \log^2 n} \\ &\leqslant n^{7/6} \cdot e^{-\frac{1}{2}\log^2 n} \leqslant n^{-6},\end{aligned}$$

当 s 足够大时，利用 $e^x \geqslant 1 + x$ （对于任意 x）。因此，根据马尔可夫不等式，

$$\mathbf{Pr}\left[\widehat{\Phi}_j^{\Delta} \leqslant 1 \;\middle|\; \mathfrak{F}^{t_0}, \widetilde{\Phi}_j^{t_0} \leqslant n^{7/6}\right] \geqslant 1 - n^{-6}.$$

When $\{\widehat{\Phi}_j^T \leqslant 1\}$ 成立，要么 $\ddot{\Phi}_j^T \leqslant 1 \leqslant Cn$，要么对于某个指示变量，有以下情况之一成立：

$$\mathbf{1}_{\cap_{\rho\in[\beta_{j-1},T_{t_0}(\Delta)]}\mathcal{K}_{j-1}^{\rho}} = 0 \quad \text{or} \quad \mathbf{1}_{\cap_{\rho\in[0,\Delta]}\ddot{\Phi}_j^{\rho} > \frac{C}{2}n} = 0.$$

在这两种情况下，都可以推出存在某个轮次 $r \in [0, \Delta]$，使得 $\widetilde{\Phi}^{T_{t_0}(r)}_j \leqslant Cn$，因为当 $\ddot{\Phi}_j^r \leqslant \frac{C}{2}n$ 时，$\widetilde{\Phi}^{T_{t_0}(j}$ $\leqslant \frac{C}{2}n + n \leqslant Cn$）。

由于每轮最多分配 $k_j = e^{v^{j+1}} \cdot \log^3 n$ 个球，因此在这 Δ 轮中最多可以分配 $n\log^5 n$ 个球。这意味着存在一个步骤 $s \in [\beta_{j-1} + n\log^6 n, \beta_j]$，使得 $\widetilde{\Phi}_j^s \leqslant Cn$。因此，由于 $\beta_{j-1} + n\log^6 n + n\log^5 n \leqslant \beta_j$，我们有

$$\mathbf{Pr}\left[\bigcup_{s\in[\beta_{j-1}+n\log^6 n,\beta_j]} \left\{\widetilde{\Phi}_j^s \leqslant Cn\right\} \;\middle|\; \mathfrak{F}^{t_0}, \widetilde{\Phi}_j^{t_0} \leqslant n^{7/6}\right] \geqslant 1 - n^{-6}.$$

最后，结合 (5.10)，得到

$$\mathbf{Pr}\left[\bigcup_{s\in[\beta_{j-1}+n\log^6 n,\beta_j]} \left\{\widetilde{\Phi}_j^s \leqslant Cn\right\}\right] \geqslant \left(1 - n^{-6}\right) \cdot \left(1 - n^{-11}\right) \geqslant 1 - n^{-5}. \qquad \square$$

5.3 完成关键引理的证明（引理 5.13）

引理 5.13 的证明借鉴了定理 4.2 证明中的一些思路。不过，在那里我们可以更宽松地对整个时间区间（由 $n\cdot$polylog(n) 步组成）使用并集界，从而确保势函数以高概率在各处都足够小。这里我们无法承受误差概率损失一个多项式因子，因为归纳步骤必须应用 $j_{\max} = \Theta(\log\log\ n)$ 次。为了解决这个问题，我们将时间区间划分为长度为 $n/\log^2 n$ 的连续区间。然后，我们将证明在每个这样的区间结束时，势函数都以高概率足够小；最后使用一个简单的平滑性论证，说明势函数以高概率在所有 步骤中都足够小。

引理 5.13 (**归纳步骤**). 考虑具有任意(a, b)偏置采样分布的MEMORY 过程，其中常数为 $a, b \geqslant 1$。那么，对于任意步骤 $t \geqslant 1$ 和 $1 \leqslant j \leqslant j_{\max} - 1$，定义$\beta_j := t + 2jn\log^6 n$，并令 $C \geqslant 6$ 如(5.1)中所定义。然后，假设

$$\mathbf{Pr}\left[\bigcap_{s\in[\beta_{j-1},t+n\log^8 n]} \{\Phi_{j-1}^s \leqslant 2Cn\}\right] \geqslant 1 - \frac{(\log n)^{11(j-1)}}{n^4},$$

那么以下结论也成立，

$$\mathbf{Pr}\left[\bigcap_{s\in[\beta_j,t+n\log^8 n]} \{\Phi_j^s \leqslant 2Cn\}\right] \geqslant 1 - \frac{(\log n)^{11j}}{n^4}.$$

证明。 考虑任意步骤 $t_0 \in [\beta_{j-1} + n\log^6 n,\ \beta_j]$ 并回顾第 31 页的内容： k_j 是归纳第 j 层中折叠过程的阶段数。我们的目标是证明，对于所有 $u \in [t_0, t+n\log^8 n]$，均有 $\Phi_j^u \leqslant 2Cn$。我们将这些步骤分组为至多 $q \leqslant \log^{10} n$ 个时期，并按照步骤 t_0 起的顺序排列。对于每个 $i \in [q]$，第 i 个时期持续随机的 N_i 步，满足 $N_i \in [n/\log^2 n - k_j \cdot \frac{v}{\alpha_2},\ n/\log^2 n]$；但最后一个时期可能例外，其长度不受相同下界约束，但包含的步数少于 $n/\log^2 n$。特别地，令 r_i 表示第 i 个时期的起始轮次，那么对于任意满足 $\Delta = n/\log^2 n$ 的轮次 $r \in [r_i, r_i + \Delta]$，我们定义

$$X_i^r := \begin{cases} X_i^{r-1} & \text{if we have allocated more than } \frac{n}{\log^2 n} - k_j \text{ balls in rounds } [r_i, r), \\ \overline{\Phi}_j^r & \text{else if } \exists \rho \in [r_i, r) \text{ such that } \overline{\Phi}_j^\rho \geqslant 5n, \\ 5n + n^{1/3} & \text{otherwise.} \end{cases}$$

这意味着随机变量 X_i^r 在分配 $n/\log^2 n$ 个球之前就停止了，因此当 $\Phi^{r_i+\Delta} = \mathcal{O}(n)$ 时，根据平滑性论证和断言 5.9，对于时期内的所有步骤 s，我们都有 $\Phi_j^s = \mathcal{O}(n)$，即 $s \in [T_{t_0}(r_i)\ T_{t_0}(r_{i+1}))$。注意，根据 X_i^r 的定义，第三个分支条件只能在时期开始时的某些轮次得到满足；这是因为一旦第二个条件被激活，就只能满足第二个或第一个条件。

沿用定理 A.10 的记号，对于任意 $r \in [r_i, r_i + \Delta]$，我们将坏事件 B_i^r 定义为以下事件的补事件

$$\left(\bigcap_{u \in [T_{t_0}(r_i), T_{t_0}(r))} \left\{ \widetilde{\Psi}_j^u \leqslant Cn^{12} \right\} \right) \bigcap \left(\bigcap_{u \in [T_{t_0}(r_i), T_{t_0}(r))} \mathcal{K}_{j-1}^u \right) \subseteq \bigcap_{u \in [T_{t_0}(r_i), T_{t_0}(r))} \left\{ \Psi_j^u \leqslant Cn^{12} \right\}.$$

~现在我们将对该坏事件发生的概率进行界定。根据引理 5.11，对于任意步骤 $u \in [\beta_{j-1} + n\log^6 n, t + n\log^8 n]$，都有 $\mathbf{E}[\,\Psi_j^u\,] \leqslant Cn$。利用马尔可夫不等式，并对步骤 $u \in [\beta_{j-1} + n\log^6 n, t + n\log^8 n]$ 应用并集界，可得

$$\begin{aligned} \mathbf{Pr}\left[\bigcap_{u \in [\beta_{j-1} + n\log^6 n, t + n\log^8 n]} \left\{ \widetilde{\Psi}_j^u \leqslant Cn^{12} \right\} \right] &\geqslant 1 - n^{-11} \cdot \left(n\log^8 n + (2j-1) n \log^6 n \right) \\ &\geqslant 1 - n^{-9}. \end{aligned} \tag{5.12}$$

根据该引理关于 $j-1$ 的假设，可得

$$\mathbf{Pr}\left[\bigcap_{u \in [\beta_{j-1}, t + n\log^8 n]} \mathcal{K}_{j-1}^u \right] \geqslant 1 - \frac{(\log n)^{11(j-1)}}{n^4}. \tag{5.13}$$

因此，对 (5.13) 和 (5.12) 应用并集界（因为 $[T_{t_0}(r_i)\ T_{t_0}(r)) \subseteq [\beta_{j-1} + n\log^6 n,\ t + n\log^8 n]$），

$$\mathbf{Pr}\left[\, \neg B_i^r \,\right] \geqslant 1 - n^{-9} - \frac{(\log n)^{11(j-1)}}{n^4} \geqslant 1 - \frac{2(\log n)^{11(j-1)}}{n^4}.$$

我们将使用以下结论，为 X_i^r 确立定理 A.10 的前提条件。

断言 5.14. 考虑 $i \in [q]$ 的第 i 个时期。于是，对于随机变量 X_i^r、任意轮次 $r \in [r_i, r_i + \Delta]$ 以及任意滤过 $\mathfrak{F}^{r-1}$，有

$$\begin{aligned} &(i) \quad \mathbf{E}\left[\, X_i^r \mid \mathfrak{F}^{r-1}, \neg B_i^{r-1} \,\right] \leqslant X_i^{r-1}, \\ &(ii) \quad \left(\left| X_i^r - X_i^{r-1} \right| \mid \mathfrak{F}^{r-1}, \neg B_i^{r-1} \right) \leqslant 2n^{1/3}. \end{aligned}$$

