arXiv:2611.32837v1 [cs.CC] 24 Nov 2026

间隔汉明距离消歧的通信复杂度

Yogesh Dahiya MikaG¨o¨os Nathaniel Harms Hamed
Hatami PooyaHatami Shachar Lovett Weiqiang Yuan

摘要

我们研究间隔汉明距离问题：Alice 和 Bob 分别收到一个 n 位字符串，并需要判断它们的汉明距离是否至多为 $n/3$ ，或至少为 $2n/3$。由于双方可以通过比较少数几个共享随机坐标来估计汉明距离，这个承诺问题的公开硬币随机通信复杂度为 $O(1)$。对于汉明距离介于 $n/3$ 和 $2n/3$ 之间的输入，所需输出未定义，即“有歧义”。

我们证明，将这个问题消歧为一个全函数的任何方式，其随机通信复杂度均为 $\Omega(n^{2/3})$。作为推论，我们得到随机通信复杂度与伪确定性通信复杂度之间存在 $O(1)$ 与 $\Omega(n^{2/3})$ 的分离。这个证明最初由一个 AI 模型发现，随后由作者重新整理。

1 引言

记两个字符串 $x, y \in \{\pm 1\}^n$ 之间的汉明距离为 $\mathrm{dist}(x, y)$，并令 $\gamma \in (0, 1/2)$ 为一个固定参数。 γ-间隔汉明距离问题 是这样一个承诺问题：Alice 和 Bob 分别接收 x 和 y，并且必须区分以下情况

$$\mathrm{dist}(x, y) \le \gamma n \qquad \text{and} \qquad \mathrm{dist}(x, y) \ge (1 - \gamma)n.$$

该问题的随机（公币）通信复杂度为 $O_\gamma(1)$，因为参与者可以通过在一个常数大小的共享随机坐标集合上比较各自的比特，足够准确地估计 $\mathrm{dist}(x, y)$。

承诺问题的消歧 是将其扩展为一个全布尔函数，也就是说，该全函数在所有满足承诺的输入上都与原问题一致。我们证明， γ-间隔汉明距离问题的每个消歧都具有较大的随机复杂度。

定理 1.1。令 $\gamma \in (0, 1/2)$ 为固定值。每个满足以下条件的全函数 $F\colon \{\pm 1\}^n \times \{\pm 1\}^n \to \{0, 1\}$

$$F(x, y) = \begin{cases} 0 & \text{if } \mathrm{dist}(x, y) \le \gamma n, \\ 1 & \text{if } \mathrm{dist}(x, y) \ge (1 - \gamma)n \end{cases}$$

的随机通信复杂度为$\Omega_\gamma(n^{2/3})$。

此前，该问题已知的最佳下界为 $\Omega(\log \log n)$，这是由 [BHH+26] 得出的。文献 $\gamma(n)$ 猜想下界为线性 Ω [GHR+26]。

伪确定性通信。我们的结果部分受到通信复杂度中伪确定性研究的启发。在求解搜索问题时，只需以高概率保证正确的随机化算法，在不同的独立运行中仍可能输出不同的结果（尽管这些结果都是正确的）。这种行为可能并不理想，因此我们可以要求随机化算法在不同运行中的输出保持一致（且以高概率成立）。这样的算法称为伪确定性算法。伪确定性由 Gat 和 Goldwasser [GG11] 引入（Huynh 和 Nordström ¨om [HN12] 也独立地在证明复杂度的背景下提出了这一概念），此后已在广泛的计算模型中得到研究，包括学习算法 [OS18, ILPS22], 、顺序与并行算法 [GGR13, GG17], 、平均情形与近似算法 [Hol17, DPV18],、交互式证明 [GGH18], 、低空间算法 [GL19]，以及流式算法 [GGMW20]。这一研究方向的核心问题是，伪确定性在多大程度上能够与一般随机化计算区分开来。

形式化地说，通信搜索问题是一个关系 $\mathcal{R} \subseteq \{\pm 1\}^n \times \{\pm 1\}^n \times \mathcal{Z}$：Alice 和 Bob 分别接收 x 和 y，并必须输出某个 z，使得 $(x, y, z) \in \mathcal{R}$。针对 $\mathcal{R}$ 的伪确定性协议 Π 是一种随机化协议，对于每个 (x, y)，都存在一个固定值 $h(x, y)$，使得 $(x, y, h(x, y)) \in \mathcal{R}$，并且 $\Pr[\Pi(x, y) = h(x, y)] \geq \frac{2}{3}$。我们将 $h(x, y)$ 称为规范输出。

虽然我们的目标是比较伪确定性协议与随机化协议，但对于随机化协议如何才算解决一个搜索问题，目前并没有标准定义。对于函数，标准定义要求算法以至少 $2/3$ 的概率输出唯一正确答案，而常数 $2/3$ 的取值本质上是任意的，因为我们可以高效地将成功概率放大到任意其他常数。对于搜索问题，我们并非总能放大成功概率（因为算法在不同次执行中可能输出不同的正确答案），因此不能简单地选取一个固定的成功概率常数。论文 [GHR+26] 更详细地讨论了这些细微之处，并综述了文献中的各种定义。不过，在本文中，如果对于每个 (x, y)，都有 $\Pr[(x, y, \Pi(x, y)) \in \mathcal{R}] \geq \frac{2}{3}$，我们就称随机化协议 Π解决$\mathcal{R}$。这之所以合理，是因为我们研究的问题确实允许高效地放大成功概率。

