arXiv:2304.02240v1 [cs.LG] 5 Apr 2023

L可复现学习中的列表与证书复杂度

Peter Dixon[1], A.Pavan [2], Jason VanderWoude[3], 和 N. V. Vinodchandran[4]

[1]内盖夫本-古里安大学
tooplark@gmail.com[2]爱荷华州立大学*pavan@cs.iastate.edu*[3]内布拉斯加大学林肯分校
jasonvw@huskers.unl.edu[4]内布拉斯加大学林肯分校
vinod@cse.unl.edu

2023年4月6日

摘要

我们研究可复现学习算法。理想情况下，我们希望设计出这样的算法：即使不同次运行从未知数据分布中观察到不同的样本集，也能在多次运行中输出相同的规范模型。一般而言，这种强意义上的可复现性无法实现。因此，我们考虑两种可行的可复现性概念，称为列表可复现性和证书可复现性。直观而言，这些概念刻画了（不可）复现的程度。我们针对某些学习问题设计了在列表复杂度和证书复杂度方面最优的算法，并证明了相应的匹配不可能性结果。

1 引言

科学中的可复制性和可复现性是至关重要的问题。科学结果和实验必须具备可复制性/可复现性这一基本要求，是科学发展与演进的核心。近年来，随着多个科学学科转向数据驱动型研究，这些问题日益受到重视；数据驱动型研究通过数据民主化和经济实惠的计算资源，推动了指数级进步。可复制性问题受到了广泛关注，涉及从大众媒体出版物（例如《经济学人》的“科学如何走向错误”，2013 [eco13]）到科学出版平台（例如，参见 [JP05、Bak16]），再到国家科学、工程与医学科学院（NASEM）等专业和科学机构的众多实体。NASEM发布的一份共识研究报告 [NAS19]深入讨论了可复制性和可复现性所面临的新兴挑战。

确保算法的可复现性/可复制性，一种广泛采用的方法是公开数据集、算法和代码。近年来，各类会议纷纷举办可复制性研讨会，以推广最佳实践并鼓励研究人员共享代码（例如，参见 [PVLS+21] 和[MPK19]）。其背后的一个假设是，使用相同的输入数据、计算方法和代码，就能获得一致的结果。然而，仅靠这种做法并不足以确保可复制性，因为现代方法所采用的计算本身就涉及随机性。

对随机变量进行计算会导致高度的不可复制性，尤其是在机器学习任务中。机器学习算法从某个（有时未知的）分布中观察样本，并输出一个假设或模型。这类算法本质上不可复制。由于算法在两次运行中看到的样本集不同，同一算法的两次不同运行会输出不同的模型。理想情况下，为了实现“完美可复制性”，我们希望设计出这样的算法：即使多次运行时观察到的未知分布样本集各不相同，也能输出相同的典范假设。

我们首先指出，学习中无法实现完美可复制性，因为输出不可避免地依赖于数据样本。我们通过一个简单的硬币偏差估计学习任务来说明这一点：给定来自偏差未知的硬币的 n 次独立投掷，输出对 b 的估计值。我们很容易论证，不存在这样的算法：它以概率 $\geq 2/3$ 输出典范估计 v_b，使得 $|v_b - b| \leq \varepsilon$。考虑一列偏差为 $b_1 < b_2 < \cdots < b_m$ 的硬币，其中每个 $b_{i+1} - b_i \leq \eta$（对于某个较小的 η），但 $b_m - b_1 \geq 2\varepsilon$。对于两个相邻的偏差，$\mathcal{D}^n_{i+1}$ 与 $\mathcal{D}^n_i$ 之间的统计距离（记作 d_{TV}）为 $\leq m\eta$，其中 $\mathcal{D}^n_i$ 是对 i^{th} 硬币进行 n 次独立投掷所得的分布。令 v_{i+1} 和 v_i 分别为算法在偏差为 b_{i+1} 和 b_i 时输出的典范估计。由于 A 在来自分布 $\mathcal{D}^n_i$ 的样本上以至少 $2/3$ 的概率输出 v_i，且 $\mathsf{d}_{\mathrm{TV}}(D^n_i, D^n_{i+1}) \leq n\eta$，因此 $A(\mathcal{D}^n_{i+1})$ 必须以至少 $2/3 - n\eta$ 的概率输出 v_i。我们取 $\eta = 1/10n$。由于 $A(\mathcal{D}^n_i)$ 必须以至少 $2/3$ 的概率输出一个典范值，这意味着 $v_i = v_{i+1}$。因此，对于所有偏差，b_1、.....、b_m、A 都应输出相同的值；但这会导致矛盾，因为 b_1 和 b_m 相距 2ε。然而，很容易看出，存在一种偏差估计算法，只需进行 $n = O(1/\varepsilon^2)$ 次投掷，就能输出两个典范估计值之一：将偏差估计到 $\varepsilon/2$ 以内，再将结果四舍五入到最接近的 ε 的整数倍。我们的工作正是从这两个观察出发。尽管可能无法设计出完美可复制的学习算法，但可以设计出将其“不可复制性”降至最低的算法。

受此启发，我们在机器学习算法的背景下研究了两种称为列表可复制性和证书可复制性的可复制性概念。这些概念量化了（不可）复制性的程度。我们所考虑的概念植根于伪确定性文献 [GG11,Gol19,GL19]，该领域一直活跃于随机算法和计算复杂性理论中。第 2 节讨论了这一联系以及与我们工作相关的其他研究。

1.1 我们的结果

我们考虑可复现学习的两个概念：列表可复现学习与证书可复现学习。在列表可复现性中，学习算法应以高概率输出大小为 $\leq k$的模型列表中的某一个模型。这意味着当该学习算法多次运行时，我们最多看到 k个不同的模型/假设。值 k 被称为该算法的列表复杂度。列表复杂度为1的算法是完全可复制的。因此，列表复杂度可被视为（非）可复制性的程度。在此设定下，目标是设计使列表复杂度 k最小化的学习算法。

在证书可复现性中，学习算法可以访问一个 ℓ位的随机字符串，称为可复现性证书（即该字符串独立于样本以及算法可能使用的其他随机性）。我们要求，对于大多数证书，算法必须输出一个规范模型 h，该模型只能依赖于这个 ℓ位的随机字符串 r。因此，一旦固定证书 r，多次运行能够访问 r 的算法都将以高概率输出相同的模型。我们称 ℓ 为证书复杂度。请注意，证书复杂度为零的算法具有完美的可复制性。因此，ℓ是衡量算法（不可）复制程度的另一项指标。在这一设定下，目标是设计能够最小化 ℓ 的学习算法。

机器学习任务中的一项关键资源是算法所观察到的样本数量，即样本复杂度。高效的学习算法会尽可能少地使用样本。在这项工作中，我们

以样本复杂度、列表复杂度和证书复杂度来衡量学习算法的效率。本工作开启了对在证书/列表复杂度以及样本复杂度方面都高效的学习算法的研究。我们针对某些学习任务建立了正面结果和不可能性结果。

估计 d 枚硬币的偏差。我们的第一组结果关于硬币偏差估计问题的高效可复现算法。我们考虑这样一个问题：通过观察 d 枚硬币各自的 n次抛掷，同时估计它们的偏差，我们称之为 d硬币偏差估计问题。任务是$\vec{}$ $\vec{}$输出一个偏差向量 $\vec{v}$，使得 $\|b-\vec{v}\|_\infty \leq \varepsilon$，其中 b 是真实偏差向量。我们证明，该问题存在一个 $(d+1)$-列表可复制学习算法，其样本复杂度（硬币抛掷次数）为$n = O(\frac{d^2}{\varepsilon^2}\cdot\log\frac{d}{\delta})$ 次/枚硬币。我们还为该问题设计了一个 $\lceil\log\frac{d}{\delta}\rceil$-证书可复现算法，其样本复杂度为 $O(\frac{d^2}{\varepsilon^2\delta^2})$ 次/枚硬币。这里 $(1-\delta)$ 是成功概率。

我们也证明了上述上界在列表复杂度与证书复杂度方面的最优性。我们证明，不存在针对 d-硬币偏差估计问题的 d-列表可复制学习算法。由此可得其证书复杂度的下界为$\Omega(\log d)$。为建立这一不可能性结果，我们使用了KKM/Sperner引理的一个版本。

PAC学习。我们在PAC学习背景下，为列表/证书可复现算法建立了可能性与不可能性结果。我们证明，任何能够使用 d 非自适应统计查询进行学习的概念类，都可以通过一个（$d+1$）列表可复制PAC学习算法进行学习，其样本复杂度为$O(\frac{d^2}{\nu^2}\cdot\log\frac{d}{\delta})$，其中 ν是统计查询参数。我们还证明，这类概念类都存在一个 $\lceil\log\frac{d}{\delta}\rceil$-证书可复现PAC学习算法，其样本复杂度为 $O(\frac{d^2}{\nu^2\delta^2}\cdot\log\frac{d}{\delta})$。最后，我们在PAC学习模型中建立了紧确结果。特别地，我们证明，在均匀分布下， d维阈值的概念类不具有a d-列表可复制学习算法。由于我们可以在均匀分布下使用 d 非自适应统计查询学习 d维阈值，因此可以得到一个在均匀分布下的 $d+1$-列表可复制PAC算法。这为均匀分布下阈值的PAC学习的列表复杂度给出了匹配的上下界。

2 先前工作与相关工作

近年来，可复现性与可复制性的形式化研究发展势头强劲。尽管可复现性与可复制性这两个术语非常接近，且经常被互换使用，但仍有研究努力对二者加以区分；据此，我们的概念属于可复制性定义 [PVLS+21]。

在随机算法的背景下，可复现性/可复制性的各种概念已得到研究。Gat 与 Goldwasser 的工作 [GG11] 形式化并定义了伪确定性算法的概念。若对于任意输入 x，存在一个典范值v_x，使得 $\Pr[A(x) = v_x] \geq 2/3$，则随机算法 A 是伪确定性的。Gat 与 Goldwasser 为代数计算问题设计了多项式时间伪确定性算法，例如寻找二次非剩余以及寻找多元多项式的非根 [GG11]。后续工作在其他算法场景中研究了伪确定性概念，例如并行计算、流式与次线性算法、交互式证明，以及它与复杂性理论的联系 [GG, GGH18, OS17, OS18, AV20, GGMW20, LOS21, DPVWV22]。

在算法领域，主要研究了伪确定性的两种一般化形式：多重伪确定性算法 [Gol19] 和影响位算法 [GL19]。如果对于每个输入 x，都存在一个大小至多为 k 的集合 S_x，使得输出

$A(x)$ 以高概率属于集合 S_x，那么随机算法 A 就是 ℓ-伪确定性的。当 $k = 1$时，我们便得到伪确定性。若对于每个输入 x，对于大多数长度为 ℓ 的字符串 r，都存在一个规范值 $v_{x,\,r}$，使得算法 A在输入 x 和 r 时，以高概率输出 $v_{x,\,r}$，那么该随机算法 A 就是 ℓ-影响位算法。这些字符串 r 称为影响位字符串。同样，当 $\ell = 0$时，我们便回到伪确定性。这些研究的重点主要是探讨随机搜索算法的可复现性。

最近，研究者开始在学习算法的背景下探索伪确定性及其推广形式，以形式化可复制性的概念。特别是，Impagliazzo、Lei、Pitassi 和 Sorrell [ILPS22] 提出了 ρ-可复制性的概念。若 $\Pr[A(S_1, r) = A(S_2, r)] \geq 1 - \rho$，则称学习算法 A ρ-可复制，其中 S_1 和 S_2 是从分布 $\mathcal{D}$ 中抽取的样本，而 r 是学习算法 A的内部随机性。他们为许多学习任务设计了可复现算法，包括统计查询、近似重频项、求中位数和学习半空间。