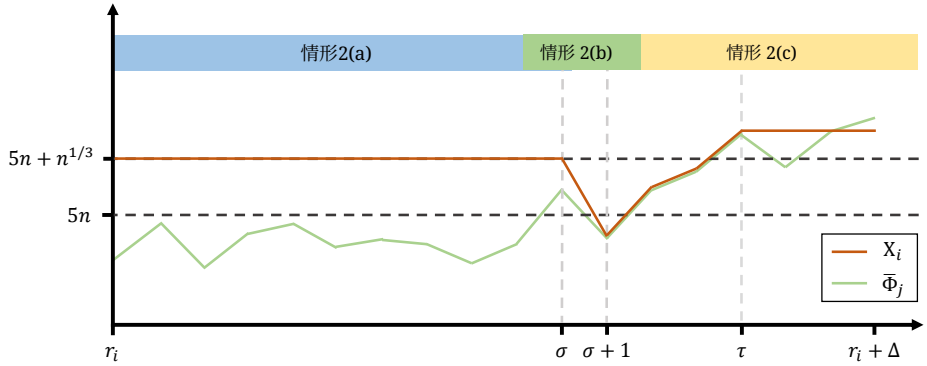

图6： 断言 5.14 证明中情况 2 的三个子情形示意图。

断言 *5.14* 的证明. 令 $\tau := \inf\{r \geqslant r_i\colon\ T_{t_0}(r) - T_{t_0}(r_i) \geqslant {}^{n\ \text{对数}2\ n} - k_j\}$。于是，对于任意轮次$r-1 \geqslant \tau$，条件都显然满足，因为 $X_i^r = X_i^{r-1}$。

对于 $r-1 < \tau$，回顾一下，根据引理 5.5，对于任意轮次 $r \in (r_i,\ r_i + \Delta]$ 且 $\overline{\Phi}_j^{r-1} \geqslant 4n$，

$$\begin{aligned}\mathbf{E}\left[\overline{\Phi}_j^r \,\middle|\, \mathfrak{F}^{r-1}, \mathcal{K}_{j-1}^{r-1}, \overline{\Phi}_j^{r-1} \geqslant 4n\right] &\leqslant \overline{\Phi}_j^{r-1}\cdot\left(1-\frac{1}{2n}\right) - \overline{\Phi}_j^{r-1}\cdot\frac{1}{2n} + 2\\ &\leqslant \overline{\Phi}_j^{r-1}\cdot\left(1-\frac{1}{2n}\right),\end{aligned} \tag{5.14}$$

并考虑以下情况：

情况 1 [$\overline{\Phi}_j^{r_i} \geqslant 5n+n^{1/3}$]： 根据断言 5.9，对于所有 $r \in (r_i, r_i+\Delta]$，我们有 $\Phi_j^{r-1} \geqslant 0.99\cdot(5n+n^{1/3}) \geqslant 4n$（因为在这些轮次中我们最多分配 $\Delta \leqslant n/\log^2 n$ 个球），因此，第一个陈述由 (5.14) 得出；对于第二个陈述，由于在给定 $\neg B_i^{r-1}$ 的条件下，断言 5.8 的前提成立，因此也可得出。

情况 2： 否则，令 $\sigma := \inf\{r \geqslant r_i\colon\ \Phi_j^r \gtrless 5n\}$。我们考虑以下三个子情形（见图 6）：

- **情形2(a)** **[$r-1 < \sigma$]：** 这里 $X_i^r = X_i^{r-1}$（利用了 $\Phi_j^{r_i} < \overline{5n} + n^{1/3}$），因此这两个陈述显然成立。
- **情形2(b)** **[$r-1 = \sigma$]：** </style}我们有 $\Phi_j^{r-1} \geqslant 5n$ 和 $\Phi_j^{r-2} < 5n$。根据断言 5.8，由于$\neg B_i^{r-1}$成立，可得 $\Phi_j^{r-1} \leqslant \Phi_j^{r-2} + n^{1/3} < 5n + n^{1/3}$。此外，根据定义，$X_i^{r-1} = 5n + n^{1/3}$ 和 $X_i^r = \Phi_j^r$，因此由 (5.14) 可得，

$$\mathbf{E}\left[\, X_i^r \mid \mathfrak{F}^{r-1}, \neg B_i^{r-1}\,\right] = \mathbf{E}\left[\, \overline{\Phi}_j^r \,\middle|\, \mathfrak{F}^{r-1}, \neg B_i^{r-1}\,\right] \leqslant \overline{\Phi}_j^{r-1} < X_i^{r-1},$$

这就证明了第一个结论。对于第二个结论，我们有

$$\left|X_i^r - X_i^{r-1}\right| = \left|\overline{\Phi}_j^r - 5n - n^{1/3}\right| \leqslant \left|\overline{\Phi}_j^{r-1} - 5n - n^{1/3}\right| + \left|\overline{\Phi}_j^r - \overline{\Phi}_j^{r-1}\right| \leqslant 2n^{1/3},$$

其中，在第二个不等式中我们用到了断言 5.8。

- **情形 2(c)** **[$r-1 > \sigma$]：** 这里，$X_i^{r-1} = \Phi_j^{r-1}$ 和 $X_i^r = \Phi_j^r$。由于 $\Phi_j^\sigma \gtrless 5n$，根据断言 5.9（因为 $r - \sigma \leqslant n/\log^2 n$），我们还有

$$\overline{\Phi}_j^{r-1} \geqslant 0.99\cdot\overline{\Phi}_j^\sigma \geqslant 0.99\cdot 5n \geqslant 4n,$$

因此，由 (5.14) 可得第一个结论。第二个结论成立，因为在给定 $\neg B_i^r$ 的条件下，断言 5.8 的前提条件成立。 ◊

根据断言 5.14，我们现已验证，对于任意满足相关坏事件 B_i^{r-1} 不发生的滤过 $\mathfrak{F}^{r-1}$， X_i^r 都满足定理 A.10 的前提条件。因此，对于 $N=\Delta$、$\lambda=$ n 对数^{10}n 和 $D=2n^{1/3}$，我们有

$$\mathbf{Pr}\left[\,X_i^{r_i}\geqslant X_i^{r_{i+1}}+\lambda\,\right]\leqslant\exp\left(-\frac{n^2/\log^{20}n}{10\cdot\Delta\cdot(4n^{2/3})}\right)+\frac{2(\log n)^{11(j-1)}}{n^4}\leqslant\frac{3(\log n)^{11(j-1)}}{n^4}.$$

对至多 $\log^{10}n$ 个区间 $i\in[q]$ 应用并集界，可得

$$\mathbf{Pr}\left[\bigcup_{i\in[q]}\left\{X_i^{r_{i+1}}\geqslant X_1^{r_1}+i\cdot\frac{n}{\log^{10}n}\right\}\,\middle|\,\mathfrak{F}^{t_0}\right]\leqslant\log^{10}n\cdot\frac{3(\log n)^{11(j-1)}}{n^4}\leqslant\frac{1}{2}\cdot\frac{(\log n)^{11j}}{n^4}.$$

在 $\Phi_j^{t_0}\leqslant Cn$ （这意味着 $X_1^{r_1}\leqslant Cn$，因为 $C\geqslant 6$）和 $q\leqslant\log^{10}n$ 的条件下，我们得到

$$\mathbf{Pr}\left[\bigcap_{i\in[q]}\left\{X_i^{r_{i+1}}\leqslant Cn+n\right\}\,\middle|\,\mathfrak{F}^{t_0},\Phi_j^{t_0}\leqslant Cn\right]\geqslant 1-\frac{1}{2}\cdot\frac{(\log n)^{11j}}{n^4}.$$

由于 $\Phi_j^{r_{i+1}}\leqslant\max\left\{X_i^{r_{i+1}},5n+n^{1/3}\right\}\leqslant Cn+n$，我们有

$$\mathbf{Pr}\left[\bigcap_{i\in[q]}\left\{\overline{\Phi}_j^{r_{i+1}}\leqslant Cn+n\right\}\,\middle|\,\mathfrak{F}^{t_0},\Phi_j^{t_0}\leqslant Cn\right]\geqslant 1-\frac{1}{2}\cdot\frac{(\log n)^{11j}}{n^4}.$$

由于第 i 个时期至多包含 $n/\log^2 n$ 个分配，我们有

$$\mathbf{Pr}\left[\bigcap_{s\in[t_0,t+n\log^8 n]}\bigcup_{u\in[s,s+n/\log^2 n]}\left\{\Phi_j^u\leqslant(C+1)n\right\}\,\middle|\,\mathfrak{F}^{t_0},\Phi_j^{t_0}\leqslant Cn\right]\geqslant 1-\frac{1}{2}\cdot\frac{(\log n)^{11j}}{n^4}.$$

对每个步骤 s 应用断言 5.9 的平滑性论证，可得

$$\mathbf{Pr}\left[\bigcap_{s\in[t_0,t+n\log^8 n]}\left\{\Phi_j^s\leqslant\frac{C+1}{0.99}\cdot n\right\}\,\middle|\,\mathfrak{F}^{t_0},\Phi_j^{t_0}\leqslant Cn\right]\geqslant 1-\frac{1}{2}\cdot\frac{(\log n)^{11j}}{n^4}.$$

由于 $2C\geqslant\frac{C+1}{0.99}$（因为 $C\geqslant 6$）且 $t_0\leqslant\beta_j$，

$$\mathbf{Pr}\left[\bigcap_{s\in[\beta_j,t+n\log^8 n]}\left\{\Phi_j^s\leqslant 2Cn\right\}\,\middle|\,\mathfrak{F}^{t_0},\Phi_j^{t_0}\leqslant Cn\right]\geqslant 1-\frac{1}{2}\cdot\frac{(\log n)^{11j}}{n^4}.\tag{5.15}$$

现在，下一步是得到一个不带条件化的界。为此，我们定义停止时间 $\tau:=\inf\{t_0\geqslant\beta_{j-1}+n\log^5 n\colon\ \Phi_j^{t_0}\leqslant Cn\}$。然后由并集界，

$$\begin{aligned}\mathbf{Pr}\left[\,\tau\leqslant\beta_j\,\right]&\geqslant\mathbf{Pr}\left[\left(\bigcup_{s\in[\beta_{j-1}+n\log^5 n,\beta_j]}\left\{\widetilde{\Phi}_j^s\leqslant Cn\right\}\right)\cap\bigcap_{s\in[\beta_{j-1}+n\log^5 n,\beta_j]}\left\{\Phi_{j-1}^s\leqslant 2Cn\right\}\right]\\&\geqslant\mathbf{Pr}\left[\bigcup_{s\in[\beta_{j-1}+n\log^5 n,\beta_j]}\left\{\widetilde{\Phi}_j^s\leqslant Cn\right\}\right]-\mathbf{Pr}\left[\neg\bigcap_{s\in[\beta_{j-1}+n\log^5 n,\beta_j]}\left\{\Phi_{j-1}^s\leqslant 2Cn\right\}\right]\\&\geqslant 1-n^{-5}-\frac{(\log n)^{11(j-1)}}{n^4}\geqslant 1-\frac{2(\log n)^{11(j-1)}}{n^4},\end{aligned}$$