分离随机化通信复杂度与伪确定性通信复杂度的问题由 [GIPS21] 提出。最近才获得了首批分离结果：$O(1)$ 与 $\Omega(\log \log n)$ [BHH+26]，以及 $O(\log n)$ 与 $\Omega(\sqrt{n})$ [GHR+26, Wat26], 。定理 1.1 改进了这两项结果，给出了 $O(1)$ 与 $\Omega(n^{2/3})$ 之间的分离。这是因为 γ-间隔汉明距离问题是一个关系问题（对于距离介于 γn 和 $(1-\gamma)n$ 之间的情况，该关系 $\mathcal{R}$ 允许输出 0 或 1），而该问题的伪确定性算法等价于一个计算消歧的随机化算法。

此外，作为主要结果的一个应用，我们得到了汉明距离估计中 $O(1)$ 与 $\Omega(n^{2/3})$ 之间的分离：对于固定的 $\varepsilon \in (0, 1/2)$，参与者必须输出一个满足以下条件的数 t

$$|\operatorname{dist}(x, y) - t| \leq \varepsilon n. \tag{1}$$

通过 $O_\varepsilon(1)$ 个共享随机坐标估算汉明距离，只需 $O_\varepsilon(1)$ 通信即可解决这个问题。现在假设 Π 是一个伪确定性协议，其规范输出为 $h(x, y)$，并考虑全函数

$$F(x, y) := \mathbf{1}[h(x, y) \geq n/2].$$

注意，F 满足定理 1.1 的条件，其中 $\gamma = \frac{1/2-\varepsilon}{2}$。此外，运行 Π 和

当且仅当其输出至少为 $n/2$ 时输出 1，以误差至多为 $1/3$ 计算 F。因此我们得到如下结果。

推论 1.2。 对于每个固定的 $\varepsilon \in (0,\ 1/2)$，搜索问题 (1) 具有 $O_\varepsilon(1)$ 的随机通信复杂度，以及 $\Omega_\varepsilon(n^{2/3})$ 的伪确定性通信复杂度。

奇偶决策树。 对于 $z \in \{\pm1\}^n$，令 $|z| \coloneqq |\{i :\ z_i = -1\}|$。对于 $\gamma \in (0,\ 1/2)$，γ-间隔多数问题是承诺问题 $f\colon\ \{\pm1\}^n \to \{0,1,*\}$，定义为

$$f(z) = \begin{cases} 0 & \text{if } |z| \le \gamma n, \\ 1 & \text{if } |z| \ge (1-\gamma)n, \\ * & \text{otherwise.} \end{cases} \tag{2}$$

γ-间隔汉明距离问题是其 XOR-提升 $F(x,\ y) = f(x \odot y)$，其中 $\odot$ 表示按坐标相乘，即 XOR的 $\{\pm1\}$ 表示。

对 $x \odot y$ 的每次奇偶查询，都可以通过交换 x 和 y 在查询集合上的奇偶性，使用两个比特的通信来模拟。因此，对于 γ-间隔多数的任何消歧，其随机奇偶决策树都可以产生一个针对 γ-间隔汉明距离的某个消歧的随机通信协议，通信开销至多加倍。根据 定理 1.1，我们得到以下结果。

推论 1.3。 对于任意固定的 $\gamma \in (0, 1/2)$，*γ-Gap Majority* 问题的每个消歧版本都具有 $\Omega_\gamma(n^{2/3})$ 的随机奇偶决策树复杂度。

这改进了 Gavinsky 对同一问题 $\gamma(\sqrt{n})$ 的 Ω [Gav25] 下界。事实上，我们对定理 1.1 的证明沿用了他的迭代策略。

人工智能方法论。 一个人工智能模型首先生成了一个 $\Omega(\sqrt{n})$ 下界的证明；经过进一步的通用提示后，它将下界改进为 $\Omega(n^{2/3})$。作者从这些人工智能生成的证明中提取了关键思想，并据此以他们认为最自然、最直观的形式撰写了论证。本说明中的论述由作者起草，人工智能工具用于编辑、排版和拼写检查。

记号。 我们用 $\mathrm{R}_\varepsilon(F)$ 表示 F 在每个输入上的错误率至多为 ε 时的公开硬币随机通信复杂度。我们用 $\mathcal{U}_m$ 表示 $\{\pm1\}^m$上的均匀分布，并用 $\mathrm{Ber}(\beta)$ 表示 $\{\pm1\}$ 上的分布：该分布以 β 的概率取值为 1，以 $1-\beta$ 的概率取值为 -1。

我们用 H 表示通常的香农熵。有限集合 Ω 上分布 μ 的最小熵 及其亏损 分别定义为（所有对数均以 2 为底）

$$H_\infty(\mu) \coloneqq -\log \|\mu\|_\infty \qquad \text{and} \qquad D_\infty(\mu) \coloneqq \log(|\Omega|\,\|\mu\|_\infty).$$