近期另一条研究脉络将可复现学习与差分隐私学习联系起来。在一项近期的开创性工作中，Bun、Livny 和 Moran [BLM20] 证明，每个 Littlestone 维数有限的概念类都可以由近似差分隐私算法学习。结合 Alon、Livny、Malliaris 和 Moran [ALMM19], 更早的工作，这确立了在线可学习性与差分隐私 PAC 可学习性之间的等价关系。颇为出人意料的是， [BLM20] 的证明采用了“全局稳定性”这一概念，它与学习背景下的伪确定性概念相似。他们将学习算法 A 定义为关于分布 D 的 (n, η)- 全局稳定算法，条件是存在一个假设 h，使得 $\Pr_{S\sim D^n}(A(S) = h) \geq \eta$。他们证明，任意 Littlestone 维数为 d的概念类都有一个(n, η)-全局稳定学习算法，其中 $m = \tilde{O}(2^{2^d}/\alpha)$且 $\eta = \tilde{O}(2^{-2^d})$，而 h （相对于未知假设）的误差为 $\leq \alpha$。随后，他们证明全局稳定学习器意味着存在差分隐私学习器。在引言中讨论的全局稳定学习概念，就是我们在 $\eta = 2/3$时所说的完美可复制性。因此，正如引言中所讨论的，由此可知，即使对于估计硬币偏差这样简单的任务，也不可能设计出具有 $\eta > 1/2$ 的全局稳定算法。Ghazi、Kumar 和 Manurangsi [GKM21] 将全局稳定性推广为伪全局稳定性和列表全局稳定性。伪全局稳定性与 Grossman 和 Liu [GL19] 提出的影响位算法概念非常相似，只需将其转换到学习背景下即可。类似地，列表全局稳定性与 Goldreich 的多重伪确定性概念相似 [Gol19]。还有已知结果表明，伪全局稳定性与 ρ-可复制性这两个概念在参数方面仅相差多项式因子 [ILPS22, GKM21]。 [GKM21] 的工作利用这些概念设计用户级隐私学习算法。

我们的列表可复制性概念受到多重伪确定性概念的启发，而证书可复现性概念受到影响位算法概念的启发。在学习场景中，我们的列表可复制性概念类似于列表全局稳定性概念，而证书可复现性概念类似于伪全局稳定性概念，后者又类似于 ρ-可复制性概念（见 [ILPS22]）。

我们引入了用于衡量（非）可复制性程度的列表复杂度与证书复杂度概念。我们的目标是在最小化样本复杂度的同时，设计具有最优列表复杂度与证书复杂度的学习算法。早期工作并未专注于最小化这些量。 [BLM20、GKM21] 的工作将可复现算法作为设计差分隐私算法的中间步骤。 [ILPS22] 的工作没有考虑在其算法中降低证书复杂度，也没有研究列表可复制性

在各个计算领域中研究不同的可复现性/可复制性概念，是一个新兴课题。在 [EKK+23] 中，作者探讨了随机老虎机背景下的可复制性。他们的概念

与 [ILPS22] 中研究的概念类似。在 [AJJ+22]中，作者探讨了在使用非精确预言机（初始化/梯度预言机）进行优化的背景下的可复现性。这些工作的设置和重点与我们的不同。

3 主要引理

在本节中，我们陈述几个建立在 [VWDP+22] 和 [DLPES02] 工作基础上的引理，这些引理将对本文其余部分中的算法构造和不可能性结果有所帮助。

3.1 划分和舍入

在本小节中，我们定义一个将反复使用的确定性舍入函数。该舍入函数基于 [VWDP+22] 的工作中定义的隔离划分概念。

我们将使用以下记号。我们用diam_∞ 表示集合相对于 ℓ_∞范数的直径，并用 $\overline{B}_\epsilon^\infty(\vec{p})$表示以 $\vec{p}$ 为中心、相对于 ℓ_∞ 范数且半径为 ϵ 的闭球。也就是说，在 $\mathbb{R}^d$ 中，我们有 $\overline{B}_\epsilon^\infty(\vec{p}) = \prod_{i=1}^d [p_i - \epsilon, p_i + \epsilon]$。

令 $\mathcal{P}$ 为 $\mathbb{R}^d$的一个划分。对于点 $\vec{p} \in \mathbb{R}^d$，我们用 $N_\epsilon(\vec{p})$ 表示该集合

$$\{X \in \mathbb{P} \mid B_\epsilon(\vec{p}) \cap X \neq \emptyset\}$$

定义3.1. 令 $\mathcal{P}$ 是 $\mathbb{R}^d$的一个划分。我们称 $\mathcal{P}$ 是(k, ϵ)-隔离的，如果对于每个点 $\vec{p} \in \mathbb{R}^d$，$|N_{\epsilon(\vec{p})}| \leq k$。

以下定理引自 [VWDP+22]。

定理 3.2. 对于每个 $d \in \mathbb{N}$，存在一个 $(d+1, \frac{1}{2d})$-隐蔽划分，其中该划分的每个成员都是一个单位超立方体。此外，该划分可被高效计算：给定任意点$\vec{x} \in \mathbb{R}^d$，可在关于d的多项式时间内计算出 $\vec{x}$ 所属划分成员的描述。

定义3.3（（k（d） ϵ（d））-确定性舍入）。确定性舍入方案是一族函数 $\mathcal{F} = \{f_d\}_{d\in\mathbb{N}}$，其中 f_d：$\mathbb{R}^d \to \mathbb{R}^d$。我们称 $\mathcal{F}$ 为（k（d） ε（d））-确定性舍入方案，如果(1)$\forall \vec{x} \in \mathbb{R}^d, d_{max}$（$\vec{x}, f_d$（$\vec{x}$））$\leq 1$[2] (2) $\forall \vec{x} \in \mathbb{R}^d, |\{f_d$（$\vec{z}$）：$\vec{z} \in B_{\varepsilon(d)}$（$\vec{x}$）$\}| \leq k$（$d$）。

以下观察出自 [VWDP+22]。

观察3.4（舍入方案与划分的等价性）。一个 ($k(d)$ $\epsilon(d)$)-确定性舍入方案会针对每个 $d \in \mathbb{N}$诱导出一个($k(d)$ $\epsilon(d)$)-隔离划分，将 $\mathbb{R}^d$ 划分为其中每个成员的直径至多为2。反过来，一个划分序列 $\langle \mathcal{P}_d \rangle_{d=1}^\infty$，其中 $\mathcal{P}_d$ 是 ($k(d)$ $\epsilon(d)$)-隔离的，并且仅包含直径至多为 1的成员，会诱导出一个($k(d)$ $\epsilon(d)$)-确定性舍入方案。

3.2 用于列表可复现性的通用舍入算法

在本小节中，我们将设计一个确定性算法，该算法将用作我们的列表可复现算法中的子程序。

[1]关于可复制性和可复制性的讨论，请参见 [PVLS+21]。[2]界限 1 并非关键。我们可以使用任意常数，并相应地缩放参数。

引理 **3.5** $d \in \mathbb{N}$ $\epsilon \in (0, \infty)$ $\epsilon_0 = \frac{\epsilon}{2d}$ $f_\epsilon : \mathbb{R}^d \to \mathbb{R}^d$。设 $\hat{x}(\cdot)$。设 $\hat{x}$。存在一个可高效计算的函数 f_ϵ，具有以下两个性质：

1. 对于任意 $x \in \mathbb{R}^d$ 和任意 $\hat{x} \in \overline{B}^\infty_{\epsilon_0}(x)$，都有 $f_\epsilon(\hat{x}) \in \overline{B}^\infty_\epsilon(x)$。
2. 对于任意 $x \in \mathbb{R}^d$ 该集合 $\left\{f_\epsilon(\hat{x}) : \hat{x} \in \overline{B}^\infty_{\epsilon_0}(x)\right\}$ 的基数至多为 $d+1$。

非正式地说，这两个条件是：(1) 若 $\hat{x}$ 是 x 的一个 ϵ_0 近似，则 $f_\epsilon(\hat{x})$ 是 x 的一个 ϵ 近似；(2) f_ϵ 将 x 的每一个 ϵ_0 近似映射到至多 $d+1$ 个可能值之一。

证明。令 $\mathcal{P}$ 为定理3.2 中的一个$(d+1, \frac{1}{2d})$-secluded划分，并令 $f : \mathbb{R}^d \to \mathbb{R}^d$ 根据观察 3.4 得到的相关确定性舍入函数。划分 $\mathcal{P}$ 由单位立方体 $[0,1)^d$ 的平移副本组成，并具有如下性质：对于任意点 $\vec{p} \in \mathbb{R}^d$，以 $\vec{p}$ 为中心、边长为 $1/d$ 的闭立方体（即 $\overline{B}^\infty_{1/2d}(\vec{p})$）至多与 $\mathcal{P}$中的 $d+1$ 个成员/立方体相交。相关舍入函数 $f : \mathbb{R}^d \to \mathbb{R}^d$将 $\mathbb{R}^d$ 中的每个点映射到 $\mathcal{P}$ 中包含该点的唯一立方体的中心点。这意味着 f 具有以下两个性质（它们与我们希望 f_ϵ具备的两个性质密切相关）：(1) 对于每个$a \in \mathbb{R}^d$，$\|f(a) - a\|_\infty \le \frac{1}{2}$（因为每个点都被映射到其所在单位立方体的中心）；以及(2) 对于任意点 $p \in \mathbb{R}^d$，集合 $\left\{f(a) : a \in \overline{B}^\infty_{1/2d}(p)\right\}$的基数至多为 $d+1$（因为 $\overline{B}^\infty_{1/2d}(p)$ 至多与 $\mathcal{P}$中的 $d+1$ 个成员相交）。定义 $f_\epsilon : \mathbb{R}^d \to \mathbb{R}^d$为 $f_\epsilon(a) = \epsilon \cdot f(\frac{1}{\epsilon}a)$。$f_\epsilon$ 的高效可计算性源于 f 的高效可计算性（即能够高效计算 $\mathcal{P}$ 中包含给定点的单位立方体的中心）。

要看出 f_ϵ 具有性质 (1)，令 $x \in \mathbb{R}^d$ 和 $\hat{x} \in \overline{B}^\infty_{\epsilon_0}(x)$。那么我们有以下结论（证明随后给出）：

$$\begin{aligned}
\left\|\tfrac{1}{\epsilon} \cdot f_\epsilon(\hat{x}) - \tfrac{1}{\epsilon}x\right\|_\infty &= \left\|f(\tfrac{1}{\epsilon}\hat{x}) - \tfrac{1}{\epsilon}x\right\|_\infty \\
&\le \left\|f(\tfrac{1}{\epsilon}\hat{x}) - \tfrac{1}{\epsilon}\hat{x}\right\|_\infty + \left\|\tfrac{1}{\epsilon}\hat{x} - \tfrac{1}{\epsilon}x\right\|_\infty \\
&\le \left\|f(\tfrac{1}{\epsilon}\hat{x}) - \tfrac{1}{\epsilon}\hat{x}\right\|_\infty + \tfrac{1}{\epsilon}\|\hat{x} - x\|_\infty \\
&\le \tfrac{1}{2} + \tfrac{1}{\epsilon}\epsilon_0 \\
&= \tfrac{1}{2} + \tfrac{1}{2d} \le 1
\end{aligned}$$

第一行根据 f_ϵ 的定义，第二行是三角不等式，第三行是范数的缩放，第四行利用了 f 的性质，即点映射后的距离不超过 $\frac{1}{2}$，并结合假设 $\hat{x} \in \overline{B}^\infty_{\epsilon_0}(x)$，第五行利用了 ϵ_0 的定义，第六行利用了 $d \ge 1$ 这一事实。

将两边按 ϵ 缩放并利用范数的缩放性质，上述内容给出 $\|f_\epsilon(\hat{x}) - x\|_\infty \le \epsilon$，这证明了引理的性质 (1)。

