arXiv:2304.02063v1 [cs.DS] 4 Apr 2023

睁大双眼进行集合覆盖

Anupam Gupta* Gregory Kehne† Roie Levin‡

摘要

在随机集合覆盖问题（Grandoni 等人，FOCS '08）中，给定全集 $\mathcal{U}$ 上的一个集合 $\mathcal{S}$，其中包含 m 个集合，全集大小为 N，以及定义在 $\mathcal{U}$ 的元素上的一个分布 D。算法从 D 中逐个抽取 n 个元素，并且必须在每个元素到达时购买一个集合来覆盖它；目标是最小化这一过程中购买的集合的总成本。通用算法预先将每个元素 $u \in \mathcal{U}$ 映射到一个集合 $S(u)$，使得如果 $U \subseteq \mathcal{U}$ 是通过从分布 D 中抽取 n 次形成的，那么算法就承诺输出 $S(U)$。Grandoni 等人为这一随机集合覆盖问题给出了一个 $O(\log mN)$-竞争的通用算法。

我们通过针对更一般的先知版本给出一个简单的、多项式时间的 $O(\log mn)$-竞争通用算法，在单方面改进这一结果。在该版本中，U 由从 n 个不同分布 D_1、……、D_n 中抽取形成。此外，我们证明无需完整的先验知识即可获知这些分布：事实上，每个分布只需一个样本即可。我们还针对两阶段先知设定和带样本在线设定证明了类似的结果。

我们通过从单样本先知设定到随机顺序设定的通用归约得到这些结果；这一归约适用于一大类最小化问题，其中包括所有覆盖问题。我们利用这一框架，为非度量设施选址和集合多重覆盖给出随机顺序算法；借助该框架，这些算法可自动转化为通用先知算法。

*卡内基梅隆大学计算机科学系，匹兹堡，宾夕法尼亚州 15213。电子邮箱：anupamg@cs.cmu.edu。研究部分由美国国家科学基金会资助，项目编号为 CCF-1955785、CCF-2006953 和 CCF-2224718。

†哈佛大学工程与应用科学学院，波士顿，马萨诸塞州 02138。电子邮箱：gkehne@g.harvard.edu。研究部分由 Siebel 学者计划资助。

‡特拉维夫大学数学科学学院统计与运筹学系，特拉维夫。电子邮箱：roiel@tauex.tau.ac.il。研究部分由富布赖特以色列博士后奖学金、以色列科学基金会项目（编号 2233/19）以及美以双边科学基金会项目（编号 2018352）资助。

1 引言

在 SetCover 问题中，给定一个集合系统（$\mathcal{U}$，$\mathcal{S}$），其中 $\mathcal{U}$是大小为 N的底集，而 $\mathcal{S}$ 是一个子集集合，满足 $|\mathcal{S}| = m$。此外，还给定一个大小为 n的子集 $U \subseteq \mathcal{U}$。目标是选取一个最小规模（或更一般地，成本最低）的子集集合 $\mathcal{S}' \subseteq \mathcal{S}$，使得 $\mathcal{S}'$ 中的集合之并为 U。针对这一问题，人们已提出许多多项式时间算法，其近似比达到 $\ln n$（例如参见 [Chv79、Joh74、Lov75、WS11]）；除非 P =NP [Fei98、DS14]，否则这一结果已是最优的。

可以将一个解 $\mathcal{S}'$ 理解为一个映射 $\mathfrak{S}$: $\mathcal{U} \to \mathcal{S}$ 将每个元素映射到一个覆盖它的集合（若存在多个选择，则任意打破平局）。在这种情况下，$\mathfrak{S}(U) = \bigcup_{u \in U}\{\mathfrak{S}(u)\}$ 就是该解 $\mathcal{S}'$。在开创性的研究中，Jia 等人定义了集合覆盖问题的通用变体，其目标是在不了解 $U \subseteq \mathcal{U}$ 实际取值的情况下，先验且不依赖具体实例地构造 $\mathfrak{S}$（因此构造时仅使用 $\mathcal{U}$和 $\mathcal{S}$）[JLN+05]。我们希望找到一个映射 $\mathfrak{S}$，使最坏情况下的比值 $\max_{U \subseteq \mathcal{U}} c(\mathfrak{S}(U))/c(\mathrm{Opt}(U))$ 最小；该比值是 $\mathfrak{S}(U)$ 的成本与针对 U 的最优集合覆盖的成本之比。如果一个通用算法满足该比值不超过 α，则称其为 α-竞争的，或称其达到竞争比 α。Jia 等人证明了该问题的 $\tilde{\Theta}(\sqrt{n})$ 界 [JLN+05]。

为克服这一多项式障碍，Grandoni 等人 [GGL+13] 研究了集合覆盖的 随机 变体，其中还假设 U 中的元素是从一个已知分布 D中独立同分布地抽取的。现在的目标是最小化期望竞争比 $E_U[c(\mathfrak{S}(U))]/E_U[c(\mathrm{Opt}(U))]$。在这一假设下，他们证明了可以获得竞争比在指数意义上更优的 $O(\log(mN))$-竞争算法，并且除 log log 因子外，这是可能达到的最佳结果。

在本文中，我们改进、推广并简化了 [GGL+13] 的结果。首先，我们将竞争比改进为 $O(\log(mn))$，当 $n \ll N$ 时，这一结果可以在指数意义上更小。对于多项式时间算法而言，这基本上是可能达到的最佳结果，因为当 $n \ll m$ [GGL+13] 时存在 $\Omega(\log m/\log\log m)$ 下界，并且除非 P = NP [Fei98, DS14]，不存在近似比为 $o(\log n)$ 的多项式时间算法。我们还将结果推广到先知设定，其中 U 由来自一列非同一分布 D^1、...、D^n 的抽样组成。事实上，我们证明不需要完全了解这些分布，只需从每个分布中获得一个样本即可。我们还展示了对另外两种相关模型的扩展，即两阶段先知设定和带样本的在线设定，以及对若干推广 SetCover 和覆盖的问题的扩展。现在，我们将对这些结果作出更正式的概述。

1.1 我们的结果

我们的主要贡献是将先知设定归约到随机顺序在线设定。在随机顺序 SetCover中，元素由对手选择，并逐个揭示。算法必须选择一个集合来覆盖该元素，且决策不可撤销。由于 LearnOrCover算法 [GKL21] 是一个用于随机顺序 $O(\log(mn))$-竞争算法SetCover，我们立即得到：

定理 1.1. 存在一个多项式时间的 $O(\log(mn))$-竞争通用算法，适用于 1- sampleprophet SetCover。

使用类似的技术，我们也为以下两种模型得到了定理。在两阶段先知模型中，算法可以在游戏开始前的第一阶段以折扣价购买集合。在第二阶段，U 按照通常的先知设定随机抽取，而在看到实现值后购买的任何集合都需要支付全价。

定理 1.2。存在一个多项式时间通用算法，适用于两阶段先知 SetCover，其竞争比相对于最优在线策略为 $O(\log(mn))$。
在带样本的在线设定中，对手选定一个未知的元素集合 U，并向算法揭示其中均匀随机抽取的 α 比例。此后，剩余元素将按对手指定的顺序逐一揭示。算法必须在元素到达时立即购买集合以覆盖该元素，且决策不可撤销。

定理1.3. 对于每个 $O < \alpha \leq 1$，都存在一个多项式时间的 $O(\log(mn)/\alpha)$ 竞争比通用算法，用于带样本在线 SetCover。
事实上，我们的归约适用于一类更一般的最小化问题，我们称之为可增广整数规划(AIPs)。对于这类问题，随着z 增大，将任意不可行解增广为可行解的代价会降低。 z 增大[1]。因此，我们将设计先知/两阶段先知/带样本在线算法来解决 AIPs 的任务，归约为设计随机顺序算法的任务。为展示这些归约，我们还给出了随机顺序下针对 $O(\log mn)$-竞争比的算法，用于 NonMetricFacilityLocation和 SetMultiCover 。这为回答 GKL 提出的一个开放问题取得了部分进展：对于带盒约束的覆盖整数规划，随机顺序下是否存在一个 [GKL21], 曾提出这一问题：是否存在一个 $O(\log mn)$-竞争比算法?

1.2 技术与概述

定理 1.1 至 1.3 的证明采用了一个共同的模板。主要思路是将设置 $\mathcal{X}$ 归约为随机顺序 SetCover，具体如下:

1. 生成一个模拟实例 $I \sim \mathcal{X}$。2. 通过人为打乱顺序，在模拟输入 I 上模拟随机顺序 $\mathcal{A}$ 算法 SetCover。3. 在真实输入 $\mathcal{I} \sim \mathcal{X}$ 上求解 $\mathcal{X}$：首先购买 $\mathcal{A}$所购买的解 z，然后使用包含任意尚未覆盖的未覆盖元素 v^t的最便宜集合 v^t 进行覆盖。

我们的思路是，将为覆盖 $\mathcal{A}$ 遗漏的元素而购买的“备份”集合的费用，计入 $\mathcal{A}$实际作出的决策，并利用 $\mathcal{A}$上的性能保证对其进行界定。这一视角使我们能够给出事后看来十分简单的证明；在我们开展工作之前，人们还不知道如何获得这样的结果。我们将在第 3 至 5节分别处理每种模型。

在第 6 和 7 节中，我们给出 NonMetricFacilityLocation和SetMultiCover的随机顺序算法。这体现了我们归约方法的普适性，也展示了 [GKL21] 提出的“学习或覆盖”框架在纯覆盖问题之外的多功能性。这些是本文技术性最强的章节。

这些结果建立在 LearnOrCover由 [GKL21] 为集合覆盖设计的框架之上。每当一个元素到达时若尚未被覆盖， LearnOrCover(a) 从集合上的分布中采样，并且 (b) 利用随机元素未被覆盖这一事实进行学习，以更新该分布。[GKL21] 表明，该算法要么在学习方面取得进展，即了解应从中进行采样的最优分布；要么在采样方面取得进展，因为该分布已经足够好。他们使用了两部分势函数，分别衡量学习和覆盖方面的进展。

非度量设施选址通常被视为集合覆盖的一种扩展，因为两者之间存在标准的

[1]GGLS [研究过这类问题08] ，但没有为其命名。从直觉上看，这些问题与 *covering* 问题有关；我们将“覆盖”一词保留用于可行区域向上封闭的问题。详情请见第2 节。

归约（[Vyg05，第 3 节。1] 以及 [KT84]）。然而，这些归约在随机顺序模型中并不成立。第一个从 NonMetricFacilityLocation到 SetCover的归约属于民间常识，但会使集合数量呈指数级膨胀。由于一般来说，随机顺序在线集合覆盖至少会损失一个 $\Omega(\log m)$ 因子，这种膨胀难以承受[GKL21]。第二个归约为每个设施-客户对引入一个新集合和一个新元素；因此，按随机顺序到达的客户会变成一批新集合和新元素。然而，元素以随机顺序的批次到达的在线集合覆盖问题通常比真正的均匀随机顺序更难，因为[GKL21] 证明了该问题存在双对数下界。因此，这两种归约在随机顺序设定中都面临障碍，需要采用新的方法。

随机顺序 FacilityLocation 面临的主要挑战是如何考虑连接成本。这使得学习设施分布的任务变得复杂，因为满足到达客户需求的成本会随时间变化。我们的方法可以看作是在一个随时间动态演变的集合系统上运行 LearnOrCover：每个设施是一个集合，每个客户是一个元素；如果开放某个设施能显著降低某个客户的连接成本，那么该客户对应的元素就包含在该设施对应的集合中。我们沿用高层次的学习/覆盖思路，但需要使用更复杂的势函数来衡量学习进展。

最后，我们的随机顺序 SetMultiCover算法建立在 [GKL21] 针对随机顺序 CIPs 提出的稍复杂一些的算法之上。这涉及若干技术挑战。首先，[GKL21] 的两部分势函数要求：如果某个变量在当前维护的分布中具有较高概率，那么它就应在期望意义上有助于覆盖尚未见过的约束。然而，重数约束禁止算法对任何变量进行多次采样，即使该变量的概率很高也是如此。我们证明，尽管如此，仍可通过逐步“遗忘”已达到上限的坐标来绕过这一困难。有趣的是，我们的乘法权重更新规则并不依赖于新到达约束的边际增补成本，而 [GKL21]中则依赖于此。

1.3 相关工作

术语 先知不等式 通常用于在线最大值查找的语境：赌徒依次从一列已知分布中抽取数字，目标是在最高数字出现时停止。先知不等式指的是此类玩家的表现界限，以能够预见未来的“先知”的表现为参照。[KS78] 展示了一种该游戏策略，其期望收益至少达到先知收益的 $1/2$ （参见 [HK92]，了解更多综述）。秘书问题 [F+89] 是一种相关的最大值查找游戏，赌徒看到的是按任意顺序排列的数字，而这些数字的顺序是随机的；目标同样是在最高数字出现时停止。[AKW14] 通过将先知问题归约为秘书问题（某类算法），给出了一种 $1/e$-竞争策略，适用于先知问题及其扩展；之后 $1/2$ 由 [RWW20]将这一界限进一步改进为 。我们的主要结果可以视为 [AKW14]针对最大化问题提出的先知问题到秘书问题归约的最小化对应形式。

自由顺序先知不等式，即赌徒可以自适应地选择打开箱子的顺序，已有研究涉及 [LLP+21, PT22, BC22]。我们的主要结果表明，对于覆盖问题，受限顺序先知问题的难度至多是其自由顺序对应问题的两倍。

此前的工作 [DEH+18] 声称将先知集合覆盖问题归约为随机（即独立同分布）集合覆盖问题的通用算法。然而，其证明（见 [EB17] 的第 9.5 节）存在问题，我们在附录 C 中详细说明了这一点；此后该主张已被撤回 [Sed]。

受算法能够访问历史数据这一情形的启发， [KNR22、KNR20] 最近在最大值查找（即秘书问题）和匹配的背景下提出了带样本在线模型。[AFGS22] 研究了该模型中的斯坦纳树、设施选址和负载均衡。

从近似算法的角度研究两阶段（以及更一般的多阶段）随机优化已有大量工作（例如， [SS06a、CCP05、GPRS04、GPRS05、GRS07]）；综述请参见[SS06b、BL11] 。我们的两阶段先知模型是一种混合型随机在线模型，其中第二阶段是完全在线的博弈；据我们所知，此前尚未研究过这一模型。

最后，利用我们的归约框架以及 [KNR23],的 3-竞争随机顺序算法，我们可以直接得到一个 6-竞争通用算法，用于求解先知度量设施选址问题，且每个分布仅有一个样本。对于所有分布均相同这一特殊情形， [GGLS08]此前的工作也蕴含了类似结果。

2 预备知识

本文中的所有对数均以 e 为底。在以下定义中，设 $x, y \in \mathrm{R}_+^n$ 为向量。 x 与 y 的标准点积记作 $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$。 我们用 $\max(x, y)$ 表示逐坐标最大值。我们采用 KL 散度的加权推广形式。给定权重函数 c，定义

$$\mathrm{KL}_c\left(x \,||\, y\right) := \sum_{i=1}^{n} c_i \left[x_i \log\left(\frac{x_i}{y_i}\right) - x_i + y_i \right].$$

可增广整数规划。 一个 覆盖 整数规划（CIP）通常定义为一种整数规划（IP），其可行解集合具有向上闭合性。我们定义以下更一般的问题类别，称为 可增广整数规划 （AIPs）。这些问题曾在 [GGLS08]，中进行过研究，但当时没有明确命名。

设 V为一个请求集合。对于任意请求子集 $V' \subseteq V$，令 $\textsc{Sols}(V') \subseteq \mathbb{Z}^m$ 表示对 V'而言可行的解所组成的子集。接下来，对于任意请求子集 V'、任意解 $z \in \textsc{Sols}(V')$ 以及任意请求集合 W，定义增补成本

$$\textsc{Aug}\left(W \;\middle|\; z, V'\right) = \min_{w} \left\{ c(w) \;\middle|\; \text{s.t. } \max(w, z) \in \textsc{Sols}(V' \cup W) \right\},$$

若不存在这样的 w，则为 ∞。当其存在时，令 $\textsc{Backup}(W \mid z, V')$ 表示一个最小化解 s.