对于 $\rho \in [-1,1]$，一个取值于 $\{\pm1\}^m \times \{\pm1\}^m$ 的随机变量（X，Y）称为 ρ-相关的，如果每个边缘分布都是均匀的，X_i，Y_i 中的各对彼此独立，并且对于每个 i 都有 $\mathbb{E}[X_iY_i] = \rho$。对于 $x \in \{\pm1\}^m$，我们用 $N_\rho(x)$ 表示在给定 $X = x$ 条件下 Y 的条件分布：在每个坐标中，$Y_i = x_i$ 以概率 $(1+\rho)$ 取值为 $/2$，并以概率 $(1-\rho)$ 取值为 $/2$。

2 证明概述

我们通过反证法证明 定理 1.1。假设 F 具有一个通信成本较小的随机化协议。必要时，将 $F(x,\ y)$ 替换为 $1-F(x,\ -y)$，我们不妨假设

$$\mathop{\mathbb{E}}_{\mathcal{U}_n\times\mathcal{U}_n} F \geq \frac{1}{2}.$$

我们遵循 Gavinsky 的迭代策略 [Gav25]。通过利用 F的低成本协议，我们逐步扩展坐标集合 K，并迫使 Alice 和 Bob 的输入在 K中的大多数坐标上保持一致，同时维持 $\mathbb{E}\,F \geq 1/2$。最终，与这些被迫坐标一致的每一对输入 K 的汉明距离 $< \gamma n$，因此 F 必定为 0，这与所维持的性质 $\mathbb{E}\,F \geq 1/2$相矛盾。

更准确地说，我们维护输入上的一个乘积分布 $\mu\times\nu$、一个已处理坐标集合 K 以及一个参考串 $a\in\{\pm1\}^K$，并满足以下两个性质：

(i) $\mathbb{E}_{\mu\times\nu}\,F \geq 1/2$;

(ii) 对于 $\mu\times\nu$支撑集中的每个 (x,y)，$\mathrm{dist}(x_K,a)$ 和 $\mathrm{dist}(y_K,a)$ 都很小（因此 $\mathrm{dist}(x_K,y_K)$ 也很小）。

我们从均匀输入和 K 空集开始。每一轮分两个步骤向 K中加入新的坐标：

1. 植入大致一致的区块（引理 *4.1*）。我们选择一个大的块 I 之外 K，并在其上选取一个参考串 w，将两个输入都限制为与 w 在 I上接近，同时保持$\mathbb{E}\,F > 1/4$不变。然后我们将 I 添加到 K 中，并将 a 扩展 w。

2. 恢复期望（引理 *4.2*）。我们使用针对 F 的低成本协议，找出一个概率足够大的矩形（协议的一片叶节点），使得在该矩形上 $\mathbb{E}\,F \geq 1/2$。对这个矩形进行条件化即可恢复期望。

最终，支撑集中的每一对输入至多在 γn 个坐标上不同。于是，在整个支撑集上 $F=0$，这与 (i) 矛盾。

为了说明这一思路，先假设 X 和 Y 均为均匀分布。我们将它们替换为一个 ρ-相关对：先选取一个随机集合 $I\subseteq[n]$，以概率 ρ独立地纳入每个坐标，再对满足 $X_I=Y_I$ 的 X，Y进行均匀采样。下面的定理表明，对于 $\rho=\Theta(d^{-1/2})$，这只会使协议的接受概率发生很小的常数变化。

定理 2.1 (Girish–Sinha–Tal–Wu，[GSTW23，定理 1。5])。对于每个随机化协议Π，其输出属于 $\{0,1\}$，且通信比特数至多为 d，以及任意 $\rho\in[-1,1]$,

$$\left|\Pr_{X,Y\sim\mathcal{U}_m}[\Pi(X,Y)=1]-\Pr_{X\sim\mathcal{U}_m,\,Y\sim N_\rho(X)}[\Pi(X,Y)=1]\right| \leq O\big(|\rho|\sqrt{d}\big).$$

由于该协议的错误率较低，F 的期望也只会发生小幅变化。通过取平均，我们可以固定一个大小为 $\Omega(\rho n)$ 的块 I 以及一个公共值 $w=X_I=Y_I$，同时保持 $\mathbb{E}\,F>1/4$。

当我们限制在一个矩形上以恢复期望时，输入不再是均匀的，因此定理 2.1 不能直接应用。为处理这一问题，我们使用耦合（引理 3.1）：任何随机

字符串 X 都可以与均匀随机串 U 进行耦合，使得 X 和 U 在期望汉明距离意义下接近。分别对 Alice 和 Bob 的输入应用这一方法，可得到均匀的代理字符串 U 和 V，于是我们可以对（U，V）应用上述相关性论证，而不是直接对输入应用。由于每个输入都接近其代理字符串，因此两个输入在块 I 上都接近 $w = U_I = V_I$。

该耦合还保证，在固定（I，w）并舍弃一个很小的例外集后，剩余坐标上的输入仍具有较高的最小熵。此外，在给定（I，w）的条件下，Alice 和 Bob 的输入相互独立。

3 与均匀分布的耦合

下面的引理使用 Marton 提出的序贯最大耦合，将一个 μ 上具有高最小熵的分布 $\{\pm1\}^m$ 与 $\{\pm1\}^m$ 上的均匀分布进行耦合 [Mart86]。所得的有序对（X，U），其中 $X \sim \mu$ 和 U 均匀，具有以下两个性质：（a） X 和 U 的期望汉明距离较小；（b）对于坐标的任意集合 I，揭示 U_I 不会显著增加 $X_{\bar{I}}$ 各个取值的概率，但在一个很小的例外集上除外。

