arXiv:2305.05590v1 [cs.GT] 9 May 2023

正则分布的描述复杂度

Renato Paes Leme
Google Research
renatoppl@google.com

Balasubramanian Sivan
Google Research
balusivan@google.com

Yifeng Teng
Google Research
yifengt@google.com

Pratik Worah
Google Research
pworah@google.com

摘要

经济学中，Myerson 分布正则性条件是一项标准假设。本文研究如何在较小的统计距离内描述正则分布的复杂度。我们的主要结果是，要在 ϵ L´evy 距离内描述支撑集为 $[0,1]$ 的正则分布，所需的比特数为 $\tilde{\Theta}(\epsilon^{-0.5})$，且这一数量既必要又充分。我们通过证明可以利用 $\tilde{O}(\epsilon^{-0.5})$ 次累积分布函数查询，近似学习该正则分布，得出这一结果。作为推论，我们证明，在 ϵ L´evy 距离内学习支撑集为 $[0,1]$ 的正则分布类，其定价查询复杂度为 $\tilde{\Theta}(\epsilon^{-2.5})$。要学习两个正则分布的混合分布，则需要 $\tilde{\Theta}(\epsilon^{-3})$ 次定价查询。

1 引言

Myerson 正则分布是指其累积分布函数为 F，且分位数空间中的收益曲线 $R(q) = q \cdot F^{-1}(1-q)$ 为凹函数的分布。对于可由概率密度函数表示的分布（即不存在点质量的分布），这等价于虚拟价值函数的单调不减性

$$\phi(v) = v - \frac{1-F(v)}{f(v)}. \tag{1}$$

Myerson 正则性（或简称正则性）是经济学中的一个标准条件，最初由 Myerson 在其关于最优拍卖的经典论文中提出 [Mye81]。此后，它在多种经济场景的设计与分析中发挥了根本作用，例如：双边交易 [MS83], 先验无关机制设计 [BK96, HR09, RTCY12, FHH13, SS13, AB20],基于样本的拍卖 [DRY10, CR14, FILS15, HMR15, MR15], 机制设计中的近似与收益管理 [CLMN14, PLPV16, AHN+19, GLMN21], ...

除了在经济学中被广泛应用之外，它们还涵盖了各种重要的概率分布：

- 具有对数凹概率密度函数的概率分布，即概率密度函数具有 $f(x) = \exp(-g(x))$的形式，其中 g是凸函数，例如均匀分布、指数分布和正态分布。
- 满足单调风险率条件（MHR）的分布。这一概念源自可靠性理论。如果一个随机变量表示某个故障事件发生前的等待时间（例如灯泡熄灭），那么 MHR 就意味着：在故障尚未发生的条件下，故障在任意时刻发生的概率会随时间弱递增。

- 等收益分布，即在支撑集的每个点定价都会带来相同的收益。例如，考虑支撑集为 $[1,\infty)$、累积分布函数为 $F(x) = 1 - 1/x$ 的分布。这类分布已知是极端正则分布，常用于推导收益管理中的下界。

- 机器学习中出现的某些分布，例如：设 x 是一个随机特征向量，取自 $\mathbb{R}^d$ 中的均匀分布或高斯分布，且该分布限制在一个 d 维凸集 K 中，并设 $w \in \mathbb{R}^d$ 是一个固定的权重向量 $w \in \mathbb{R}^d$。那么，点积 $\langle w, x\rangle$ 的分布由 Prekopa-Leidner 定理可知是正则的 [Pr´e73]。

描述正则分布 本文探讨描述一个正则分布需要多少比特的信息。我们得到了一个精确界，并据此推导出学习正则分布的新紧确保证。

由于正则分布是连续对象，且其支撑集可能无界，因此我们需要作出一些假设，以使问题定义明确。

- **有界支撑集**：我们假设该分布的支撑集位于 $[0,1]$ 中。若不作此假设，即使表示支撑在 $[n, n+1]$ 上的均匀分布这一子类，也需要无限多个比特。

- **L´evy 距离近似**我们允许表示中的误差为 ϵ，并以 L´evy 距离衡量；这允许价值和概率都存在 ϵ 误差。形式化地说：

$$\text{Lévy}(F, G) := \inf\{\epsilon \text{ s.t. } F(v-\epsilon) - \epsilon \le G(v) \le F(v+\epsilon) + \epsilon, \forall v\} \tag{2}$$

现在我们可以正式陈述主要问题：

定义 1. 如果存在一个满足 $|\mathcal{C}'| \le 2^b$ 的概率分布类 $\mathcal{C}'$，使得对于每个概率分布 $F \in \mathcal{C}$，都存在 $F' \in \mathcal{C}'$，满足 $L\acute{e}vy(F, F') \le \epsilon$，那么我们称可以用 b 比特和 ϵ 误差描述一个概率分布类 $\mathcal{C}$。

我们的主要定理是：

定理1.1. 可以用 $\tilde{O}(1/\epsilon^{0.5})$ 比特，在 ϵ *L´evy* 距离内描述取值有界于 $[0,1]$ 的正则分布类。此外，该结果在多对数因子范围内是紧的。

将其与支撑集为 $[0,1]$ 的一般分布类别进行对比很有帮助。对于这类分布，需要 $\tilde{\Omega}(1/\epsilon)$ 比特，而 $O(1/\epsilon)$ 比特就足够了。对于充分性部分，我们可以用以下数值表示累积分布函数（CDF）F：$\lfloor F(k\epsilon)/\epsilon \rfloor$，其中 $k = 0, 1, \ldots, \lceil 1/\epsilon \rceil$。给定这些数值后，我们可以构造一个累积分布函数如下的分布：

$$\hat{F}(x) = \epsilon \lfloor F(k\epsilon)/\epsilon \rfloor \text{ for } x \in [k\epsilon, (1+k)\epsilon)$$

该分布在 L´evy 距离下与 F 相差不超过 ϵ。为了证明至少需要 $\Omega(1/\epsilon)$ 比特，只需构造一个包含 $2^{\Omega(1/\epsilon)}$ 个分布的集合，使得其中任意一对在 L´evy 距离下的差异至少为 ϵ。给定表示 $k = 1/(2\epsilon)$ 的比特 $b_0, \ldots, b_k \in \{0,1\}$，构造一个满足以下条件的分布：

$$F_b(x) = 2(k + b_k)\epsilon \text{ for } x \in [2k\epsilon, 2(1+k)\epsilon), k = 0, 1, \ldots, 1/(2\epsilon)$$