定义 2.1（AIP）。可增广整数线性规划（AIP）是指增补成本单调的整数线性规划，即对于任意 $V' \subseteq V'' \subseteq V$，以及任意满足 $z' \in \textsc{Sols}(V')$ 且$z'' \in \textsc{Sols}(V'')$ 的 $z' \le z''$，对于任意请求集合 $W \subseteq V$，均有 $\textsc{Aug}\left(W \mid z'', V''\right) \le \textsc{Aug}\left(W \mid z', V'\right)$。

观察 2.2（AIPs 是次可加的）。对于任意 $A, B \subseteq V$，我们有 $\textsc{Opt}(A \cup B) \le \textsc{Opt}(A) + \textsc{Opt}(B)$。

证明。我们有

$$\textsc{Opt}(A \cup B) \le \textsc{Opt}(A) + \textsc{Aug}\left(B \mid \textsc{Opt}(A), A\right) \le \textsc{Opt}(A) + \textsc{Opt}(B).$$

第一个不等式成立，因为先构造一个满足 A 的解，再增补它以满足B，其成本只会高于 $\textsc{Opt}$（$A \cup B$）。第二个不等式源于 AIPs的增补成本单调性，即 $z' = \vec{0}$、 $z'' = \textsc{Opt}$（A）、 $V' = \emptyset$、 $V'' = A$和 $W = B$。 □

请注意， NonMetricFacilityLocation的标准整数规划（IP）形式为 $\{x_f\}_f$ 中的设施设置指示变量，并为设施与客户之间的连接设置$\{y_{fc}\}_{f,\ c}$，这是一个 AIP，但不是 CIP。同样，SetMultiCover是一个 AIP ，但不是 CIP。

在线模型。我们简要介绍本文讨论的各种模型。

1. 在线 AIP 是一种 AIP ，其中一些约束预先给定，另一些则随时间依次揭示。算法必须维护一个单调递增的解，使其满足迄今为止揭示的所有约束。2. 先知 AIP 实例是一种在线 AIP 实例，其中约束 v^1、.....、 v^n 是从已知分布 D^1、.....、 D^n中抽取的。3. k 样本先知 AIP 实例是一种在线 AIP，其中约束 v^1、.....、 v^n 是从未知分布 D^1、.....、 D^n中抽取的，但在在线序列开始之前，算法会获得来自每个分布的 k 个样本。4. 自由顺序（分别为k样本）先知 AIP 实例是一种（分别为 k样本）先知 AIP实例，其中算法可以自适应地决定从已知分布（分别为未知但已抽样 k 次的分布） D^1、.....、 D^n中抽样的顺序。5. 两阶段先知 AIP 实例由一个先知 AIP 实例和一个正数 λ构成。算法可以在在线序列开始前购买一个初始解 z_1 ，而在线序列期间进行的任何购买 z_2 都要承担 λ的加价成本。6. 带样本在线 AIP 实例是一种在线 AIP 实例；对手确定输入后，算法会预先获得该序列中均匀随机抽取的 α 比例部分。

3 通用先知算法

在本节中，我们通过归约到随机顺序集合覆盖来证明定理 1.1 。回顾一下， [AKW14] 为最大化问题给出了这样一种归约。我们的结果则是针对最小化问题所做的互补尝试。

事实上，我们的归约在两个方面更加强大：

- 它适用于更一般的 AIP，不局限于集合覆盖。因此，要为某个AIP构造先知算法，只需构造一个随机顺序算法。
- 我们要求的性质甚至比随机顺序更弱。事实上，我们可以将先知设定归约为自由顺序先知设定，在该设定中，算法可以自由选择对分布 D_1、.....、 D_n 进行采样的顺序。随机顺序算法是自由顺序算法的一个特殊类别。

定理 3.1。设 $\mathcal{I}$ 是先知 AIP的一个实例类。若算法 $\mathcal{A}$是一个自由顺序先知AIP 算法，在类 $\mathcal{I}$ 上使用 k 个样本可达到竞争比 Δ ，则存在一个固定顺序先知 AIP 算法 $\mathcal{A}'$，在类 $\mathcal{I}$ 上使用 $k+1$ 个样本可达到竞争比 2Δ 。

特别是，随机顺序算法属于零样本自由顺序先知算法，并且LearnOrCover的 [GKL21] 是一种随机顺序集合覆盖算法。因此，我们可以将定理 1.1 作为推论得出。

定理3.1的证明。设 $\mathcal{A}$ 是一个 k-样本先知 AIP 采用自定义/自适应顺序的 π 算法，其期望竞争比为 Δ。定义 $\mathcal{A}'$ 为算法1：

归约 1 $(k+1)$-SAMPLE PROPHET TO k-SAMPLE FREE ORDER PROPHET

1: 训练算法 $\mathcal{A}$，每个 $D^1, \ldots, D^n$ 使用 k 个样本。2: 令 $\textsc{MockRun} = \{\widehat{v}^1, \ldots, \widehat{v}^n\}$ 分别为从 $D^1, \ldots, D^n$ 中抽取的一个样本。3: 对于 $\tau = 1, 2, \ldots, n$，执行以下操作：4: 按照 $\mathcal{A}$ 指定的（可能是自适应的）顺序 π，τ^{th} 个样本。5: 将 π^τ 输入 $\mathcal{A}$。6: 令 $\widehat{z}$ 为 $\mathcal{A}$ 的输出。7: 初始化 $z \leftarrow \widehat{z}$。8: 对于 $t = 1, 2, \ldots, n$，执行以下操作：9: 抽取 $v^t \sim D^t$。10: 如果 z 不满足 v^t，则 11: 更新 $z \leftarrow$ 为 $\max(z, \textsc{backup}(v^t \mid z, \textsc{MockRun} \cup \{v^1, \ldots, v^{t-1}\}))$。

12: 返回 z。

显然，$\mathcal{A}'$ 使用 $k+1$ 个样本，因为这些样本足以模拟 $\mathcal{A}$：算法 $\mathcal{A}$ 需要预先获得 k 个样本，而 $\mathcal{A}'$ 还需要再使用一个样本来模拟从每个分布中进行的真实抽样。接下来我们界定 $\mathcal{A}$ 的成本。

两组样本 $v_1, \ldots, v_n$ 和 $\widehat{v}^1, \ldots, \widehat{v}^n$ 的分布相同，因此

$$\mathbb{E}[c\left(\textsc{Opt}\left(v_1, \ldots, v_n\right)\right)] = \mathbb{E}[c\left(\textsc{Opt}\left(\widehat{v}^1, \ldots, \widehat{v}^n\right)\right)].$$

因此，解的期望成本 $\widehat{z}$ 由 $\mathcal{A}$ 购买，至多为 $\Delta \cdot \mathbb{E}[c\left(\textsc{Opt}(v_1, \ldots, v_n)\right)]$，根据 $\mathcal{A}$的保证。剩下的问题是界定第 11 行中的备用购买成本。为此，考虑每一对请求 $v^t, \widehat{v}^t \sim D^t$，其中 v^t 是在第 9 行抽取的，而 $\widehat{v}^t$ 是第 4 行所指定模拟运行的一部分。我们将这些请求对称为 配对请求。

我们将证明，在模拟 $\mathcal{A}$ 的过程中，请求 v^t 的期望增补成本不超过其配对请求 $\widehat{v}^t$ 的期望增补成本。为此，设 z（v）为请求 v 到达所在轮次开始时 z 的状态。设 $\widehat{z}$（$\widehat{v}$）为算法 $\mathcal{A}$ 的解在模拟 $\mathcal{A}$ 中 $\widehat{v}$ 到达所在轮次开始时的状态（即按照由 $\mathcal{A}$ 选定的顺序 π）。最后，设 $\textsc{MockRun}_{<\widehat{v}}$ 为根据顺序 π，在 $\widehat{v}$ 之前到达的 $\textsc{MockRun}$ 的请求集合。现在：

$$\begin{aligned}\mathbb{E}\left[\textsc{Aug}\left(v^t \mid z(v^t), \textsc{MockRun} \cup \{v^1, \ldots, v^{t-1}\}\right)\right] &\leq \mathbb{E}\left[\textsc{Aug}\left(v^t \mid \widehat{z}(\widehat{v}^t), \textsc{MockRun}_{<\widehat{v}^t}\right)\right] \\ &= \mathbb{E}\left[\textsc{Aug}\left(\widehat{v}^t \mid \widehat{z}(\widehat{v}^t), \textsc{MockRun}_{<\widehat{v}^t}\right)\right]\end{aligned}$$

根据 AIP在定义 2.1中的增补单调性性质，该不等式成立，因为$z(v^t) \geq \widehat{z}(\widehat{v}^t)$。等式成立是因为 v^t 和 $\widehat{v}^t$ 同分布。对 t求和，并注意到如果 $\mathcal{A}$到达时尚未满足，则在第 $\textsc{Aug}\left(\widehat{v}^\tau \mid \widehat{z}(\widehat{v}^\tau)\right.$ $\left.\textsc{MockRun}_{<\widehat{v}^\tau}\right)$ 轮必须支付至少 τ这一事件中，$\widehat{v}^\tau$，可得总备份成本满足

$$\begin{aligned}\sum_t \mathbb{E}\left[\textsc{Aug}\left(v^t \mid z(v^t), \textsc{MockRun} \cup \{v^1, \ldots, v^{t-1}\}\right)\right] &\leq \sum_t \mathbb{E}\left[\{\textsc{Aug}\left(\widehat{v}^t \mid \widehat{z}(\widehat{v}^t), \textsc{MockRun}_{<\widehat{v}^t}\right)\right] \\ &\leq \mathbb{E}[c(\widehat{z})],\end{aligned}$$

根据假设 $\Delta \cdot \mathbb{E}[c\left(\textsc{Opt}(v_1, \ldots, v_n)\right)]$，可将其上界定为 $\mathcal{A}$。因此，算法 1总共至多支付 $2 \cdot \Delta \cdot \mathbb{E}[c(\textsc{Opt})]$。 □

我们注意到，该证明只要求单个请求满足增补成本的单调性。

4 两阶段先知算法

回顾两阶段先知设定。开始时，我们可以访问分布 D_1、...、 D_t 的样本，以及一些 $\lambda > 0$。

- 阶段 1：算法可以购买一个部分解 z_0 ，并承担成本 $c(z_0)$。
- 阶段 2：请求 $v_t \sim D_t$ 逐个到达，算法必须立即扩充其解以满足这些请求。如果 z_1 是算法在第二阶段购买的解的部分，则算法还需承担额外成本 $\lambda \cdot c(z_1)$。

请注意，对于 $0 < \lambda \leq 1$ ，算法应始终等待在线购买集合，因此这就归结为上述先知设定。对于 $\lambda > 1$ ，我们可以不失一般性地假设 λ 是整数（代价是一个较小的常数因子）。

我们的目标是与 $\textsc{Opt}_{\textsc{Online}}$竞争，即为该两阶段问题的最优在线算法所购买的解。我们记作 $\textsc{Opt}_{\textsc{Online}} = \max(z_0^*, z_1^*)$，其中预先购买的解 z_0^*是确定性的，而在线序列期间购买的解 z_1^* 取决于从这些分布中抽取的实现值。

本节的主要结果是，两阶段设定并不比随机顺序设定更难。

定理 4.1. 设 $\mathcal{I}$ 是先知问题的一个实例类。若算法 $\mathcal{A}$ 是一个随机顺序AIP 算法，并且在类 $\mathcal{I}$上达到竞争比 Δ ，则存在一个适用于类 $\mathcal{I}$ 的两阶段先知 AIP算法 $\mathcal{A}'$ ，其相对于使用 λ 个样本的最优在线策略达到竞争比 2Δ 。

证明。 令 $\mathcal{A}$为实例类 $\mathcal{I}$ 的随机顺序算法，其期望竞争比为Δ。定义 $\mathcal{A}'$ 为算法 2:

算法 2 2-STAGE PROPHET TO RANDOM ORDER

第一阶段:

</style>对于 $i = 1, \ldots, \lambda$ **do**
  令 $\textsc{MockRun}_i \leftarrow \{\hat{v}_i^1, \ldots, \hat{v}_i^n\}$ 分别为来自 D^1、.....、 D^n 的一个样本。
令 $\textsc{MockRun} = \bigcup_i \textsc{MockRun}_i$。
对于 $\tau = 1, 2, \ldots, \lambda \cdot n$ 执行
  $\pi^\tau \leftarrow$ 以随机顺序处理 $\textsc{MockRun}$ 的 τ^{th} 请求 π。
  将 π^τ 输入到 $\mathcal{A}$。
令 z_0 为 $\mathcal{A}$ 的输出。购买 z_0。

第二阶段:

初始化 $z_1 \leftarrow 0$。
对于 $t = 1, 2, \ldots, n$ **do**
  绘制 $v^t \sim D^t$。
  令 $z \leftarrow \max(z_0, z_1)$。
  如果 z 未满足 v^t ，则
    令 $w \leftarrow \textsc{backup}(v^t \mid \max(z_0, z_1)\ \textsc{MockRun} \cup \{v^1, \ldots, v^{t-1}\})$。
    更新 $z_1 \leftarrow \max(z_1, w)$。
返回 $\max(z_0, z_1)$。

证明分两步进行。令 $Z := \mathbb{E}[c(z_0^*) + \lambda \cdot c(z_1^*)]$ 表示 $\textsc{Opt}_{\textsc{Online}}$的期望成本。首先，我们根据 Z界定由 $\mathcal{A}$ 计算出的 z_0 的期望成本；然后，我们根据 z_0的成本界定备份的总成本，即 z_1。

对于 z_0的第一个界，我们沿用 [GPRS04]的"增强采样"论证。假设 z_0^*是最优第一阶段解，而 $z_1^*, \ldots, z_\lambda^*$ 是最优在线策略在分别输入序列 $\textsc{MockRun}_i$ 时购买的第二阶段解。根据观察 AIP 中的次可加性 2.2，

$$c\left(\textsc{Opt}\left(\textsc{MockRun}\right)\right) \le c(z_0^*) + \sum_{i=1}^{\lambda} c(z_i^*) = c(z_0^*) + \mathop{\mathbb{E}}_{i\sim[\lambda]}[\lambda \cdot c(z_i^*))],$$

其中，$i \sim [\lambda]$ 表示 i 是从 $[\lambda]$中均匀抽取的。对 $\textsc{MockRun}$的抽取取期望，可得 $\mathbb{E}[c(\textsc{Opt}(\textsc{MockRun}))] \le Z$。由于算法 $\mathcal{A}$是 Δ-竞争的，我们立即得到 $\mathbb{E}[c(z_0)] \le \Delta \cdot Z$.

接下来需要界定 z_1的成本。证明的后半部分与定理 3.1的证明相似。令 $z_0(v$表示请求 $\mathcal{A}$ 到达之前，客户 v所持有的解的状态。

固定一个索引 $t \in [n]$。不失一般性，将相应的 $\textsc{MockRun}$ 样本$\widehat{v}_1^t, \ldots, \widehat{v}_\lambda^t$ 重新排序，使其与它们在 π中的相对顺序一致。定义 $\textsc{MockRun}_{<v}$为以下客户集合：客户$\textsc{MockRun}$ 按照顺序 v 在 π之前到达。根据定义 2.1，由于 $z_0(\widehat{v}_1^t) \le \ldots \le z_0(v^t)$ 对随机变量的每一种实现都成立，因此对于每个请求 v,

$$\begin{aligned}\textsc{Aug}\left(v \;\middle|\; z_0(\widehat{v}_1^t), \textsc{MockRun}_{<\widehat{v}_1^t}\right) \ge \cdots &\ge \textsc{Aug}\left(v \;\middle|\; z_0(\widehat{v}_\lambda^t), \textsc{MockRun}_{<\widehat{v}_\lambda^t}\right) \\ &\ge \textsc{Aug}\left(v \mid z_0(v^t), \textsc{MockRun} \cup \{v^1, \ldots, v^{t-1}\}\right).\end{aligned}$$

然后，对 $z_0(\widehat{v}_1^t) \ldots, z_0(v^t)$ 的随机序列以及 $\widehat{v}_1^t, \ldots, \widehat{v}_\lambda^t, v^t \sim D^t$ 的同分布抽样取期望，

$$\begin{aligned}\mathbb{E}\left[\textsc{Aug}\left(\widehat{v}_1^t \;\middle|\; z_0(\widehat{v}_1^t), \textsc{MockRun}_{<\widehat{v}_1^t}\right)\right] \ge \cdots &\ge \mathbb{E}\left[\textsc{Aug}\left(\widehat{v}_\lambda^t \;\middle|\; z_0(\widehat{v}_\lambda^t), \textsc{MockRun}_{<\widehat{v}_\lambda^t}\right)\right] \\ &\ge \mathbb{E}\left[\textsc{Aug}\left(v^t \;\middle|\; z_0(v^t), \textsc{MockRun} \cup \{v^1, \ldots, v^{t-1}\}\right)\right].\end{aligned}$$

求和可得

$$\lambda \cdot \mathbb{E}\left[\textsc{Aug}\left(v^t \;\middle|\; z_0(\widehat{v}^t), \textsc{MockRun} \cup \{v^1, \ldots, v^{t-1}\}\right)\right] \le \sum_{i=1}^{\lambda} \mathbb{E}\left[\textsc{Aug}\left(\widehat{v}_i^t \;\middle|\; z_0(\widehat{v}_i^t), \textsc{MockRun}_{<\widehat{v}_i^t}\right)\right].$$

最后，再对 $t \in [n]$ 求和，我们得到算法 2 的第二阶段成本上界为

$$\begin{aligned}\lambda \cdot \mathbb{E}[c(z_1)] &= \sum_{t=1}^{n} \lambda \cdot \mathbb{E}\left[\textsc{Aug}\left(v^t \;\middle|\; z_0(v^t), \textsc{MockRun} \cup \{v^1, \ldots, v^{t-1}\}\right)\right] \\ &\le \sum_{t=1}^{n}\sum_{i=1}^{\lambda} \mathbb{E}\left[\textsc{Aug}\left(\widehat{v}_i^t \;\middle|\; z_0(\widehat{v}_i^t), \textsc{MockRun}_{<\widehat{v}_i^t}\right)\right] \\ &\le \mathbb{E}[c(z_0)],\end{aligned}$$

其中，最后一个不等式成立，是因为在 $\widehat{v}_i^t$ 到达时尚未满足的轮次中，$\mathcal{A}$ 至少支付 $\textsc{Aug}\left(\widehat{v}_i^t \;\middle|\; z_0(\widehat{v}_i^t), \textsc{MockRun}_{<\widehat{v}_i^t}\right)$。

我们得出结论，该算法的总支出为 $\mathbb{E}[c(z_0) + \lambda \cdot c(z_1)] \le 2\,\mathbb{E}[c(z_0)] \le 2\Delta \cdot Z$。

□

5 带样本在线

本节中，我们展示了从带样本在线 AIPs 到随机顺序版本的一般归约。

定理 5.1。设 $\mathcal{I}$ 是一个 AIP 的实例类。如果算法 $\mathcal{A}$ 是一个随机顺序 AIP算法，在类 $\mathcal{I}$上的竞争比为 Δ，那么存在一个适用于类 $\mathcal{I}$ 的带样本在线算法 $\mathcal{A}'$，其竞争比为 Δ/α。

定理 *5.1*的证明。设 $\mathcal{A}$ 是实例类 $\mathcal{I}$ 上的随机顺序算法，其期望竞争比为 Δ。定义 $\mathcal{A}'$ 为算法 3：

算法 3 ONLINE-WITH-A-SAMPLE TO RANDOM ORDER

1：令 $\textsc{Samples} = \{\widehat{v}^1, \ldots, \widehat{v}^{\alpha \cdot n}\}$ 为预先给定的样本。2：对于 $\tau = 1, 2, \ldots, \alpha \cdot n$，执行以下操作。3：将 τ^{th} 样本按随机顺序排列。4：将 π^τ 输入 $\mathcal{A}$。5：令 $\widehat{z}$ 为 $\mathcal{A}$的输出。6：初始化 $z \leftarrow \widehat{z}$。7：对于 $t = 1, 2, \ldots, n$，执行以下操作。8：抽取 $v^t \sim D^t$。9：如果 v^t 未被 z 满足，则 10：更新 $z \leftarrow \max(z, \textsc{backup}(v^t \mid z, \textsc{Samples} \cup \{v^1, \ldots, v^{t-1}\}))$。

11：返回 z。

不失一般性，假设 n 是 $1/\alpha$的倍数。我们设想按照以下过程生成 SAMPLES。首先，将 v^1、.....、v^n 随机划分为 $\alpha \cdot n$组，每组大小为 $1/\alpha$。从每个子集中均匀随机地选取一个代表请求，将其纳入 SAMPLES。对于每个请求 v，令 $r(v)$为包含 v的组中的代表请求，并令 $\textsc{Samples}_{<v}$ 为按照随机顺序 π在 v 之前到达的样本集合。现在：

$$\begin{aligned}\mathbb{E}\left[\textsc{Aug}\left(v^t \mid z(v^t), \textsc{Samples} \cup \{v^1, \ldots, v^{t-1}\}\right)\right] &\leq \mathbb{E}\left[\textsc{Aug}\left(v^t \mid \widehat{z}(r(v^t)), \textsc{Samples}_{<r(v^t)}\right)\right] \\ &= \mathbb{E}\left[\textsc{Aug}\left(r(v^t) \mid \widehat{z}(r(v^t)), \textsc{Samples}_{<r(v^t)}\right)\right].\end{aligned}$$

根据定义2.1中的增广条件的增补单调性，不等式成立，因为 $z(v^t) \geq \widehat{z}(r(v^t))$。等式成立，因为在运行开始时，$v^t$ 和 $r(v^t)$服从相同的分布。对 t求和，并注意到当 $\widehat{v}^\tau$ 在到达时未得到满足时，$\mathcal{A}$ 在第 τ轮至少需要支付 $\textsc{Aug}\left(\widehat{v}^\tau \mid \widehat{z}(v^\tau)\right)$，可得总备份成本满足

$$\begin{aligned}\sum_{t=1}^{n} \mathbb{E}\left[\textsc{Aug}\left(v^t \mid z(v^t), \textsc{Samples} \cup \{v^1, \ldots, v^{t-1}\}\right)\right] &\leq \sum_{t=1}^{n} \mathbb{E}\left[\textsc{Aug}\left(r(v^t) \mid z(r(v^t)), \textsc{Samples}_{<r(v^t)}\right)\right] \\ &= \sum_{\tau=1}^{\alpha \cdot n} \frac{1}{\alpha} \cdot \mathbb{E}[\textsc{Aug}\left(\widehat{v}^\tau \mid z(\widehat{v}^\tau), \textsc{Samples}_{<\widehat{v}^\tau}\right)] \\ &\leq \frac{1}{\alpha}\, \mathbb{E}[c(\widehat{z})].\end{aligned}$$

上面的等式源于构成 SAMPLES的每个代表元在第一个求和式中都出现$1/\alpha$次。根据 $\Delta/\alpha \cdot \mathbb{E}[c\left(\textsc{Opt}(v_1, \ldots..., v_n)\right)]$ 的假设，我们可以对上面的表达式进行如下界定，因此，算法 1 的总开销至多为 $\Delta/\alpha \cdot \mathbb{E}[c(\textsc{Opt})]$。

□

最后补充一点，类似的论证也能证明定理 5.1 在稍有不同的“带样本在线”设定下同样成立：在线序列中的每个请求都以概率 α独立采样。更一般地，对于 $\alpha \leq 1$，也可以在定理 α 的样本数量与定理 $1/\alpha$ 的近似比之间得到类似的权衡；定理3.1 和定理4.1 也同样如此。

6 在线设施选址

本节将我们的框架应用于在线设施选址问题。

6.1 设施选址

在 FacilityLocation中，输入包括一组 n 客户和 m 设施。每个设施 f 都有开启成本c_f，每对客户与设施（v，f）都有连接成本c_{fv}。目标是开启若干设施，并将每个客户恰好连接到一个已开启的设施，使总成本最小化。离线（单位需求） FacilityLocation的标准整数规划表述如下 [CCZ14]:

$$\begin{aligned} \min \sum_f c_f \cdot x_f + \sum_{f,v} c_{fv} \cdot y_{fv} & \qquad (6.1) \\ \sum_j y_{fv} \geq 1 \qquad \forall v & \\ y_{fv} \leq x_f \qquad \forall f, v & \\ x_f, y_{fv} \in \{0,1\}. & \end{aligned}$$

在随机顺序在线版本中，m 设施是预先已知的，而 n 未知客户按随机顺序在线到达；每当一位客户到达时，算法必须决定是否开启任何新设施，然后将该客户连接到一个已开启的设施。决策不可撤销，也就是说，客户到达后不能更改其所连接的设施。

我们注意到， FacilityLocation适用于我们的框架:

观察 6.1。 FacilityLocation是一个 AIP。

这使我们能够将随机顺序 FacilityLocation的算法转换为适用于先知、两阶段先知和带样本设定的算法。

6.2 度量设施选址

随机顺序 FacilityLocation在连接成本c_{fv} 满足三角不等式时，已有充分研究。Meyerson 在开创性工作中为此问题提出了一个 8-近似算法 [Mey01], ，该算法最近已改进为 3-近似[KNR23]。根据观察6.1 ，并将该算法作为黑箱代入定理3.1、4.1 和5.1，我们得到:

推论 6.2。对于 MetricFacilityLocation问题，在单样本先知设定中存在一个 6-竞争算法，在两阶段先知设定中存在一个 6-竞争算法，而在带样本的在线设定中存在一个 $3/\alpha$-竞争算法。

下一节中，我们将研究不满足度量假设的更一般问题。

6.3 非度量设施选址

当连接成本不满足三角不等式时，这个问题会更加复杂。具体而言， NonMetricFacilityLocation会在所有服务成本均为 $c_{fv} \in \{0, \infty\}$ 的特殊情形下恢复 SetCover 。本节中，我们将为随机顺序 NonMetric-FacilityLocation 给出一种算法，该算法是最优的，即使在集合覆盖这一特殊情形下也是如此。我们将证明：