引理 3.1（均匀耦合）。设 μ 是 $\{\pm1\}^m$ 上的一个分布，且具有 $D_\infty(\mu) \le D$。存在 μ 和 $\mathcal{U}_m$ 的一个耦合（X，U），满足以下性质：

(a) $\mathbb{E}\big[\text{距离}(X,U)\big] \le \sqrt{mD}$。

(b) 对每个 $I \subseteq [m]$ 和 $s > 0$，以下结论成立。对于 $u \in \{\pm1\}^I$，令 ν_u 为 $(X_{\bar{I}} \mid U_I = u)$ 的分布。于是

$$\Pr_{(X,U)}\Big[\nu_{U_I}(X_{\bar{I}}) > 2^{-(m-|I|)+D+s}\Big] \le 2^{-s}. \tag{3}$$

证明。对于每个 $i \in [m]$，令

$$a_i(x_{<i}) \coloneqq \mathbb{E}_{X\sim\mu}\big[X_i \;\big|\; X_{<i} = x_{<i}\big].$$

取 ind $T_1, \ldots, T_m$，它们各自在 $[-1,1]$ 上均匀分布；依次对 $i = 1, \ldots, m$，令

$$X_i = 1 \iff T_i \le a_i(X_{<i}), \qquad U_i = 1 \iff T_i \le 0.$$

由于 $\Pr[T_i \le a] = (1+a)/2$，我们有 $X \sim \mu$ 和 $U \sim \mathcal{U}_m$。此外，

$$\Pr[X_i \ne U_i \mid X_{<i}] = \frac{|a_i(X_{<i})|}{2}.$$

部分 (a)。在给定 $X_{<i} = x_{<i}$ 的条件下，比特 X_i 服从分布 $\mathrm{Ber}\big(\frac{1+a_i(x_{<i})}{2}\big)$。根据熵的链式法则以及对于 $a \in [-1,1]$ 的基本界 $H\big(\mathrm{Ber}(\frac{1+a}{2})\big) \le 1 - a^2/4$，

$$H(X) = \sum_{i=1}^m H(X_i \mid X_{<i}) \le m - \frac{1}{4}\sum_{i=1}^m \mathbb{E}\big[a_i(X_{<i})^2\big].$$

重新整理并利用 $H(X) \ge H_\infty(X) = m - D_\infty(\mu) \ge m - D$，我们得到

$$\frac{1}{4}\sum_{i=1}^m \mathbb{E}\big[a_i(X_{<i})^2\big] \le m - H(X) \le D. \tag{4}$$

因此，根据柯西–施瓦茨不等式和（4），

$$\mathbb{E}\big[\mathrm{dist}(X,U)\big] = \frac{1}{2}\sum_{i=1}^{m}\mathbb{E}\big|a_i(X_{<i})\big| \le \sqrt{mD}.$$

（b）部分。 固定 $I \subseteq [m]$，并令 $Z := (U_I,\ X_{\bar{I}})$。由于 U_I 是均匀的，

$$\mathbb{E}\big[2^{m-|I|}\,\nu_{U_I}(X_{\bar{I}})\big] = 2^{m-|I|}\sum_{u,y}\Pr[Z=(u,y)]\,\frac{\Pr[Z=(u,y)]}{\Pr[U_I=u]} = 2^m\sum_{z}\Pr[Z=z]^2.$$

由于 $2^m\,\|\mu\|_\infty \le 2^D$，马尔可夫不等式将（3）归结为证明 Z的碰撞概率满足

$$\mathrm{col}(Z) := \sum_{z}\Pr[Z=z]^2 \le \|\mu\|_\infty . \tag{5}$$

我们对 m 上的归纳法证明（5）。对于 $a \in \mathrm{supp}(X_1)$，令 $p_a := \Pr[X_1 = a]$，并令 μ_a 为（$X_{>1} \mid X_1 = a$）的分布。于是 $\|\mu_a\|_\infty \le \|\mu\|_\infty / p_a$。

注意，对于任意具有正概率的 $(X_1,\ U_1) = (a,\ b)$，在该条件下，$(X_{>1},\ U_{>1})$ 这对变量就是关于 μ_a 的上述耦合。因此根据归纳法，$Z_{>1}$ 的条件分布的碰撞概率至多为 $\|\mu\|_\infty / p_a$。

在 $1 \notin I$ 的情况下，我们有 $Z_1 = X_1$，因此

$$\mathrm{col}(Z) = \sum_a p_a^2\,\mathrm{col}(Z_{>1} \mid X_1 = a) \le \sum_a p_a^2 \cdot \frac{\|\mu\|_\infty}{p_a} = \|\mu\|_\infty .$$

如果 $1 \in I$，则 $Z_1 = U_1$ 是均匀的。根据碰撞概率的凸性和归纳假设，

$$\begin{aligned}\mathrm{col}(Z) &= \frac{1}{2}\,\mathbb{E}_{U_1}\big[\mathrm{col}(Z_{>1} \mid U_1)\big] \le \frac{1}{2}\,\mathbb{E}_{X_1,U_1}\big[\mathrm{col}(Z_{>1} \mid X_1, U_1)\big] \\ &\le \frac{\|\mu\|_\infty}{2}\,\mathbb{E}_{X_1}\left[\frac{1}{p_{X_1}}\right] \le \|\mu\|_\infty ,\end{aligned}$$