对于正则分布，我们能够构造一种更简洁的表示，只需使用比特数的平方根。不过，我们需要一种更复杂的采样过程。

超越正则分布 虽然正则分布在拍卖理论中很常见，但实践中遇到的许多分布都不是正则的。为了将我们的结果推广到正则性之外，我们定义了非正则系数这一概念，用来衡量一个分布与正则分布有多接近：

$$\textsc{IrregCoeff}(F) = \inf\left\{\beta \geq 0 \text{ s.t. } qF^{-1}(1-q) + \beta\int_{1-q}^{1} F^{-1}(x)dx \text{ is concave}\right\} \tag{3}$$

0-非正则分布就是通常意义上的正则分布，而 ∞-非正则分布则是一般分布。 $\beta > 0$ 的 β-非正则类包含非正则分布，这些分布足够接近正则性，因而具有较低的描述复杂度：

定理 1.2. 可以使用 $\tilde{O}(1/\epsilon^{0.5})$ 比特，在 ϵ *L´evy* 距离内描述有界于 $[0,1]$ 的 $O(1)$-非正则分布类。

.....的概念 β-不规则性与 α-强正则性的概念相重合，其中 Cole 和 Rough- garden [CR14] 在符号为负时成立。该定义原本旨在当 α取正值时，在正则分布与 MHR 分布之间进行插值。当考虑 α取负值时，我们得到了一种衡量某个概率分布与正则性接近程度的方法。一个分布在 α的意义下是 [CR14]-强正则的，当且仅当：

$$f'(v)\cdot(1-F(v)) \geq -(2-\alpha)f^2(v)$$

因此，当它是（$-\beta$）-强正则时，它是 β-非正则的，即：

$$f'(v)\cdot(1-F(v)) \geq -(2+\beta)f^2(v) \tag{4}$$

应用：定价查询复杂度 我们的首个应用是确定在 L ϵ-误差下学习正则分布的定价查询复杂度´（以 [PLSTW23] 距离衡量）。定价查询复杂度模型由 提出，旨在学习这样一种经济场景：唯一可行的机制是公布价格。在这类场景中，我们只能观察向不同代理人报出的价格是否促成了交易。在拍卖形式为第一价格拍卖时，这一概念也很有用，因为我们无法获知真实出价，但仍能知道竞标者是否选择了高于保留价的出价。在这类场景中，我们利用之前各期观察到的二元交易/未交易结果，来优化未来拍卖中的价格。

在该模型中，学习算法可以通过定价查询与分布 F 交互：每次查询中，算法选择一个价格 p_t，并抽取一个样本 v_t ，该样本来自 F ，算法只能获知 $\{+1,-1\}$ 的符号 $v_t - p_t$。算法的目标是估计某类分布的若干参数，例如均值、中位数或垄断价格。Paes Leme、Sivan、Teng 和 Worah [PLSTW23] 为若干重要参数给出了匹配的上界和下界，但在估计学习正则分布的累积分布函数所需的定价查询复杂度方面仍存在差距。他们给出了一个 $\tilde{O}(1/\epsilon^3)$ 上界和一个 $\Omega(1/\epsilon^{2.5})$ 下界。

我们通过给出在多对数因子范围内匹配下界。该结果通过计算 $O(1/\epsilon^{0.5})$ 描述获得，并使用 $O(1/\epsilon^2)$ 次定价查询，利用 Chernoff界评估每个点：

定理 1.3. 在 $\tilde{O}(1/\epsilon^{2.5})$ 范围内，学习正则分布的累积分布函数时，定价查询复杂度存在一个上界，误差以 ϵ *L´evy* 距离衡量。除多对数因子外，此结果是紧的。

应用：混合分布 我们称累积分布函数为 F 的概率分布是k个概率分布 $F_1, \ldots, F_k$的混合分布，当且仅当存在权重 $w_1, \ldots, w_k \geq 0$满足 $\sum_i w_i = 1$，使得 $F(v) = \sum_i w_i F_i(v)\ \forall v$。在许多应用中，将概率分布表示为其他概率分布的混合分布很有用。例如，Sivan 和 Syrgkanis [SS13]设计了针对正则分布混合分布的拍卖，其表现取决于混合分布中概率分布的数量。我们可以进一步询问，表示一般概率分布需要多少个正则分布。描述复杂度上界自动蕴含以下推论：

推论 1.4. 存在一个不规则分布，无法在 *L´evy* 距离下由 $o(1/\epsilon^{0.5})$ 个正则分布的混合分布进行 ϵ-逼近。

这直接源于以下事实： k 个正则分布的混合分布可以用 $O(k/\epsilon^{0.5})$ 描述，而表示一般分布则需要 $\Omega(1/\epsilon)$ 比特。

在机器学习应用中，将概率分布表示为高斯混合分布也很有用。我们的结果还意味着，表示一个正则分布可能需要许多高斯分布，因为~一个高斯分布可以用 $O(1)$ 比特表示。

推论 1.5. 存在一个正则分布，无法通过由 $o(1/\epsilon^{0.5})$ 个高斯分布构成的混合分布，在 *L´evy* 距离下达到 ϵ 近似。

学习正则分布的混合分布 描述两个正则分布的混合分布需要 $\tilde{O}(1/\epsilon^{0.5})$ 比特，因为描述每个分布需要 $\tilde{O}(1/\epsilon^{0.5})$ 比特，另外还需要 $\tilde{O}(1)$ 比特来描述权重。据此，人们可能会猜测，学习混合分布的定价查询复杂度也是 $\tilde{O}(1/\epsilon^{2.5})$。我们以如下颇为令人惊讶的结果作为本文结尾：

定理 1.6。令 $\mathcal{C}$ 为一类支撑集位于 $[0,1]$ 、且可写成两个正则分布的混合分布的分布。若要以 *L´evy* 距离将该类中某个分布的累积分布函数估计到 ϵ 以内，则至少需要 $\Omega(1/\epsilon^3)$ 次定价查询。