要看出 f_ϵ 具有性质（2），令 $x \in \mathbb{R}^d$。我们有以下等式组：

$$\begin{aligned}
\left\{f_\epsilon(\hat{x}) : \hat{x} \in \overline{B}^\infty_{\epsilon_0}(x)\right\} &= \left\{\epsilon \cdot f(\tfrac{1}{\epsilon}\hat{x}) : \hat{x} \in \overline{B}^\infty_{\epsilon_0}(x)\right\} \\
&= \left\{\epsilon \cdot f(a) : a \in \overline{B}^\infty_{\frac{1}{\epsilon}\epsilon_0}(x)\right\} \\
&= \left\{\epsilon \cdot f(a) : a \in \overline{B}^\infty_{\frac{1}{2d}}(x)\right\}
\end{aligned}$$

第一行来自 f_ϵ 的定义，第二行来自重新缩放，第三行来自 ϵ_0 的定义。

由于 f 在$\overline{B}^\infty_{\frac{1}{2d}}(x)$上至多取 $d+1$ 个不同的值，该集合的基数至多为 $d+1$，这证明了引理的性质（2）。 □

3.3 证书可复现性的通用舍入算法

为设计证书可复现学习算法，我们将使用下述引理中陈述的一种通用随机化过程。该引理采用一种随机舍入方案。类似的随机舍入方案已在若干先前工作 [SZ99、DPV18、Gol19、GL19、ILPS22]中使用过。

引理 3.6. 设 $d \in \mathbb{N}$, $\epsilon_0 \in (0,\infty)$ 和 $0 < \delta < 1$。存在一个可高效计算的确定性函数 f: $\{0,1\}^\ell \times \mathbb{R}^d \to \mathbb{R}^d$，满足以下性质。对于任意 $x \in \mathbb{R}^d$,

$$\Pr_{r\in\{0,1\}^\ell}\left[\exists x^* \in \overline{B}_\varepsilon^\infty(x)\ \ \forall \hat{x} \in \overline{B}_{\varepsilon_0}^\infty(x) : f(r,\hat{x}) = x^*\right] \geq 1-\delta$$

其中 $\ell = \lceil \log \frac{d}{\delta} \rceil$ 且 $\varepsilon = (2^\ell+1)\epsilon_0 \leq \frac{2\varepsilon_0 d}{\delta}$。

证明. 将 $\mathbb{R}^d$ 的每个坐标划分为宽度为 $2\varepsilon_0$ 的区间。计算函数 f 的算法执行以下简单的随机舍入：

函数 f：选择一个随机整数 $r \in \{1\dots 2^\ell\}$。注意，r 可用 ℓ 位表示。考虑 $\hat{x}$ 的 i^{th} 个坐标中的第一个，记为 $\hat{x}[i]$。将 $\hat{x}[i]$ 舍入到最近的 $k*(2\varepsilon_0)$，使得 $k \bmod 2^\ell \equiv r$。

现在我们将证明 f 满足所需的性质。

首先，我们证明近似保证。令 x' 表示将 $\hat{x}$ 的每个坐标舍入后在 $\mathbb{R}^d$ 中得到的点。满足 $k \bmod 2^\ell \equiv r$ 的 k 相隔 $2^\ell \cdot 2\varepsilon_0$。因此，$x'[i]$ 的舍入误差至多为 $2^\ell\epsilon_0$。也就是说，对于每个 i，$1 \leq i \leq d$，都有 $|x'[i] - \hat{x}[i]| \leq 2^\ell\epsilon_0 = \frac{\varepsilon_0 d}{\delta}$。由于 $\hat{x}$ 是一个 ε_0-近似（即每个坐标 $\hat{x}[i]$ 与真实值 $x[i]$ 的差距不超过 ε_0），那么 x' 的每个坐标与 $x[i]$ 的差距不超过 $(2^\ell+1)\varepsilon_0$。因此，x' 是 $x[i]$ 的 $(2^\ell+1)\varepsilon_0$-近似。因此，对于任意 r 的选择，$x' \in \overline{B}_\varepsilon^\infty(x)$。

现在我们证明，对于 $r \in \{1\dots 2^\ell\}$ 的 $\geq 1-\delta$ 比例，存在 x^*，使得每个 $\hat{x} \in \overline{B}_{\varepsilon_0}^\infty(x)$ 都会被舍入到 x^*。我们针对每个坐标进行论证，并应用并集界。固定一个 x 和一个坐标 i。对于 $x[i]$，考虑其周围的 ε_0 区间。

考虑来自 $\{1\dots 2^\ell\}$ 的 r。选定这个 r 后，我们将 $\hat{x}[i]$ 舍入到最近的 $k*(2\varepsilon_0)$，使得 $k \bmod 2^\ell \equiv r$。令 p_1^r、p_2^r、$\dots p_j^r \dots$ 为这类点的集合：更准确地说，$p_j = (j2^l + r) * 2\varepsilon_0$。注意，$\hat{x}[i]$ 会被舍入到某个 j 的一个 p_j。令 m_j^r 表示 p_j^r 和 p_{j+1}^r 之间的中点。也就是说，$m^r = (p_j^r + p_{j+1}^r)/2$。如果 $x[i]$ 接近某个 m_j^r，我们就称 r 对于 $x[i]$ 是“坏的”。也就是说，若 $|x[i] - m_j^r| < \varepsilon_0$，则称 r 是“坏的”。注意，对于一个坏的 r，存在 $\hat{x}_1$ 和 $\hat{x}_2$，属于 $\overline{B}_{\varepsilon_0}^\infty(x)$，使得它们的 i^{th} 坐标分别舍入到 p_j^r 和 p_{j+1}^r。关键在于，如果 r 对于 $x[i]$“不坏”，那么对于每个 $x' \in \overline{B}_{\varepsilon_0}^\infty(x)$，都存在一个规范的 p^*，使得 $x'[i]$ 被舍入到 p^*。如果 r 对于 x 是坏的，我们就称 r 对于 x 是坏的；这意味着至少存在一个 i、$1 \leq i \leq d$，使得 r 对于 $x[i]$ 是坏的。由此可知，如果 r 对于 x 不坏，那么就存在一个规范的 x^*，使得每个 $x' \in \overline{B}_{\varepsilon_0}^\infty(x)$ 都会被舍入到 x^*。

由此，目标是界定随机选取的 r 对 x而言是坏值的概率。为此，我们首先界定 r 对 $x[i]$而言是坏值的概率。我们将证明，几乎至多存在一个对 $x[i]$而言是坏值的 r。假设存在两个数 r_1和 r_2，它们对 $x[i]$而言都是坏值。这意味着，对于某些 j_1和 j_2，$|x[i] - m_{j_1}^{r_1}| < \varepsilon_0$且$|x[i] - m_{j_2}^{r_2}| < \varepsilon_0$。因此，根据三角不等式，$|m_{j_1}^{r_1} - m_{j_2}^{r_2}| < 2\varepsilon_0$。然而，注意到$|p_{j_1}^{r_1} - p_{j_2}^{r_2}|$ 是 $|(j_1-j_2)2^\ell + (r_1-r_2)|2\epsilon_0$。由于 $r_1 \neq r_2$，这个值至少为 $2\varepsilon_0$。这意味着 $m_{j_1}^{r_1}$与 $m_{j_2}^{r_2}$之差的绝对值至少为 2ε，从而导致矛盾。

因此，r 对 $x[i]$ 不利的概率至多为 $\frac{1}{2^\ell}$，并且由并集界，r 对 x不利的概率至多为$\frac{d}{2^\ell} \leq \delta$。这就完成了证明。 □

3.4 Sperner引理/KKM引理的推论

以下结果是Sperner引理/KKM引理的一些立方变体的推论，这些变体最初在 [DLPES02] 中发展，并在 [VWDP+22]中进一步扩展。该结果的陈述和证明与 [VWDP+22] 中的定理9.4（第二最优性定理）非常相似，只是这里针对 $[0,1]^d$而不是 $\mathbb{R}^d$进行陈述。

引理 3.7。令 $\mathcal{P}$ 是 $[0,1]^d$ 的一个划分，使得对于每个成员 $X \in \mathcal{P}$，都有$\mathrm{diam}_\infty(X) < 1$。那么存在 $\vec{p} \in [0,1]^d$，使得对于所有 $\delta > 0$，$\overline{B}_\delta^\infty(\vec{p})$ 都至少与 $\mathcal{P}$ 的 $d+1$ 个成员相交。

4 学习硬币偏差的可复制性

在本节中，我们建立了估计 d 硬币偏差的可复制性结果。

定义4.1。 d-硬币偏差估计问题是以下问题：设计一个算法A（可能是随机化的），给定 $\epsilon \in (0,\infty)$、$\delta \in (0,1]$,以及 n对一个具有偏差向量 $\vec{b} \in [0,1]^d$的有序集合中的 d枚有偏硬币进行的独立投掷（每枚硬币），输出 $\vec{v}$，使得 $\|\vec{b}-\vec{v}\|_\infty \le \varepsilon$的概率为 $\ge 1-\delta$。

定义4.2. 我们称算法 A 对于 d-硬币偏差估计问题是 k-列表可复现的，如果对于任意偏差向量 $\vec{b} \in [0,1]^d$ 以及参数 ε、δ，存在集合 $L \subseteq \overline{B}_\varepsilon^\infty(\vec{b})$和整数 n，使得$|L| \le k$和 A在输入 ε、δ以及根据偏差向量 $\vec{b}$进行的 n 次独立抛掷（每枚硬币）时，以概率 $\ge 1-\delta$输出估计值 $\vec{v} \in L$。 n 是 A 的样本复杂度，而 k 是A 的列表复杂度。

定义 4.3。我们称用于 d-硬币偏差估计问题的算法 A 是 ℓ-证书可复现的，如果对于任意偏差向量 $\vec{b} \in [0,1]^d$以及参数 ε、δ：在输入 ϵ、δ、$r \in \{0,1\}^\ell$上，以及根据 b 进行的 n 次独立硬币 $\vec{}$ 抛掷（每枚硬币），满足以下条件：

$$\Pr_{r\in\{0,1\}^\ell}\left[\exists \vec{v_r} \in \overline{B}_\varepsilon^\infty(\vec{b}) : \Pr[A \text{ outputs } \vec{v_r}] \ge 1-\delta\right] \ge 1-\delta$$

在上文中，内部概率是对 A的内部随机性以及硬币抛掷取的。算法 A 还将 r 作为一项输入（除其他输入之外）。我们将 n 称为样本复杂度，并将随机比特数 ℓ 称为 A的证书复杂度。

我们注意到，从粗略意义上说，我们可以将列表可复现算法转换为证书可复现算法，反之亦然。然而，此类转换会导致样本复杂度劣化，而这正是本文所关注的问题。

在下文中，用于 d-硬币偏差估计问题的算法的输出记为$\mathcal{D}_{A,\vec{b},n}$。

4.1 可复现算法

我们给出了 d-硬币偏差估计问题的两种算法。第一种是 $d+1$-列表可复现的，第二种是 $\lceil \log \frac{d}{\delta} \rceil$-证书可复现的。

定理 4.4. 存在一种 $(d+1)$-列表可复现算法，可解决 d-Coin Bias Estimation Problem，其样本复杂度为 $n = O(\frac{d^2}{\varepsilon^2} \cdot \log \frac{d}{\delta})$（每枚硬币）。