其中最后一个不等式由$\mathbb{E}[1/p_{X_1}] = |\mathrm{supp}(X_1)| \le 2$得到。由此完成了（5）的证明，从而也完成了（b）部分的证明。 □

4 定理 1.1 的证明

在本节中，$d := \mathrm{R}^{1/100}(F)$，并且 $\rho := c_0/\sqrt{d}$，其中 $c_0 > 0$是一个足够小的绝对常数，使得定理 2.1 中的界至多为 $1/100$。对于定义在 μ 上的分布 $\{\pm1\}^n$ 和集合 $J \subseteq [n]$，我们将 μ_J 在 μ 上的边缘分布 J记为。

引理 4.1（植入大致一致的区块）。设 $\mu \times \nu$ 是定义在 $\{\pm1\}^n \times \{\pm1\}^n$ 上的乘积分布，满足 $\mathbb{E}_{\mu\times\nu} F \ge 1/2$。设 $J \subseteq [n]$ 是一个包含 $m \ge 100/\rho$ 个坐标的集合，使得 μ_J 和 ν_J 的亏损至多为 D，并令 $b := 20\rho\sqrt{mD}$。

那么，存在一个大小至少为 $\rho m/2$ 的集合 $I \subseteq J$、一个字符串 $w \in \{\pm1\}^I$ 以及一个乘积分布 $\widetilde{\mu} \times \widetilde{\nu}$，其支撑集包含于 $\mu \times \nu$ 的支撑集内，使得：

（*i*）每个 $(x, y) \in \mathrm{supp}(\widetilde{\mu} \times \widetilde{\nu})$ 都满足 $\mathrm{dist}(x_I, w) \le b$ 以及 $\mathrm{dist}(y_I, w) \le b$；

(*ii*) $\mathbb{E}_{\widetilde{\mu}\times\widetilde{\nu}} F > 1/4$;

(iii) $\widetilde{\mu}_{J\setminus I}$ 和 $\widetilde{\nu}_{J\setminus I}$ 的亏损至多为 $D+7$。

证明。 应用 引理 3.1 于 μ_J和 ν_J，得到耦合（X_J， U_J）和（Y_J， V_J），其中 U_J和V_J 在 $\{\pm1\}^J$上均匀分布。考虑如下协议 Π，其代价为 d，其中 Alice 和 Bob 分别接收 u、 $v \in \{\pm1\}^J$:

- Alice 从在 $U_J = u$条件下的耦合中采样 X_J，然后从在 X_J条件下的 μ中采样 $X_{[n]\setminus J}$。
- Bob 对 $V_J = v$ 执行相同操作，以获得 Y，使用独立的随机性 s.
- 他们在（X， Y）上运行针对 F的协议，其代价为 d，错误率至多为 $1/100$。

我们在来自两个不同分布的输入（u， v）上运行 Π。首先，令（U'_J， V'_J）相互独立且均匀分布，并令 X' 和 Y' 为 Alice 和 Bob 在输入（U'_J， V'_J）上采样得到的字符串；于是（X'， Y'）$\sim \mu\times\nu$。其次，令（U_J， V_J）为 ρ 相关，并令 X 和 Y 为在输入（U_J， V_J）上采样得到的字符串。注意，在这种情况下， X 和 Y并不独立。

根据定理 2.1， Π在两次运行中的接受概率相差至多 $1/100$。由于 F 的原始协议在每个输入上的错误概率至多为 $1/100$，因此我们得到

$$\mathbb{E}\, F(X,Y) \geq \mathbb{E}\, \Pi(U_J,V_J) - \frac{1}{100} \geq \mathbb{E}\, \Pi(U'_J,V'_J) - \frac{2}{100} \geq \mathbb{E}\, F(X',Y') - \frac{3}{100} \geq \frac{47}{100}. \tag{6}$$

一个 ρ-相关对 $(U_J,\ V_J)$ 可以按如下方式采样：首先选择一个随机集合 $I \subseteq J$，使每个坐标以概率 ρ 独立地包含在其中；然后均匀采样 $V_I = U_I$；最后独立且均匀地采样 $U_{J\setminus I}$ 和 $V_{J\setminus I}$。

记 $R := (I,\ U_I)$，令 p_R 和 q_R 分别为在给定 R 时 $X_{J\setminus I}$ 和 $Y_{J\setminus I}$ 的条件分布，并设 $M_I := 2^{-(m-|I|)+D+5}$。考虑以下事件

$$\mathcal{A} := \big\{\mathrm{dist}(X_I,U_I) \leq b,\ p_R(X_{J\setminus I}) \leq M_I\big\},$$
$$\mathcal{B} := \big\{\mathrm{dist}(Y_I,V_I) \leq b,\ q_R(Y_{J\setminus I}) \leq M_I\big\}.$$

如果 R 的值 $(I,\ w)$ 满足以下条件，我们就称其为良好值

$$|I_0| \geq \rho m/2 \qquad \text{and} \qquad \mathbb{E}\big[F(X,Y)\,\mathbf{1}_{\mathcal{A}\cap\mathcal{B}} \;\big|\; R=(I_0,w)\big] > \frac{1}{4}. \tag{7}$$

我们先证明存在一个良好值，然后利用它构造 $\widetilde{\mu}$ 和 $\widetilde{\nu}$.