定理 6.3。存在一个 $O(\log mn)$-竞争算法，用于随机顺序 NonMetric-FacilityLocation。

根据观察 6.1 以及定理 3.1、4.1 和5.1 我们可以直接得到 ：

推论 6.4。对于随机顺序 NonMetricFacilityLocation，在单样本先知设定中存在一个 $O(\log mn)$-竞争算法，在两阶段先知设定中存在一个 $O(\log mn)$-竞争算法，并且在带样本的在线设定中存在一个 $O(\log mn/\alpha)$-竞争算法。

我们展示了 LearnOrCover算法 [GKL21] 在随机顺序集合覆盖中的适用性，并通过将其调整到随机顺序 FacilityLocation来加以说明，后者并非纯覆盖问题。挑战在于决定每一轮要开放哪些设施；一旦决定完成，就可以假设该轮到达的客户始终连接到成本最低的开放设施。

Let $\mathcal{C}^t$ 表示截至第 t 轮已购买的设施。对于每个客户 v，定义

$$f^t(v) \in \operatorname*{argmin}_f(\mathbb{I}\{f \notin \mathcal{C}^t\} \cdot c_f + c_{fv})$$
$$\kappa_v^t := \min_f(\mathbb{I}\{f \notin \mathcal{C}^t\} \cdot c_f + c_{fv})$$

表示能够以尽可能低的成本连接该客户的设施（必要时包括开放该设施），并定义相应的边际成本。这是第 t 轮结束时的成本。注意，在我们更一般的记号中， k_v^t 对应于 $\textsc{Aug}\left(v^t \mid \mathcal{C}^t, \{v^1, \ldots, v^{t-1}\}\right)$。此外，定义

$$\Gamma^t(v) := \{f : c_{fv} \le \kappa_v^t/2\}$$

表示这样一个设施集合：若开放其中的设施，就能将连接 v 的边际成本至少降低一半。接下来，如果 $f \in \Gamma^t(v)$，我们就称设施 f 在时间 t 覆盖客户 v 。
我们将证明，算法 4 对于随机顺序 $O(\log mn$，其期望近似比为 NonMetricFacilityLocation。我们的方法可以看作是在一个 LearnOrCover集合覆盖算法，但所处理的是一个动态变化的集合系统。设施对应集合，客户对应元素；在时间 t ，若 $f \in \Gamma^t(v)$，则客户对应的元素属于设施对应的集合。

采用猜测并加倍方法，我们可以假设算法可以访问一个上界 β 使得$\mathrm{LP}_{\mathrm{OPT}} \le \beta \le 2 \cdot \mathrm{LP}_{\mathrm{OPT}}$；这里 $\mathrm{LP}_{\mathrm{OPT}}$ 是给定 NonMetricFacilityLocation实例的线性规划松弛 (6.1) 的最优解成本。我们将事件 Ξ^t记为：到达客户端 v^t 满足第 5 行上的条件，即 $\kappa_{v^t}^{t-1} \ge \beta/t$ ，并执行算法 4 的主要部分。若不满足该条件，我们称 v^t 为抢先连接；这对应事件 $\neg\Xi^t$。

通过第 10 行，我们保持以下不变量：

不变量 1. 对所有时间步 t，均有 $\langle c, x^t \rangle = \beta$。

算法 4 LearnOrCoverNMFL

令 $\mathcal{F}' \leftarrow \{f:\ \beta/m \le c_f \le \beta\}$ 并令 $m' \leftarrow |\mathcal{F}'|$。初始化 $x_f^0 \leftarrow \frac{\beta}{c_f \cdot m'} \cdot \mathbb{I}\{f \in \mathcal{F}'\}$。对于 $t = 1, 2 \ldots, n$，执行 按随机顺序选取 $v^t \leftarrow t^{th}$ 客户，并令 $\mathcal{R}^t \leftarrow \emptyset$。如果 $\kappa_{v^t}^{t-1} \ge \beta/t$，则 对每个设施 f，以 $\min(\kappa_{v^t}^{t-1} \cdot x_f^{t-1}/\beta, 1)$ 的概率添加 $\mathcal{R}^t \leftarrow \mathcal{R}^t \cup \{f\}$。更新 $\mathcal{C}^t \leftarrow \mathcal{C}^{t-1} \cup \mathcal{R}^t$。如果 $\sum_{f \in \Gamma^{t-1}(v^t)} x_f^{t-1} < 1$，则 对每个设施 f，更新 $x_f^t \leftarrow x_f^{t-1} \cdot \exp\left\{\mathbb{I}\{f \in \Gamma^{t-1}(v^t)\} \cdot \kappa_{v^t}^{t-1}/c_f\right\}$。令 $Z^t = \langle c, x^t \rangle / \beta$ 并对 $x^t \leftarrow x^t/Z^t$ 进行归一化。否则 $x^t \leftarrow x^{t-1}$。否则 $x^t \leftarrow x^{t-1}$。$\mathcal{C}^t \leftarrow \mathcal{C}^t \cup \{f^{t-1}(v^t)\}$ 并将 v^t 连接到 $f^{t-1}(v^t)$。

返回 $\mathcal{C}$。

我们先定义记号。令 $\text{Opt} = (x^*, y^*)$ 为一个最优分数解。令 $U^t = \{v^{t+1}, \ldots, v^n\}$ 表示第 t 轮结束时仍未覆盖的客户（其中 $U^0 = U$ 是客户全集）。令 $X^t(v) := \sum_{f \in \Gamma^{t-1}(v)} x_f^{t-1}$ 表示在时刻 $t-1$ 覆盖 v 的设施 f 的分数权重。我们定义 $\rho^t := \sum_v \kappa_v^t$，并考虑以下势函数:

$$\Phi(t) := C_1 \cdot \underbrace{\left(\mathrm{KL}_c\left(x^* \mid\mid x^t\right) + 2 \cdot \sum_{v \in U^t} \sum_f c_{fv} \cdot y_{fv}^* \right)}_{\Phi_L(t)} + C_2 \cdot \underbrace{\beta \cdot \log\left(\frac{\rho^t}{\beta} + \frac{1}{n}\right)}_{\Phi_C(t)},$$

其中常数 C_1 和 C_2 将在后文确定。我们将 Φ_L 称为势函数的“学习”部分，将 Φ_C 称为势函数的“覆盖”部分。

我们的势函数 $\Phi(t)$ 与用于分析集合覆盖 LearnOrCover的 [GKL21] 中使用的势函数相似，也可以分解为“学习”和“覆盖”两部分。其学习部分也包含 KL 散度项，但我们的势函数更复杂，因为我们还将分数解 支付的连接成本计入其中。

引理 6.5（关于 Φ 的界）。初始势函数有界于 $\Phi(0) = O(\beta \cdot \log mn)$，且对所有 t，$\Phi(t) \ge -\beta \cdot \log n$。

现在我们证明势函数在每一轮次中都能充分下降。我们分别界定势函数中每一项的下降幅度。

引理 6.6（C hange in Φ_L）。对于事件 Ξ^t holds 发生的轮次，Φ_L 中的期望变化 is

$$\begin{aligned} &\mathop{\mathbb{E}}_{v^t, \mathcal{R}^t}\left[\Phi_L(t) - \Phi_L(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t\right] \\ &\le \mathop{\mathbb{E}}_{v \sim U^{t-1}}\left[\frac{e^2 - 1}{2} \cdot \kappa_v^{t-1} \cdot \min\left(X^t(v),\ 1\right) - \kappa_v^{t-1}\right]. \end{aligned} \tag{6.2}$$

当到达的 v^t 已抢先连接，且 Ξ^t 不成立时，

$$\Phi_L(t) - \Phi_L(t-1) \le 0. \tag{6.3}$$

注意，上述陈述中的期望变化仅取决于到达的未覆盖客户 v^t 的随机性，而不取决于随机选取的设施 $\mathcal{R}^t$。另一方面，我们可以如下界定 Φ_C 中的变化。

引理 6.7（ Φ_C 的变化）。对于满足 Ξ^t 的所有轮次 t， Φ_C 的期望变化为

$$\mathop{\mathbb{E}}_{v^t,\mathcal{R}^t}\left[\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t\right] \le -\frac{1-e^{-1}}{4} \cdot \mathop{\mathbb{E}}_{u\sim U^{t-1}}\left[\kappa_u^{t-1} \cdot \min\left(X^t(v),\ 1\right)\right]. \tag{6.4}$$

对于 Ξ^t 不成立的轮次，

$$\Phi_C(t) - \Phi_C(t-1) \le 0. \tag{6.5}$$

我们将引理的证明6.5 至6.7 推迟到附录A，现在说明如何将它们结合起来证明该定理。

定理 6.3 的证明。设 $c(\textsc{Alg}(t))$ 表示算法 4 截至并包括时间 t 所支付的成本；此外，设 $c_{\textsc{pre}}(\textsc{Alg}(t))$ 表示算法 4 截至并包括第 t 轮，在第15行抢先连接客户所支付的成本；设 $c_{\textsc{LoC}}(\textsc{Alg}(t))$ 表示算法 4 在第6行和第15行截至并包括第 t 轮所支付的成本，其中这些轮次满足 v^t 未被抢先连接且 Ξ^t 成立。（$\textsc{Pre}$ 表示抢先连接， $\textsc{LoC}$ 表示 $\textsc{LearnOrCover}$。）

我们可以对算法在第 15 行购买的廉价设施与客户之间的连接给出一个简单的上界，条件是 v^t 被抢先连接（也就是说， Ξ^t 不成立）。这些连接的总成本至多为

$$\begin{aligned} c_{\textsc{pre}}(\textsc{Alg}(n)) &= c_{\textsc{pre}}(\textsc{Alg}(n)) - c_{\textsc{pre}}(\textsc{Alg}(0)) \\ &= \sum_t \left(c_{\textsc{pre}}(\textsc{Alg}(t)) - c_{\textsc{pre}}(\textsc{Alg}(t-1))\right) \\ &\le \sum_{t\in[n]} \frac{\beta}{t} = O(\beta \cdot \log n). \end{aligned} \tag{6.6}$$

现在我们考虑算法主体在 Ξ^t 成立的各轮次中产生的每轮成本。在每一轮 t，第 $\mathcal{R}^t$ 行采样设施的期望成本为6 $\kappa_{v^t}^{t-1} \cdot \langle c,\ x^{t-1}\rangle/\beta = \kappa_{v^t}^{t-1}$ （根据不变式 1）。算法在第 $\kappa_{v^t}^{t-1}$ 15 行最多还需支付 $2 \cdot \kappa_{v^t}^{t-1}$，因此每轮的总期望成本至多为 。

结合引理 6.6 和6.7，并令常数 $C_1 = 2$ 和 $C_2 = 4e(e+1)$，可得

$$\begin{aligned} &\mathop{\mathbb{E}}_{v^t,\mathcal{R}^t}\left[\Phi(t) - \Phi(t-1) \,\middle|\, v^1,\dots,v^{t-1},\mathcal{R}^1,\dots,\mathcal{R}^{t-1},\Xi^t\right] \\ &= \mathop{\mathbb{E}}_{v^t,\mathcal{R}^t}\left[\begin{array}{rl} & C_1 \cdot (\Phi_L(t) - \Phi_L(t-1)) \\ + & C_2 \cdot (\Phi_C(t) - \phi_C(t-1)) \end{array} \,\middle|\, v^1,\dots,v^{t-1},\mathcal{R}^1,\dots,\mathcal{R}^{t-1},\Xi^t\right] \\ &\le -\mathop{\mathbb{E}}_{v^t,\mathcal{R}^t}\left[2 \cdot \kappa_{v^t}^{t-1} \,\middle|\, v^1,\dots,v^{t-1},\mathcal{R}^1,\dots,\mathcal{R}^{t-1},\Xi^t\right], \end{aligned}$$

这会抵消 表示每轮中 $c_{\textsc{LoC}}$ 的期望变化。因此，我们得到不等式 y

$$\mathop{\mathbb{E}}_{v^t,\mathcal{R}^t}\left[\Phi(t) - \Phi(t-1) + c_{\textsc{LoC}}(\textsc{Alg}(t)) - c_{\textsc{LoC}}(\textsc{Alg}(t-1)) \,\middle|\, v^1,\dots,v^{t-1},\mathcal{R}^1,\dots,\mathcal{R}^{t-1}\right] \le 0, \tag{6.7}$$

这里用到了：当 Ξ^t 不成立时， Φ_L 和 Φ_C 的变化至多为 0。

对所有 $1 \le t \le n$反复应用 (6.7)，我们得到

$$\begin{aligned} \mathop{\mathbb{E}}_{v,\mathcal{R}}[\Phi(n) - \Phi(0) + c_{\text{LoC}}(\text{ALG}(n)) - c_{\text{LoC}}(\text{ALG}(0))] &\le 0 \\ \mathop{\mathbb{E}}_{v,\mathcal{R}}[c_{\text{LoC}}(\text{ALG}(n))] &\le \Phi(0) + c_{\text{LoC}}(\text{ALG}(0)) - \mathop{\mathbb{E}}_{v,\mathcal{R}}[\Phi(n)] \\ &\le O(\beta \cdot \log mn) + 0 - (-\beta \cdot \log n)\,, \end{aligned} \tag{6.8}$$

其中，(6.8) 来自以下事实： $c_{\text{LoC}}(\text{ALG}(0)) = 0$，以及 Φ 的上界，该上界已在引理 6.5 中确立。

最后，由 (6.6) 和 (6.8) 可得

$$\begin{aligned} \mathop{\mathbb{E}}_{v,\mathcal{R}}[c(\text{ALG}(n))] &= \mathop{\mathbb{E}}_{v,\mathcal{R}}[c_{\text{PRE}}(\text{ALG}(n))] + \mathop{\mathbb{E}}_{v,\mathcal{R}}[c_{\text{LoC}}(\text{ALG}(n))] \\ &\le \beta + O(\beta \cdot \log n) + O(\beta \cdot \log mn), \end{aligned} \tag{6.9}$$

正如所需。 □

至此，我们对随机顺序 FacilityLocation 的讨论告一段落。我们的结果确定了该问题在随机顺序模型下的度量版本和非度量版本的可近似性，因此也适用于先知、两阶段先知和带样本在线模型。