为什么使用 Levy 距离? 衡量两个概率分布之间距离的方法有多种，例如总变差距离（TV）、Wasserstein 距离、柯尔莫哥洛夫距离和 L´evy 距离。我们注意到，L´evy 距离的界会自动推出 Wasserstein 距离的界。柯尔莫哥洛夫距离和 TV 距离是更强的概念，但使用定价查询不可能在其中任何一种距离下获得近似结果。详情请参见 [PLSTW23]。两个概率分布 F 和 G 之间的柯尔莫哥洛夫距离定义为：

$$\mathrm{Kol}(F,G) := \inf\{\epsilon \text{ s.t. } F(v) - \epsilon \leq G(v) \leq F(v) + \epsilon, \forall v\} \tag{5}$$

如果一个分布在 $[0,1]$, 中是一个确定性价值，那么要在柯尔莫哥洛夫距离下获得任何有意义的近似，就必须精确估计这个价值，而这无法通过定价查询实现。

研究 L´evy 距离的另一个原因是，文献中已在研究学习收益最优拍卖的样本复杂度时考虑过这种度量，相关示例可参见 Brustle 等人 [BCD20]和 Cherapanamjeri 等人 [CDIZ22]。更深入地理解在 L´evy 距离下估计分布的复杂度，有望改进拍卖学习复杂度方面的研究结果。

相关工作 我们的研究总体上属于算法经济学中关于样本复杂度的主题，旨在理解描述或优化某种经济情境所需的最少信息量。探索这一问题有多种方式。例如，Dhangwatnotai 等人 [DRY10]和 Fu 等人 [FHH13] 探讨能否仅使用分布的单个样本来优化拍卖，以及在多大程度上可以做到这一点。另一方面，我们也可以探讨优化拍卖所需的特定分布样本数量（Cole 和 Roughgarden[CR14] 以及 Morgenstern 和 Roughgarden [MR15]），或计算最优保留价格所需的样本数量（Huang 等人 [HMR15]）。

与我们论文最接近的是 Paes Leme 等人的论文 [PLSTW23]，该论文研究了一种受限查询模型。我们无法获取分布的样本，而只能报出一个价格，并观察从该分布中重新抽取的样本对应的价值是高于还是低于所报价格。这一设定的动机来自两个重要场景：（i）采用标价机制、我们只能观察购买或不购买决策的情形；（ii）非真实拍卖（如第一价格拍卖）中的学习问题，此时出价并非价值的无偏样本，但我们仍能观察竞标者是否决定报出高于保留价格的出价。在这种设定下，Paes Leme 等人 [PLSTW23] 给出了学习单调风险率分布、正则分布和一般分布的垄断价格所需样本数的紧确界，并证明，学习保留价格所需的样本数严格少于学习正则分布的整个累积分布函数所需的样本数。

然而，[PLSTW23]对于学习正则分布的累积分布函数所需的定价查询数量，仍留下了一个差距：他们证明了下界 $\Omega(1/\epsilon^{2.5})$，以及上界 $\tilde{O}(1/\epsilon^{3})$。我们通过给出一个定价查询复杂度为 $\tilde{O}(1/\epsilon^{2.5})$ 的算法，解决了这一问题。

我们的工作也与计算机科学中关于查询复杂度的研究方向相关，该方向关注完成某项任务所需的黑箱查询的最少次数。这种方法已被应用于学习理论、并行计算、量子计算、布尔函数分析和优化等领域。由于相关文献过于广泛，无法在此一一综述，我们请读者参阅 Kothari、Lee、Newman 和 Szegedy 撰写的书 [KL+23]。

2 正则分布的描述复杂度

正则分布的描述复杂度的一个 $\Omega(1/\epsilon^{0.5})$ 下界，隐含于 $1/\epsilon^{2.5}$中给出的定价查询复杂度的 $\Omega($ [PLSTW23]) 下界。我们甚至可以修改他们的例子，将结果加强到适用于具有非递减概率密度函数的概率分布类（该类是单调风险率分布和正则分布的子类）。

定理2.1。 存在一个由 $2^{\Omega(1/\epsilon^{0.5})}$ 个支撑在 $[0,1]$上的、具有非递减概率密度函数的概率分布组成的集合，使得对于任意给定的一对分布，其 *L´evy* 距离至少为 $\Omega(\epsilon)$。

证明。 设 $n = 1/(4\epsilon^{0.5})$，且 $b = (b_0, \ldots, b_{n-1}) \in \{0,1\}^n$ 是一个比特序列。现在定义一个概率密度函数 $f_b(x)$，使得对于 $v \in [4\sqrt{\epsilon}k, 4\sqrt{\epsilon}(k+1)]$，我们有 $f_b(v) = 2v$，如果 $b_k = 0$ 并且

$$f_b(v) = \begin{cases} 2v, & \text{if } v \in [v_0 - 2\sqrt{\epsilon}, v_0 - \sqrt{\epsilon}] \cup [v_0 + \sqrt{\epsilon}, v_0 + 2\sqrt{\epsilon}]; \\ 2v_0, & \text{if } v \in (v_0 - \sqrt{\epsilon}, v_0 + \sqrt{\epsilon}). \end{cases} \qquad \text{for } v_0 = (4k+2)\sqrt{\epsilon}$$

若 $b_k = 1$ （见图 1）。首先，我们注意到所有这类函数的概率密度函数都是单调非递减的。现在，如果两个函数在某一比特位上不同 b_t，那么通过比较它们在 $v_0 = (4k+2)\sqrt{\epsilon}$ 附近的累积分布函数，L´évy 距离必定至少为 $\Omega(\epsilon)$。特别地，设 b、$b' \in \{0,1\}^n$ 是仅在索引 k 处不同的两个序列，且 $b_k = 0$ 和 $b'_k = 1$。设 F、G 分别是由序列 b 和 b' 构造的分布的累积分布函数。那么，由于 $F(v_0) < G(v_0 - 0.1\epsilon) - 0.1\epsilon$，$F$ 和 G 之间的 L´évy 距离为 $\Omega(\epsilon)$。 □