存在一个良好的 R。我们对 $\mathcal{A}$或 $\mathcal{B}$失败的概率进行界定。由于 $(X_J,\ U_J)$ 与I 相互独立，且每个坐标以概率 I 属于 ρ，根据引理 3.1(a),

$$\mathbb{E}\big[\mathrm{dist}(X_I,U_I)\big] = \rho\, \mathbb{E}\big[\mathrm{dist}(X_J,U_J)\big] \leq \rho\sqrt{mD},$$

因此，由于 $b = 20\rho\sqrt{mD}$，根据马尔可夫不等式可得 $\Pr[\mathrm{dist}(X_I,U_I) > b] \leq 1/20$（这里可以假设 $D>0$）。对每个固定的 I应用 引理 3.1 (b) 以及 $s=5$，可得 $\Pr[p_R(X_{J\setminus I}) > M_I] \leq 1/32$。根据并集界，并对 (Y_J,V_J) 应用相同的论证，我们得到

$$\Pr[\mathcal{A}^c] \leq \frac{1}{20} + \frac{1}{32} \qquad \text{and} \qquad \Pr[\mathcal{B}^c] \leq \frac{1}{20} + \frac{1}{32}.$$

最后，根据切尔诺夫界，$\Pr[|I| < \rho m/2] \le e^{-\rho m/8} < 1/100$。

令 $\mathcal{G} := \{|I| \ge \rho m/2\} \cap \mathcal{A} \cap \mathcal{B}$。由并集界，$\Pr[\mathcal{G}^c] < \frac{1}{5}$，因此由-(6)，

$$\mathbb{E}\big[F(X,Y)\,\mathbf{1}_{\mathcal{G}}\big] \ge \frac{47}{100} - \frac{1}{5} > \frac{1}{4}.$$

对 R取平均，存在某个值（I_0，w），使得 R满足 $\mathbb{E}\big[F(X,Y)\,\mathbf{1}_{\mathcal{G}} \mid R = (I_0, w)\big] > \frac{1}{4}$，且该值满足良好性条件（7）。

完成证明。固定 R的一个良好值(I_0, w)。给定 $R = (I_0, w)$后，输入X 和 Y 相互独立。令

$$\alpha := \Pr\big[\mathcal{A} \mid R = (I_0, w)\big], \qquad \beta := \Pr\big[\mathcal{B} \mid R = (I_0, w)\big],$$

并令 $\widetilde{\mu}$ 和 $\widetilde{\nu}$ 分别为在给定 $R = (I_0, w)$和 $\mathcal{A}$的条件下 X 的分布，以及在给定 $R = (I_0, w)$和$\mathcal{B}$的条件下 Y的分布。

性质（i）由 $\mathcal{A}$ 和 $\mathcal{B}$的定义可知成立，因为$U_{I_0} = V_{I_0} = w$。对于（ii），注意到

$$\frac{1}{4} < \mathbb{E}\big[F(X,Y)\,\mathbf{1}_{\mathcal{A}\cap\mathcal{B}} \mid R = (I_0, w)\big] = \alpha\beta\,\mathbb{E}_{\widetilde{\mu}\times\widetilde{\nu}}\,F \le \min\Big\{\alpha,\ \beta,\ \mathbb{E}_{\widetilde{\mu}\times\widetilde{\nu}}\,F\Big\},$$

这也给出了 $\alpha,\ \beta > 1/4$。

对于 (iii)，令 p 为在 $R = (I_0, w)$ 条件下的 $X_{J\setminus I_0}$ 分布。于是每个 $z \in \{\pm 1\}^{J\setminus I_0}$ 都满足

$$\widetilde{\mu}_{J\setminus I_0}(z) \le \frac{p(z)\,\mathbf{1}_{\{p(z)\le M_{I_0}\}}}{\alpha} < 4M_{I_0} = 2^{-(m-|I_0|)+D+7},$$

$\widetilde{\nu}$也同样成立。 □

引理 4.2（恢复期望）。令 ξ 为任意满足 $\mathbb{E}_\xi\,F \ge 1/4$的分布。则存在一个矩形 $A \times B$，使得

$$\xi(A\times B) \ge \frac{2^{-d}}{5} \qquad \textit{and} \qquad \mathbb{E}_\xi[F \mid A\times B] \ge \frac{1}{2}.$$

证明。固定一个关于 F、成本为 d 且错误率至多为 $1/100$ 的协议的随机性，从而得到一个在 ξ下错误概率至多为 $1/100$的确定性协议。令 $\mathcal{L}$ 为其输出 $\mathbf{1}$ 的叶节点集合；回顾每个叶节点都是一个矩形 $A \times B$。不妨设对所有 $L \in \mathcal{L}$ 都有 $\mathbb{E}_\xi[F \mid L] \ge 1/2$（否则我们可以翻转该协议在 L 上的输出，同时降低错误概率）。由于错误概率至多为$1/100$ 且 $\mathbb{E}_\xi\,F \ge 1/4$，我们有 $\sum_{L\in\mathcal{L}} \xi(L) \ge \frac{1}{4} - \frac{1}{100} \ge \frac{1}{5}$。由于 $|\mathcal{L}| \le 2^d$，存在一个叶节点 L，使得 $\xi(L) \ge 2^{-d}/5$。