算法 1 $(d+1)$　　如定理 4.4 所示，用于 d-硬币偏差估计问题的 -list 可复现算法

$\epsilon > 0\delta \in (0,1]$ 对具有偏差的 d 硬币进行采样访问

$\vec{b} \in [0,1]^d$

输出b 并以算法表现为一个属于$d[0,1]^d$的确定性的 (ϵ, δ)-近似的 $\vec{\ }$

算法：$\epsilon_0 \stackrel{\text{def}}{=} \frac{\epsilon}{2d}\ \delta_0 \stackrel{\text{def}}{=} \frac{\delta}{d}\ n \stackrel{\text{def}}{=} O\left(\frac{\ln(1/\delta_0)}{\epsilon_0^2}\right) = O\left(\frac{d^2 \ln(d/\delta)}{\epsilon^2}\right)$ 对某个常数令 f_ϵ：$\mathbb{R}^d \to \mathbb{R}^d$ 如定理3.5 中所述。令 g： $\mathbb{R}^d \to [0,1]^d$ 为将坐标限制在单位区间内的函数（即

$$g(\vec{y}) \stackrel{\text{def}}{=} \left\langle \begin{cases} 0 & y_i < 0 \\ y_i & y_i \in [0,1] \\ 1 & y_i > 1 \end{cases} \right\rangle_{i=1}^{d}$$

）从每枚硬币中抽取 n 个样本，并令 $\vec{a}$ 为经验偏差。返回 $g(f(\vec{a}))$

证明。注意，当 $\varepsilon \geq 1/2$时，一个在每个分量中输出带有 $1/2$ 的向量的平凡算法即可奏效。因此，最有趣的情况是 $\varepsilon < 1/2$。我们的列表可复现算法如算法 1 所述。

因此，我们将通过展示每种可能的偏差都出现在一个具有有限个可能性质的集合中来证明$L_{\vec{b}}$ 。

因此，$[0,1]^d$ 我们将通过展示每种可能的偏差都出现在一个具有有限个可能性质的集合中来证明$L_{\vec{b}}| \leq d+1$，（2）$L_{\vec{b}} \subseteq \overline{B}_\epsilon^\infty(\vec{b})$（以及特定于问题的限制，即$L_{\vec{b}} \subseteq [0,1]^d$），（3）当可以使用偏差为 $\vec{b}$ 的硬币时，该算法以至少 $1-\delta$ 的概率返回一个属于 $L_{\vec{b}}$ 的值。

沿用算法1 的记号。设 $L_{\vec{b}} = \left\{ g(f_\epsilon(\vec{x})) \colon \vec{x} \in \overline{B}_{\epsilon_0}^\infty(\vec{b}) \right\}$。根据定理 3.5， f_ϵ 在 $\overline{B}_{\epsilon_0}^\infty(\vec{b})$上至多取 $d+1$ 个值（这意味着 $g \circ f_\epsilon$ 在此球上也至多取 $d+1$ 个值），由此证明了 $|L_{\vec{b}}| \leq d+1$。这证明了性质 (1)。

接下来我们陈述以下观察结果，它表明算法 1 的坐标限制函数 g 不会降低近似质量。证明相对直接，但较为繁琐。

观察 4.5。使用算法 1 的记号，若 $\vec{y} \in \overline{B}_\epsilon^\infty(\vec{b})$，则 $g(\vec{y}) \in \overline{B}_\epsilon^\infty(\vec{b})$。

证明。设 $\vec{z} = g(\vec{y})$。我们必须证明对每个 $i \in [d]$，都有 $z_i \in [b_i - \epsilon, b_i + \epsilon]$。注意到

$$z_i = \begin{cases} 0 & y_i < 0 \\ y_i & y_i \in [0,1] \\ 1 & y_i > 1 \end{cases}$$

因此，我们分三种情况进行。

情况 1： $y_i < 0$。在这种情况下， $z_i = 0$ ，因此因为 $b_i \in [0,1]$ ，我们有 $z_i \leq b_i + \epsilon$。此外，因为 $y_i \in [b_i - \epsilon, b_i + \epsilon]$，我们有 $b_i \leq y_i + \epsilon < 0 + \epsilon = z_i + \epsilon$，所以从等式两边减去 ϵ 得到 $z_i > b_i - \epsilon$。因此，我们如预期得到 $z_i \in [b_i - \epsilon, b_i + \epsilon]$ 。

情况 2： $y_i > 1$。这种情况与情况 1 对称。

情形 3： $y_i \in [0,1]$。在此情形下， $z_i = y_i \in [b_i - \epsilon, b_i + \epsilon]$ ，因此我们完成了。 □

现在我们建立性质（2）。根据定理 3.5，我们知道对于每个 $\vec{x} \in \overline{B}^{\infty}_{\epsilon_0}(\vec{b})$，都有 $f_\epsilon(\vec{x}) \in \overline{B}^{\infty}_{\epsilon}(\vec{b})$；根据定理 4.5， g 保持了这一质量，并且我们有 $g(f_\epsilon(\vec{x})) \in \overline{B}^{\infty}_{\epsilon}(\vec{b})$。这表明 $L_{\vec{b}} \subseteq \overline{B}^{\infty}_{\epsilon}(\vec{b})$，从而证明了性质（2）。

根据 Chernoff 界，对于单个有偏硬币， $n = O\left(\frac{\ln(1/\delta_0)}{\epsilon_0^2}\right)$ 个该硬币的独立样本足以使得以至少 $1 - \delta_0$ 的概率，经验偏差落在真实偏差的 ϵ_0 范围内。因此，根据并集界，若我们对 d 枚硬币中的每一枚各取 n 个样本，则至少有一个硬币的经验偏差不在真实偏差的 ϵ_0 范围内的概率至多为 $d \cdot \delta_0 = \delta$ 。因此，通过对每枚硬币取 n 个样本，我们以至少 $1 - \delta$ 的概率使得经验偏差 $\vec{a}$ 属于 $\overline{B}^{\infty}_{\epsilon_0}(\vec{b})$。在这种情况发生时，根据 $L_{\vec{b}}$ 的定义，算法返回的值 $g(f_\epsilon(\vec{a}))$ 属于集合 $L_{\vec{b}}$。这就证明了性质（3）。 □

定理 4.6. 存在一个 $\lceil \log \frac{d}{\delta} \rceil$-证书可复现算法，用于 d-硬币 偏差 估计 问题，其每个硬币的样本复杂度为 $n = O(\frac{d^2}{\varepsilon^2 \delta^2})$。

证明。令 ε 和 δ 为算法的输入参数，令 $\vec{b}$ 为偏差向量。设 $\varepsilon_0 = \frac{\varepsilon\delta}{2d}$。算法 A 首先使用标准估计算法估计每枚硬币的偏差，误差上限为 ε_0 ，概率误差参数为 $\frac{\delta}{d}$ 。注意到，这可以通过对每枚硬币进行 $O(\frac{d^2}{\varepsilon^2\delta^2})$ 次投掷来完成。令 $\vec{v}$ 为输出向量。由此可知，以至少 $1 - \delta$ 的概率， $\vec{v} \in \overline{B}^{\infty}_{\varepsilon_0}(\vec{b})$。然后，它运行引理 3.6 中描述的确定性函数 f，输入为 $r \in \{0,1\}^\ell$，其中包含 $\ell = \lceil \log \frac{d}{\delta} \rceil$ 和 $\vec{v}$，并输出该函数的值。引理 3.6 保证，对于 r 的 $1 - \delta$ 比例，所有 $\vec{v} \in \overline{B}^{\infty}_{\varepsilon_0}(\vec{b})$都会被$f$舍入为相同的值。因此，算法 A 满足证书可复制性的要求。证书复杂度为$\lceil \log \frac{d}{\delta} \rceil$。

□

请注意， ℓ-证书可复现会导出一个 2^ℓ-列表可复现算法。因此，定理 4.6 给出了一个用于 d-硬币偏差估计问题的 $O(\frac{d}{\delta})$-列表可复现算法，其样本复杂度为 $O(\frac{d^2}{\varepsilon^2\delta^2})$。然而，这显然是次优的，而定理 4.4 给出的算法具有小得多的列表复杂度和样本复杂度。

4.2 一个不可能性结果

定理 4.7. 对于 $k < d + 1$，不存在针对 d-Coin BiasEstimation Problem 的 k-列表可复制算法。

在证明该定理之前，我们需要一个引理，其证明见附录。
引理 4.8。对于偏差 $\vec{a}$、$\vec{b} \in [0,1]^d$，我们有 $\mathrm{d}_{\mathrm{TV}}\left(\mathcal{D}_{A,\vec{a},n}, \mathcal{D}_{A,\vec{b},n}\right) \leq n \cdot d \cdot \|\vec{b} - \vec{a}\|_\infty$。

证明。我们可以将该模型视为算法 A ，其可以访问来自一个分布的一次抽样。对于一组偏差为 b 的硬币，给出每枚硬币一次翻转样本的分布是伯努利分布 $\prod_{i=1}^{d}\mathrm{Bern}(b_i)$ 的 d 重乘积（为简洁起见，我们将其记作$\mathrm{Bern}(\vec{b})$，因此给出每枚硬币 n 次独立翻转的分布是该分布的 n 重乘积（使用 [可15] 写为 $\mathrm{Bern}(\vec{b})^{\otimes n}$）。

比较偏差为 b 与偏差为 $\vec{a}$时， d 枚硬币 n 次独立抛掷的分布，对于每个 $i \in [d]$，我们有

$$\mathrm{d_{TV}}\left(\mathrm{Bern}(b_i), \mathrm{Bern}(a_i)\right) = |b_i - a_i|$$

因此，根据 [Can15] 的C.1.2和C.1.3，我们有

$$\mathrm{d_{TV}}\left(\mathrm{Bern}(\vec{b}), \mathrm{Bern}(\vec{a})\right) \leq \sum_{i=1}^{d} |b_i - a_i| \leq d \cdot \|\vec{b} - \vec{a}\|_\infty$$

and

$$\mathrm{d_{TV}}\left(\mathrm{Bern}(\vec{b})^{\otimes n}, \mathrm{Bern}(\vec{a})^{\otimes n}\right) \leq n \cdot d \cdot \|\vec{b} - \vec{a}\|_\infty.$$

因为 A 是该分布一次抽取的随机函数，根据 [Can15] 的D.1.2，我们有A不能用于增大全变差距离，因此

$$\mathrm{d_{TV}}\left(\mathcal{D}_{A,\vec{a},n}, \mathcal{D}_{A,\vec{b},n}\right) \leq \mathrm{d_{TV}}\left(\mathrm{Bern}(\vec{b})^{\otimes n}, \mathrm{Bern}(\vec{a})^{\otimes n}\right)$$

至此完成证明。 □

定理 4.7 的证明。固定任意 $d \in \mathbb{N}$，并选取 ϵ 和 δ 分别为$\epsilon < \frac{1}{2}$和 $\delta \leq \frac{1}{d+2}$。

反设这样的算法 A 对某个 $k < d+1$确实存在。这意味着，对于每个可能的阈值 $\vec{t} \in [0,1]^d$，存在某个假设集合 $L_{\vec{t}} \subseteq \mathcal{H}$ ，具有三个性质：(1) $L_{\vec{t}}$的每个元素都是对 $h_{\vec{t}}$的 ϵ-近似；(2) $|L_{\vec{t}}| \leq k$；(3) 以至少 $1-\delta$的概率， A返回 $L_{\vec{t}}$中的一个元素。

反设对于某个 $k < d+1$，这样的算法确实存在。这意味着，对于每个可能的偏差 $\vec{b} \in [0,1]^d$，存在某个集合 $L_{\vec{b}} \subseteq \overline{B}_\epsilon^\infty(\vec{b})$（不一定唯一，但考虑其中一个固定的集合），并且 $|L_{\vec{b}}| \leq k$使得以至少至少 $\frac{1}{k} \cdot (1-\delta)$ 的概率 $\geq \frac{1}{k} \cdot (1 - \frac{1}{d+2}) = \frac{1}{k} \cdot \frac{d+1}{d+2} \geq \frac{1}{k} \cdot \frac{k+1}{k+2}$， A返回 $L_{\vec{b}}$ 中的一个元素。根据简单的平均论证，这意味着 $L_{\vec{b}}$ 中至少存在一个元素，该元素被 A 返回的概率至少为$\frac{1}{k} \cdot \frac{k+1}{k+2}$。令 f: $[0,1]^d \to [0,1]^d$为一个函数，该函数将每个偏差 b 映射到 $L_{\vec{b}}$ 中的这样一个元素。