其中，在第三个不等式中，我们使用了引理 5.12 和 (5.13)（因为 $\beta_j \leqslant t + n\log^5 n$）。
最后，结合 (5.15) 即可得到结论，

$$\begin{aligned}
&\mathbf{Pr}\left[\bigcap_{s\in[\beta_j,t+n\log^8 n]} \left\{\Phi_j^s \leqslant 2Cn\right\}\right] \\
&\geqslant \sum_{t_0=\beta_{j-1}+n\log^5 n}^{\beta_j} \mathbf{Pr}\left[\bigcap_{s\in[\beta_j,t+n\log^8 n]} \left\{\Phi_j^s \leqslant 2Cn\right\} \;\middle|\; \mathfrak{F}^{t_0}, \Phi_j^{t_0} \leqslant Cn\right] \cdot \mathbf{Pr}\left[\tau = t_0\right] \\
&\geqslant \left(1 - \frac{1}{2}\cdot\frac{(\log n)^{11j}}{n^4}\right)\cdot\left(1 - \frac{2(\log n)^{11(j-1)}}{n^4}\right) \\
&\geqslant 1 - \frac{(\log n)^{11j}}{n^4}.
\end{aligned}$$

□

5.4 利用引理 5.13 证明主定理（定理 1.1）

定理 1.1 （重述）。考虑采用MEMORY 过程，其采样分布有 (a, b) 偏差，其中常数 $a, b \geqslant 1$ 任意。于是存在常数 $\kappa := \kappa(a, b) > 0$，使得对每个步骤 $m \geqslant 1$,

$$\mathbf{Pr}\left[\operatorname{Gap}(m) \leqslant \kappa \cdot \log\log n\right] \geqslant 1 - n^{-3}.$$

证明。首先考虑 $m \geqslant n\log^8 n$ 的情形，并令 $t = m - n\log^8 n$。我们将对势函数 Φ_j（$j = 1, \ldots, j_{\max} - 1$）进行归纳。基础情形由定理 4.2 得出，因为对于所有步骤 $s \in [t, t + n\log^8 n]$，都有 $\Phi_0^s \leqslant \Gamma^s$。因此，

$$\mathbf{Pr}\left[\bigcap_{s\in[t,m]} \{\Phi_0^s \leqslant 2Cn\}\right] \geqslant 1 - n^{-4}.$$

对于 归纳步骤，我们使用引理 5.13。经过 $j_{\max} = \Theta(\log\log n)$ 次应用后，我们得到

$$\mathbf{Pr}\left[\bigcap_{s\in[t+\beta_{j_{\max}-1},m]} \left\{\Phi_{j_{\max}-1}^s \leqslant 2Cn\right\}\right] \geqslant 1 - \frac{(\log n)^{11\cdot(j_{\max}-1)}}{n^4} \geqslant 1 - n^{-3}.$$

当该事件发生时，轮次 m 的差距不可能大于某个常数$\kappa := \kappa(\alpha_2, v) > 0$，否则

$$\begin{aligned}
2Cn \geqslant \Phi_{j_{\max}-1}^m &\geqslant \exp\left(\alpha_2 \cdot v^{j_{\max}-1} \cdot \frac{4v^2}{\alpha_2^2}\right) \\
&= \exp\left(4 \cdot v^{j_{\max}} \cdot \frac{v}{\alpha_2}\right) = \exp\left(4 \cdot \frac{\alpha_2}{2v}\log n \cdot \frac{v}{\alpha_2}\right) = \exp(2\cdot\log n) = n^2,
\end{aligned}$$

这会导致矛盾。因此，对于某个常数 $\kappa := \kappa(\alpha_2, v)$ $\kappa(a, b) > 0$，$\operatorname{Gap}(m) \leqslant z_{j_{\max}-1} + \frac{2v^2}{\alpha_2^2} = \frac{5v}{\alpha_2}\cdot(\log_v\left(\frac{\alpha_2}{2v}\log n\right) - 1) + \frac{2v^2}{\alpha_2^2} \leqslant \kappa \cdot \log\log n$，所以

$$\mathbf{Pr}\left[\operatorname{Gap}(m) \leqslant \kappa \cdot \log\log n\right] \geqslant 1 - n^{-3}.$$

另一种情况是 $m < n\log^8 n$，此时上述分析中的某些 β_j 会为负。为了解决这一问题，考虑一个修改后的过程。该修改后的过程从第 $n\log^8 n - m$ 轮开始，初始负载向量为空。对于任意时间 $t \in [n\log^8 n, -m]$，它都会向每个 n 箱分配一个权重为 $\frac{1}{n}$ 的球。对于 $t \geqslant 1$，该过程与原始的 Memory 过程完全相同。由于每个步骤 $t < 0$ 的负载向量都完全平衡，因此 $\dot{\Psi}_j^t = \dot{\Phi}_j^t = 0$（以及 $\Phi_j^t = \Psi_j^t = n$）是确定性的。我们的证明仅依赖于势函数的上界，因此这些条件显然成立，故上述分析适用于修改后的过程。此外，对于 $t \geqslant 1$，修改后的过程与原始过程的相对负载表现相同，因此结论得证。

□

6 Pr在宽松设定下 Memory 和 d-弱 Memory 的证明 s

本节将 Memory 的有界差距结果推广到具有完全支撑的任意抽样分布，并证明带有加权球的 d-Weak-Memory 的 $\mathcal{O}(\log n)$ 差距界。

6.1 一般偏置抽样

回顾一下，定理 1.1 证明了：对于任意常数 $a,\ b \geqslant 1$，在任意 $(a,\ b)$-偏置分布上的 Memory 过程，以高概率具有 $\mathcal{O}(\log\log\ n)$ 的差距。以下结果表明，对于具有全支撑的任意 抽样分布，该差距以高概率保持有界，且界与 m 无关。

我们先给出下面这个简单引理，以便像在定理 4.1 的证明中那样应用定理 4.4。

引理 6.1。 考虑任意满足条件 $\mathcal{C}_1$ 的概率向量 p，其中参数为 $\delta \leqslant 1/2$ 和 $\varepsilon \in (0,1)$，且对应于 (δ,ε)。那么， p 也满足条件 $\mathcal{C}_1$，对应于 $(1/2,\varepsilon')$，其中 $\varepsilon' = \varepsilon/n$。

证明。 我们将证明 p 满足参数为 $\varepsilon' = \varepsilon/n$ 时关于 $(1/2,\ \varepsilon')$ 的条件 $\mathcal{C}_1$。我们考虑指标 $k \in [n]$ 的以下几种情况：

- **情况 A** $[1 \leqslant k \leqslant \delta n]$**:**

$$\sum_{i=1}^{k} p_i \leqslant (1-\varepsilon)\cdot\frac{k}{n} \leqslant \left(1-\frac{\varepsilon}{n}\right)\cdot\frac{k}{n}.$$

- **情况 B** $[\delta n < k \leqslant n/2]$**:**

$$\begin{aligned}\sum_{i=1}^{k} p_i &\leqslant 1-\left(1+\varepsilon\cdot\frac{\delta}{1-\delta}\right)\cdot\left(1-\frac{k}{n}\right)\\ &= \frac{k}{n}+\frac{\varepsilon\delta}{1-\delta}\cdot\left(\frac{k}{n}-1\right)\\ &\overset{(a)}{\leqslant} \frac{k}{n}-\frac{\varepsilon\delta}{1-\delta}\cdot\frac{k}{n}\\ &= \frac{k}{n}\cdot\left(1-\frac{\varepsilon\delta}{1-\delta}\right)\\ &\overset{(b)}{\leqslant} \frac{k}{n}\cdot\left(1-\frac{\varepsilon}{n}\right),\end{aligned}$$

  根据 (a) 中的 $k \leqslant n/2$ 以及 (b) 中的 $\delta/(1-\delta) \geqslant 1/n$，因为 $\delta \geqslant 1/n$。

- **情况 C** $[n/2 < k]$**:**

$$\sum_{i=k}^{n} p_i \geqslant \left(1+\varepsilon\cdot\frac{\delta}{1-\delta}\right)\cdot\frac{n-k+1}{n} \geqslant \left(1+\frac{\varepsilon}{n}\right)\cdot\frac{n-k+1}{n}.$$ □

定理 1.3 (重述)。 考虑具有任意抽样分布 $s = (s_i)^{i\in[n]}$ 的MEMORY 过程，该分布满足 $s_{\min}$: $= \min_{i\in[n]} s_i > 0$。那么对于每个步骤 $m \geqslant 1$，我们有

$$\mathbf{Pr}\left[\mathrm{Gap}(m) \leqslant \frac{n^8}{s_{\min}^{10}}\right] \geqslant 1-n^{-1}.$$

证明。 首先，我们注意到，每个满足 $s\ \ s_{\min} > 0$ 的分布都是关于 $a = 1/(s_{\min}n)$ 和 $b \leqslant n(1-(n-1)s_{\min})$ 的 $(a,\ b)$-偏置分布。另外，由于对于任意 $a' \geqslant a$ 和 $b' \geqslant b$，每个 $(a,\ b)$-偏置分布都是 $(a',\ b')$-偏置分布，因此只需

考虑满足 $a = 1/(s_{\min} n)$ 和 $b \leqslant n(1-(n-1)s_{\min})$ 的 $(a,\ b)$-偏置分布，并且我们可以假设 a 和 b 都足够大。

我们按照定理 4.1 证明中的相同方法进行。在断言 4.9 中，我们证明了当选择 $d = \left\lceil \frac{2b}{(ab-1)^2 (b-1)^2 (1-\varepsilon)} \right\rceil$ 时，$(\widehat{p}_i)_{i\in[n]}$ 满足 $\mathcal{C}_1$，其中参数为 $\delta = M/n = \frac{a-1}{ab-1}$, $\varepsilon = 1/2$。由于 $(a,\ b)$ 不一定是常数，我们使用引理 6.1 得出 $(\widehat{p}_i)_{i\in[n]}$ 对（常数） $\delta' = 1/2$ 和 $\varepsilon' = 1/n$ 满足 $\mathcal{C}_1$。这样我们便可以应用定理 4.4，该定理要求分位数 δ 为常数。注意，根据 (4.20)，我们也可以取 $K = 2bd^3$。