□

备注 4.3. 应用引理 4.2，使用乘积分布 $\mu\times\nu$ 并对 $A\times B$ 进行条件化，得到乘积分布 $\mu(\,\cdot \mid A) \times \nu(\,\cdot \mid B)$。由于 $\mu(A)\ \nu(B) \ge 2^{-d}/5$，条件化会使每个概率至多乘以 $5\cdot 2^d$。因此，对于每个 $J \subseteq [n]$，在 J上的边缘分布的亏损至多增加 $d + \log 5 < d + 3$。

定理 定理 *1.1*的证明。 如概述中所示，假设 $\mathbb{E}_{\mathcal{U}_n\times\mathcal{U}_n} F \ge 1/2$。注意 $d \ge 1$，并反设 $d \le c_\gamma n^{2/3}$ 对于一个充分小的常数 $c_\gamma > 0$。

对于每一步 $t \in \{0,\dots,T\}$（其中 T将在下文确定），我们维护一个乘积分布 $\mu_t \times \nu_t$，满足 $\mathbb{E}_{\mu_t\times\nu_t} F \ge 1/2$；一个包含植入坐标 $K_t \subseteq [n]$ 的集合及其参考串 $a_t \in \{\pm 1\}^{K_t}$；以及数值 Δ_t和 D_t，使得支撑集中的每个 $(x,\ y)$ 都满足 $\mathrm{dist}(x_{K_t},\ a_t)$，$\mathrm{dist}(y_{K_t},\ a_t) \le \Delta_t$，并且 μ_t和 ν_t在 $J_t := [n]\setminus K_t$ 上的边缘分布的亏损至多为 D_t。记 $m_t := |J_t|$。初始时，$K_0 = \varnothing$、$\mu_0 = \nu_0 = \mathcal{U}_n$以及 $\Delta_0 = D_0 = 0$。

只要 $m_t > \gamma n/2$，我们就应用 引理 4.1，令 $J = J_t$，随后应用 引理 4.2，并将新的块 I 添加到 K_t 中，从而得到 K_{t+1}；同时，我们将 a_t 扩展 w，从而得到 a_{t+1}。由于两个引理都只会缩小支撑集，K_t 上的距离界得以保持；根据 引理 4.1 （iii）和注 4.3，

$$m_{t+1} \le (1-\rho/2)\, m_t, \qquad D_{t+1} \le D_t + d + 10, \qquad \Delta_{t+1} \le \Delta_t + 20\rho\sqrt{m_t D_t}.$$

因此，该过程会在 $T \le \lceil (2/\rho)\ln(2/\gamma)\rceil = O_\gamma(1/\rho)$ 轮次后停止。此外，$D_t \le t(d+10)$ 且 $m_t \le n$，所以

$$\Delta_T \le 20\rho\sqrt{n(d+10)}\,T^{3/2} = O_\gamma\Big(\sqrt{n(d+10)/\rho}\Big) = O_\gamma\big(\sqrt{n}\,d^{3/4}\big).$$

由于 $\rho m_t > c_0\gamma n/(2\sqrt{d}) \ge c_0\gamma n^{2/3}/(2\sqrt{c_\gamma})$在停止之前，条件 $\rho m_t \ge 100$ 满足引理 4.1对于充分大的 n 成立。

通过选择 c_γ 充分小，我们得到 $\Delta_T \le \gamma n/4$。因此，最终支撑集中的每一对 (x,y) 都满足

$$\mathrm{dist}(x,y) \le 2\Delta_T + m_T \le \gamma n,$$

从而迫使 $F(x,y) = 0$。这与 $\mathbb{E}_{\mu_T\times\nu_T} F \ge 1/2$ 矛盾。因此，$\mathrm{R}^{1/100}(F) = \Omega_\gamma(n^{2/3})$。 □

致谢

H.H. 谨向 Amir Yehudayoff 和 Arkadev Chattopadhyay 致谢，感谢他们早期就解决这一问题的尝试所进行的讨论。

M.G. 和 W.Y. 获得瑞士教育、研究与创新国务秘书处（SERI）根据合同编号 MB22.00026 提供的资助。H.H. 和 N.H. 获得 NSERC 探索基金资助。Y.D. 和 S.L. 获得西蒙斯研究员奖 #929894 的资助。S.L. 还获得美国国家科学基金会奖项 CCF-2425349 的资助。

参考文献

[BHH+26]Ari Blondal、Hamed Hatami、Pooya Hatami、Chavdar Lalov 和 Sivan Tretiak。Borsuk–Ulam 与大间隔半空间的可复现学习。载于 第*58*届*ACM*计算理论年会论文集 *(STOC)*，第529–540页，2026年。 doi:10.1145/3798129.3800771。