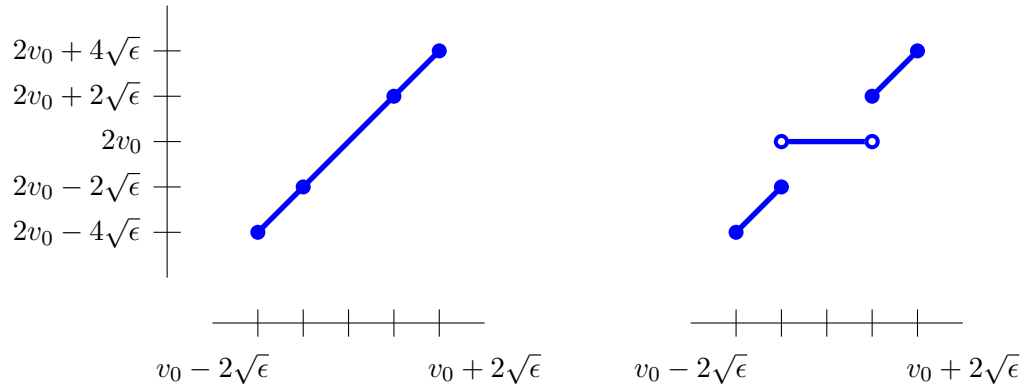

图 1：位于区域 $[v_0 - 2\sqrt{\epsilon}, v_0 + 2\sqrt{\epsilon}]$ 中的位 $b_k = 0$ 的概率密度函数 f_b（左）和 $b_k = 1$（右）。

上述定理表明，要在 ϵL´évy 距离内描述一个正则分布，至少需要 $\Omega(\epsilon^{-0.5})$ 比特。本节余下部分的重点是给出描述复杂度的匹配上界 $\tilde{O}(\epsilon^{-0.5})$。

2.1 学习平滑分布

我们首先假设分布是平滑的，也就是说，它没有点质量，并且由 C^1 类概率密度函数描述，即导数 $f'(v)$ 存在且连续。在此假设下，正则性可以写成以下条件：

$$f'(v) \cdot (1 - F(v)) \geq -2f^2(v) \tag{6}$$

这对应于虚拟价值函数的导数非负。我们先证明引理 2.2 ，给出学习平滑分布的充分条件。在第 2.2 节中，我们将介绍一种简化版算法，用于学习具有凸概率密度函数的分布（$f'(v) \geq 0$）。介绍完这一特殊情形下的主要思路后，我们将在第 2.3 节转而学习光滑正则分布。最后，在第 2.4 节中，我们通过保持正则性的磨光论证，去掉平滑性假设。

引理 2.2。 对于任意未知分布 F和 m 个点 $x_1 < x_2 < \cdots < x_m$，使得对于每个 i，

- 我们可以确定 $F_i \in [0, 1]$，使得 $F(x_i) \in [F_i - \epsilon, F_i + \epsilon]$，并且
- 或者是 $x_{i+1} - x_i \leq \epsilon$，或者存在 $f_i \leq f_i$使得 $f(x) \in [f_i, f_i]$对每个 $x \in [x_i, x_{i+1}]$都成立，并满足 $(x_{i+1} - x_i)(f_i - f_i) \leq \epsilon$。

^我们可以构造一个分布 F，它与 F 在 $[x_1, x_m]$ 上的距离小于 $O(\epsilon)$ L´evy 距离。

证明。 我们可以假设 $F_1 \leq F_2 \leq \cdots \leq F_m$。否则，我们可以将 F_i的价值按递增顺序排序，而引理中的性质仍将保持不变。要看出这一点，首先注意，如果$F_i > F_{i+1}$，那么 $F(x_i)$ 和 $F(x_{i+1})$ 都在范围 $[F_i - \epsilon, F_{i+1} + \epsilon]$ 内。这是因为 $F(x_i) \leq F(x_{i+1})$，并且 $F(x_i) \in [F_i - \epsilon, F_i + \epsilon]$，$F(x_{i+1}) \in [F_{i+1} - \epsilon, F_{i+1} + \epsilon]$。于是$F(x_i) \in [F_i - \epsilon, F_{i+1} + \epsilon] \subseteq [F_{i+1} - \epsilon, F_{i+1} + \epsilon]$，而 $F(x_{i+1}) \in [F_i - \epsilon, F_{i+1} + \epsilon] \subseteq [F_i - \epsilon, F_i + \epsilon]$。因此，如果我们交换 F_i 和 F_{i+1}，引理中的性质仍然成立。

ˆˆ 考虑以下分布 F：对于每个 x_i， $F(x_i) = F_i$ 以及

$$\hat{F}(x) = F_i + \frac{F_{i+1} - F_i}{x_{i+1} - x_i}(x - x_i) \text{ for } x \in (x_i, x_{i+1}).$$

换言之， F 由在若干点处估计 F，并用线性函数填充其间的曲线来定义。

现在我们证明，在 $[x_1, x_m]$上， F 与 F 的距离不超过 $O(\epsilon)$ Lˊévy 距离。我们通过证明这一点对每个区间 $[x_i, x_{i+1}]$都成立来完成证明。

若 $x_{i+1} - x_i \le \epsilon$，则 F 且 $\hat{F}$ 在 $[x_i, x_{i+1}]$ 上的 Lˊévy 距离不超过 ϵ，因为对任意 $x \in [x_i, x_{i+1}]$

$$F(x - \epsilon) - \epsilon \le F(x_i) - \epsilon \le F_i \le \hat{F}(x) \le F_{i+1} \le F(x_{i+1}) + \epsilon \le F(x + \epsilon) + \epsilon.$$