由于 $\frac{1}{k} \cdot \frac{k+1}{k+2} > \frac{1}{k+1}$，令 η 使得 $0 < \eta < \frac{1}{k} \cdot \frac{k+1}{k+2} - \frac{1}{k+1}$。

函数 f 诱导出对 $[0,1]^d$ 的划分 $\mathcal{P}$，其中 $\mathcal{P}$ 的成员是 f 的纤维（即 $\mathcal{P} = \{f^{-1}(\vec{y}) : \vec{y} \in \mathrm{range}(f)\}$）。根据定义，对于任意成员 $X \in \mathcal{P}$，存在某个 $\vec{y} \in \mathrm{range}\ f$，使得对于所有 $\vec{b} \in X$，

$f = \vec{y}$ k ϵ $f \in (\vec{b})$。根据 -伪确 $L_{\vec{b}} \subseteq \overline{B}_\epsilon^\infty\ \vec{b}$ $\vec{y} \in \overline{B}_\epsilon^\infty\ \vec{b}$ $\vec{b} \in \overline{B}_\epsilon^\infty\ \vec{y}$ $X \subseteq \overline{B}_\epsilon^\infty\ \vec{y}$
定性 -近似的定义，我们有 $(\vec{b})$ ( )，这表明 ( )，并且由对称性可得 ( )。这表明 ( )，因此 $\mathrm{diam}_\infty(X) \leq 2\epsilon < 1$。

设 $r = \frac{\eta \cdot d}{n}$。由于 $\mathcal{P}$ 的每个成员的 ℓ_∞ 直径均小于 1，根据定理 3.7，存在一点 $\vec{p} \in [0,1]^d$，使得 $\overline{B}_r^\infty(\vec{p})$ 至少与 $\mathcal{P}$ 的 $d+1 > k$ 个成员相交。设 $\vec{b}^{(1)}, \ldots, \vec{b}^{(d+1)}$ 为分属 $\mathcal{P}$ 的不同成员、且均属于 $\overline{B}_r^\infty(\vec{p})$ 的点。根据 $\mathcal{P}$ 的定义，这意味着对于不同的 j、$j' \in [d+1]$， $f(\vec{b}^{(j)}) \neq f(\vec{b}^{(j')})$。

现在，对于每个 $j \in [d+1]$，因为 $\|\vec{p} - \vec{b}^{(j)}\|_\infty \leq r$，根据定理 4.8，我们有 $\mathrm{d_{TV}}(\mathcal{D}_{A,\vec{p},n}, \mathcal{D}_{A,b^{(j)},n}) \leq n \cdot d \cdot r = \eta$。然而，这导致了矛盾，因为具有访问权限的 A

使用有偏硬币 $\vec{b}^{(j)}$返回 $f(\vec{b}^{(j)})$ 至少为 $\frac{1}{k}\cdot\frac{k+1}{k+2}$，根据全变差距离，必然有$A$使用有偏硬币 $\vec{p}$返回 $f(\vec{b}^{(j)})$ 的概率至少为$\frac{1}{k}\cdot\frac{k+1}{k+2}-\eta>\frac{1}{k+1}$；记作，$\Pr^{\mathcal{D}_{A,b}\vec{j}}()_n(\left\{f(\vec{b}^{(j)})\right\})\geq\frac{1}{k}\cdot\frac{k+1}{k+2}$，且 $\mathrm{d_{TV}}(^{\mathcal{D}}{}_{A,b^{(j)}},n,\mathcal{D}_A,\vec{p},n)\leq\eta$，因此 $\Pr^{\mathcal{D}_A,\vec{p},n}(\left\{f(\vec{b}^{(j)})\right\})\geq\frac{1}{k}\cdot\frac{k+1}{k+2}-\eta>\frac{1}{k+1}$。这是一个矛盾，因为一个分布不可能有 $d+1\geq k+1$ 个不相交事件，而每个事件的概率都大于 $\frac{1}{k+1}$。

□

我们以指出上述不可能性结果意味着 d-硬币偏差估计问题的证书复杂度存在下界，来结束本节。因此，不存在适用于 d-硬币偏差估计问题的 $\lfloor\log(d)\rfloor$-证书可复现算法。特别地，任何 l-证书可复现算法都需要 $l=\Omega(\log d/$

$delta$）。因此，我们针对 d-硬币偏差估计问题的算法具有最优的列表复杂度和证书复杂度。

5 PAC学习中的可复制性

在本节中，我们 es 建立 PAC 模型中的可复制性结果。首先，我们定义 PAC学习模式 l.

设 $\mathcal{H}$ 为定义在 X上的布尔函数（假设）类，且 $\mathcal{D}$ 为 X上的分布。对于函数 $f\in\mathcal{H}$，令 $\mathcal{D}_f$ 为通过根据 $\mathcal{D}$对元素 $x\in X$进行采样而获得的 $X\times\{0,1\}$ 上的分布，并输出 $\langle x,f(x)\rangle$。对于假设 h，其相对于 $\mathcal{D}_f$的误差记为 $^{e}{}_{\mathcal{D}_f}(h)$，其值为 $\Pr_{\langle x,f(x)\rangle\in\mathcal{D}_f}(f(x)\neq h(x))$。

定义 5.1. 若存在具有以下性质的学习算法 A，则假设类（或概念类）$\mathcal{H}$ 是 PAC 可学习的，样本复杂度为 m：对于任意 $f\in\mathcal{H}$，对于任意 $\mathcal{D}$ 上的分布 X，对于所有 $0<\delta$、$\epsilon<1$、A，在输入 δ、ϵ 以及根据 $\mathcal{D}_f$ 独立同分布抽取的 S 上，其中 $|S|\leq m$：算法输出一个假设h，使得 $^{e}{}_{\mathcal{D}_f}(h)\leq\varepsilon$。其成立的概率为 $\geq(1-\delta)$。

我们证明，任何能够通过统计查询学习算法学习的假设类，都存在一个可重复的 PAC 学习算法。首先，我们定义使用统计查询 [Kea98] 进行学习的概念。

定义 5.2. 统计预言机 $STAT(\mathcal{D}_f,\nu)$ 接收一个实值函数 ϕ：$X\times\{0,1\}\to(0,1)$ 作为输入，并返回一个估计值 v，满足

$$|v-E_{\langle x,y\rangle\in\mathcal{D}_f}[\phi(x,y)]|\leq\nu$$

定义5.3。我们称算法 A 通过统计查询学习概念类 $\mathcal{H}$：对于每个分布 $\mathcal{D}$ 和每个函数 $f\in\mathcal{H}$，对于每个 $0<\varepsilon<1$，存在 ν，使得算法 A 在输入 ε，并以 $STAT(\mathcal{D}_f,\nu)$ 作为预言机时，输出一个假设 h，使得 $^{e}{}_{D_f}(h)\leq\varepsilon$。如果概念类所提出的所有查询均为非自适应的，则称该概念类是非自适应可学习的。

5.1 可复现 PAC 学习的概念

现在，我们定义 PAC 模型中的列表式和证书式可复现学习的概念。

定义 5.4. 设 $\mathcal{H}$ 为一个假设类。我们称 $\mathcal{H}$ 是 k-列表可复制可学习的，如果存在一个算法 A 具有以下性质。对于每一个 $f\in\mathcal{H}$ 以及每一个分布 $\mathcal{D}$（在 X上），对于每一个$0<\epsilon<1$和 $0<\delta\leq1$，存在一个列表 L，其大小至多为 k，由 ε-近似假设组成，使得 A 在输入 δ、ϵ 和样本 $S\in\mathcal{D}_f^m$上，$\Pr_{S\sim\mathcal{D}_f^m}[A(S)\in L]\geq1-\delta$。我们称 m 为该学习算法的样本复杂度，并称 k 为其列表复杂度。

接下来我们定义证书可复现性。这与 [?] 中给出的可复制性概念非常接近。不过，我们主要关注的是算法实现完全可复现所需的随机性数量。

定义5.5.设 $\mathcal{H}$ 是一个概念类。我们称 $\mathcal{H}$ 是 ℓ-证书可复现可学习的，如果满足以下条件。存在一个学习算法 A，使得对于每个 $f \in \mathcal{H}$，对于 X 上的每个分布 $\mathcal{D}$，A 获取以下输入：ϵ、δ、$r \in \{0,1\}^\ell$ 和 $S \sim \mathcal{D}_f^m$。

$$\Pr_{r\in\{0,1\}^\ell}\left[\exists h_r : \Pr_{s\sim\mathcal{D}^m}[A(s;r) = h_r] \geq 1-\delta\right] \geq 1-\delta$$

我们称 m 为样本复杂度，并称随机比特数 ℓ 为 A的证书复杂度。

该定义可以进一步解释如下。对于算法 A，我们称 r 是“可复制性证书”，如果 A 以 m 个来自 $\mathcal{D}_f$的独立样本以及 r 作为输入，输出一个典范的 ε-近似假设 h_r，其概率为 $\geq 1-\delta$。那么上述定义要求 $1-\delta$ 比例的 $r \in \{0,1\}^l$是可复制性证书。

5.2 可复现算法

定理 5.6. 设 $\mathcal{H}$ 是一个可用 d 非自适应统计查询学习的概念类，则 $\mathcal{H}$是 $(d+1)$-列表可复现可学习的。此外，$(d+1)$-列表可复现算法的样本复杂度 $n = n(\nu,\delta)$ 等于 $O(d^2 \log \frac{d}{\delta} \cdot \frac{1}{\nu^2})$，其中 ν 是每个统计查询预言机的近似误差参数。

证明。该证明与定理 4.4 的证明非常相似。我们的可复现算法 B 的工作方式如下。令 ε和 δ 为输入参数，$\mathcal{D}$ 为一个分布，且 $f \in \mathcal{H}$。令 A 为针对 $\mathcal{H}$ 的统计查询学习算法，该算法输出一个近似误差为 $e_{\mathcal{D}_f}(h) = \varepsilon$ 的假设 h。令 $STAT(D_f,\nu)$ 为该算法的统计查询预言机。令 $\phi_1,\ldots,\phi_d$ 为 A 所提出的统计查询。

Let $\vec{b} = \langle E_{\langle x,y\rangle\in\mathcal{D}_f}[\phi_1(\langle x,y\rangle)\ldots, E_{\langle x,y\rangle\in\mathcal{D}_f}[\phi_d(\langle x,y\rangle)\rangle$。设 $\varepsilon_0 = \frac{\nu}{2d}$。算法 B首先估计值 $b[i] = E_{\langle x,y\rangle\in\mathcal{D}_f}[\phi_i(\langle x,y\rangle)]$, $1 \leq i \leq d$，对每个查询的近似误差不超过 ϵ_0，且成功概率为$1-\delta/d$。注意，这可以通过一种简单的经验估计算法实现，总共使用 $n = O(\frac{d^2}{\nu^2} \cdot \log d/\delta)$ 个样本。设 $\vec{v}$为估计得到的向量。由此可知，以至少 $1-\delta$ 的概率，$\vec{v} \in \overline{B}_{\varepsilon_0}^{\infty}(\vec{b})$。

现在，算法 B 在输入 $\vec{v}$上计算引理3.5中的确定性函数 f_ϵ。设 $\vec{u}$ 为输出向量。现在，算法 B 模拟统计查询算法 A，并以 $\vec{u}[i]$ 作为查询 ϕ_i的答案。根据引理3.5，$\vec{u} \in \overline{B}_{\nu}^{\infty}(\vec{b})$。因此，该算法所输出假设的误差至多为 ϵ。由于 A 是确定性算法，可能输出的数量仅取决于函数 f_ε的输出数量；更精确地说，可能输出的数量是集合 $\{f_\varepsilon(\vec{v}) : v \in \overline{B}_{\varepsilon_0}^{\infty}(\vec{b})\}$的大小，根据引理3.5，该大小几乎为 $d+1$。因此，算法 B 的可能输出总数至多为 $d+1$，其概率至少为 $1-\delta$。