与定理 4.1 的证明中一样，应用定理 4.4 可得，存在一个常数$c > 0$，使得对于任意满足 d 整除 s 的步骤 s，平滑参数为 $\alpha \leqslant \min\{\frac{1}{d}, \frac{\varepsilon'\delta}{d\cdot 8K}\}$ 的双曲余弦势函数 Γ 满足

$$\mathbf{E}\left[\Gamma^s\right] \leqslant cn.$$

根据马尔可夫不等式，我们有

$$\mathbf{Pr}\left[\Gamma^s \leqslant cn^3\right] \geqslant 1 - n^{-2}.$$

因此，由于差距在 d 步的区间内最多增加 d，我们有，对于每个步骤 t,

$$\mathbf{Pr}\left[\mathrm{Gap}(t) \leqslant d + \frac{\log(cn^3)}{\alpha}\right] \geqslant 1 - n^{-2}.$$

回想一下，根据开篇段落，我们可以假设 $a,\ b > 100$。最后，由于 $\varepsilon' = 1/n$ 和 $\delta = \frac{a-1}{ab-1} \geqslant 1/n$，我们可以选择 $\alpha = \Omega(\frac{1}{n^2 bd^4})$。因此，由于 $d = \Theta\left(\frac{b(ab-1)^2}{b-1}()^2\right)$，我们有

$$d + \frac{\log(cn^3)}{\alpha} = \mathcal{O}\left(bd^4 \cdot n^2 \log n\right) = \mathcal{O}\left(\frac{b^5(ab-1)^{10}}{(b-1)^{10}} \cdot n^3\right) = \mathcal{O}\left(a^{10}b^5 \cdot n^3\right) = \mathcal{O}\left(\frac{n^8}{s_{\min}^{10}}\right),$$

这就证明了该结论。 □

6.2 加权 2-弱记忆

我们考虑 2-弱记忆的加权情形。按照 [27], ，每个球的权重将从定义在 $[0,\infty$ 上的固定分布 W 中独立抽取，并满足:

$$\mathbf{E}[W] = 1. \tag{6.1}$$

$$\mathbf{E}[e^{\lambda W}] < \infty \text{ for some constant } \lambda > 0. \tag{6.2}$$

显然，当 $\mathbf{E}[W] = \Theta(1)$ 时，通过缩放 W，我们总能实现 $\mathbf{E}[W] = 1$。满足上述条件（缩放后）的分布的具体例子包括几何分布、指数分布、二项分布和泊松分布。

我们将用到以下引理:

引理 6.2 （ [18]中的引理 2.4）。 对于满足 （6.1）和（6.2）的任意随机变量 W，存在 $S := S(\lambda) \geqslant \max(1,\ 1/\lambda)$，使得对于任意 $\alpha \in (0,\ \min(\lambda/2,\ 1)$ 以及任意 $\kappa \in [-1, 1]$,

$$\mathbf{E}\left[e^{\alpha\cdot\kappa\cdot W}\right] \leqslant 1 + \alpha\cdot\kappa + S\alpha^2\cdot\kappa^2.$$

现在我们可以陈述并证明本小节的结论。

定理 1.5 (重述)。 考虑采用均匀采样分布的2-弱记忆 过程。此外，假设每个球的权重独立地从某个分布W 中抽取，且满足$\mathbf{E}[W] = 1$ 和$\mathbf{E}[e^{\lambda W}] < \infty$，其中 $\lambda > 0$为常数。那么存在一个常数$\kappa := \kappa(W) > 0$，使得对于每个步骤 $m \geqslant 1$,

$$\mathbf{Pr}\left[\mathrm{Gap}(m) \leqslant \kappa \cdot \log n\right] \geqslant 1 - n^{-2}.$$

证明。 考虑 2-弱记忆过程中的任意偶数步骤 $2t \geqslant 0$，并令 W_1 和 W_2 分别为第 $(2t+1)$ 个球和第 $(2t+2)$ 个球的权重。那么，其概率为：

- $q_{11} = \frac{1}{n} \cdot \frac{i}{n}$：我们将两个球都分配到负载最高的第 i 个箱子。
- $q_{01} = \frac{i-1}{n} \cdot \frac{1}{n}$：我们将第二个球分配到负载最高的第 i 个箱子。
- $q_{10} = \frac{1}{n} \cdot \big(1 - \frac{i}{n}\big)$：我们将第一个球分配到负载最高的第 i 个箱子。
- $q_{00} = 1 - q_{11} - q_{01} - q_{10}$：两个球都不分配到负载最高的第 i 个箱子 .

现在，我们考虑势函数经过两个步骤后的期望变化

$$\begin{aligned}
\mathbf{E}\left[\, \Phi^{2t+2} \,\middle|\, \mathfrak{F}^{2t} \,\right] &= \Phi^{2t} \cdot \Big(q_{00} \cdot \mathbf{E}\left[e^{-\alpha(W_1+W_2)/n} \right] + q_{10} \cdot \mathbf{E}\left[e^{\alpha W_1(1-1/n)} \right] \cdot \mathbf{E}\left[e^{-\alpha W_2/n} \right] \\
&\qquad + q_{01} \cdot \mathbf{E}\left[e^{-\alpha W_1/n} \right] \cdot \mathbf{E}\left[e^{\alpha W_2(1-1/n)} \right] + q_{11} \cdot \mathbf{E}\left[e^{\alpha(W_1+W_2)(1-1/n)} \right] \Big) \\
&\overset{(a)}{\leqslant} \Phi^{2t} \cdot \Big(1 + \alpha \cdot \Big(- q_{00} \cdot \frac{2}{n} + q_{10} \cdot \Big(1 - \frac{2}{n}\Big) \\
&\qquad + q_{01} \cdot \Big(1 - \frac{2}{n}\Big) + q_{11} \cdot \Big(2 - \frac{2}{n}\Big) \Big) + 6S \cdot \frac{\alpha^2}{n} \Big) \\
&= \Phi^{2t} \cdot \Big(1 + \alpha \cdot \Big(\frac{i-1+n-i+2i}{n^2} - \frac{2}{n} \Big) + 6S \cdot \frac{\alpha^2}{n} \Big) \\
&= \Phi^{2t} \cdot \Big(1 + \alpha \cdot \Big(\frac{2i-1}{n^2} - \frac{1}{n} \Big) + 6S \cdot \frac{\alpha^2}{n} \Big),
\end{aligned}$$

根据（a）中的引理 6.2，对于某个常数 $S := S(\lambda) \geqslant \max(1, 1/\lambda)$。

同理，我们有

$$\mathbf{E}\left[\, \Psi^{2t+2} \,\middle|\, \mathfrak{F}^{2t} \,\right] \leqslant \Psi^{2t} \cdot \Big(1 + \alpha \cdot \Big(\frac{2i-1}{n^2} - \frac{1}{n} \Big) + 6S \cdot \frac{\alpha^2}{n} \Big).$$

注意到 $p_i = \frac{2i-1}{n^2}$ 是双选概率分配向量，根据 [18, 命题 2.3]，它满足条件 $\mathcal{C}_1$，其中 $\delta = \frac{1}{4}$ 和 $\varepsilon = \frac{1}{2}$。因此，在偶数下标处应用定理 4.4，可得存在常数 $c > 0$，使得对于 $\alpha = \frac{1}{384S}$

$$\mathbf{E}\left[\Gamma^{2t+2} \right] \leqslant \Gamma^{2t} \cdot \Big(1 - \frac{\alpha}{384n} \Big) + c \cdot \alpha,$$

这意味着对于任意步骤 $2t \geqslant 0$,

$$\mathbf{E}\left[\Gamma^{2t} \right] \leqslant 384c \cdot n.$$

利用马尔可夫不等式,

$$\mathbf{Pr}\left[\mathrm{Gap}(2t) \leqslant \frac{1}{\alpha} \cdot (3 \log n + \log(384c)) \right] \geqslant 1 - n^{-2}.$$

令 $t := \lceil m/2 \rceil$。若 $m = 2t$，则结论成立；否则，第 $(2t)$ 个球以高概率为 $\mathcal{O}(\log n)$，因为矩母函数是常数。因此，在最后这一步中，差距的变化不超过 $\mathcal{O}(\log n/n)$，从而我们得出 $\kappa := 4/\alpha$ 的结论。

□

7 下界的证明

我们的第一个结果给出了 Memory 过程差距的下界。

定理 1.2（重述）。 考虑具有均匀采样分布的MEMORY 过程。那么存在一个常数 $\kappa > 0$，使得对于每个步骤 $m \geqslant n$,

$$\mathbf{Pr}\left[\operatorname{Gap}(m) \geqslant \kappa \cdot \log\log n\right] \geqslant 1 - n^{-1}.$$

从宏观上看，这一定理的证明沿用了 [3] 在轻负载情形下为双重选择的差距给出下界时所采用的分层归纳论证。然而，对于重负载情形下的 Memory 过程，我们还需要一些额外论证来启动归纳，并处理缓存引入的相关性。

7.1 定理1.2的证明

我们先给出证明中使用的一些一般定义和记号。回顾第 4.1 节中的双曲余弦势函数定义:

$$\Gamma^t = \sum_{i=1}^{n} e^{\alpha y_i^t} + e^{-\alpha y_i^t}.$$

证明的起始点是第 t_0 轮: $= m - n$。为方便起见，我们可以假设 m 是 n 的倍数。这一假设不失一般性，因为若 m 不是 n 的倍数，我们只需将定理应用于不超过 m 且为 n 倍数的最大值 $\widetilde{m} \leqslant m$，即可推出相同的差距界，因为任意箱的（归一化）负载在 n 步内最多只会减少 1。

现在，我们将区间 $[t_0, m]$ 划分为连续的阶段，编号为 1 至 $j_{\max}$: $= \varepsilon \cdot \log\log n$，其中常数 ε 足够小: $= 1/(2\log C)$，且 C: $= 100$。每个阶段持续 ℓ: $= n/j_{\max}$ 步。此外，对于每个阶段 $j \geqslant 1$，我们定义箱子集合

$$B_j := \left\{ i \in [n] \colon x_i^{t_0 + j\cdot\ell} \geqslant \frac{t_0}{n} + L + j \right\},$$