若 $x_{i+1} - x_i > \epsilon$，我们证明一个更强的结论：在 $[x_i, x_{i+1}]$上， F 与ˆF的柯尔莫哥洛夫距离不超过 $O(\epsilon)$，即对任意 $x \in [x_i, x_{i+1}]$， $|F(x) - F(x)| \le O(\epsilon)$。根据引理的陈述，这对每个 x_i都成立。现在我们证明，这对每个 $x \in (x_i, x_{i+1})$也成立。设ˆ$f_i = \frac{F_{i+1} - F_i}{x_{i+1} - x_i}$ 为 F 在(x_i, x_{i+1})上的概率密度函数。~~那么~~

$$\begin{aligned} |\hat{F}(x) - F(x)| &= \left|(F_i + (x - x_i)\hat{f}_i) - (F(x_i) + \int_{x_i}^{x} f(t)dt)\right| \\ &\le |F_i - F(x_i)| + \left|(x - x_i)\hat{f}_i - \int_{x_i}^{x} f(t)dt\right| \\ &\le \epsilon + \int_{x_i}^{x} |\hat{f}_i - f(t)|dt \le \epsilon + \int_{x_i}^{x_{i+1}} |\hat{f}_i - f(t)|dt. \end{aligned} \tag{7}$$

这里，第一行根据累积分布函数和概率密度函数的定义；第二行根据 $|a + b| \le |a| + |b|$ˆ 对任意 a， b 均成立；第三行根据 F_i 的定义。现在我们来界定 f_i。由于对每个 $x \in (x_i, x_{i+1})$，都有 $f(x) \le f_i$，因~~此~~

$$(x_{i+1} - x_i)\hat{f}_i - 2\epsilon = (F_{i+1} - \epsilon) - (F_i + \epsilon) \le F(x_{i+1}) - F(x_i) = \int_{x_i}^{x_{i+1}} f(t)dt \le (x_{i+1} - x_i)\overline{f}_i.$$

同理，由于对每个 $x \in (x_i, x_{i+1})$，都有 $f(x) \ge f_i$，因此

$$(x_{i+1} - x_i)\hat{f}_i + 2\epsilon = (F_{i+1} + \epsilon) - (F_i - \epsilon) \ge F(x_{i+1}) - F(x_i) = \int_{x_i}^{x_{i+1}} f(t)dt \ge (x_{i+1} - x_i)\underline{f}_i.$$

ˆ 结合以上两个不等式，我们可以将 f_i 界定为

$$\underline{f}_i - \frac{2\epsilon}{x_{i+1} - x_i} \le \hat{f}_i \le \overline{f}_i + \frac{2\epsilon}{x_{i+1} - x_i}.$$

因此，对于任意 $t \in (x_i, x_{i+1})$，当 $f(t) \in [f_i, f_i]$时，我们有 $|\hat{f}_i = f(t)| \le ~~f_i - f_i~~ + \frac{2\epsilon}{x_{i+1} - x_i}$。将此界应用于（7），可得

$$\begin{aligned} |\hat{F}(x) - F(x)| &\le \epsilon + \int_{x_i}^{x_{i+1}} |\hat{f}_i - f(t)|dt \\ &\le \epsilon + (x_{i+1} - x_i)\left(\overline{f}_i - \underline{f}_i + \frac{2\epsilon}{x_{i+1} - x_i}\right) \\ &= 3\epsilon + (x_{i+1} - x_i)(\overline{f}_i - \underline{f}_i) \le 4\epsilon. \end{aligned}$$

□

2.2 学习具有凸累积分布函数的平滑分布

本节中，我们假设给定一个平滑分布 F 以及一个预言机，它允许我们对其累积分布函数 $F(\cdot)$ 和概率密度函数 $f(\cdot)$ 的值进行采样。我们先讨论凸累积分布函数 ($f'(v) > 0$) 的特殊情况，以描述我们算法的简化版本。

为了获得直观的理解，我们可以根据引理 2.2进行如下"启发式"计算：如果 Δx 表示查询点之间的距离，而 Δf 表示这些点之间概率密度函数的变化量，我们的目标是得到 $\Delta x \cdot \Delta f \leq \epsilon$。近似 $\Delta f \approx f'(x) \cdot \Delta x$，可得$\Delta x \leq \sqrt{\epsilon / f'(x)}$。因此，我们将尝试以 $\sqrt{f'(x)/\epsilon}$ 的采样率选取点。

我们将在下面的证明中将此形式化，并界定为保证这一采样率所需的查询次数。请注意，我们的证明是算法式的，之后很容易将其转换为定价查询复杂度界。

定理 2.3. 设 F 是一个具有凸概率密度函数的平滑分布。那么， $\tilde{O}(\epsilon^{-0.5})$ 次预言机查询就足以在 ϵ *L ́evy* 距离内学习该分布，其中查询对象为 F 和 f 。

证明。 我们将证明，可以找到一组 $O(1/\epsilon^{0.5}$ 个点，满足引理 2.2 的条件。首先注意，对于凸累积分布函数，其概率密度函数 f 是单调的。

步骤 1 我们令 $K = \log(1/\epsilon)$，并使用二分搜索找出点 $x_0, x_1, \ldots, x_{2K+3}$，使得对于每个 $k = 0, \ldots, K+1$，都有 $|x_{2k+1} - x_{2k+2}| \leq \epsilon$；并且对于区间 $I_k = [x_{2k}, x_{2k+1}]$，有:

- $f(v) \leq 1$ 对于 $v \in I_0$
- $2^{k-1} \leq f(v) \leq 2^k$ 对于 $v \in I_k$, $k = 1, ..., K$
- $f(v) \geq 2^K$ 对于 $v \in I_{K+1}$

~ ~识别每个点需要 $O(\log(1/\epsilon)) = O(1)$ 次查询才能 $f(\cdot)$。鉴于我们有 $O(1)$ 个这样的点，目前为止我们总共使用了 $O(1)$。我们注意到，最后一个区间 I_{K+1} 已经满足引理 2.2的条件，因为其长度 $|I_{K+1}|$ 至多为 ϵ:

$$1 \geq \int_{v \in I_{K+1}} f(v)dv \geq 2^K |I_{K+1}| = \frac{1}{\epsilon}|I_{K+1}|.$$