[DPV18] Peter Dixon、A. Pavan 和 N. V. Vinodchandran。论伪确定性近似算法。载于第43届国际数学计算机科学基础研讨会 (MFCS 2018)，LIPIcs第117卷，第61:1–61:11页，2018年。doi:10.4230/LIPIcs.MFCS.2018.61.[Gav25] Dmytro Gavinsky。无歧义奇偶查询复杂度。《随机结构与算法》，66(3):e70010，2025年。doi:10.1002/rsa.70010.[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 TR11-136，计算复杂性电子讨论会，2011年。URL: https://eccc.weizmann.ac.il/report/ 2011/136/.[GG17]Shafi Goldwasser 和 Ofer Grossman。伪确定性 NC 中的二分图完美匹配。载于第44届国际自动机、语言与程序设计学术讨论会 (ICALP 2017)，LIPIcs第80卷，第87:1–87:13页，2017年。doi: 10.4230/LIPIcs.ICALP.2017.87.[GGH18] Shafi Goldwasser、Ofer Grossman 和 Dhiraj Holden。伪确定性证明。载于第9届理论计算机科学创新会议 (ITCS 2018)，LIPIcs第94卷，第17:1–17:18页，2018年。doi:10.4230/LIPIcs.ITCS.2018.17.[GGMW20] Shafi Goldwasser、Ofer Grossman、Sidhanth Mohanty 和 David P. Woodruff。伪确定性流式处理。载于第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年。doi:10.4230/LIPIcs.ITCS. 2020.79.[GGR13] Oded Goldreich、Shafi Goldwasser 和 Dana Ron。论伪确定性算法的可能性与局限性。载于第4届理论计算机科学创新会议论文集 (ITCS 2013)，第127–138页，2013年。doi:10.1145/2422436. 2422453.[GHR+26] Mika G¨o¨os、Nathaniel Harms、Artur Riazanov、Anastasia Sofronova、Dmitry Sokolov 和 Weiqiang Yuan。伪确定性通信复杂度。载于第58届ACM计算理论年会论文集 (STOC)，第2030–2039页，2026年。doi:10.1145/3798129.3800907.[GIPS21] Shafi Goldwasser、Russell Impagliazzo、Toniann Pitassi 和 Rahul Santhanam。论 NP 搜索问题的伪确定性查询复杂度。载于第36届计算复杂性会议 (CCC)，LIPIcs第200卷，第36:1–36:22页，2021年。doi: 10.4230/LIPIcs.CCC.2021.36.[GL19] Ofer Grossman 和 Yang P. Liu。对数空间中的可复现性与伪确定性。载于第30届ACM-SIAM离散算法研讨会论文集 (SODA 2019)，第606–620页，2019年。doi:10.1137/1.9781611975482.38.[GSTW23]Uma Girish、Makrand Sinha、Avishay Tal 和 Kewen Wu。异或函数通信协议的傅里叶增长。载于第64届IEEE计算机科学基础研讨会论文集 (FOCS)，第721–732页，2023年。doi:10.1109/FOCS57990.2023.00047。

[HN12] Trinh Huynh 和 Jakob Nordstr¨om。论简洁证明的价值：将通信复杂度的困难性放大为证明复杂度中的时间-空间权衡。在第*44*届计算理论研讨会（*STOC*）论文集，第233–248页，美国纽约州纽约市，2012年。ACM。doi:10.1145/2213977.2214000。

[Hol17] Dhiraj Holden。《关于无条件次指数时间伪确定性算法用于 BPP 搜索问题的说明》，2017 年。网址：https://arxiv.org/abs/1707.05808，arXiv：1707.05808。

[ILPS22] Russell Impagliazzo、Rex Lei、Toniann Pitassi 和 Jessica Sorrell。《学习中的可复现性》。载于第 *54* 届 *ACM* 计算理论研讨会年度会议论文集（*STOC 2022*），第 818–831 页，2022 年。doi:10.1145/3519935.3519973。

[Mar86] Katalin Marton。《膨胀引理的一个简单证明》。*IEEE* 信息论汇刊，32(3)：445–446，1986 年。doi:10.1109/tit.1986.1057176。

[OS18] Igor Carboni Oliveira 和 Rahul Santhanam。《学习与近似的伪去随机化》。载于近似、随机化与组合优化：算法与技术（*APPROX/RANDOM 2018*），第 116 卷*LIPIcs*，第 55:1–55:19 页，2018 年。doi:10.4230/LIPIcs.APPROX-RANDOM.2018.55。

[沃特26] 托马斯·沃森。通信复杂性中的伪确定性与 MA $\neq \mathrm{NP}^{\mathrm{BPP}}$。技术报告，arXiv，2026年。doi:10.48550/ARXIV.2608.26425。

Yogesh Dahiya加利福尼亚大学圣迭戈分校，美国加利福尼亚州拉霍亚电子邮箱：ydahiya@ucsd.edu

Mika GöösEPFL，瑞士洛桑电子邮箱：mika.goos@epfl.ch

Nathaniel Harms不列颠哥伦比亚大学，加拿大温哥华电子邮箱：nharms@cs.ubc.ca

Hamed Hatami麦吉尔大学，蒙特利尔′，加拿大电子邮箱：hatami@cs.mcgill.ca

Pooya Hatami俄亥俄州立大学，美国俄亥俄州哥伦布电子邮箱：hatami.2@osu.edu

Shachar Lovett美国加利福尼亚大学圣迭戈分校，拉霍亚，加利福尼亚州，美国电子邮箱地址：slovett@ucsd.edu

Weiqiang YuanEPFL，洛桑，瑞士，以及哥本哈根大学，哥本哈根，丹麦电子邮箱地址：weiqiang.yuan@di.ku.dk