7 具有重数约束的集合多重覆盖

现在，我们将框架的第二个应用用于带重数约束的集合多重覆盖问题，简称为 SetMultiCover。这推广了单位成本集合覆盖，并引入了非覆盖约束；具体而言，这些约束施加于决策变量的重数。集合多重覆盖问题可形式化地写成以下整数规划：

$$\begin{aligned} \min_z \quad & \langle 1, z \rangle \\ \text{s.t.} \quad & Az \ge b \\ & z \in \{0,1\}^m, \end{aligned} \tag{7.1}$$

其中， A 的各项取值属于 $\{0,1\}$。因此，不失一般性，可假设 $b \in \mathbb{Z}^n_+$。当 $b = 1$时，我们可以将 SetCover 作为其特例。我们再次发现， SetMultiCover适用于我们的框架：

观察7.1。 SetMultiCover是一个 AIP。

这使我们能够将随机顺序 SetMultiCover的算法转化为适用于先知、两阶段和带样本设置的算法。

在随机顺序 SetMultiCover中， A 的各行按随机顺序逐一揭示，且算法必须维护一个单调递增的解 z ，使其满足截至目前已揭示的所有约束。我们证明：

定理 7.2. 存在一种用于随机顺序 SetMulti-Cover 的 $O(\log mn)$-竞争算法。

根据观察 7.1 和定理 3.1、 4.1 和 5.1 ，我们可直接得到：

推论 7.3. 对于 SetMultiCover，在单样本先知设定下，存在一个具有 $O(\log mn)$ 竞争比的算法；在两阶段先知设定下，存在一个具有 $O(\log mn)$ 竞争比的算法；在带样本的在线设定下，存在一个具有 $O(\log(mn)/\alpha)$ 竞争比的算法。

我们再次展示如何将 [GKL21] 中的思路应用于这个更一般的设定。这次，我们将其针对随机顺序 CIP的更一般算法加以扩展。我们将用到第2节给出的预备知识，采用的方法与第6节类似。

令 z^t 表示第 t轮的整数解。令 d_i^t 表示第 t轮开始时 i 的覆盖不足量；也就是说，$d_i^t := \max(0, b_i - \langle a_i, z^{t-1}\rangle)$，其中 z^t 表示第t轮结束时的整数解。现在我们定义

$$\rho^t := \sum_{i\in[n]} d_i^t.$$

我们采用猜测并加倍方法来确定 β，使得 $\mathrm{LP}_{\mathrm{OPT}} \le \beta \le 2\cdot \mathrm{LP}_{\mathrm{OPT}}$，其中$\mathrm{LP}_{\mathrm{OPT}}$ 是针对（ x^* 7.1）的最优分数解的成本。我们将维护一个针对（ x^t 7.1）的解，其成本为 β。

算法 5 LearnOrCoverSMC

初始化 $x_j^0 \leftarrow \frac{\beta}{m}$ 对每个 j，并设置 $z_j^0 \leftarrow 0$.
对于 $t = 1, 2, \ldots$ 执行
  随机顺序中的第 $i \leftarrow t$ 个约束。
  如果 i 到达时未被覆盖，则
    令 $\mathcal{T}^t := \{j : z_j^{t-1} = 0, a_{ij} = 1\}$。 // 覆盖 i 的未购买坐标
    令 $X_i^{t-1} := \sum_{j\in\mathcal{T}^t} x_j^{t-1}$. // 未购买项的分数覆盖率 j
    $d_i^t := b_i - \langle a_i, z^{t-1}\rangle$。 // 整数未覆盖
    对于每个 j，从 $z_j^t \leftarrow$ 中抽取服从 $\mathrm{Ber}(d_i^t \cdot x_j^{t-1}/\beta)$ 的样本。
    如果 $X_i^{t-1} \le d_i^t$，则 // i 被分数地覆盖不足
      为 $j \in \mathcal{T}^t$ 设置 $z_j^t \leftarrow 1$，使用 $x_j^{t-1} \ge 1/e$。
      对每个 j，更新 $x_j^t \leftarrow x_j^{t-1} \cdot \exp\{\mathbb{I}\{j \in \mathcal{T}^t\}\}$。
      令 $Z^{(t)} := \langle 1, x^t\rangle/\beta$，并重新归一化 $x^t \leftarrow x^t/Z^t$。
    如果 i 仍未被覆盖，则
      $x^t \leftarrow x^{t-1}$ 和 $z^t \leftarrow z^{t-1}$。
      $z_j^t \leftarrow 1$ 针对任意 d_i^t 个 $j \in \mathcal{T}^t$。 // 购买备用集合
  else
    以及 $z^t \leftarrow z^{t-1}$。
返回 z

与算法 4类似，我们的算法会对未被覆盖的约束作出不同反应。在算法 4 中，判据是到达客户端无法以低成本连接；而在此处，如果元素到达时覆盖不足，我们会执行一个 LearnOrCover步骤，这意味着 $d_{it}^t > 0$位于第 4行。

定理 7.4。 对于具有重数约束的集合多重覆盖，算法 5的期望竞争比为 $O(\log mn)$。

我们的势函数为

$$\Phi(t) := C_1 \cdot \underbrace{\mathrm{KL}\left(x^* \,||\, x^t\right)}_{\Phi_L(t)} + C_2 \cdot \underbrace{\beta \cdot \log\left(\frac{\rho^t}{\beta} + \frac{1}{m}\right)}_{\Phi_C(t)},$$

这里我们再次将 Φ_L 视为势函数的学习部分，将 Φ_C 视为覆盖部分。稍后我们将确定常数 C_1 和 C_2。首先，我们来界定该势函数的值：

引理 7.5（ Φ的界）。初始势函数满足 $\Phi(0) = O(\beta \cdot \log mn)$，且对所有轮次 t，均有 $\Phi(t) \geq -\beta \cdot \log(n)$。

令 U^t表示在第 t轮尚未到达的约束 i；也就是说， $U^t = \{i^t, i^{t+1}, \ldots, i^n\}$。现在我们来讨论每轮学习部分的期望变化。

引理 7.6（ Φ_L的变化）。对于 i^t 到达时未被覆盖的轮次， Φ_L的期望变化为

$$\mathop{\mathbb{E}}_{i^t, \mathcal{R}^t}\left[\Phi_L(t) - \Phi_L(t-1) \mid x^{t-1}, U^{t-1}, d^t_{i^t} > 0\right] \leq \mathop{\mathbb{E}}_{i \sim U^{t-1}}[(e-1) \cdot \min(X_i^{t-1},\, d_i^t) - d_i^t]. \tag{7.2}$$

当 i^t 到达时已覆盖时，

$$\Phi_L(t) - \Phi_L(t-1) \leq 0. \tag{7.3}$$

接下来，我们来界定 $\Phi_C(t)$的期望变化。

引理 7.7（ Φ_C的变化）。对于每一轮，只要 i^t 到达时未覆盖， Φ_C 的期望变化为

$$\mathop{\mathbb{E}}_{i^t, \mathcal{R}^t}\left[\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, d^t_{i^t} > 0\right] \leq -\frac{\gamma}{2} \cdot \mathop{\mathbb{E}}_{i \sim U^{t-1}}\left[\min\left(X_i^{t-1},\, d_i^t\right)\right], \tag{7.4}$$

其中 γ 是一个固定常数。在 i^t 到达时已覆盖的轮次中，

$$\Phi_C(t) - \Phi_C(t-1) \leq 0. \tag{7.5}$$

我们再次将引理的证明 7.5 至7.7 推迟到附录A，现在说明如何将它们结合起来，以界定算法 5 的期望成本：

定理 7.4 的证明。 结合引理 7.6 和引理 7.7，选择 $C_1 = (e+2)$ 和 $C_2 = 2 \cdot (e + 2(e-1)/\gamma$，并回想 $d^t_{t^t} = 0$ 在 t^t 到达时已覆盖的轮次中，因此我们有

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i^t, \mathcal{R}^t}\left[\Phi(t) - \Phi(t-1) \,\middle|\, i^1, \ldots, i^{t-1}, \mathcal{R}^1, \ldots, \mathcal{R}^{t-1}\right] \\
&= \mathop{\mathbb{E}}_{i^t, \mathcal{R}^t}\left[\begin{array}{rl} & C_1 \cdot (\Phi_L(t) - \Phi_L(t-1)) \\ + & C_2 \cdot (\Phi_C(t) - \Phi_C(t-1)) \end{array} \,\middle|\, i^1, \ldots, i^{t-1}, \mathcal{R}^1, \ldots, \mathcal{R}^{t-1}\right] \\
&\leq -\mathop{\mathbb{E}}_{i^t, \mathcal{R}^t}\left[(e+2) \cdot d_i^t \,\middle|\, i^1, \ldots, i^{t-1}, \mathcal{R}^1, \ldots, \mathcal{R}^{t-1}\right]
\end{aligned} \tag{7.6}$$

在所有轮次中 t。每轮算法在第 $e \cdot d_i^t$ 行最多购买10 个坐标（因为这只会在 $X_i^{t-1} \leq d_i^t$的情况下发生），并且根据第 d_i^t 行中的期望，抽取8个集合，并购买

最多 d_i^t 个坐标（第15行），总计期望购买不超过 $(e+2)\cdot d_i^t$ 个集合。因此，由（7.6）可知，对于所有轮次 t,

$$\mathop{\mathbb{E}}_{\substack{i^t\sim U^{(t)}\\ \mathcal{R}\sim x^{(t)}}}\left[\Phi(t)-\Phi(t-1)+c(\textsc{Alg}(t))-c(\textsc{Alg}(t-1))\;\middle|\; i^{(1)},\ldots,i^{(t-1)},\mathcal{R}^{(1)},\ldots,\mathcal{R}^{(t-1)}\right]\le 0. \quad (7.7)$$

对所有7.7反复应用（ $1\le t\le n$ 可得

$$\mathop{\mathbb{E}}_{i,\mathcal{R}}[\Phi(n)-\Phi(0)+c(\textsc{Alg}(n))-c(\textsc{Alg}(0))]\le 0$$

$$\mathop{\mathbb{E}}_{i,\mathcal{R}}[c(\textsc{Alg}(n))]\le c(\textsc{Alg}(0))+\Phi(0)-\mathop{\mathbb{E}}_{i,\mathcal{R}}[\Phi(n)],$$

因此，注意到 c（$\textsc{Alg}(0)$）$=0$ 并应用引理7.5 可得

$$\mathop{\mathbb{E}}_{i,\mathcal{R}}[c(\textsc{Alg}(n))]\le O(\beta\cdot\log mn), \quad (7.8)$$

如所需。 □

至此，我们完成了对 SetMultiCover的讨论。再次强调，我们的结果确定了该问题在随机顺序、先知、两阶段先知和带样本在线模型中的近似性。这为 [GKL21] 关于一般箱约束 CIPs 是否也能得到相同结果的问题提供了部分答案；理解一般情形仍是一个引人深思的开放问题。

8 结论

本文表明，随机集合覆盖问题的求解可以比[GGL+13]“更加不依赖信息”，只需关于生成输入的过程的粗略建议。我们不禁想尝试放宽集合覆盖的带建议在线模型，允许建议存在有界误差。我们在附录B中讨论了这一挑战，因为一些自然的放宽模型候选方案具有较强的下界。

我们将确定定理 α 中的紧确依赖关系作为一个有趣的开放问题提出1.3。我们猜想，该依赖关系应为 $O(\log(mn)\log(1/\alpha))$。定理1.3 意味着，当$\alpha=\Theta(1)$ 时，存在一个 $O(\log mn)$ 竞争比算法；而当 $\alpha=1/\mathrm{poly}(n)$ 时， $O(\log m\log n)$ 竞争比算法[BN09] 已是最优可能结果；这一猜想在这两个极端之间平滑过渡。

致谢 Roie Levin 感谢 Guy Even 提出关于两阶段先知模型的问题，也感谢 Niv Buchbinder 的有益讨论。

A 延后证明

下面给出支撑　　我们第6 节和第7节随机顺序算法所用引理的证明　　.

A.1 设施选址

引理 6.5（ Φ的界）。初始势函数有界为 $\Phi(0)=O(\beta\cdot\log mn)$，并且对所有 t， $\Phi(t)\ge-\beta\cdot\log n$ 。

证明。首先，我们声称，对于 x^* 支持集中的所有 f，都有 $c_f \le \beta$。为此，考虑 (6.1) 的一个最优分数解 (x^*, y^*)，并反设存在某个设施 $f \in \text{support}(x^*)$，使得 $x^*_f > \beta$。该设施 f 为某些客户提供了分数连接；令 $\epsilon := \min(y^*_{f,v} : y^*_{f,v} > 0)$ 表示为所有此类客户提供的最小连接量。最后，考虑如下扰动解：设置 $x'_{\hat{f}} = x^*_{\hat{f}} - \epsilon$，对 $f \neq f$ 设置 $x'_f = \min(x^*_f + \epsilon, 1)$ ˆ，并对所有 f, v 设置 $y'_{fv} = \min(y^*_{f,v} + \epsilon, 1)$。该解 (x', y') 对 ˆ (6.1) 仍然可行，且成本严格小于 β；因此不存在这样的 f。

既然已确立 $c_f \le \beta$对所有 f都成立，我们知道 $\text{support}(x^*) \subseteq \text{support}(x^0)$。接下来，我们将初始的 KL-散度项界定为

$$\begin{aligned}\mathrm{KL}_c\left(x^* \,||\, x^0\right) &= \sum_f c_f x^*_f \log\left(x^*_f \frac{c_f m'}{\beta}\right) + \sum_f c_f (x^0_f - x^*_f) \\ &\le \beta(\log m + 1),\end{aligned}$$

其中我们用到了上文的 $x^*_f c_f \le \beta$、$\langle c, x^*\rangle = \mathrm{LP}_{\mathrm{OPT}} \le \beta$，以及 $\langle c, x^0\rangle = \beta$。

接下来考虑 ϕ_L 中的第二项。这是分数松弛（6.1）目标函数的后半部分，在最优解处求值；因此它至多为 $\mathrm{LP}_{\mathrm{OPT}} \le \beta$。