步骤 2 对于每个满足 $k = 0$ 的区间 $I_k, ...,$ K，将 I_k 划分为长度为 $\frac{1}{2^k}\sqrt{\epsilon}$ 的区间 $I_{k,1}, I_{k,2}, \cdots$。每个区间 I_k 添加的端点数量至多为 $\frac{|I_k|}{\sqrt{\epsilon}/2^k} = 2^k \epsilon^{-1/2}|I_k|$，总数至多为

$$\begin{aligned}
\sum_{k=0}^{K} 2^k \epsilon^{-1/2}|I_k| &= \epsilon^{-1/2}|I_0| + \sum_{k=1}^{K} 2^k \epsilon^{-1/2}|I_k| \\
&\leq \epsilon^{-1/2} + \sum_{k=1}^{K} \epsilon^{-1/2} \int_{v \in I_k} 2f(v)dv \\
&\leq \epsilon^{-1/2} + \epsilon^{-1/2} \int_{v \in [0,1]} 2f(v)dv \\
&= \epsilon^{-1/2} + 2\epsilon^{-1/2} = 3\epsilon^{-1/2}.
\end{aligned}$$

因此，如果我们对前两个步骤中各区间的所有端点查询 f 和 F，所需查询总数至多为 $O(\epsilon^{-1/2})$。

步骤 3 对步骤 2 中划分得到的每个区间 $I_{k,j} = [v_{k,j}, \bar{v}_{k,j} + 2^{-k}\sqrt{\epsilon}] \subseteq I_k$，定义 $f_{k,j} = f(v_{k,j})$ 和 $f_{k,j} = f(v_{k,j} + 2^{-k}\sqrt{\epsilon})$。若 $f_{k,j} - f_{k,j} = t_{k,j}2^k\sqrt{\epsilon}$，则将区间划分为 $\tilde{t}_{k,j}$：$\max(1, \lceil t_{k,j} \rceil)$ 个长度为 $\tilde{t}_{k,j}^{-1}2^{-k}\sqrt{\epsilon}$ 的区间，并对新添加的端点查询 F。由于任意相邻点 v, v 的距离为 $v - v \le \tilde{t}_{k,j}^{-1}2^{-k}\sqrt{\epsilon}$，因此该区间满足引理 2.2 中的第二个条件：

$$(\bar{v} - \underline{v})(\bar{f} - \underline{f}) \le \frac{1}{\tilde{t}_{k,j}2^k}\sqrt{\epsilon} \cdot t_{k,j}2^k\sqrt{\epsilon} \le \epsilon$$

这使我们能够将分布学习到 ϵ L´evy 距离精度。

最后，我们只需界定这一步所需的查询次数。在长度为 $2^{-k}\sqrt{\epsilon}$ 的 $I_{k,j}$ 中，最多需要 $\lfloor t_{k,j} \rfloor$ 次查询。现在，若 $I_k = [x_{2k}, x_{2k+1}]$，则有：

$$2^{k-1} \ge f(x_{2k+1}) - f(x_{2k}) = \sum_j (f(v_{k,j+1}) - f(v_{k,j})) = \sum_j t_{k,j}2^k\sqrt{\epsilon},$$

这意味着 $\sum_j t_{k,j} \le 1/(2\sqrt{\epsilon})$。因此在这一步中，对所有 k，在 F 上的查询总次数为 $\tilde{O}(\epsilon^{-1/2})$。

□

2.3 学习光滑正则分布

与凸函数不同，正则分布的概率密度函数既可以增加也可以减少，因此我们无法像前一个引理那样轻易地划分区间。然而，正则性限制了概率密度函数的下降速率（公式（6））。我们的第一个引理表明，正则分布的概率密度函数不可能多次下降至原来的二分之一。我们将在下一个引理中将这一直观认识形式化：

引理 2.4。设 F 是概率密度函数为 f 的光滑正则分布。固定一个整数 ℓ。现在假设 S 是一组形如 $[v_i, v_i']$ 的互不相交区间，满足：

- 对于每个区间 $[v_i, v_i']$ 我们有 $2^{-\ell} \le 1 - F(v_i') \le 1 - F(v_i) < 2^{-\ell+1}$
- 对于每个 $v \in [v_i, v_i']$，我们有 $f(v_i') \le f(v) \le f(v_i)$，以及 $f(v_i') \le \frac{1}{2}f(v_i)$。

那么，区间的数量被一个常数 $|S| \le 32$ 所界定。

证明。我们将计算 S 中区间的最大数量。不失一般性，我们可以假设 $f(v_i') = \frac{1}{2}f(v_i)$，因为我们总可以取一个满足最后一个条件取等号的子区间。

考虑一个 int区间 $[v_i, v_i']$，属于集合 S，其中 $f(v_i) \in [2^{k-1}, 2^k)$。与任何 $v \in [v_i, v_i']$ 一样，

$$f'(v) \ge -\frac{2f^2(v)}{1 - F(v)} \ge -\frac{2f^2(v_i)}{1 - F(v_i')} \ge -\frac{2 \cdot (2^k)^2}{2^{-\ell}} = -2^{2k+\ell+1}. \tag{8}$$

这里，第一个不等式根据正则分布的定义得出；第二个和第三个不等式则是因为 f 被限制在 $f(v_i')$ 之间，因此 $f(v) \le f(v_i) \le 2^k$，且 $1 - F(v) \ge 1 - F(v_i) \ge 2^{-\ell}$。因此

$$-\frac{1}{2} \cdot 2^{k-1} \ge f(v_i') - f(v_i) = \int_{v_i}^{v_i'} f'(v)dv \ge -2^{2k+\ell+1}(v_i' - v_i). \tag{9}$$

这里，第一个不等式由 $f(v_i') = \frac{1}{2}f(v_i)$ 和 $f(v_i) \ge 2^{k-1}$ 得出；第二个不等式由不等式（8）$v_i' - v_i \ge \frac{1}{2}2^{k-1-(2k+\ell+1)} = 2^{-k-\ell-3}$）得出。根据不等式（9），可以给出该区间长度的下界。因此，该区间内累积分布函数的变化可以如下给出下界：

$$F(v_i') - F(v_i) = \int_{v_i}^{v_i'} f(v)dv \ge f(v_i')(v_i' - v_i) \ge 2^{k-2} \cdot 2^{-k-\ell-3} = 2^{-\ell-5}. \tag{10}$$