对于某个将在下文定义的常数 L，我们定义事件

$$\mathcal{E}_j := \left\{ |B_j| \geqslant \frac{n}{(\log n)^{C^j}} \right\}.$$

我们还定义事件 $\mathcal{F}$: $= \{\Gamma^{t_0} \leqslant 6Cn\}$ 和 $\mathcal{L} = \{y^{t_0}{}_{\min} \geqslant -\widetilde{c} \cdot \log n\}$，其中 $c, \widetilde{c} > 0$为某些常数。证明定理 1.2 的关键步骤是建立关于事件 $\mathcal{E}_j$的以下两个引理，其中 $0 \leqslant j \leqslant j_{\max}$。

引理 7.1。 我们有 $\mathcal{L} \cap \mathcal{E}_0 \subseteq \mathcal{F}$以及 $\mathbf{Pr}\left[\mathcal{F}\right] \geqslant 1 - n^{-2}$。
引理 7.2。 对于任意 $j \in [1, j_{\max}]$，我们有 **概率** $\left[\mathcal{E}_j \,\middle|\, \mathfrak{F}^{t_0+\ell\cdot\ (j-1)} \cap_{k=0}^{j-1} \mathcal{E}_k,\ \mathcal{F}\right] \geqslant 1 - n^{-\omega(1)}$。

在证明这两个引理之前，我们先完成定理 1.2 的证明。

定理 *1.2* 的证明（假设引理 *7.1* 和引理 *7.2* 成立）。首先注意，对于任意 $1 \leqslant j \leqslant j_{\max}$，若事件 $\mathcal{E}_j$ 成立，则

$$|B_j| \geqslant \frac{n}{(\log n)^{C^j}} \geqslant \frac{n}{(\log n)^{C^{j_{\max}}}} \geqslant \frac{n}{(\log n)^{\sqrt{\log n}}} \geqslant \sqrt{n}. \tag{7.1}$$

因此，如果 $\mathcal{E}_{j_{\max}}$ 成立，则存在一个箱 $i \in [n]$ 满足 $x_i^m \geqslant \frac{t_0}{n} + L + j_{\max}$，于是

$$y_i^m = x_i^m - \frac{m}{n} \geqslant \frac{t_0}{n} + L + j_{\max} - \frac{m}{n} = L + j_{\max} - 1 = \Omega(\log\log n),$$

由 $|L| = \mathcal{O}(1)$ 可知。接下来需要求 $\mathbf{Pr}[\mathcal{E}_{j_{\max}}]$ 的下界。利用引理 7.2，

$$\begin{aligned}\mathbf{Pr}\left[\mathcal{E}_{j_{\max}} \mid \mathfrak{F}^{t_0}, \mathcal{F}\right] &\geqslant \mathbf{Pr}\left[\bigcap_{k=0}^{j_{\max}} \mathcal{E}_k \,\middle|\, \mathfrak{F}^{t_0}, \mathcal{F}\right] \\ &= \mathbf{Pr}\left[\mathcal{E}_{j_{\max}} \,\middle|\, \mathfrak{F}^{t_0}, \mathcal{F}, \bigcap_{k=0}^{j_{\max}-1} \mathcal{E}_k\right] \cdot \mathbf{Pr}\left[\bigcap_{k=0}^{j_{\max}-1} \mathcal{E}_k \,\middle|\, \mathfrak{F}^{t_0}, \mathcal{F}\right] \\ &\geqslant (1 - n^{-\omega(1)})^{j_{\max}} \\ &= 1 - n^{-\omega(1)}.\end{aligned}$$

此外，由引理 7.1 可知，由于 $\mathcal{F}$ 关于 $\mathfrak{F}^{t_0}$ 可测，且 $\mathbf{Pr}[\mathcal{F}] \geqslant 1 - n^{-2}$，

$$\begin{aligned}\mathbf{Pr}[\mathcal{E}_{j_{\max}}] &\geqslant \mathbf{Pr}[\mathcal{E}_{j_{\max}} \cap \mathcal{F}] \\ &\geqslant \mathbf{Pr}[\mathcal{F}] \cdot \mathbf{E}\left[\mathbf{Pr}\left[\mathcal{E}_{j_{\max}} \mid \mathfrak{F}^{t_0},\ \mathcal{F}\right]\right] \\ &\geqslant (1 - n^{-2}) \cdot \left(1 - n^{-\omega(1)}\right) \\ &\geqslant 1 - n^{-1},\end{aligned}$$

至此，证明完成。 □

接下来只需证明引理 7.1 和引理 7.2。首先证明引理 7.1，它涉及事件 $\mathcal{E}_0$。

引理 *7.1* 的证明。 由定理 4.3，存在一些常数 $c, \alpha > 0$，使得对于 $\Gamma := \Gamma(\alpha)$ 和 $\mathcal{F} := \left\{\Gamma^{t_0} \leqslant 6cn\right\}$，有

$$\mathbf{Pr}[\mathcal{F}] \geqslant 1 - n^{-2}.$$

此外，为方便起见，在下文中我们可以假设 c 在 $\mathcal{F}$ 的定义中满足 $c \geqslant 1$。根据 Γ 的定义，事件 $\mathcal{F}$ 意味着对于某个常数 $\widetilde{c} > 0$，有 $y^{t_0}_{\min} \geqslant -\widetilde{c} \cdot \log n$，即 $\mathcal{L} \supseteq \mathcal{F}$。通过一阶矩论证，$\mathcal{F}$ 还意味着存在常数 $c_1 > 0$ 和 $c_2 > 0$，使得在时间 t_0，至少有 $c_1 \cdot n$ 个箱子的负载位于 $[-c_2, +c_2]$。因此，根据抽屉原理，存在一个负载阈值 $L \in [-c_2, c_2]$，使得满足 $y_i^{t_0} = L$（等价地，$x_i^{t_0} = \frac{t_0}{n} + L$）的箱子数量 i 至少为 $c_1/(2c_2+1) \cdot n$。注意，L 可以为正、为负（或为零）；我们只需它属于区间 $[-c_2, c_2]$。定义 $B_0 := \left\{i \in [n] \colon x_i^{t_0} = \frac{t_0}{n} + L\right\}$；于是对于常数 $c_3 := c_1/(2c_2+1)$，$|B_0| \geqslant c_3 \cdot n \geqslant \frac{n}{(\log n)^C}$。 □

接下来，我们证明从 $j-1$ 到 j 的、更复杂的归纳步骤：

引理 *7.2* 的证明。 为了证明这个关键不等式，我们将从步骤 $t_0 + \ell \cdot (j-1)$ 开始进行分析（即阶段 j 的第一步），并假设该步骤的负载（和缓存）配置任意，且满足 $\mathcal{L} \cap \bigcap_{k=0}^{j-1} \mathcal{E}_k$。

考虑阶段 j 中的任意步骤 s。由于 $\mathcal{E}_{j-1}$ 成立，我们有

$$\left|\left\{i \in [n] \colon x_i^s \geqslant \frac{t_0}{n} + L + (j-1)\right\}\right| \geqslant \frac{n}{(\log n)^{C^{j-1}}}.$$

此外，我们可以假设

$$\left|\left\{i \in [n] \colon x_i^s \geqslant \frac{t_0}{n} + L + j\right\}\right| < \frac{n}{(\log n)^{C^j}}, \tag{7.2}$$

si否则，$\mathcal{E}_j$ 成立，结论即得。结合前两个不等式可得 ，

$$\left|\left\{i \in [n] \colon x_i^s = \frac{t_0}{n} + L + (j-1)\right\}\right| \geqslant \frac{n}{(\log n)^{C^{j-1}}} - \frac{n}{(\log n)^{C^j}} \geqslant \frac{n}{2(\log n)^{C^{j-1}}}. \tag{7.3}$$

对于步骤 s 时缓存中的箱 b^s，我们可以推导出$x^s_{b^s} \geqslant x^s_{\min} \geqslant x^{t_0}_{\text{最小}} \geqslant \frac{t_0}{n} - \widetilde{c}\cdot\log n$，因为我们假设事件 $\mathcal{L}$ 成立。

回想一下，在某一轮中，若采样箱的负载严格大于缓存箱的负载，我们就将球分配到缓存箱（且不更新缓存）。此外，如果采样箱的负载至少与缓存箱的负载相同，那么我们必然会增加一个具有该负载的箱子的负载。基于此，定义停止时间

$$\rho := \min\left\{r \geqslant s:\ x^r_{b^s} \geqslant \varphi(j)\right\},$$

其中 $\varphi(j) := \frac{t_0}{n} + L + (j-1)$。注意，

$$\begin{aligned}
&\mathbf{Pr}\left[\rho \leqslant \varphi(j) - x^s_{b^s} \,\middle|\, \mathfrak{F}^s,\ \cap_{k=0}^{j-1}\mathcal{E}_k,\ \mathcal{F}\right] \\
&\quad\geqslant \mathbf{Pr}\left[\bigcap_{r=s}^{\varphi(j)-x^s_{b^s}} \left\{x^r_{i^r} \geqslant x^s_{b^s} + (r-s)\right\} \,\middle|\, \mathfrak{F}^s,\ \cap_{k=0}^{j-1}\mathcal{E}_k,\ \mathcal{F}\right] \\
&\quad\geqslant \prod_{r=s}^{\varphi(j)-x^s_{b^s}} \mathbf{Pr}\left[x^r_{i^r} \geqslant x^s_{b^s} + (r-s) \,\middle|\, \mathfrak{F}^{r-1},\ \cap_{k=0}^{j-1}\mathcal{E}_k,\ \mathcal{F}\right] \\
&\quad\geqslant \prod_{\lambda=x^s_{b_s}-\frac{t_0}{n}}^{L+j-1} p(\lambda).
\end{aligned} \tag{7.4}$$

其中

$$p(\lambda) := \mathbf{Pr}\left[x^{\frac{t_0}{n}+\lambda-x^s_{b_s}}_{i^{\frac{t_0}{n}+\lambda-x^s_{b_s}}} \geqslant \frac{t_0}{n} + \lambda \,\middle|\, \mathfrak{F}^{r-1},\ \cap_{k=0}^{j-1}\mathcal{E}_k,\ \mathcal{F}\right];$$