□

我们注意到，我们可以模拟一个进行 d 次自适应查询的统计查询算法，从而得到一个 2^d-列表可复制学习算法。这可以通过将每个查询舍入为两个可能的值来实现（近似因子增加2）。该算法的样本复杂度将为 $O(\frac{d}{\nu^2} \cdot \log \frac{1}{\delta})$。利用自适应数据分析 [BNS+21]中的技术，样本复杂度可以改进为 $\tilde{O}(\frac{\sqrt{d}}{\nu^2})$。

接下来，我们为那些允许通过统计查询学习算法进行学习的假设类设计一种证书可复现算法。以下定理的证明见附录。

定理 5.7. 设 $\mathcal{H}$ 是一个可通过 d 非自适应统计查询学习的概念类，则 $\mathcal{H}$ 是 $\lceil \log \frac{d}{\delta} \rceil$-证书可复现可学习的。此外，该算法的样本复杂度 $n = n(\nu, \delta)$ 等于 $O(\frac{d^2}{\nu^2\delta^2} \cdot \log \frac{d}{\delta})$，其中 ν 是每个统计查询预言机的近似误差参数。

证明。该证明与定理 4.6 的证明非常相似。我们的可复制算法 B 的工作方式如下：设 ε 和 δ 为输入参数，$\mathcal{D}$ 为一个分布，且 $f \in \mathcal{H}$。设 A 为针对 $\mathcal{H}$ 的统计查询学习算法，它输出一个近似误差为 $e_{\mathcal{D}_f}(h) = \varepsilon$ 的假设 h。设 $STAT(D_f, \nu)$ 为该算法的统计查询预言机。设 ϕ_1，……，ϕ_d 为 A 所做的统计查询。

设 $\vec{b} = \langle E_{\langle x、y\rangle \in \mathcal{D}_f}[\phi_1(\langle x, y\rangle)] \ldots\ldots E_{\langle x、y\rangle \in \mathcal{D}_f}[\phi_d(\langle x, y\rangle)] \rangle$。令 $\varepsilon_0 = \frac{\nu\delta}{2d}$。算法 B 首先将值 $b[i] = E_{\langle x、y\rangle \in \mathcal{D}_f}[\phi_i(\langle x, y\rangle)]$、$1 \le i \le d$ 估计到近似误差 ϵ_0 以内，且每个查询的成功概率为 $1 - \delta/d$。注意，这可由一个简单的经验估计算法完成，该算法总共使用 $n = O(\frac{d^2}{\nu^2\delta^2} \cdot \log d/\delta)$ 个样本。设 $\vec{v}$ 为所估计的向量。由此可得，$\vec{v} \in \overline{B}^{\infty}_{\varepsilon_0}(\vec{b})$ 以至少 $1 - \delta$ 的概率成立。现在，算法 B 对引理 3.6 中所述的确定性函数 f 进行求值，输入为 $r \in \{0,1\}^{\ell}$，其中 $\ell = \lceil \log \frac{d}{\delta} \rceil$ 且 $\vec{v}$。根据引理 3.6，对于至少 $1-\delta$ 比例的 r，函数 f 会输出一个典范的 $\vec{v^*} \in \overline{B}^{\infty}_{\nu}(\vec{b})$。现在，算法 B 模拟统计查询算法 A，并以 $\vec{v*}[i]$ 作为对查询 ϕ_i 的回答。由于 A 是确定性算法，因此我们的算法 B 是证书可复制的。最后，注意证书复杂度为 $\lceil \log \frac{d}{\delta} \rceil$。 □

与之前一样，我们考虑统计查询算法进行 d 自适应查询的情况。以下定理的证明见附录。

定理5.8。设 $\mathcal{H}$ 是一个可通过 d 自适应统计查询学习的概念类，则 $\mathcal{H}$ 是 $\lceil d \log \frac{d}{\delta} \rceil$-证书可复现学习的。此外，该算法的样本复杂度等于 $O(\frac{d^3}{\nu^2\delta^2} \cdot \log \frac{d}{\delta})$，其中 ν 是每个统计查询预言机的近似误差参数。

证明思路。证明使用了与之前类似的论证。主要区别在于，我们会对每次查询进行评估，其近似误差为 $\frac{\nu\delta}{d}$，概率误差为 d/δ。这要求每次查询耗费 $O(\frac{d^2}{\nu^2\delta^2} \cdot \log \frac{d}{\delta})$。每次进行此类评估时，我们都会使用一组全新的证明随机性。注意，每次查询的证明长度为 $\lceil \log d/\delta \rceil$。因此，总证书复杂度为 $\lceil d \log \frac{d}{\delta} \rceil$。 □

5.3 PAC模型中的不可能性结果

在本节中，我们针对均匀分布，建立了 PAC 模型中 d-阈值估计问题的匹配上界和下界。我们证明，该问题存在一个 $(d+1)$-列表可复现算法，但不存在 d-列表可复现算法。

问题5.9（d-阈值估计问题）。固定某个 $d \in \mathbb{N}$。令 $X = [0,1]^d$。对于每个值 $\vec{t} \in [0,1]^d$（恰好与 X 相同），令 $h_{\vec{t}}$：$X \to \{0,1\}$ 为定义如下的函数

$$h_{\vec{t}}(\vec{x}) = \begin{cases} 1 & \text{for each } i \in [d], \text{ it holds that } x_i \le t_i \\ 0 & \text{otherwise} \end{cases}.$$

这是一个用于判断每个坐标是否小于或等于 $\vec{t}$ 指定阈值的函数。令 $\mathcal{H}$ 表示由所有此类阈值函数组成的假设类：$\mathcal{H} = \left\{ h_{\vec{t}} \mid \vec{t} \in [0,1]^d \right\}$。

我们首先指出，在一般的 PAC 模型中，列表可复现算法是不可能实现的。这是由已知结果得出的。

观察 5.10。令 k 为任意常数。即使在 $d=1$ 时，PAC模型中也不存在适用于 d-阈值估计 问题 的 k-列表可复现算法。

证明。根据 [BLM20]和 [ALMM19], 的工作可知，当且仅当存在一个常数 k ，使得概念类在 PAC 模型中具有一个 k-列表可复现算法时，任意类才具有有限的Littlestone 维度。由于概念类 d-阈值估计问题即使在 $d=1$时也具有无限的 Littlestone 维度，因此定理得证。 □

上述结果排除了通用PAC 模型中的列表可复现算法。在本节的其余部分，我们研究均匀分布下 PAC模型中 d-阈值估计问题的可复现可学习性。我们为列表复杂度建立了匹配的上界和下界。

定理5.11。在均匀分布下的 PAC 模型中，存在一个 $d+1$-列表可复现算法，用于 d-阈值估计问题

证明。众所周知且易于看出， d-阈值估计问题在均匀分布下是可学习的，只需进行 d 非自适应统计查询。因此，根据定理 5.6， d-阈值估计问题存在一个（$d+1$）列表可复现算法。 □

接下来，我们通过证明在均匀分布下，PAC 模型中不存在 d-列表可复现算法，来确立上述结果是紧的。

定理5.12。对于 $k < d+1$，不存在 k-列表可复现算法，用于 d-阈值估计 问题 在 PAC 模型中。

对于 $k < d+1$，不存在任何算法能够以 k-列表可复现地学习 d-阈值估计问题，在PAC 模型中的证明类似于定理4.7的证明。其原因在于，采-样 d个偏差为 $\vec{b}$ 的有偏硬币，与从 $[0,1]^d$中均匀随机抽取一个点 $\vec{x}$ 并在其上评估阈值函数 $h_{\vec{b}}$ 类似——这相当于询问所有硬币是否都是正面/1。不过，这两个模型有所不同，因为在 d-硬币偏差估计问题的样本模型中，算法可以看到每枚硬币是正面还是反面，但在 d-阈值估计问题的 PAC 模型中无法获得这一信息。反过来，在d-阈值估计问题的PAC 模型中，算法可以获得从 $[0,1]^d$ 中随机抽取的样本，但在 d-硬币偏差估计问题的样本模型中，算法无法获得这一信息。

此外， d-阈值估计问题的不可能性结果还有以下额外复杂性。在 d-硬币偏差估计问题中，根据定义，我们称由偏差向量 $\vec{a}$ 参数化的一组 d 枚硬币是由偏差向量 b 参数化的一组 d 枚硬币的 ϵ-近似，当且仅当 $\|b-\vec{a}\|_\infty \le \epsilon$；在应用 [VWDP+22] 的结果时，我们使用了这个范数。然而，PAC 模型中的 ϵ-近似概念与此大不相同。在 d-阈值估计问题中，可能存在假设 $h_{\vec{a}}$ 和 $h_{\vec{b}}$，使得 $\|b-\vec{a}\|_\infty > \epsilon$，但对于定义域 X 上的某个分布 $\mathcal{D}_X$，我们有 $\mathrm{e}^{\mathcal{D}_X}(h_{\vec{a}}, h_{\vec{b}}) \le \epsilon$。例如，如果 $\mathcal{D}_X$ 是 $X=[0,1]^d$ 上的均匀分布，且 $\vec{a}=\vec{0}$ 和 $\vec{b}$ 是第一个标准基向量 $\vec{b}=\langle 1、0、.....、0\rangle$，并且 $\epsilon=\frac{1}{2}$，那么 $\|\vec{b}-\vec{a}\|_\infty = 1 > \epsilon$，但 $\mathrm{e}^{\mathcal{D}_X}(h_{\vec{a}}, h_{\vec{b}}) = 0 \le \epsilon$，因为 $h_{\vec{a}}(\vec{x}) \ne h_{\vec{b}}(\vec{x})$ 当且仅当 $\vec{x}$ 的最后 $d-1$ 个坐标全为 0，且第一个坐标为 >0；然而，从 $X=[0,1]^d$ 上的均匀分布中采样到这样的 $\vec{x}$ 的概率为 0。

因此，我们不能像证明定理 4.7 时那样直接划分 $[0,1]^d$，而必须采用更巧妙的方法。事实证明，我们可以找到一个子集 $[\alpha,1]^d$，使得由这个立方体 $[\alpha,1]^d$ 相对两侧面上的向量参数化的假设之间具有较高的 PAC 误差。根据

$\mathsf{e}^{\mathcal{D}_X}$ 的三角不等式可知，这样的两个假设不可能都由同一个第三方假设近似。下面的引理阐述的正是这一点。

引理 5.13。设 $d \in \mathbb{N}$ 和 $\alpha = \frac{d-1}{d}$。设 $\vec{s}$、$\vec{t} \in [\alpha, 1]^d$ 满足：存在一个坐标 $i_0 \in [d]$，使得 $s_{i_0} = \alpha$ 和 $t_{i_0} = 1$（即 $\vec{s}$ 和 $\vec{t}$ 位于这个立方体的相对面上）。设 $\epsilon \leq \frac{1}{8d}$。那么不存在点 $\vec{r} \in X$，使得 $\mathsf{e}_{\mathrm{unif}}(h_{\vec{s}}, h_{\vec{r}}) \leq \epsilon$ 且 $\mathsf{e}_{\mathrm{unif}}(h_{\vec{t}}, h_{\vec{r}}) \leq \epsilon$ 同时成立（即不存在一个假设能同时近似 $h_{\vec{s}}$ 和 $h_{\vec{t}}$，且误差不超过 ϵ）。

证明。令 $\vec{q} = \left\langle \begin{cases} s_i & i = i_0 \\ t_i & i \neq i_0 \end{cases} \right\rangle_{i=1}^{d}$ 作为 $\vec{s}$ 的代理。我们需要以下断言。

断言 5.14。对于每个 $\vec{x} \in X$，以下条件等价：

1. $h_{\vec{q}}(\vec{x}) \neq h_{\vec{t}}(\vec{x})$
2. $h_{\vec{q}}(\vec{x}) = 0$ *and* $h_{\vec{t}}(\vec{x}) = 1$
3. $x_{i_0} \in (q_{i_0}, t_{i_0}] = (\alpha, 1]$ *and for all* $i \in [d] \setminus \{i_0\}$, $x_i \in [0, t_i]$.