这里，第一个不等式由概率密度函数在该区间内单调不增时的 $f(v) \geq f(v_i)$ 得出；第二个不等式由 $f(v_i') = \frac{1}{2} f(v_i) \geq 2^{k-2}$ 和 $v_i' - v_i \geq 2^{-k-\ell-3}$ 得出。由于区间 $[v_i, v_i')$ 两两不相交， $F(v_i') - F(v_i)$ 的总和至多为 $2^{-\ell+1} - 2^{-\ell} = 2^{-\ell}$。因此，此类区间至多有 $2^5 = O(1)$ 个。 □

引理 2.4 表明，在分位数之比不超过 2 的区间内，概率密度函数最多只能下降 32 次，每次下降幅度为$\frac{1}{2}$ 倍。一个有用的推论是，在这种情况下，两个点之间的概率密度函数最多只能下降 $1 - 2^{-33}$ 倍。

推论 2.5. 设 F 是一个光滑正则分布， ℓ 是一个整数。对于任意两个满足 $2^{-\ell} \leq 1 - F(x') < 1 - F(x) \leq 2^{-\ell+1}$ 的点 $x < x'$，有 $f(x') > 2^{-33} f(x)$。

证明. 若 $f(x) \geq 2^{33} f(x')$，由于 f 是连续的， $[x, x')$ 可以划分为 33 个区间 $[x, v_1) = [v_0, v_1)$， $[v_1, v_2)$， $[v_2, v_3)$， $\cdots$， $[v_{31}, v_{32})$， $[v_{32}, x') = [v_{32}, v_{33})$，使得对于任意区间 $[v_i, v_{i+1})$，$f(v_{i+1}) \leq \frac{1}{2} f(v_i)$，这与引理 2.4 矛盾。 □

利用引理 2.4我们还可以得到学习算法的以下关键组成部分。

引理 2.6. 设 F 为光滑正则分布， ℓ 为整数， $\epsilon > 0$ 为参数。对于任意区间 $[x, x]$，其中 $2^{-\ell} \leq 1 - F(x) < 1 - F(x) \leq 2^{-\ell+1}$，使用 $\tilde{O}(1)$ 次查询 f，我们可以找到区间内的点序列 $x = x_1 \leq x_2 \leq \cdots \leq x_m = x$，使得：对于任意子区间$[x_i, x_{i+1})$，或者

- $x_{i+1} - x_i \leq \epsilon$;
- $f(x_{i+1}) < 2^{-33}$，这意味着对于任意点 $x \in [x_i, x_{i+1})$， $f(x) \leq 1$;
- 对于任意两点 $x, x' \in [x_i, x_{i+1})$，$2^{-36} f(x) < f(x') < 2^{36} f(x)$.

证明。 考虑以下二分搜索算法：对于任意区间 $[a, b]$（在输出序列中以 $[a, b] = [x, x - \epsilon]$ 开始且仅包含 x、$x - \epsilon$、 x ）， _ _ _

1. 如果 $b - a \leq \epsilon$，停止搜索 $[a, b]$；2. 如果 $f(b) < 4f(a)$，停止搜索 $[a, b]$；3. 如果 $f(b) < 2^{-33}$，停止搜索 $[a, b]$；4. 否则，将 $\frac{a+b}{2}$ 添加到输出序列中，然后同时搜索 $[a, \frac{a+b}{2}]$ 和 $[\frac{a+b}{2}, b]$。

— — —

我们证明，算法中的前三种情况都是有效的停止条件。

在情况 1 中，我们无需搜索区间 $[a, b]$，因为端点已经足够接近，满足引理所要求的性质。

在情况 2 中，考虑区间 $[a, b]$，其中 $f(b) \leq 4f(a)$，以及 $[a, b]$ 中的任意 $x < x'$。根据推论 2.5，$f(x') < 2^{36} f(x)$。如果 $f(x') \geq 2^{36} f(x)$，由于 $f(b) < 4f(a)$，则至少存在 33 个区间 $[a, v_1)$、$[v_1, v_2)$、$\cdots$、 $[v_{j-2}, v_{j-1})$、 $[v_{j-1}, x)$、 $[x', v_{j+1})$、 $[v_{j+1}, v_{j+2})$、 $\cdots$、 $[v_{31}, v_{32})$、 $[v_{32}, b)$，使得对于上述列出的任一区间，其两个端点之间的概率密度函数至少减少至原来的 1/2。这也与引理 2.4 相矛盾，因为在 $[x, x]$ 中，概率密度函数最多只能减半 32 次。同理， $f(x') > 2^{-36} f(x)$，否则在 x 和 x' 之间至少会有 33 个区间，其中概率密度函数减半。

在情况 3 中，如果 $f(b) < 2^{-33}$，根据推论 2.5 不存在任何 $x \in [a, b)$ 使得$f(x) > 1 > 2^{33} f(b)$。因此对于任意 $x \in [a, b)$， $f(x) \leq 1$。

因此，算法的前 3 种情况都是有效的停止条件。现在我们分析有多少个点被添加到序列中。首先注意，对于任意 $v \le x - \epsilon$, $f(v) < \frac{2^{33}}{\epsilon}$：否则，若$f(v) \ge \frac{2^{33}}{\epsilon}$，由于 $f(x) > 2^{-33}f(v) \ge \frac{1}{\epsilon}$对每个$x \in [v, v+\epsilon]$都成立，根据推论 2.5，我们有 $F(v+\epsilon)-F(v) = \int_v^{v+\epsilon} f(t)dt > \frac{1}{\epsilon}((v+\epsilon) - v) = 1$，这是不可能的。

考虑以下集合 S，其中包含所有距离叶节点一层的已搜索区间 $[a_i, b_i]$：换句话说，算法搜索 $[a_i, \frac{a_i+b_i}{2}]$ 和 $[\frac{a_i+b_i}{2}, b]$，但不会进一步深入搜索任一区间。若按任一端点的递增顺序列出 S 中的所有区间$[a_1, b_1]$、$[a_2, b_2]$、$\cdots$、$[a_s, b_s]$，则 f 从 a_i 到 b_i 增加了 4 倍，但从 b_i 到 a_{i+1} 至少减半的情况至多发生 32 次，这是引理 2.4 的结论。因此，除至多 32 次外，对于所有情况都有 $f(b_{i+1}) > 2f(b_i)$。将 S 划分为至多 32 个区间子序列 $S_d = ([a_{d+1}, b_{d+1}], [a_{d+2}, b_{d+2}], \cdots, [a_{d+j}, b_{d+j}])$，使得对每个 $i \in [j-1]$，都有 $f(b_{d+i+1}) > 2f(b_{d+i})$。由于对于每个$[a_i, b_i] \in S$，$2^{-33} \le f(b_i) < \frac{2^{33}}{\epsilon}$，而 S_d 至多包含 $O(\log \frac{1}{\epsilon}) = \tilde{O}(1)$ 个区间，因此 S 也包含 $O(1)$ 个区间。

请注意，搜索树中的每个区间都满足以下 4 个条件之一：

1. 区间属于 S;

2. 区间是 S 中某个区间的子节点;

3. 在搜索树中，该区间是 S 中某个区间的祖先;

4. 在搜索树中，该区间是 S 中某个区间的祖先的子节点.