需要记住的重要一点是，对 $p(\lambda)$ 给出下界，意味着我们是在给采样到负载至少为 $\frac{t_0}{n}+\lambda$ 的箱子的概率给出下界（条件是之前的 $j-1$ 层“好”事件均已发生）。回想一下，由于我们假设 $\mathcal{F} := \left\{\Gamma^{t_0} \leqslant 6cn\right\}$ 成立，因此可推知对任意 $\lambda \in \mathbb{R}$，

$$\left|\left\{i \in [n]\colon y_i^{t_0} < \lambda\right\}\right| \leqslant 6cn\cdot e^{\alpha\lambda},$$

这意味着对任意未来步骤 $r \geqslant t_0$，

$$\left|\left\{i \in [n]\colon x_i^r < \frac{t_0}{n} + \lambda\right\}\right| \leqslant \left|\left\{i \in [n]\colon x_i^{t_0} < \frac{t_0}{n} + \lambda\right\}\right| \leqslant 6cn\cdot e^{\alpha\lambda}.$$

因此，对于任意 $r \geqslant s$，

$$\mathbf{Pr}\left[x^r_{i^r} \geqslant \frac{t_0}{n} + \lambda \,\middle|\, \mathfrak{F}^{r-1},\ \cap_{k=0}^{j-1}\mathcal{E}_k, \mathcal{F}\right] \geqslant 1 - 6c\cdot e^{\alpha\lambda}. \tag{7.5}$$

其次，根据 $\mathcal{F}$ 蕴含 $|B_0| \geqslant c_3\cdot n$ 这一事实，我们还可以得到另一种估计。对于任意 $\lambda \leqslant L$，

$$\begin{aligned}
\mathbf{Pr}\left[x^r_{i^r} \geqslant \frac{t_0}{n} + \lambda \,\middle|\, \mathfrak{F}^{r-1},\ \cap_{k=0}^{j-1}\mathcal{E}_k,\ \mathcal{F}\right] &\geqslant \mathbf{Pr}\left[i^r \in B_0 \,\middle|\, \mathfrak{F}^{r-1},\ \cap_{k=0}^{j-1}\mathcal{E}_k,\ \mathcal{F}\right] \\
&\geqslant \frac{|B_0|}{n} \geqslant c_3 > 0.
\end{aligned} \tag{7.6}$$

最后，对于任意 $1 \leqslant \lambda < j$，我们有

$$\mathbf{Pr}\left[x^r_{i^r} \geqslant \frac{t_0}{n} + L + \lambda \,\middle|\, \mathfrak{F}^{r-1}, \cap_{k=0}^{j-1}\mathcal{E}_k, \mathcal{F}\right] \geqslant \frac{|B_\lambda|}{n} \geqslant (\log n)^{-C^\lambda}. \tag{7.7}$$

现在，我们将依次应用 (7.5)、(7.6) 和 (7.7)，以求得 (7.4) 的下界。为简单起见，假设 L 为正（L 为负的情形类似，而且我们可以得到更强的概率下界）。回到 (7.4) 中负载阈值从 $x^s_{b_s}$ 到 $L+j$ 的乘积，我们将负载值分为三部分：(*i*) 从 $x^s_{b_s}$ 到 $-c_4$；(*ii*) 从 $-c_4+1$ 到 L；以及 (*iii*) 从 $L+1$ 到 $L+j$，其中常数 $c_4>0$ 是满足以下条件的最小整数：

$$6c\cdot\sum_{z=c_4}^{\infty}e^{-\alpha(z-1)}\leqslant 1/2.$$

因此，

$$\begin{aligned}\mathbf{Pr}\left[\,\rho\leqslant\varphi(j)-x^s_{b^s}\;\middle|\;\mathfrak{F}^s,\,\cap_{k=0}^{j-1}\mathcal{E}_k,\,\mathcal{F}\,\right]&=\prod_{\lambda=x^r_{b_s}-\frac{t_0}{n}}^{-c_4}p(\lambda)\cdot\prod_{\lambda=-c_4+1}^{L}p(\lambda)\cdot\prod_{\lambda=L+1}^{L+j-1}p(\lambda)\\&\geqslant\prod_{\lambda=x^r_{b_s}-\frac{t_0}{n}}^{-c_4}\left(1-6c\cdot e^{\alpha\lambda}\right)\cdot c_3^{L+c_4}\cdot\prod_{\lambda=1}^{j-1}(\log n)^{-C^\lambda},\end{aligned}$$

其中，我们利用（7.5）给出了第一个乘积的下界，利用（7.6）给出了第二个乘积的下界，并利用（7.7）给出了第三个乘积的下界。进一步估计这个下界可得，

$$\begin{aligned}\mathbf{Pr}\left[\,\rho\leqslant\varphi(j)-x^s_{b^s}\;\middle|\;\mathfrak{F}^s,\,\cap_{k=0}^{j-1}\mathcal{E}_k,\,\mathcal{F}\,\right]&\geqslant\left(1-6c\cdot\sum_{z=c_4}^{\infty}e^{-\alpha(z-1)}\right)\cdot c_3^{L+c_4}\cdot(\log n)^{-\sum_{\lambda=1}^{j-1}C^\lambda},\\&\geqslant\left(1-6c\cdot\sum_{z=c_4}^{\infty}e^{-\alpha(z-1)}\right)\cdot c_3^{L+c_4}\cdot(\log n)^{-C^{j+1/3}},\\&\geqslant 4\cdot(\log n)^{-C^{j+2/3}},\end{aligned}$$

其中，我们利用了这样一个事实：$C=100$，且倒数第二行中的第一和第二个因子都是严格大于 0 的常数。

注意，$\varphi(j)-x^s_{b_s}\leqslant c\log n+L+j-1\leqslant 2c\log n$，因此，

$$\mathbf{Pr}\left[\,\rho\leqslant 2c\log n\;\middle|\;\mathfrak{F}^t,\mathcal{E}_0\,\right]\geqslant 4\cdot(\log n)^{-C^{j+2/3}}.\tag{7.8}$$

因此，从阶段 j 中的任意步骤 s 开始，以至少 $4\cdot(\log n)^{-C^{j+2/3}}$ 的概率，在至多再经过 $2c\log n$ 个步骤后，我们将达到这样一种情形：缓存中的箱的负载至少为 $x^r_{b_r}\geqslant\varphi(j)=\frac{t_0}{n}+L+j-1$。

回顾一下，根据（7.2），我们假设至少有 $2^{n/(\log n)^{Cj}}$ 个箱的负载为 $\frac{t_0}{n}+L+(j-1)$（见（7.3）），这意味着下一个采样箱的负载至少为 $\frac{t_0}{n}+L+(j-1)$ 的概率为 $2^{-n/(\log n)^{Cj}}$，因此有二个箱的负载达到 $\frac{t_0}{n}+L+j$。

设 $X_1,\ X,\ \ldots\ldots,\ X_z$ 为相互独立的伯努利随机变量，其成功概率为 $z:=n/(j_{\max}\cdot(2c\log n+1))$

$$\mathbf{Pr}\left[\,\rho\leqslant 2c\log n\;\middle|\;\mathfrak{F}^t,\mathcal{E}_0\,\right]\cdot\frac{1}{2(\log n)^{Cj}}\geqslant 4\cdot(\log n)^{-C^{j+2/3}}\cdot\frac{1}{2(\log n)^{Cj}}\geqslant 2\cdot(\log n)^{-C^{j+3/4}},$$

其中，第一个不等式由 (7.8) 得出，最后一个不等式在 n 足够大时成立，因为 $C=100$。令 $X=\sum_{i=1}^{z}X_i$。同样，由于我们固定了 $C=100$，并且可以假设 n 足够大，因此有

$$\mathbf{E}\left[\,X\,\right]:=\frac{n}{j_{\max}(2c\log n)}\cdot\frac{2}{(\log n)^{Cj+3/4}}\geqslant 2\cdot\frac{n}{(\log n)^{Cj+1}}.$$

由于 $j \leqslant j_{\max}$，且 $j_{\max} \leqslant \varepsilon \cdot \log\log n$，$\varepsilon = 1/(2\log C)$，我们有 $\mathbf{E}[X] = \Omega(\mathrm{poly}(n))$（由 (7.1) 得出），因此根据 Chernoff界，

$$\mathbf{Pr}\left[X \geqslant \frac{n}{(\log n)^{C^{j+1}}}\right] \geqslant \mathbf{Pr}[X \geqslant 1/2 \cdot \mathbf{E}[X]] \geqslant 1 - n^{-\omega(1)}.$$

这意味着，在阶段 j 的某一步骤中，至少有 $\frac{n}{(\log n)^{C^{j+1}}}$ 个箱子的负载至少为 $\frac{t_0}{n} + L + j$，由此完成归纳步骤。因此，我们证明了该断言。 □

7.2 d-Weak-Memory 的下界

我们还给出一个简单论证，证明运行段长度为常数 d 的 d-Weak-Memory 的差距下界。

引理 7.3. 对于具有常数 $d > 0$ 且满足 $m = \frac{1}{400d} n \log n$ 的 d-Weak-Memory过程，我们有

$$\mathbf{Pr}\left[\mathrm{Gap}(m) \geqslant \frac{1}{400d} \cdot \log n\right] \geqslant 1 - n^{-2}.$$

证明。在 d-弱 Memory 中，缓存每隔 d 步就会重置，因此球会通过单选进行分配。因此，在 m 步内，有 m/d 个球通过单选进行分配。例如，根据 [27, 第 4] 节（另见 [20, 引理 A.9]），当有 $cn\log n$ 个球通过单选进行分配时，对于任意常数 $c > 0$，最大负载以至少 $1 - n^{-2}$ 的概率不小于 $(c + \sqrt{c/10})\log n$。因此，我们得到

$$\mathbf{Pr}\left[\max_{i\in[n]} y_i^m \geqslant \left(\frac{1}{400d^2} + \frac{1}{200d}\right) \cdot \log n\right] \geqslant 1 - n^{-2}.$$

因此，由于在第 m 步，平均值为 $(\log n)/(400d)$，

$$\mathbf{Pr}\left[\mathrm{Gap}(m) \geqslant \frac{1}{400d} \cdot \log n\right] \geqslant 1 - n^{-2}.$$ □