此外，上述等价条件还意味着以下结论：

4. $h_{\vec{s}}(\vec{x}) \neq h_{\vec{t}}(\vec{x})$.

证明。

(2) $\Longrightarrow$(1)：这很简单。

(1) $\Longrightarrow$(2)：

注意，由于 $q_{i_0} = s_{i_0} = \alpha < 1 = t_{i_0}$，对于所有 $i \in [d]$，我们都有 $q_i \leq t_i$。若 $h_{\vec{t}}(\vec{x}) = 0$，那么对于某个 $i_1 \in [d]$，必定有 $x_{i_1} > t_{i_1}$；但由于 $t_{i_1} \geq q_{i_1}$，也会有 $x_{i_1} > q_{i_1}$，因此$h_{\vec{q}}(\vec{x}) = 0$，这就与 $h_{\vec{q}}(\vec{x}) = h_{\vec{t}}(\vec{x})$ 矛盾。因此 $h_{\vec{t}}(\vec{x}) = 1$，并且由于 $h_{\vec{q}}(\vec{x}) \neq h_{\vec{t}}(\vec{x})$，我们有 $h_{\vec{q}}(\vec{x}) = 0$。

(1) $\Longleftrightarrow$ (3)：我们将 $[0, 1]^d$ 划分为三个集合，并分别考察这三种情况。

情况 1： $x_{i_0} \in (q_{i_0}, t_{i_0}] = (\alpha, 1]$ 并且对于所有 $i \in [d] \setminus \{i_0\}$, $x_i \in [0, t_i]$。在这种情况下， $q_{i_0} < x_{i_0}$因此$h_{\vec{q}}(\vec{x}) = 0$并且对于所有 $i \in [d]$ $x_i \leq t_i$，因此 $h_{\vec{t}}(\vec{x}) = 1$，因此 $h_{\vec{q}}(\vec{x}) \neq h_{\vec{t}}(\vec{x})$。

情况 2： $x_{i_0} \notin (q_{i_0}, t_{i_0}] = (\alpha, 1]$ 并且对于所有 $i \in [d] \setminus \{i_0\}$, $x_i \in [0, t_i]$。在这种情况下，由于 $x_{i_0} \in [0, 1]$且 $x_{i_0} \notin (\alpha, 1]$，我们有 $x_{i_0} \leq \alpha = q_{i_0} \leq t_{i_0}$，并且对于所有其他 $i \in [d] \setminus \{i_0\}$，也有 $x_i \leq t_i = q_i$（根据 $\vec{q}$的定义）。因此 $h_{\vec{q}}(\vec{x}) = 1 = h_{\vec{t}}(\vec{x})$。

情况 3：对于某些 $i_1 \in [d] \setminus \{i_0\}$、 $x_{i_1} \notin [0, t_{i_1}]$。在此情况下，由于 $x_{i_1} \in [0, 1]$，我们有 $x_{i_1} > t_{i_1} = q_{i_1}$。因此 $h_{\vec{q}}(\vec{x}) = 0 = h_{\vec{t}}(\vec{x})$。

因此， $h_{\vec{q}}(\vec{x}) \neq h_{\vec{t}}(\vec{x})$ 当且仅当 $x_{i_0} \in (q_{i_0}, t_{i_0}] = (\alpha, 1]$ 且对所有 $i \in [d] \setminus \{i_0\}$,$x_i \in [0, t_i]$.

(1,2, 3) $\Longrightarrow$ (4)：根据 (2)，我们有 $x_{i_0} > q_{i_0}$，并且由于 $q_{i_0} = s_{i_0}$ 根据 $\vec{q}$ 的定义，可得$x_{i_0} > s_{i_0}$ 这意味着 $h_{\vec{s}}(\vec{x}) = 0$。根据 (3)， $h_{\vec{t}}(\vec{x}) = 1$这给出 $h_{\vec{s}}(\vec{x}) \neq h_{\vec{t}}(\vec{x})$。 □

有了这一结论，接下来我们将证明以下两个不等式：

$$2\epsilon < \mathrm{e}_{\mathrm{unif}}(h_{\vec{q}}, h_{\vec{t}}) \leq \mathrm{e}_{\mathrm{unif}}(h_{\vec{s}}, h_{\vec{t}}).$$

对于这两个不等式中的第二个，注意到根据上述结论的 (1) $\implies$ (4) 部分，由于 $h_{\vec{q}}(\vec{x}) \neq h_{\vec{t}}(\vec{x})$ 蕴含 $h_{\vec{s}}(\vec{x}) \neq h_{\vec{t}}(\vec{x})$，我们有

$$\begin{aligned}
\mathrm{e}_{\mathrm{unif}}(h_{\vec{q}}, h_{\vec{t}}) &= \Pr_{\vec{x} \sim \mathrm{unif}(X)}[h_{\vec{q}}(\vec{x}) \neq h_{\vec{t}}(\vec{x})] \\
&\leq \Pr_{\vec{x} \sim \mathrm{unif}(X)}[h_{\vec{s}}(\vec{x}) \neq h_{\vec{t}}(\vec{x})] \\
&= \mathrm{e}_{\mathrm{unif}}(h_{\vec{s}}, h_{\vec{t}}).
\end{aligned}$$

现在，对于上述不等式中的第一个，我们将使用该断言的 (1) $\iff$ (3) 部分，使用我们的假设 $\vec{t} \in [\alpha, 1]^d$（这意味着对每个 $i \in [d]$，都有 $[0, t_i] \subseteq [0, \alpha]$），使用假设 $\epsilon \leq \frac{1}{8d}$，并使用 ??。利用这些，我们得到如下结果：

$$\begin{aligned}
&\mathrm{e}_{\mathrm{unif}}(h_{\vec{q}}, h_{\vec{t}}) \\
&= \Pr_{\vec{x} \sim \mathrm{unif}(X)}[h_{\vec{q}}(\vec{x}) \neq h_{\vec{t}}(\vec{x})] \\
&= \Pr_{\vec{x} \sim \mathrm{unif}(X)}[x_{i_0} \in (\alpha, 1] \ \wedge \ \forall i \in [d] \setminus \{i_0\}, \, x_i \in [0, t_i]] \\
&= \Pr_{x_{i_0} \sim \mathrm{unif}([0,1])}[x_{i_0} \in (\alpha, 1]] \cdot \prod_{\substack{i=1 \\ i \neq i_0}}^{d} \Pr_{x \sim \mathrm{unif}([0,1])}[x \in [0, t_i]] \\
&\geq \Pr_{x_{i_0} \sim \mathrm{unif}([0,1])}[x_{i_0} \in (\alpha, 1]] \cdot \prod_{\substack{i=1 \\ i \neq i_0}}^{d} \Pr_{x \sim \mathrm{unif}([0,1])}[x \in [0, \alpha]] \\
&= (1 - \alpha) \cdot \alpha^{d-1} \\
&> \frac{1}{4d} \\
&\geq 2\epsilon.
\end{aligned}$$

因此，我们得到所需的两个不等式：

$$2\epsilon < \mathrm{e}_{\mathrm{unif}}(h_{\vec{q}}, h_{\vec{t}}) \leq \mathrm{e}_{\mathrm{unif}}(h_{\vec{s}}, h_{\vec{t}}).$$

这几乎完成了证明。如果存在某个点 $\vec{r} \in X$ 使得 $\mathrm{e}_{\mathrm{unif}}(h_{\vec{s}}, h_{\vec{r}}) \leq \epsilon$ 且 $\mathrm{e}_{\mathrm{unif}}(h_{\vec{t}}, h_{\vec{r}}) \leq \epsilon$，那么由 $\mathrm{e}_{\mathrm{unif}}$ 的三角不等式可得

$$\mathrm{e}_{\mathrm{unif}}(h_{\vec{s}}, h_{\vec{t}}) \leq \mathrm{e}_{\mathrm{unif}}(h_{\vec{s}}, h_{\vec{r}}) + \mathrm{e}_{\mathrm{unif}}(h_{\vec{t}}, h_{\vec{r}}) \leq 2\epsilon$$

但这将与上述不等式矛盾，因此不存在这样的 $\vec{r}$。 □

有了上述引理，我们现在可以证明定理 5.12。

证明定理 5.12。固定任意 $d \in \mathbb{N}$，并选择 ϵ 和 δ 作为 $\epsilon \leq \frac{1}{4d}$ 和 $\delta \leq \frac{1}{d+2}$。我们将使用常数 $\alpha = \frac{d-1}{d}$，并考虑立方体 $[\alpha, 1]^d$。我们还将只考虑 X 上的均匀分布。

假设为了矛盾，对某个 $k < d + 1$，这样的算法 A 确实存在。这意味着，对于每个可能的阈值 $\vec{t} \in [0, 1]^d$，存在某个假设集合 $L_{\vec{t}} \subseteq \mathcal{H}$，具有三个性质：(1) $L_{\vec{t}}$ 的每个元素都是对 $h_{\vec{t}}$ 的 ϵ-近似，(2) $|L_{\vec{t}}| \leq k$，以及 (3) 以至少 $1 - \delta$ 的概率，A 返回 $L_{\vec{t}}$ 的一个元素。

根据平凡的平均论证，这意味着至少存在一个 $L_{\vec{t}}$ 中的元素，它被 A 以至少 $\frac{1}{k}\cdot(1-\delta) \geq \frac{1}{k}\cdot(1-\frac{1}{d+2}) = \frac{1}{k}\cdot\frac{d+1}{d+2} \geq \frac{1}{k}\cdot\frac{k+1}{k+2}$ 的概率返回。令 f: $[\alpha,1]^d \to [0,1]^d$ 为一个函数，它将每个阈值 $\vec{t}\in[\alpha,1]^d$ 映射到 $L_{\vec{t}}$ 中的这样一个元素。这与定理 4.7 的证明略有不同，因为我们只在可能阈值中一个非常特定的子集上定义函数 f。其原因在定理 5.12 陈述后的讨论中已有提及。

由于 $\frac{1}{k}\cdot\frac{k+1}{k+2} > \frac{1}{k+1}$，令 η 使得 $0<\eta<\frac{1}{k}\cdot\frac{k+1}{k+2}-\frac{1}{k+1}$。

函数 f 诱导出 $\mathcal{P}$ 的一个划分 $[\alpha,1]^d$，其中 $\mathcal{P}$ 的成员是 f 的纤维（即 $\mathcal{P}=\{f^{-1}(\vec{y}):\vec{y}\in$ 的范围（f）$\}$）。对于任意成员 $W\in\mathcal{P}$ 和任意坐标 $i\in[d]$，集合 w_i： $\vec{w}\in W$ 不可能同时包含值 α 和 1——如果确实如此，那么将存在两个点 $\vec{s}$、$\vec{t}\in W$，使得 $s_i=\alpha$ 和 $t_i=1$；但由于它们都属于 W，因此存在某个 $\vec{y}\in[0,1]^d$，使得 $f(\vec{s}) = \vec{y} = f(\vec{t})$。然而，根据划分的定义， $h_{\vec{y}}$ 必须同时是 $h_{\vec{s}}$ 和 $h_{\vec{t}}$ 的 ϵ-近似（在 PAC 模型中），但根据定理 5.12，这是不可能的。

因此，划分 $\mathcal{P}$ 是 $[\alpha,1]^d$ 的一个“非张成”划分，正如定理 3.7的证明中那样，故存在某个点 $\vec{p}\in[\alpha,1]^d$，使得对每个半径 $r>0$，都有$\overline{B}_r^\infty(\vec{p})$与 $\mathcal{P}$中至少 $d+1$个成员相交。