现在来看最后一项 $\beta \cdot \log\left(\sum_v \kappa^0_v/\beta + 1/n\right)$。为了证明它至多为 $O(\beta \cdot \log n)$，只需证明对所有客户 v，$\kappa^0_v \le \beta$ 都成立。考虑将（6.1）松弛为仅服务客户 v的问题，即在时刻 0 为 v 分数地寻找成本最低的增补。由于这是（6.1）的松弛问题，其最优解的成本至多为 β。最后，该松弛问题会整数地选择设施 $f_v := \mathrm{argmin}_f\, c_f + c_{fv}$；因此 $\kappa^0_v \le \beta$。（由于增补成本只会降低，事实上对所有 v 以及 t，$\kappa^t_v \le \kappa^0_v \le \beta$。）

我们最后给出 Φ 的下界。Φ_L 中的两项均非负，因此 $\Phi_C \ge -\beta \cdot \log n$，因为 $\rho^t \ge 0$。

□

引理 6.6（Φ_L 的变化）。对于事件 Ξ^t 成立的轮次，Φ_L 的期望变化为

$$\begin{aligned}&\mathop{\mathbb{E}}_{v^t, \mathcal{R}^t}\left[\Phi_L(t) - \Phi_L(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t\right] \\ &\le \mathop{\mathbb{E}}_{v \sim U^{t-1}}\left[\frac{e^2 - 1}{2} \cdot \kappa^{t-1}_v \cdot \min\left(X^t(v),\, 1\right) - \kappa^{t-1}_v\right].\end{aligned} \tag{6.2}$$

当到达的 v^t 被抢先连接且 Ξ^t 不成立时，

$$\Phi_L(t) - \Phi_L(t-1) \le 0. \tag{6.3}$$

证明。不等式（6.3）显然成立：当第 5 行未执行时，x^t 不会发生变化，因此 KL 项保持不变。同时，分数最优项只会减小。

我们的主要任务是证明（6.2）。我们将证明分为几种情况。设 Λ^t 为事件：$X^t(v^t) < 1$。如果 Λ^t 不成立，则在第 12 行中，我们将向量设为 $x^t = x^{t-1}$，因此 KL 项的变化同样为零。这意味着不等式（6.2）显然成立，因为 $(e^2 - 1)/2 > 1$。此后我们只考虑 Λ^t 成立的情况。

回顾一下，相对熵的期望变化仅取决于到来的未覆盖元素v^t。从 KL 项开始并展开定义，在上下文明确时将 $v = v^t$ 记作，

$$\mathop{\mathbb{E}}_{v^t, \mathcal{R}^t}\left[\mathrm{KL}_c\left(x^* \,||\, x^t\right) - \mathrm{KL}_c\left(x^* \,||\, x^{t-1}\right) \mid x^{t-1}, U^{t-1}, \Lambda^t\right]$$

$$= \mathop{\mathbb{E}}_{v\sim U^{t-1}}\left[\sum_f c_f \cdot x_f^* \cdot \log\frac{x_f^{t-1}}{x_f^t}\,\middle|\,\Xi^t,\Lambda^t\right]$$

$$= \mathop{\mathbb{E}}_{v\sim U^{t-1}}\left[\langle c,x^*\rangle\cdot\log Z^t - \sum_{f\in\Gamma^{t-1}(v)} c_f\cdot x_f^*\cdot\log e^{\kappa_v^{t-1}/c_f}\,\middle|\,\Xi^t,\Lambda^t\right]$$

$$\le \mathop{\mathbb{E}}_{v\sim U^{t-1}}\left[\begin{array}{l}\beta\cdot\log\left(\displaystyle\sum_{f\in\Gamma^{t-1}(v)}\frac{c_f}{\beta}\cdot x_f^{t-1}\cdot e^{\kappa_v^{t-1}/c_f} + \sum_{f\notin\Gamma^{t-1}(v)}\frac{c_f}{\beta}\cdot x_f^{t-1}\right)\\ -\displaystyle\sum_{f\in\Gamma^{t-1}(v)}\kappa_v^{t-1}\cdot x_f^*\end{array}\,\middle|\,\Xi^t,\Lambda^t\right], \tag{A.1}$$

其中，在最后一步（A.1）中，我们展开了 Z^t 的定义，并使用了 $\langle c,\ x^*\rangle \le \beta$。然后，我们可以进一步将（A.1）界定为

$$\le \mathop{\mathbb{E}}_{v\sim U^{t-1}}\left[\beta\cdot\log\left(\sum_f\frac{c_f}{\beta}\cdot x_f^{t-1} + \frac{e^2-1}{2}\cdot\frac{\kappa_v^{t-1}}{\beta}\cdot X^t(v)\right) - \sum_{f\in\Gamma^{t-1}(v)}\kappa_v^{t-1}\cdot x_f^*\,\middle|\,\Xi^t,\Lambda^t\right], \tag{A.2}$$

其中，我们对 $e^a \le 1+(e^2-1)\cdot a/2$ 使用近似比 $a\in[0,2]$ (注意，κ_v^{t-1}是 v的最小边际连接成本，因此对于任意 $f\in\Gamma^{t-1}(v)$，也就是说，$c_{fv}\le\kappa_v^{t-1}/2$，有$c_f \ge \kappa_v^{t-1}/2$ 因此 $\kappa_v^{t-1}/c_f \le 2$)。最后，利用不变量 1，并结合近似式 $\log(1+y)$，我们将 (A.2) 的上界估计为

$$\le \mathop{\mathbb{E}}_{v\sim U^{t-1}}\left[\frac{e^2-1}{2}\cdot\kappa_v^{t-1}\cdot X^t(v) - \sum_{f\in\Gamma^{t-1}(v)}\kappa_v^{t-1}\cdot x_f^*\,\middle|\,\Xi^t,\Lambda^t\right]$$

$$\le \mathop{\mathbb{E}}_{v\sim U^{t-1}}\left[\frac{e^2-1}{2}\cdot\kappa_v^{t-1}\cdot\min\left(X^t(v),\,1\right) - \sum_{f\in\Gamma^{t-1}(v)}\kappa_v^{t-1}\cdot x_f^*\,\middle|\,\Xi^t,\Lambda^t\right], \tag{A.3}$$

其中 (A.3) 由事件 Λ^t 的定义可得。

现在我们讨论 Φ_L 的第二部分。每一轮分数最优项的变化为$-2\cdot\sum_f c_{fv^t}\cdot y^*_{fv^t}$。结合 (A.3) 可得

$$\mathop{\mathbb{E}}_{v\sim U^{t-1}}\left[\Phi_L(t)-\Phi_L(t-1)\,\middle|\,\Xi^t,\Lambda^t\right]$$

$$\le \mathop{\mathbb{E}}_{v\sim U^{t-1}}\left[\kappa_v^{t-1}\cdot\frac{e^2-1}{2}\cdot\min\left(X^t(v),\,1\right) - \left(\sum_{f\in\Gamma^{t-1}(v)}\kappa_v^{t-1}\cdot x_f^* + 2\cdot\sum_f c_{fv^t}\cdot y^*_{fv}\right)\,\middle|\,\Xi^t,\Lambda^t\right].$$

由于分数连接 v 从 $\Gamma^{t-1}(v)$ 外部获得的连接按定义至少需要 κ_v^{t-1} 的成本，因此我们可以将其界定为

$$\le \mathop{\mathbb{E}}_{v\sim U^{t-1}}\left[\kappa_v^{t-1}\cdot\frac{e^2-1}{2}\cdot\min\left(X^t(v),\,1\right) - \left(\sum_{f\in\Gamma^{t-1}(v)}\kappa_v^{t-1}\cdot x_f^* + 2\cdot\sum_{f\notin\Gamma^{t-1}(v)}\frac{\kappa_v^{t-1}}{2}\cdot y^*_{fv}\right)\,\middle|\,\Xi^t,\Lambda^t\right]$$

$$\le \mathop{\mathbb{E}}_{v\sim U^{t-1}}\left[\kappa_v^{t-1}\cdot\frac{e^2-1}{2}\cdot\min\left(X^t(v),\,1\right) - \kappa_v^{t-1}\,\middle|\,\Xi^t,\Lambda^t\right], \tag{A.4}$$

其中，(A.4) 之所以成立，是因为分数连接 v 在 (x^*, y^*) 中从 $\Gamma^{t-1}(v)$ 外部获得的连接至少为 $1 - \sum_{f \in \Gamma^{t-1}(v)} x^*_f$。
我们已经证明了在 Λ^t 成立和不成立两种情况下的引理陈述，由此完成证明。

□

引理 6.7（变　　化 Φ_C）。对于所有满足 Ξ^t 的轮次 t，Φ_C 的期望变化为

$$\mathop{\mathbb{E}}_{v^t, \mathcal{R}^t}\left[\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t\right] \le -\frac{1-e^{-1}}{4} \cdot \mathop{\mathbb{E}}_{u \sim U^{t-1}}\left[\kappa_u^{t-1} \cdot \min\left(X^t(v),\ 1\right)\right]. \tag{6.4}$$

对于 Ξ^t 不成立的轮次，

$$\Phi_C(t) - \Phi_C(t-1) \le 0. \tag{6.5}$$

证明。等式（6.5）再次显而易见，因为 ρ^t 关于 t单调递减。因此，我们着重证明（6.4）。

我们先针对固定的到达客户端 v，考察采样随机性下 Φ_C 的期望变化。展开定义可得，

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathcal{R}^t}\left[\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t, v^t = v\right] \\
&= \beta \cdot \mathop{\mathbb{E}}_{\mathcal{R}^t}\left[\log\left(1 - \frac{\rho^{t-1} - \rho^t}{\rho^{t-1} + \frac{\beta}{n}}\right) \,\middle|\, U^{t-1}, \Xi^t, v^t = v\right] \\
&\le -\beta \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{n}} \cdot \mathop{\mathbb{E}}_{\mathcal{R}^t}\left[\rho^{t-1} - \rho^t \,\middle|\, U^{t-1}, \Xi^t, v^t = v\right].
\end{aligned} \tag{A.5}$$

上文中，式（A.5）由近似界 $\log(1-y) \le -y$得出。展开 ρ^t的定义后，式（A.5）的上界为

$$\begin{aligned}
&\le -\frac{\beta}{\rho^{t-1} + \frac{\beta}{n}} \cdot \mathop{\mathbb{E}}_{\mathcal{R}^t}\left[\sum_{u \in U^{t-1}} \frac{\kappa_u^{t-1}}{2} \cdot \mathbb{I}\{\Gamma^{t-1}(u) \cap \mathcal{R}^t \neq \emptyset\} \,\middle|\, U^{t-1}, \Xi^t, v^t = v\right] \\
&= -\frac{\beta}{\rho^{t-1} + \frac{\beta}{n}} \sum_{u \in U^{t-1}} \frac{\kappa_u^{t-1}}{2} \cdot \mathop{\mathbb{P}}_{\mathcal{R}^t}(\Gamma^{t-1}(u) \cap \mathcal{R}^t \neq \emptyset \mid U^{t-1}, \Xi^t, v^t = v) \\
&\le -\frac{\beta}{\rho^{t-1} + \frac{\beta}{n}} \sum_{u \in U^{t-1}} \frac{\kappa_u^{t-1}}{2} \cdot (1 - e^{-1}) \cdot \min\left(\frac{\kappa_v^{t-1}}{\beta} \cdot \sum_{f \in \Gamma^{t-1}(u)} x_f^{t-1},\ 1\right) &\text{(A.6)} \\
&\le -\frac{1-e^{-1}}{2} \cdot \kappa_v^{t-1} \cdot \frac{|U^{t-1}|}{\rho^{t-1} + \frac{\beta}{n}} \cdot \mathop{\mathbb{E}}_{u \sim U^{t-1}}\left[\kappa_u^{t-1} \cdot \min\left(X^t(u),\ 1\right)\right]. &\text{(A.7)}
\end{aligned}$$

步骤（A.6）源于这样一个事实：每个设施 f 都以概率 $\min(\kappa_v^{t-1} x_f^{t-1}/\beta,\ 1)$ 独立抽样，因此任意给定的客户 $u \in U^{t-1}$ 从$\Gamma^{t-1}(u)$中获得至少一个设施的概率为

$$1 - \prod_{f \in \Gamma^{t-1}(u)}\left(1 - \min\left(\frac{\kappa_v^{t-1} x_f^{t-1}}{\beta},\ 1\right)\right) \ge 1 - \exp\left\{-\min\left(\frac{\kappa_v^{t-1}}{\beta} \sum_{f \in \Gamma^{t-1}(u)} x_f^{t-1},\ 1\right)\right\}$$

$$\overset{(**)}{\geq} (1-e^{-1}) \cdot \min\left(\frac{\kappa_v^{t-1}}{\beta} \sum_{f \in \Gamma^{t-1}(u)} x_f^{t-1},\ 1\right).$$

上文中，（$**$）由指数函数的凸性得出。接着，步骤（A.7）可通过将求和重写为期望并利用以下事实得出： $\kappa_v^{t-1}/\beta \leq 1$，这在引理 6.5 的证明中得到了论证（并被使用）。

对A.7)关于 $v^t \sim U^{t-1}$取期望，并利用$\mathbb{E}_{v\sim U^{t-1}}\left[\kappa_v^{t-1}\right] = \rho^{t-1}/|U^{t-1}|$这一事实， Φ_C 的期望变化变为

$$\begin{aligned}
&\mathop{\mathbb{E}}_{v^t,\mathcal{R}^t}\left[\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t\right] \\
&\leq -\frac{1-e^{-1}}{2} \cdot \frac{\rho^{t-1}}{\rho^{t-1} + \frac{\beta}{n}} \cdot \mathop{\mathbb{E}}_{u\sim U^{t-1}}\left[\kappa_u^{t-1} \cdot \min\left(X^t(u),\ 1\right)\right], \\
&\leq -\frac{1-e^{-1}}{4} \cdot \mathop{\mathbb{E}}_{u\sim U^{t-1}}\left[\kappa_u^{t-1} \cdot \min\left(X^t(u),\ 1\right)\right], \qquad \text{(A.8)}
\end{aligned}$$

在A.8)中，我们最后用到了事件 Ξ^t 成立这一事实。既然如此，我们知道v^t在第 t轮次并未抢先连接，因此 $\rho^{t-1} \geq \kappa_{v^t}^{t-1} \geq \beta/t \geq \beta/n$，因此 $\frac{\rho^{t-1}}{\rho^{t-1}+\frac{\beta}{n}} \geq \frac{1}{2}$。这就是所要证明的界。