8 结论

在这项工作中，我们对重负载情形 $m \geqslant n$ 下的 Memory 进行了渐近紧确的分析。我们证明，Memory 的差距为 $\Theta(\log\log n)$，其性能与双重选择相差至多一个常数因子。与双重选择不同，我们表明，即使在抽样分布可能被任意大的常数因子扭曲的异质环境中，Memory 仍能良好运行。我们还分析了其他放宽的设定，包括球带有权重的情形。在这些设定中，缓存每隔固定数量的步骤就会重置（d-弱 Memory）。对于这些情形，我们证明，差距仍为 $\mathcal{O}(\log n)$，且与 m 无关。

还有几个有趣的研究方向。其中之一是考虑更加偏斜的采样分布，例如重尾分布，如幂律分布（与 [8], 中研究的$m = n$ 类似）。人们可能会猜想，Memory 仍然能够胜过 d-Choice。特别是，我们已经证明，即使面对最不规则的采样分布，Memory 也能将差距控制在有界范围内（定理 1.3）。若能针对幂律分布等特定的、值得关注的采样分布，获得 Memory 更精确的差距界，将会很有意思。

第二个研究方向是确定差距界中的首项常数。[25] 关于 $m = n$ 的研究结果表明，即使在重负载情形下，Memory 也可能略胜于双重选择。

另一个研究方向是探讨在 d 可能是 n 的函数时，d 对 d-Weak-Memory（或 d-Reset-Memory）的影响。例如，d 的最小值是多少，才能实现

$\mathcal{O}(\log \log n)$ 的差距？目前，我们只知道 d 不能是常数，但还不知道 d 是否需要是对数级或多项式级。

最后，据我们所知，所有 Memory 的实现（包括我们的实现）都采用贪心规则，即始终将球分配到缓存和采样箱中负载最小的选项，并以同样的方式更新缓存。需要注意的是，当缓存容量较大且球带有权重时，这种做法未必最优，因为还有更复杂的策略会主动尝试让缓存中至少保留一个明显欠载的箱。

参考文献

[1] Dan Alistarh、Trevor Brown、Justin Kopinsky、Jerry Zheng Li 和 Giorgi Nadiradze。2018年。《分布式线性化数据结构》。载于第*30*届 *ACM*并行算法与体系结构年会（*SPAA'18*）。ACM，133–142。doi[2] Yossi Azar、Andrei Z. Broder、Anna Karlin、Michael Mitzenmacher 和 Eli Upfal。2020年。ACM Paris Kanellakis 理论与实践奖。https://www.acm.org/media-center/ 2021/may/technical-awards-2020.[3] Yossi Azar、Andrei Z. Broder、Anna R. Karlin 和 Eli Upfal。1999年。平衡分配。*SIAM*计算杂志29，1（1999年），180–200。doi[4] Petra Berenbrink、André Brinkmann、Tom Friedetzky 和 Lars Nagel。2012年。相关随机选择下的球入桶。并行与分布式计算杂志72，2（2012年），246–253。doi[5] Petra Berenbrink、André Brinkmann、Tom Friedetzky 和 Lars Nagel。2014年。球入非均匀桶。并行与分布式计算杂志74，2（2014年），2065–2076。doi[6] Petra Berenbrink、Artur Czumaj、Angelika Steger 和 Berthold Vöcking。2006年。平衡分配：重载情形。*SIAM*计算杂志35，6（2006年），1350–1385。doi[7] Tom Bohman、Alan Frieze 和 Eyal Lubetzky。2015年。随机三角形删除。数学进展280（2015年），379–438。doi[8] John W. Byers、Jeffrey Considine 和 Michael Mitzenmacher。2004年。两次选择力量的几何推广。载于第*16*届 *ACM*并行算法与体系结构研讨会（*SPAA'04*）。ACM，54–63。doi[9] Fan Chung 和 Linyuan Lu。2006年。集中不等式与鞅不等式：综述。*Internet Math.* 3，1（2006年），79–127。http://projecteuclid.org/euclid.im/ 1175266369[10] Devdatt P. Dubhashi 和 Alessandro Panconesi。2009年。随机化算法分析中的测度集中。剑桥大学出版社，剑桥。xvi+196页。doi[11]Paolo Giaccone、Balaji Prabhakar 和 Devavrat Shah。2002年。面向高聚合带宽交换机的简单高性能调度器。载于第*21*届 *IEEE*计算机通信会议（*INFOCOM'02*）。IEEE计算机学会，1160–1169。[12] Richard J. Gibbens、Frank P. Kelly 和 Peter B. Key。1988年。动态备用路由——建模与行为。载于第*12*届国际电信业务大会。爱思唯尔，阿姆斯特丹。

[13] Brighten Godfrey。2008年。具有结构的球与箱子：超图上的均衡分配。载于第*19*届*ACM-SIAM*离散算法年度研讨会（*SODA'08*）。ACM，511–517。[14] Brighten Godfrey 和 Ion Stoica。2005年。分布式哈希表中的异质性与负载均衡。载于第*24*届 *IEEE*计算机通信会议（*INFOCOM'05*）。IEEE，596–606。doi[15]Catherine Greenhill、Bernard Mans 和 Ali Pourmiri。2020年。动态超图上的平衡分配。载于第*24*届随机化与计算国际研讨会（*RANDOM'20*）（莱布尼茨国际信息学会议录（*LIPIcs*），第*176*卷）。Schloss Dagstuhl–Leibniz信息学中心，11:1–11:22。doi[16] Richard M. Karp、Michael Luby 和 Friedhelm Meyer auf der Heide。1996年。在分布式内存机器上高效模拟PRAM。$Algorithmica$ 16，4–5（1996年），517–542。doi[17] Adam Kirsch、Michael Mitzenmacher 和 Udi Wieder。2009/10年。更鲁棒的哈希：带暂存区的布谷鸟哈希。*SIAM*计算杂志39，4（2009/10年），1543–1561。doi[18] Dimitrios Los 和 Thomas Sauerwald。2022年。批量平衡分配：简化与推广。载于第*34*届 *ACM*并行算法与体系结构年会（*SPAA'22*）。ACM，389–399。doi[19] Dimitrios Los 和 Thomas Sauerwald。2022年。信息不完整条件下的平衡分配：两次查询的力量。载于第*13*届理论计算机科学创新会议（*ITCS'22*）（莱布尼茨国际信息学会议录（*LIPIcs*），第*215*卷）。Schloss Dagstuhl – Leibniz信息学中心，103:1–103:23。doi[20] Dimitrios Los 和 Thomas Sauerwald。2022年。带噪声选择的平衡分配。载于第*41*届*ACM-SIGOPT*分布式计算原理年会（*PODC'22*）（*PODC'22*）。ACM，164–175。doi[21] Dimitrios Los、Thomas Sauerwald 和 John Sylvester。2022年。平衡分配：缓存与装箱、成对化与稀疏化。载于第*33*届 *ACM-SIAM*离散算法年度研讨会（*SODA'22*）。SIAM，1847–1874。doi[22] Dimitrios Los、Thomas Sauerwald 和 John Sylvester。2023年。异构箱中的均衡分配：记忆的力量。载于*2023*年 *ACM-SIAM*离散算法研讨会论文集（*SODA 2023*）。SIAM，4448–4477。doi[23] Malwina J. Luczak 和 James R. Norris。2013年。马尔可夫链流体极限中对快速变量取平均：带记忆的超市模型的应用。应用概率年刊23，3（2013年），957–986。[24] Albert W. Marshall、Ingram Olkin 和 Barry C. Arnold。2011年。不等式：主序理论及其应用（第二版）。Springer，纽约。xxviii+909 页。doi[25]Michael Mitzenmacher、Balaji Prabhakar 和 Devavrat Shah。2002年。带记忆的负载均衡。载于第*43*届 *IEEE*计算机科学基础年会（*FOCS'02*）。IEEE，799–808。doi[26] Michael Mitzenmacher、Andréa W. Richa 和 Ramesh Sitaraman。2001年。两个随机选择的力量：技术与成果综述。载于随机计算手册，第*I*、*II*卷。组合优化，第9卷。Kluwer学术出版社，多德雷赫特，255–312。doi[27] Yuval Peres、Kunal Talwar 和 Udi Wieder。2015年。图上的均衡分配与（$1+\beta$）选择过程。随机结构与算法47，4（2015年），760–775。doi

[28] Tiberiu Popoviciu。1935年。关于所有根均为实根的代数方程。*Mathematica*（克卢日）9（1935年），129–145。[29] Devavrat Shah 和 Balaji Prabhakar。2002年。随机负载均衡中记忆的作用。载于*IEEE*国际信息论研讨会（*ISIT'02*）。125。[30] Kunal Talwar 和 Udi Wieder。2014年。均衡分配：重负载情形的简单证明。载于第*41*届国际自动机、语言与编程会议（*ICALP'14*）（《计算机科学讲义》第*8572*卷）。Springer，979–990。doi[31] Berthold Vöcking。1999年。非对称性如何助力负载均衡。载于第*40*届*IEEE*计算机科学基础年会（*FOCS'99*）。IEEE，131–141。doi[32] Udi Wieder。2007年。异质箱的均衡分配。载于第*19*届*ACM*并行算法与体系结构年会（*SPAA'07*）。ACM，188–193。doi[33] Udi Wieder。2017年。哈希、负载均衡与多重选择。计算机科学理论基础与趋势12，3-4（2017年），275–379。doi

A 基本工具与不等式

本节将陈述若干辅助引理，这些引理将在全文中使用 r.