与定理 4.8及其在定理 4.7的证明中的用法类似，我们可以利用以下两个事实。首先，由 $\gamma_1(\vec{s},\vec{t}) = \mathsf{e}_{\text{unif}}(h_{\vec{s}},h_{\vec{t}})$定义的函数 γ_1是连续的（关于定义域上的 ℓ_∞范数）。其次，函数 $\gamma_2(h_{\vec{s}},h_{\vec{t}}) = \mathsf{d}_{\text{TV}}(\mathcal{D}_{A,\vec{s},n},\mathcal{D}_{A,\vec{t},n})$是连续的（关于定义域上的$\mathsf{e}_{\text{unif}}$距离概念）。由此可得，复合 $\gamma_{12}(\vec{s},\vec{t}) = \mathsf{d}_{\text{TV}}(\mathcal{D}_{A,\vec{s},n},\mathcal{D}_{A,\vec{t},n})$是连续的。因此，我们可以找到某个半径 $r>0$，使得若 $\|\vec{t}-\vec{s}\|_\infty \leq r$，则$\mathsf{d}_{\text{TV}}(\mathcal{D}_{A,\vec{s},n},\mathcal{D}_{A,\vec{t},n}) \leq \eta$。

现在我们得到了与定理 4.7 的证明中相同类型的矛盾：对于特殊点 $\vec{p}$，我们有 $\mathcal{D}_{A,\,\vec{p},\,n}$ 是一个具有 $d+1\geq k+1$ 个不相交事件的分布，而这些事件的概率均大于 $\frac{1}{k+1}$。因此，不存在 k-列表可复现算法。

□

6 结论

在这项工作中，我们从算法的角度研究了机器学习中可复制性这一紧迫问题。我们观察到，绝对意义上的可复制性难以实现。因此，我们考虑了两种自然的扩展，用以刻画（非）可复制性的程度：列表可复现性与证书可复现性。我们为 d-硬币偏差估计问题以及可通过进行非自适应统计查询的统计查询算法学习的问题类，设计了具有较小列表复杂度、证书复杂度和样本复杂度的可复现算法。我们还在学习的 PAC 模型中以及针对 d-硬币偏差估计问题建立了某些不可能性结果。我们的工作引出了若干有趣的研究方向。在参数相当的情况下，列表可复现性与证书可复现性的样本复杂度之间存在差距。这一差距是否不可避免？目前，可通过非自适应统计查询与自适应统计查询学习的假设类之间，在可复制性参数上存在指数级差距。这一差距是否必要？一个一般性的问题是探索样本复杂度、列表复杂度、证书复杂度、自适应性与非自适应性之间的权衡。

7 Acknowledgements

感谢一位匿名审稿人建议提出观察 5.10。Pavan 的部分工作得到了美国国家科学基金会项目 2130536 的资助。Pavan 在西蒙斯理论研究所访问期间完成了部分工作

计算。Vander Woude 和 Vinodchandran 的部分工作得到了美国国家科学基金会项目 2130608 的资助。

参考文献

[AJJ+22] Kwangjun Ahn、Prateek Jain、Ziwei Ji、Satyen Kale、Praneeth Netrapalli和 Gil I. Shamir。《优化中的可复现性：理论框架与局限》，2022年。arXiv:2202.04598。[ALMM19] Noga Alon、Roi Livni、Maryanthe Malliaris 和 Shay Moran。隐私 PAC学习意味着有限的 Littlestone 维数。载于 Moses Charikar 和 Edith Cohen（编），《第 51 届ACMSIGACT计算理论研讨会论文集》，STOC 2019，美国亚利桑那州凤凰城，2019 年6 月23—26日，第852–860 页。ACM，2019 年。[AV20] Nima Anari 和Vijay V.Vazirani。在 NC 模型中，匹配问题与判定问题一样容易。载于 Thomas Vidick（编），《第11届理论计算机科学创新会议》，ITCS 2020，美国华盛顿州西雅图，2020 年1月 12—14日，LIPIcs 第 151 卷，第 54:1–54:25 页。SchlossDagstuhl - Leibniz-Zentrum f¨ur Informatik，2020 年。[Bak16] Monya Baker。1500名科学家揭示了可复现性的真相。《自然》，533:452–454，2016 年。[BLM20] MarkBun、RoiLivni 和 Shay Moran。隐私分类与在线预测之间的等价性。载于 Sandy Irani（编），《第 61届IEEE计算机科学基础年会》，FOCS2020，美国北卡罗来纳州达勒姆，2020年 11 月16—19日，第389–402 页。IEEE，2020 年。[BNS+21] Raef Bassily、Kobbi Nissim、Adam D. Smith、Thomas Steinke、UriStemmer 和Jonathan R. Ullman。自适应数据分析的算法稳定性。《SIAM计算杂志》，50(3)，2021 年。[Can15] Cl´ement L. Canonne。分布检验综述：你的数据很大，但它是蓝色的吗？《电子计算复杂性讨论会》，TR15-063，2015 年。URL: https://eccc.weizmann.ac.il/report/2015/063，arXiv:TR15-063。[DLPES02]Jesus A. De Loera、Elisha Peterson和 FrancisEdward Su。Sperner引理的多面体推广。《组合理论杂志》A辑，100(1):1–26，2002 年 10 月。URL: https://www.sciencedirect.com/science/article/pii/S0097316502932

747,

doi:10.1006/jcta.2002.3274.[DPV18] Peter Dixon、Aduri Pavan 和N. V. Vinodchandran。关于伪确定性近似算法。载于 Igor Potapov、PaulG.Spirakis 和 James Worrell（编），第43届国际计算机科学数学基础研讨会，MFCS 2018，2018年8月27日至31日，英国利物浦，LIPIcs第117卷，第61:1–61:11页。SchlossDagstuhl - Leibniz-Zentrum f¨ur Informatik，2018年。[DPVWV22] Peter Dixon、AduriPavan、JasonVanderWoude和 N. V. Vinodchran。伪确定-性：承诺与下界。载于Stefano Leonardi 和Anupam Gupta（编），STOC ’22：第54届ACM SIGACT计算理论年会，意大利罗马，2022年6月20日至24日，第1552–1565页。ACM，2022年。

[eco13] 科学如何出错。《经济学人》，第25–30页，2013年。[EKK+23] Hossein Esfandiari、Alkis Kalavasis、Amin Karbasi、Andreas Krause、Vahab Mirrokni和Grigoris Velegkas。可复现的老虎机，2023。arXiv:2210.01898。[GG] Shafi Goldwasser和Ofer Grossman。伪确定性NC中的二分图完美匹配。载于Ioannis Chatzigiannakis、Piotr Indyk、Fabian Kuhn和Anca Muscholl编，第44届自动机、语言与编程国际学术会议，ICALP 2017，2017年7月10日至14日，波兰华沙，LIPIcs第80卷。[GG11] Eran Gat和Shafi Goldwasser。具有唯一答案的概率搜索算法及其密码学应用。技术报告136，2011。URL: https://eccc.weizmann.ac.il/report/2011/136/。[GGH18] Shafi Goldwasser、Ofer Grossman和Dhiraj Holden。伪确定性证明。载于Anna R. Karlin编，第9届理论计算机科学创新会议，ITCS 2018，2018年1月11日至14日，美国马萨诸塞州剑桥，LIPIcs第94卷，第17:1–17:18页。Schloss Dagstuhl - Leibniz-Zentrum f¨ur Informatik，2018。[GGMW20] Shafi Goldwasser、Ofer Grossman、Sidhanth Mohanty和David P. Woodruff。伪确定性流式算法。载于Thomas Vidick编，第11届理论计算机科学创新会议，ITCS，LIPIcs第151卷，第79:1–79:25页，2020。[GKM21] Badih Ghazi、Ravi Kumar和Pasin Manurangsi。通过相关采样实现用户级差分隐私学习。载于Marc'Aurelio Ranzato、Alina Beygelzimer、Yann N. Dauphin、Percy Liang和Jennifer Wortman Vaughan编，《神经信息处理系统进展》第34卷：2021年神经信息处理系统年会，NeurIPS 2021，2021年12月6日至14日，线上，第20172–20184页，2021。[GL19] Ofer Grossman和Yang P. Liu。对数空间中的可复现性与伪确定性。载于Timothy M. Chan编，第30届ACM-SIAM离散算法年度研讨会论文集，SODA 2019，美国加利福尼亚州圣迭戈，2019年1月6日至9日，第606–620页。SIAM，2019。URL: https://doi.org/10.1137/1.9781611975482.38，doi:10.1137/1.9781611975482.38。[Gol19] Oded Goldreich。多伪确定性算法。电子计算复杂性讨论会，TR19-012，2019。URL: https://eccc.weizmann.ac.il/report/2019/012，arXiv:TR19-012。[ILPS22] Russell Impagliazzo、Rex Lei、Toniann Pitassi和Jessica Sorrell。学习中的可复现性。载于Stefano Leonardi和Anupam Gupta编，STOC '22：第54届ACM SIGACT计算理论年会，意大利罗马，2022年6月20日至24日，第818–831页。ACM，2022。URL: https://doi.org/10.1145/3519935.3519973，doi:10.1145/3519935.3519973。[JP05] Ioannidis JP。为什么大多数已发表的研究结果是错误的。《PLOS Medicine》，2(8)，2005。[Kea98] Michael J. Kearns。基于统计查询的高效抗噪学习。J. ACM，45(6):983–1006，1998。

[LOS21] Zhenjian Lu、Igor Carboni Oliveira 和 Rahul Santhanam。伪确定性算法与概率时间的结构。载于 Samir Khuller 和 Virginia Vassilevska Williams（编），STOC ’21：第 53 届 ACM SIGACT 计算理论研讨会，线上活动，意大利，2021 年 6 月 21—25 日，第 303–316 页。ACM，2021 年。[MPK19] Harshal Mittal、Kartikey Pandey 和 Yash Kant。ICLR 可复现性挑战报告（Padam：缩小训练深度神经网络时自适应梯度方法的泛化差距），2019 年。arXiv:1901.09517。[NAS19] 科学中的可复现性与可重复性。https://doi.org/10.17226/25303，2019 年。美国国家科学院、工程院和医学院。[OS17] I. Oliveira 和 R. Santhanam。亚指数时间内的伪确定性构造。载于第 49 届 ACM SIGACT 计算理论研讨会论文集（STOC 2017），加拿大魁北克省蒙特利尔，2017 年 6 月 19—23 日，第 665–677 页，2017 年。[OS18] Igor Carboni Oliveira 和 Rahul Santhanam。伪去随机化学习与近似。载于《近似、随机化与组合优化：算法与技术》，APPROX/RANDOM 2018，LIPIcs 第 116 卷，第 55:1–55:19 页，2018 年。[PVLS+21] Joelle Pineau、Philippe Vincent-Lamarre、Koustuv Sinha、Vincent Lariviere、Alina Beygelzimer、Florence d’Alche Buc、Emily Fox 和 Hugo Larochelle。提升机器学习研究的可复现性（NeurIPS 2019 可复现性计划报告）。机器学习研究期刊，22(164):1–20，2021 年。网址：http://jmlr.org/papers/v22/20-303.html。[SZ99] Michael E. Saks 和 Shiyu Zhou。BPhSPACE(S) ⊆ DSPACE(S3/2)。《计算机与系统科学杂志》，58(2):376–403，1999 年。[VWDP+22]Jason Vander Woude、Peter Dixon、Aduri Pavan、Jamie Radcliffe 和 N. V. Vinodchandran。舍入几何。CoRR，abs/2211.02694，2022 年。网址：https://doi.org/10.48550/arXiv.2211.02694，arXiv:2211.02694，doi:10.48550/arXiv.2211.02694。