□

A.2 集合多重覆盖

引理 7.5（关于 Φ的界）。初始势函数有界为 $\Phi(0) = O(\beta \cdot \log mn)$，且对所有轮次 t， $\Phi(t) \geq -\beta \cdot \log(n)$。

证明。我们先讨论上界，并依次处理势函数的各个部分。首先，$\mathrm{KL}\left(x^* \mid\mid x^0\right) = \sum_j x_j^* \log \left(\frac{x_j^*}{x_j^0}\right) = \sum_j x_j^* \log \left(m\frac{x_j^*}{\beta}\right) \leq \sum_j x_j^* \log\ m \leq \beta \log\ m$，因为对于所有 j， $x_j^* \leq \mathrm{LP}_{\mathrm{OPT}} \leq \beta$。其次，$\beta \cdot \log \left(\frac{\rho^0}{\beta} + \frac{1}{m}\right) \leq \beta \cdot \log \left(\sum_i \frac{b_i}{\beta}\right) \leq \beta \log\ n$，因为对于所有 b_i， $b_i \leq \mathrm{LP}_{\mathrm{OPT}} \leq \beta$。

现在我们证明下界。KL 散度非负， ρ^t也非负；因此 $\Phi(t) \geq \beta \cdot \log(1/m) = -\beta \cdot \log m$。

□

引理 7.6（ Φ_L的变化）。对于 i^t 到达时未被覆盖的轮次， Φ_L 的期望变化为

$$\mathop{\mathbb{E}}_{i^t,\mathcal{R}^t}\left[\Phi_L(t) - \Phi_L(t-1) \mid x^{t-1}, U^{t-1}, d_{i^t}^t > 0\right] \leq \mathop{\mathbb{E}}_{i\sim U^{t-1}}[(e-1) \cdot \min(X_i^{t-1},\ d_i^t) - d_i^t]. \qquad (7.2)$$

当 i^t 到达时已覆盖时，

$$\Phi_L(t) - \Phi_L(t-1) \leq 0. \qquad (7.3)$$

证明。 我们首先证明（7.3）。这是因为，如果 i^t 到达时已覆盖，则 n

我们将（7.2）的证明分为两种情况。若 $X_i^{t-1} \geq d_i^t$，根据第 9 行，向量 x^t 在第 t轮中未更新，因此 KL 散度的变化为 0，这意味着

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i^t,\mathcal{R}^t}\left[\mathrm{KL}\left(x^* \mid\mid x^t\right) - \mathrm{KL}\left(x^* \mid\mid x^{t-1}\right) \mid x^{t-1}, U^{t-1}, X_{i^t}^{t-1} \geq d_{i^t}^t\right] \\
&\leq \mathop{\mathbb{E}}_{i\sim U^{t-1}}[(e-1) \cdot \min(X_i^{t-1}, d_i^t) - d_i^t \mid X_i^{t-1} \geq d_i^t], \qquad \text{(A.9)}
\end{aligned}$$

从而（7.2）显然成立。接下来我们重点讨论 $X_i^{t-1} < d_i^t$ 的情况。

回想一下，相对熵的期望变化只取决于到达时未覆盖的元素i^t。展开定义可得，

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i^t,\mathcal{R}^t}\left[\mathrm{KL}\left(x^* \mid\mid x^{t+1}\right) - \mathrm{KL}\left(x^* \mid\mid x^t\right) \mid x^{t-1}, U^{t-1}, X_{i^t}^{t-1} < d_i^t\right] \\
&= \mathop{\mathbb{E}}_{i\sim U^{t-1}}\left[\sum_j x_j^* \cdot \log \frac{x_j^{t-1}}{x_j^t} \;\middle|\; X_i^{t-1} < d_i^t\right] \\
&= \mathop{\mathbb{E}}_{i\sim U^{t-1}}\left[\sum_j x_j^* \cdot \log Z^t - \sum_{j\in\mathcal{T}^t} x_j^* \cdot \log e \;\middle|\; X_i^{t-1} < d_i^t\right] \\
&\le \mathop{\mathbb{E}}_{i\sim U^{t-1}}\left[\beta \cdot \log Z^t - \sum_{j\in\mathcal{T}^t} x_j^* \;\middle|\; X_i^{t-1} < d_i^t\right] \qquad \text{(A.10)} \\
&\le \mathop{\mathbb{E}}_{i\sim U^{t-1}}\left[\beta \cdot \log Z^t - d_i^t \mid X_i^{t-1} < d_i^t\right], \qquad \text{(A.11)}
\end{aligned}$$

其中在 (A.10) 中我们使用了 $\langle 1,\ x^*\rangle \le \beta$。由于 x^* 是一个可行的分数集合覆盖，我们知道 $\sum_j a_{ij}x_j^* \ge b_i$，并且由于 $x_j^* \le 1$，这意味着 $\sum_{j\in\mathcal{T}^t} x_j^* \ge d_i^t$，得到 (A.11)。展开 Z^t，可得

$$\begin{aligned}
&= \mathop{\mathbb{E}}_{i\sim U^{t-1}}\left[\beta \cdot \log\left(\frac{1}{\beta}\sum_j x_j^{t-1} + \frac{(e-1)}{\beta}\sum_{j\in\mathcal{T}^t} x_j^{t-1}\right) - d_i^t \;\middle|\; X_i^{t-1} < d_i^t\right] \\
&= \mathop{\mathbb{E}}_{i\sim U^{t-1}}\left[\beta \cdot \log\left(1 + \frac{(e-1)}{\beta}X_i^{t-1}\right) - d_i^t \;\middle|\; X_i^{t-1} < d_i^t\right] \qquad \text{(A.12)}
\end{aligned}$$

最后，近似界 $\log(1+y) \le y$ 使我们可以将 (A.12) 上界为

$$\begin{aligned}
&\le \mathop{\mathbb{E}}_{i\sim U^{t-1}}\left[(e-1)\cdot X_i^{t-1} - d_i^t \mid X_i^{t-1} < d_i^t\right] \\
&= \mathop{\mathbb{E}}_{i\sim U^{t-1}}[(e-1)\cdot \min(X_i^{t-1}, d_i^t) - d_i^t \mid X_i^{t-1} < d_i^t]. \qquad \text{(A.13)}
\end{aligned}$$

结合 (A.9) 和 (A.13)，再利用全期望定律，即可得到该引理的结论。 □

引理 7.7（ Φ_C的变化）。在 i^t 到达时未覆盖的每一轮中， Φ_C 的期望变化为

$$\mathop{\mathbb{E}}_{i^t,\mathcal{R}^t}\left[\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, d_{i^t}^t > 0\right] \le -\frac{\gamma}{2} \cdot \mathop{\mathbb{E}}_{i\sim U^{t-1}}\left[\min\left(X_i^{t-1},\ d_i^t\right)\right], \qquad (7.4)$$

其中 γ 是一个固定常数。在 i^t 到达时已覆盖的轮次中，

$$\Phi_C(t) - \Phi_C(t-1) \le 0. \qquad (7.5)$$

我们将 需要一个额外的事实，其证明见 [GKL21，附录 A] :

事实 A.1. 给定概率 p_j 和系数 $b_j \in [0,1]$，令 $W := \sum_j b_j \,\mathrm{Ber}(p_j)$ 为相互独立的加权伯努利随机变量之和。设 $\Delta \ge (e-1)^{-1}$ 为某个常数。那么

$$\mathbb{E}[\min(W, \Delta)] \ge \gamma \cdot \min(\mathbb{E}[W], \Delta),$$

对于一个固定常数 γ，它与 p_j 和 b_j 无关。

引理的证明 7.7。 当 i^t 到达时已覆盖（7.5）显然成立，因为此时 $d_{i^t}^t = 0$，因此不会购买任何集合 j，并且 $\rho^t = \rho^{t-1}$。因此，我们着重证明在 7.4）的情形下 i^t到达时未覆盖。

在给定 $i^t = i$ 的条件下，$\log \rho^t$ 的期望变化仅取决于 $\mathcal{R}^t$:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i^t,\mathcal{R}^t}\left[\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, i^t = i\right] \\
&= \beta \cdot \mathop{\mathbb{E}}_{\mathcal{R}^t}\left[\log\left(1 - \frac{\rho^{t-1} - \rho^t}{\rho^{t-1} + \frac{\beta}{m}}\right) \,\middle|\, U^{t-1}, i^t = i\right] \\
&\le -\beta \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{m}} \mathop{\mathbb{E}}_{\mathcal{R}^t}\left[\rho^{t-1} - \rho^t \mid U^{t-1}, i^t = i\right]. \qquad \text{(A.14)}
\end{aligned}$$

上文中的（A.14）由近似式 $\log(1-y) \le -y$得出。再次展开定义，我们进一步将（A.14）上界为

$$\begin{aligned}
&= -\beta \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{m}} \mathop{\mathbb{E}}_{\mathcal{R}^t}\left[\sum_{i' \in U^{t-1}} (d_{i'}^{t-1} - d_{i'}^t)\right] \qquad \text{(A.15)} \\
&= -\beta \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{m}} \sum_{i' \in U^{t-1}} \mathop{\mathbb{E}}_{\mathcal{R}^t}[d_{i'}^{t-1} - d_{i'}^t] \\
&\le -\beta \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{m}} \sum_{i' \in U^{t-1}} \mathop{\mathbb{E}}_{\mathcal{R}^t}\left[\min\left(\sum_{j \in \mathcal{T}^t} \mathrm{Ber}(x_j^{t-1} \cdot d_i^{t-1}/\beta),\ d_{i'}^t\right)\right]. \qquad \text{(A.16)}
\end{aligned}$$

这里的（A.16）由上一行根据第 8 行所执行的随机抽样定义得出（这里是不等式，因为算法也会购买其他坐标）。该伯努利概率是良定义的，因为 $d_{i'}^{t-1} \le \mathrm{LP}_{\mathrm{OPT}} \le \beta$，且由于第 10 行保证了 $x_j^{t-1} \le 1$对所有 j 属于 $\mathcal{T}^t$。这些伯努利变量之和的期望为 $\sum_{j \in \mathcal{T}^t} x_j^{t-1} d_i^{t-1}/\beta = X_i^{t-1} d_i^{t-1}/\beta$。由于$d_i^{t-1} \ge 1$，我们可以应用事实 A.1 得到

$$\begin{aligned}
&\le -\frac{1}{\rho^{t-1} + \frac{\beta}{m}} \sum_{i' \in U^{t-1}} \gamma \cdot \min\left(\frac{d_i^{t-1}}{\beta} \cdot X_{i'}^{t-1},\ d_{i'}^{t-1}\right) \qquad \text{(A.17)} \\
&\le -\gamma \cdot d_i^{t-1} \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{m}} \sum_{i' \in U^{t-1}} \min\left(X_{i'}^{t-1}, d_{i'}^{t-1}\right) \\
&= -\gamma \cdot d_i^{t-1} \cdot \frac{|U^{t-1}|}{\rho^{t-1} + \frac{\beta}{m}} \cdot \mathop{\mathbb{E}}_{i' \sim U^{t-1}}\left[\min\left(X_{i'}^{t-1}, d_{i'}^{t-1}\right)\right] \\
&\le -\gamma \cdot d_i^{t-1} \cdot \frac{|U^{t-1}|}{2 \cdot \rho^{t-1}} \cdot \mathop{\mathbb{E}}_{i' \sim U^{t-1}}\left[\min\left(X_{i'}^{t-1}, d_{i'}^{t-1}\right)\right], \qquad \text{(A.18)}
\end{aligned}$$

其中，（A.18）由以下观察得出：i^t 到达时尚未被覆盖，因此 $\rho^{t-1} \ge d_i^t \ge 1$ 而 $\beta/m \le 1$；因此 $\frac{1}{\rho^{t-1} + \beta/m} \le \frac{1}{2 \cdot \rho^{t-1}}$。对 $i \sim U^{t-1}$取（A.18）的期望，并利用 $\mathbb{E}_{i \sim U^{t-1}}\left[d_i^{t-1}\right] = \rho^{t-1}/|U^{t-1}|$这一事实，$\log\ \rho^t$ 的期望变化为

$$\mathop{\mathbb{E}}_{i^t,\mathcal{R}^t}\left[\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}\right] \le -\frac{\gamma}{2} \cdot \mathop{\mathbb{E}}_{i \sim U^{t-1}}\left[\min\left(X_i^{t-1}, d_i^t\right)\right],$$

如所需。 □

B 放宽模型下的下界

B.1 对抗性扰动

人们很容易想尝试将带样本模型扩展到样本带有噪声的情形。本节研究一种自然的模型，并遗憾地证明，除非 P $=$NP，否则不存在能在多项式时间内实现 $o(\log m \log n)$ 竞争比的随机化算法。

我们首先描述该模型，将其称为（α，δ）-NoisySampleSetCover。对手首先确定一个长度为 n 的在线集合覆盖输入序列，该序列来自一个包含 N 个元素和 m 个集合的集合系统。随后，从输入 S_1 中均匀随机抽取一个 $\alpha \cdot n$。然后，对手选择一个大小为 $\delta \cdot n$ 的集合 S_2。算法预先获得约束集合 $S = S_1 \cup S_2$。最后，在线序列开始。

现在，当 α、$\delta =$ 为 $\Omega(1)$ 时，我们为（α，δ）-NoisySampleSetCover 构造困难实例。我们将使用 [Kor04] 中的构造作为子实例，该构造为原始在线集合覆盖问题给出了 $\Omega(\log m \log n)$ 的下界。令（$\mathcal{U}_{hard}$，$\mathcal{S}_{hard}$）表示 [Kor04]实例中的基础集合系统。令 σ_{hard} 表示提供给算法的元素序列，并将这些元素组成的无序集合称为 $U_{hard} \subseteq \mathcal{U}_{hard}$。在 [Kor04] 的构造中，$U_{hard}$ 是一个随机变量，使得 $|U_{hard}| = \Theta(\sqrt{|\mathcal{U}_{hard}|})$。

我们的构造如下。集合系统由一个包含 $N - \delta\sqrt{N}$ 个元素的集合 S_0，以及集合系统 $(\mathcal{U}_{hard}, \mathcal{S}_{hard})$ 的一个副本组成，其参数为 $n' := |\mathcal{U}_{hard}| = \delta\sqrt{N}$ 和 $m' := |\mathcal{S}_{hard}| = \text{poly}(n')$。对手预先确定在线序列：先以任意顺序揭示 S_0 的所有元素，然后按顺序 σ_{hard} 揭示 $U_{hard} \subseteq \mathcal{U}_{hard}$ 的元素。该序列的长度为 $\Theta(N)$。对手选取大小为 $\delta \cdot N \geq \delta\sqrt{N}$ 的 S_2，使其包含 $\mathcal{U}_{hard}$ 的所有元素。因此，无论样本 S_1 的实现为何，算法都无法获知 U_{hard} 的身份，必须在没有有用建议的情况下覆盖长度为 $\Theta(\sqrt{n'}) = \Theta(\delta^{1/2} N^{1/4})$ 的困难在线集合覆盖序列 σ_{hard}。由于任何多项式时间随机化在线集合覆盖算法在 σ_{hard} 上的竞争比都是 $\Omega(\log n' \log m')$，当 $\alpha, \delta = \Omega(1)$ 时，任何算法都无法在 (α, δ)-NoisySampleSetCover 上达到 $o(\log n \log m)$ 的竞争比。

B.2 RelaxedRandom Order

另一个有趣的问题是，是否可以[GKL21] 在输入顺序并非完全均匀随机、而只是近似均匀时仍然适用。

我们证明，到达顺序分布的熵并不能很好地刻画与随机顺序的距离：存在一些实例及其到达顺序上的分布，其熵几乎达到最大值 $(1-\epsilon)n\log n \approx \log(n!)$，但任何在线算法的竞争比都是 $\Omega(\log(\epsilon m)\log(\epsilon n))$.