A.1 凸性

为完整起见，我们定义舒尔凸性（见 [24]），并陈述两个基本结果：

定义 A.1 (参见 [24, 定义 A.1])。 如果对于任意非递减的 $x, y \in \mathbb{R}^n$，当 x 主序于 y 时，函数 $f : \mathbb{R}^n \to \mathbb{R}$ 就称为舒尔凸函数： $f(x) \geqslant f(y)$。

引理 A.2 (参见 [24, 命题 C.1])。 设 $g \colon \mathbb{R} \to \mathbb{R}$ 是一个凸函数。那么 $g(x_1, \ldots, x_n) := \sum_{i=1}^{n} g(x_i)$ 是舒尔凸的。

A.2 不等式

接下来给出两个简单的不等式，以确保内容完整。

引理 A.3。 对于任意 $0 < z \leqslant 1$，有

$$\frac{1}{1+z} \leqslant 1 - \frac{z}{2}.$$

证明。 对于 $z \in (0, 1]$, ，有如下推论链成立

$$\frac{1}{1+z} \leqslant 1 - \frac{z}{2} \Leftrightarrow 1 \leqslant 1 + z - \frac{z}{2} - \frac{z^2}{2} \Leftrightarrow z \leqslant 1. \qquad \square$$

引理 A.4。 对于任意 $0 < z \leqslant 1/2$，有

$$\frac{1}{1-z} \leqslant 1 + 2z.$$

证明。 对于 $z \in (0, 1]$, ，有如下推论链成立

$$\frac{1}{1-z} \leqslant 1 + 2z \Leftrightarrow 1 \leqslant 1 + 2z - z - 2z^2 \Leftrightarrow z \leqslant \frac{1}{2}. \qquad \square$$

引理 A.5。 对于任意 $r < 1$，有 $\sum_{i=0}^{\infty}(i+1) \cdot r^i = \frac{1}{1-r} + \frac{r}{1-r()^2}$。

证明。 对于第一个求和，注意到对于任意 d,

$$(1-r)\sum_{i=0}^{d}(i+1) \cdot r^i = \sum_{i=0}^{d}(i+1) \cdot r^i - \sum_{i=1}^{d+1} i \cdot r^i = 1 + \sum_{i=1}^{d} r^i + (d+1)r^{d+1}.$$

因此，令 $d \to \infty$，并利用几何级数的求和公式 $\sum_{i=1}^{\infty} r^i = \frac{r}{1-r}$，可得

$$(1-r)\sum_{i=0}^{\infty}(i+1) \cdot r^i = 1 + \frac{r}{1-r},$$

如前所述。 $\square$

A.3 概率不等式

为方便起见，我们陈述并证明以下众所周知的结果。

引理 A.6. 考虑一个随机变量序列 $(Z_i)_{i\in\mathbb{N}}$，使得存在 $0 < a < 1$ 和 $b > 0$，满足每个 $i \geqslant 1$,

$$\mathbf{E}[\,Z_i \mid Z_{i-1}\,] \leqslant Z_{i-1} \cdot a + b.$$

那么对每个 $i \geqslant 1$,

$$\mathbf{E}[\,Z_i \mid Z_0\,] \leqslant Z_0 \cdot a^i + \frac{b}{1-a}.$$

证明. 我们将通过归纳法证明，对每个 $i \in \mathbb{N}$,

$$\mathbf{E}[\,Z_i \mid Z_0\,] \leqslant Z_0 \cdot a^i + b \cdot \sum_{j=0}^{i-1} a^j.$$

对于 $i = 0$，$\mathbf{E}[\,Z_0 \mid Z_0\,] \leqslant Z_0$。假设归纳假设对某个 $i \geqslant 0$ 成立，那么由于 $a > 0$,

$$\begin{aligned}
\mathbf{E}[\,Z_{i+1} \mid Z_0\,] &= \mathbf{E}[\,\mathbf{E}[\,Z_{i+1} \mid Z_i\,] \mid Z_0\,] \leqslant \mathbf{E}[\,Z_i \mid Z_0\,] \cdot a + b \\
&\leqslant \left(Z_0 \cdot a^i + b \cdot \sum_{j=0}^{i-1} a^j\right) \cdot a + b \\
&= Z_0 \cdot a^{i+1} + b \cdot \sum_{j=0}^{i} a^j.
\end{aligned}$$

利用对于 $a \in (0,1)$，$\sum_{j=0}^{\infty} a^j = \frac{1}{1-a}$，即可得到所述结论。 □

引理 A.7。考虑任意一列随机变量$(Z_i)_{i\in\mathbb{N}}$，使得存在 $0 < \alpha < 1$ 和 $b > 0$

$$\mathbf{E}[\,Z_i \mid Z_{i-1}\,] \leqslant Z_{i-1} \cdot (1-\alpha) + b.$$

那么，假设 $Z_0 \leqslant \frac{b}{\alpha}$，则对于任意 $i \geqslant 0$,

$$\mathbf{E}[\,Z_i\,] \leqslant \frac{b}{\alpha}.$$

证明。我们将通过归纳法证明该结论。基础情形由假设 $Z_0 \leqslant \frac{b}{\alpha}$ 得出。接下来，假设该结论对 $i \geqslant 0$ 成立，则对于 $i+1$，我们有：

$$\begin{aligned}
\mathbf{E}[\,Z_{i+1}\,] &= \mathbf{E}[\,\mathbf{E}[\,Z_{i+1} \mid Z_i\,]\,] \\
&\leqslant \mathbf{E}[\,Z_i \cdot (1-\alpha) + b\,] \\
&= \mathbf{E}[\,Z_i\,] - \alpha \cdot \mathbf{E}[\,Z_i\,] + b \\
&\leqslant \mathbf{E}[\,Z_i\,] - \alpha \cdot \frac{b}{\alpha} + b = \mathbf{E}[\,Z_i\,].
\end{aligned}$$ □

引理 A.8 (上鞅的 Azuma不等式 [10, 问题 6。5])。设 $X_0, \ldots, X_n$ 为一个上鞅，对任意 $i \in [n]$ 均满足 $|X_i - X_{i-1}| \leqslant c_i$，则对任意 $\lambda > 0$,

$$\mathbf{Pr}[\,X_n \geqslant X_0 + \lambda\,] \leqslant \exp\left(-\frac{\lambda^2}{2 \cdot \sum_{i=1}^{n} c_i^2}\right).$$

A.4 带有坏事件的集中不等式

我们分析中的一个核心工具是 Chung 和 Lu [9]针对上鞅提出的集中不等式，该不等式以坏事件不发生为条件。在这种情况下，坏事件将是对所有 $s \in [t - T_r, t]$, $\Phi_1^s = \text{poly}(n)$。

我们从 [9]中的以下定义开始。考虑任意随机变量 X （在我们的情形中，它将是势函数 Γ_2），该变量可通过由有限个输出（分配的球所选定的箱子）构成的决策序列 Y_1, Y_2, $\ldots$, Y_N来求值。我们可以用决策树 T描述这一过程，它是一棵深度为 N、顶点集为 $V(T)$ 的完全有根树。T 的每条边（u, v）都关联一个概率 $p_{u,\,v}$，该概率取决于从 u 到 v所作的决策。

我们称 $f\colon V(T) \to \mathbb{R}$ 满足容许条件 P，如果对每个顶点$v \in V(T)$, $P = \{P_v\}$ 都成立。对于容许条件 P，其在 X^i上的对应坏集 B^i定义为

$$B^i = \{v \mid v \text{ 的深度为 } i, \text{ 并且对于 } v\}.\text{的某个祖先 } u, \ P_u \text{ 不成立}$$

定理 A.9 (定理 8.5，见 [9])。 对于一个滤过 $\mathfrak{F}$, $\{\emptyset, \Omega\} = \mathfrak{F}^0 \subset \mathfrak{F}^1 \subset \ldots \subset \mathfrak{F}^N = \mathfrak{F}$，假设随机变量 X^s 是 $\mathfrak{F}^s$-可测的，其中 $0 \leqslant s \leqslant N$。令 $B = B^N$ 为与以下容许条件相关联的坏集：

$$\begin{aligned}
\mathbf{E}\left[X^s \mid \mathfrak{F}^{s-1}\right] &\leqslant X^{s-1},\\
\mathbf{Var}\left[X^s \mid \mathfrak{F}^{s-1}\right] &\leqslant \sigma_s^2,\\
X^s - \mathbf{E}\left[X^s \mid \mathfrak{F}^{s-1}\right] &\leqslant a_s + M,
\end{aligned}$$

对于某些 $\sigma_s > 0$ 和 $a_s > 0$。那么，对于任意 $\lambda > 0$，我们有：

$$\mathbf{Pr}\left[X^N \geqslant X^0 + \lambda\right] \leqslant \exp\left(-\frac{\lambda^2}{2(\sum_{s=1}^N(\sigma_s^2 + a_s^2) + M\lambda/3)}\right) + \mathbf{Pr}\left[B\right].$$

特别地，我们总会用到以下简化版本：

定理 A.10 (定理 A.9 的推论)。 对于一个滤过 $\mathfrak{F}$, $\{\emptyset, \Omega\} = \mathfrak{F}^0 \subset \mathfrak{F}^1 \subset \ldots \subset \mathfrak{F}^N = \mathfrak{F}$，假设随机变量 X^s 是 $\mathfrak{F}^s$-可测的，其中 $0 \leqslant s \leqslant N$。令 $B = B^N$ 为与以下容许条件相关联的坏集：

$$\begin{aligned}
\mathbf{E}\left[X^s \mid \mathfrak{F}^{s-1}\right] &\leqslant X^{s-1},\\
|X^s - X^{s-1}| &\leqslant D,
\end{aligned}$$

对于某个 $D > 0$。那么，对于任意 $\lambda > 0$,

$$\mathbf{Pr}\left[X^N \geqslant X^0 + \lambda\right] \leqslant \exp\left(-\frac{\lambda^2}{10ND^2}\right) + \mathbf{Pr}\left[B\right].$$

证明。 我们将证明，有界差分条件 $|X^s - X^{s-1}| \leqslant D$ 蕴含定理 A.10 中的第二和第三个条件。

对于第二个条件，我们将使用方差的波波维丘斯不等式（引理 A.11），该不等式指出，对于任意满足 $a \leqslant Y \leqslant b$ 的随机变量 Y，

$$\mathbf{Var}\left[Y\right] \leqslant \frac{1}{4} \cdot (b - a)^2.$$

有界差分条件意味着

$$X^{s-1} - D \leqslant X^s \leqslant X^{s-1} + D.$$

结合两者，我们得到

$$\mathbf{Var}[\,X^s \mid \mathfrak{F}^{s-1}\,] \leqslant \frac{1}{4}\cdot\left(X^{s-1}+D-\left(X^{s-1}-D\right) \mid \mathfrak{F}^{s-1}\right)^2 = D^2.$$

对于第三个条件，

$$X^s - \mathbf{E}\left[\,X^s \mid \mathfrak{F}^{s-1}\,\right] \leqslant X^{s-1}+D-(X^{s-1}-D) \leqslant 2D.$$

最后，根据定理 A.10，结合 $a_s = 2D$、$\sigma_s = D$ 和 $M =$，即可得出结论 $0.$ □

引理 A.11 (波波维丘斯不等式 [28])。对于满足 $a \leqslant Y \leqslant b$ 的任意随机变量 Y，我们有

$$\mathbf{Var}\,[\,Y\,] \leqslant \frac{1}{4}\cdot(b-a)^2.$$