一个简单的此类实例如下：先以均匀随机顺序呈现 $(1-\epsilon)n$ 个虚拟元素，随后呈现一个困难在线集合覆盖序列 σ_{hard}，长度为 ϵn。该排列分布的熵接近最大值，但根据 $o(\log(\epsilon m)\log(\epsilon n))$ 的下界，除非 P $=$ NP，否则不存在竞争比为 [Kor04] 的随机多项式时间算法。

C 错误于 [DEH+18]

在 [DEH+18] 中，作者声称，通过将先知集合覆盖归约到已知的独立同分布集合覆盖（当所有分布

相同时的特殊情形），可以得到一个 $O(1)$-竞争比算法 MetricFacilityLo-cation 。（他们还声称，通过相同的归约，可以得到一个用于先知集合覆盖的 [EB17，定理 9。10] 其归约证明依赖于以下论断。

令 $D^1, \ldots, D^n$ 为定义在 $\mathcal{U}$元素上的一列分布，并令 $D^* = \frac{1}{n}\sum_{i=1}^{n} D^i$ 为平均分布。令 $\textsc{Opt}_{pht}$ 为当通过分别从 $D^1, \ldots, D^n$中各抽取一个元素来构造 U 时，其最优集合覆盖的期望大小。令 $\textsc{Opt}_{iid}$为当通过从 D^*中抽取 n 次来构造 U 时，最优集合覆盖的大小。我们声称 $\textsc{Opt}_{iid} \le \textsc{Opt}_{pht}$，现在证明这一结论并非普遍成立。

Co考虑全集为 (i, j)、对应于 $i \in [2]$ 和 $j \in [2]$ 的实例，以及集合系统

$$\mathcal{S} = \{\{(1,1),(2,1)\},\{(1,1),(2,2)\},\{(1,2),(2,1)\},\{(1,2),(2,2)\}\}.$$

令先知分布为 D^1 和 D^2，它们分别是在 (1, 1)、(1, 2) 和 (2, 1)、(2, 2) 上的均匀分布。令 $D^* = \frac{1}{2}(D^1 + D^2)$ 为平均分布。于是

$$\textsc{Opt}_{pht} = 1, \qquad \mathbb{E}[\textsc{Opt}_{iid}] = \frac{5}{4}.$$

更一般地，令全集为 (i, j)，其中 $i \in [n]$ and $j \in [\log\ n]$，并令集合的集合为所有序列 $\mathcal{S} = [\log\ n]^{[n]}$。那么， n 先知分布 D^i 分别是在{$(i, 1)$、$(i, 2)$.....、$(i, \log\ n)$}上的均匀分布，而平均分布与前面相同。通过球和箱子的论证可得

$$\textsc{Opt}_{pht} = 1, \qquad \mathbb{E}[\textsc{Opt}_{iid}] = \Omega\left(\frac{\log n}{\log\log n}\right),$$

这表明，在最坏情况下，所声称的不等式至少被这一乘法因子所违反。

参考文献

[AFGS22] C.J. Argue、Alan Frieze、Anupam Gupta 和 Christopher Seiler。从样本中学习在线算法。载于 *NeurIPS*，2022 年 12 月。[AKW14] Pablo Daniel Azar、Robert Kleinberg 和 S. Matthew Weinberg。信息有限条件下的先知不等式。载于 Chandra Chekuri（编），*Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms SODA 2014 Portland Oregon USA January 5-7 2014* 1358 1377 ，，，，，，

页码 –。SIAM，2014 年。[BC22] Archit Bubna 和 Ashish Chiplunkar。先知不等式：顺序选择胜过随机顺序。 *CoRR*，abs/2211.04145，2022 年。[BL11] John R Birge 和 Francois Louveaux。 *Introduction to stochastic programming*。Springer Science & Business Media，2011 年。[BN09] Niv Buchbinder 和 Joseph Naor。覆盖与装箱的在线原始-对偶算法。 *Math. Oper.*，

Res. 34 2 270 286, 2009

（）：–。[CCP05] Moses Charikar、Chandra Chekuri 和 Martin P´al。随机优化的抽样界。载于 Chandra Chekuri、Klaus Jansen、Jos´e D. P. Rolim 和 Luca Trevisan（编），*Approximation, Randomization and Combinatorial Optimization, Algorithms and Techniques, 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2005 and 9th InternationalWorkshop on Randomization and Computation, RANDOM 2005, Berkeley, CA, USA, August 22-24,* 2005 3624 257 269 2005， *Proceedings*， *Lecture Notes in Computer Science*卷，页码 –。Springer，。[CCZ14] Michele Conforti、G´erard Cornu´ejols 和 Giacomo Zambelli。 *Integer Programming*。Springer国际出版，2014 年。

[Chv79] Vasek Chvatal。集合覆盖问题的贪心启发式算法。 *Mathematics of operations research*，4 3 233 235，1979（）：–。

[DEH+18] Sina Dehghani、Soheil Ehsani、MohammadTaghi Hajiaghayi、Vahid Liaghat 和 Saeed Seddighin。在线可生存网络设计的贪心算法。载于 *45th International Colloquium on Automata Lan-*，。Schloss *guages and Programming (ICALP 2018)* 2018 Dagstuhl-Leibniz-Zentrum fuer Informatik，。[DS14] Irit Dinur 和 David Steurer。平行重复的分析方法。载于 *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing*，STOC '14，页码 624–633，美国纽约州纽约市，2014 年。ACM。[EB17] Soheil Ehsani Banafati。 *Online Decision Making via Prophet Setting*。博士论文，马里兰大学，2017 年。[F+89] Thomas S Ferguson 等。谁解决了秘书问题？ *Statistical science*，4(3):282–289，1989 年。[Fei 98] Uriel Feige。用于近似集合覆盖的 ln 阈值 n。 *J. ACM*，45(4):634–652，1998 年。[GGL+13] Fabrizio Grandoni、Anupam Gupta、Stefano Leonardi、Pauli Miettinen、Piotr Sankowski 和 Mohit Singh。闭眼进行集合覆盖。 *SIAM J. Comput.*，42(3):808–830，2013 年。[GGLS08] Naveen Garg、Anupam Gupta、Stefano Leonardi 和 Piotr Sankowski。在线组合优化问题的随机分析。载于 Shang-Hua Teng（编）， *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms SODA 2008 San Francisco California USA January*，，，，，，页码 *20-22 2008* 942 951 2008 –。SIAM，。[GKL21] Anupam Gupta、Gregory Kehne 和 Roie Levin。随机顺序在线集合覆盖与离线问题同样容易。载于 *62nd IEEE Annual Symposium on Foundations of Computer Science*、 *FOCS 2021*、 *Denver*、 *CO*、 *USA*、 *February 7-10 2022* 1253 1264 2021 ，页码 –。IEEE，。[GPRS04] Anupam Gupta、Martin P´al、R. Ravi 和 Amitabh Sinha。增强采样：随机优化的近似算法。载于 L´aszlo Babai（编）， *Proceedings of the 36th Annual ACM Symposium on Theory of Computing Chicago IL USA June 13-16 2004* 417 426 2004 ，，，，，，页码 –。ACM，。[GPRS05] Anupam Gupta、Martin P´al、R. Ravi 和 Amitabh Sinha。星期三怎么办？多阶段随机优化的近似算法。载于 Chandra Chekuri、Klaus Jansen、Jos´e D. P. Rolim 和 Luca Trevisan（编）， *Approximation*、 *Randomization and Combinatorial Optimization*、 *Algorithms andTechniques*、 *8th International Workshop on Approximation Algorithms for Combinatorial OptimizationProblems*、 *APPROX 2005 and 9th InternationalWorkshop on Randomization and Computation*、 *RANDOM 2005 Berkeley CA USA August 22-24 2005 Proceedings* 3624 *Lecture Notes in*，，，，，，卷，*Computer Science* 86 98 2005 页码 –。Springer，。[GRS07] Anupam Gupta、R. Ravi 和 Amitabh Sinha。随机网络设计的线性规划舍入近似算法。 *Math. Oper. Res.*，32(2):345–364，2007 年。[HK92] Theodore P Hill 和 Robert P Kertz。最优停止理论中先知不等式综述。 *Contemp.*：–。[JLN+05] *Math* 125 191 207, 1992 Lujun Jia、Guolong Lin、Guevara Noubir、Rajmohan Rajaraman 和 Ravi Sundaram。旅行商问题、斯坦纳树和集合覆盖的通用近似。载于 Harold N. Gabow 和 Ronald Fagin（编）， *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*、 *Baltimore*、 *MD*、 *USA*、 *May 22-24*，2005 年，页码 386–395。ACM，2005 年。[Joh74] David S Johnson。组合问题的近似算法。 *Journal of computer and system sciences* 9 3 256 278, 1974 ，（）：–。[KNR20] Haim Kaplan、David Naori 和 Danny Raz。基于样本的竞争分析与秘书问题。载于 Shuchi Chawla（编）， *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms SODA*，*Salt Lake City*，，，*2020 UT USA 2020* 2082 2095 2020 *January 5-8*，页码 –。SIAM，。[KNR22] Haim Kaplan、David Naori 和 Danny Raz。基于样本的在线加权匹配。载于 Joseph (Seffi) Naor 和 Niv Buchbinder（编）， *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms SODA 2022 VA USA 2022* 1247 1272 ，，*Virtual Conference / Alexandria*，，， *January 9 - 12*，页码 –。SIAM，2022 年。[KNR23] Haim Kaplan、David Naori 和 Danny Raz。随机顺序模型下在线设施选址的近乎紧确界。载于 Nikhil Bansal 和 Viswanath Nagarajan（编）， *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms SODA 2023 Florence Italy January 22-25 2023* 1523 1544，，，，页码 –。SIAM，2023 年。[Kor04] Simon Korman。论随机化在在线集合覆盖问题中的应用。 *Master's thesis Weizmann*，，，，，。 *Institute of Science Rehovot Israel* 2004

[KS78] Ulrich Krengel 和 Louis Sucheston。《论半鞅、渐进鞅及有限值过程》。 *Probability on Banach spaces* 4 197 266, 1978 ，：–。[KT84] Antoon WJ Kolen 和 Arie Tamir。《 *Covering problems*》。计量经济学研究所，1984 年。[LLP+21] Allen Liu、Renato Paes Leme、Martin P´al、Jon Schneider 和 Balasubramanian Sivan。《先知不等式与最优排序的变量分解》。载于 P´eter Bir´o、Shuchi Chawla 和 Federico Echenique 编， *EC '21: The 22nd ACM Conference on Economics and Computation Budapest Hungary July 18-23 2021* ，，，，，第 页。692 2021 ACM，。[Lov75] L´aszlo Lovasz。 *Discrete mathematics* 13 4 383 390, 1975 《最优整数覆盖与分数覆盖之比》。《》，（）：。[Mey01] Adam Meyerson。《在线设施选址》。载于 *Proceedings 42nd IEEE Symposium on Foundations of Computer Science* 426 431 2001 ，第 – 页。IEEE，。[PT22] Bo Peng 和 Zhihao Gavin Tang。《顺序选择先知不等式：从阈值优化到到达时间设计》。载于 *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver CO USA October 31 - November 3 2022* 171 178 2022 ，，，，，第 – 页。IEEE，。[RWW20] Aviad Rubinstein、Jack Z. Wang 和 S. Matthew Weinberg。《基于样本的最优单选先知不等式》。载于 Thomas Vidick 编， *11th Innovations in Theoretical Computer Science Conference, ITCS 2020 2020 Seattle USA* 151 *LIPIcs* 60 1 60 10 ， *January 12-14*，，， *Washington*，第 ¨ 卷，页码：–：。施洛斯·达格施图尔信息学研究所，2020 年。[Sed] Saeed Seddighin。个人通信。[SS06a] David B. Shmoys 和 Chaitanya Swamy。《随机线性规划的近似方案及其在随机整数规划中的应用》。《 *J. ACM*》，53(6):978–1012，2006 年。[SS06b] Chaitanya Swamy 和 David B. Shmoys。《两阶段随机优化问题的近似算法》。载于 S. Arun-Kumar 和 Naveen Garg 编， *FSTTCS 2006: Foundations of Software Technology and Theoretical Computer Science, 26th International Conference, Kolkata, India, December 13-15*，2006 4337 5 19 2006， *Proceedings, Lecture Notes in Computer Science* 卷，第 – 页。Springer，。[Vyg05] Jens Vygen。《 *Approximation algorithms facility location problems*》。离散数学研究所，莱茵弗里德里希-威廉大学……，2005 年。[WS11] David P Williamson 和 David B Shmoys。《 *The design of approximation algorithms*》。剑桥大学出版社，2011 年。