由于每个被搜索区间的长度至少为 ϵ，搜索树的深度为 $O(\log \frac{1}{\epsilon}) = \tilde{O}(1)$。因此，前一段中定义的每个区间 $[a_i, b_i] \in S$ 在搜索树中都有 $O(1)$ 个祖先。因此，S 中的每个区间最多可以定义上述 4 个条件中的 $O(1)$ 个区间，这意味着搜索树中有 $O(|S|) = O(1)$ 个区间。由此可知，所提出的算法使用 $O(1)$ 次查询 f，以输出一个满足引理陈述中条件的点序列。

□

现在，我们可以将针对概率密度函数单调的分布的算法推广到任意正则分布。该算法的结构几乎相同，但在查询复杂度分析中需要更加谨慎。

定理 2.7. 设 F 为光滑正则分布。那么，$\tilde{O}(\epsilon^{-0.5})$ 次对 F 和 f的预言机查询，足以在 ϵ *L´evy* 距离意义下学习该分布。

证明. 设 $L = \log \frac{1}{\epsilon}$。算法步骤如下。我们先设置“步骤 0”，以保持步骤 1、2 和 3 之间的对应关系。

步骤 0 使用二分搜索找出 $2(L+1)$ 个点 z_i，使得对于区间 $I^{(\ell)} = [z_{2\ell-1}, z_{2\ell}]$，累积分布函数满足 $1-2^{-\ell+1} \le F(v) < 1-2^{-\ell}$，且 $I^{(L+1)}$ 包含所有满足 $1-F(v) \ge 1-2^{-L} = 1-\epsilon$ 的值。并且 $z_{2\ell+1} - z_{2\ell} \le \epsilon$。对于每个区间，二分搜索需要 $O(\log \frac{1}{\epsilon}) = \tilde{O}(1)$ 次查询，因此，由于总共只有 $O(1)$ 个区间，所有区间的查询总数为 $O(1)$。我们将针对每个区间运行一个与概率密度函数单调情形类似的算法，并证明，对 F 和 f 进行 $\tilde{O}(\epsilon^{-0.5})$ 次查询，就足以学习每个 $I^{(\ell)}$ 的累积分布函数，因此总共只需要 $\tilde{O}(\epsilon^{-0.5})$ 次查询。

步骤 1 以下所有操作均应用于固定 ℓ 的区间 $I^{(\ell)}$。我们只关注 $I^{(\ell)}$ 且满足$\ell \le L$的情形，因为无需以 Levy 距离学习满足 ϵ 的 $I^{(L+1)}$。若无歧义，我们会省略上标 ℓ。

对于区间 $I^{(\ell)}=[x,\bar{x}]$，根据引理 2.6，我们可以在该区间中找到一个点序列 $x_1=x\le x_2\le\cdots\le x_m=\bar{x}$，使得：对于任意子区间 $[x_i,x_{i+1})$，以下情况之一成立：(a) $x_{i+1}-x_i\le\epsilon$；或 (b) $f(x_{i+1})<2^{-33}$，这意味着对于任意点 $x\in[x_i,x_{i+1})$，$f(x)\le 1$；或 (c) 对于任意两个点 x、$x'\in[x_i,x_{i+1})$，$2^{-36}f(x)<f(x')<2^{36}f(x)$。

按如下方式为某个 $K=O(1)$ 构造区间集合 I_0、I_1、$\cdots$、I_K、I_{K+1}。对于满足 (a) 的所有子区间 $[x_i,x_{i+1})$，将它们归入区间集合 I_{K+1}；对于满足 (b) 的所有子区间，将它们归入区间集合 I_0；对于满足 (c) 的所有子区间，如果 $2^{k-1}\le 2^{33}f(x_{i+1})<2^k$，则将 $[x_i,x_{i+1})$ 加入区间集合 I_k。不同于对单调非递减 f 的分析，对于 I_k 中的任意区间 $[x_i,x_{i+1})$，点 $x\in[x_i,x_{i+1})$ 满足 $f(x)<2^{36}f(x_{i+1})<2^{k+3}$，并且 $f(x)>2^{-36}f(x_{i+1})\ge 2^{k-70}$。由于这些集合中的区间由引理 2.6 构造而成，而该引理只使用 $O(1)$ 次查询，因此此步骤中对 f 的查询总次数为 $O(1)$。

步骤 2 对于每个包含 I_k 且具有 $0\le k\le K$ 的区间集合，将 I_k 中的所有点划分到长度为 $I_{k,1},I_{k,2},\cdots$ 的区间中。记 I_k 中所有区间的总长度为 $|I_k|$。对于每个区间集合 I_k，新增的端点数至多为 $\frac{|I_k|}{\sqrt{\epsilon}/2^{k+\ell}}=2^{k+\ell}\epsilon^{-1/2}|I_k|$，总计至多为

$$\begin{aligned}\sum_{k=0}^{K}2^k\epsilon^{-1/2}|I_k| &= \epsilon^{-1/2}|I_0|+\sum_{k=1}^{K}2^{k+\ell}\epsilon^{-1/2}|I_k|\\ &\le \epsilon^{-1/2}+\sum_{k=1}^{K}\epsilon^{-1/2}\int_{v\in I_k}2^{70+\ell}f(v)dv\\ &\le \epsilon^{-1/2}+2^{70+\ell}\epsilon^{-1/2}\int_{v\in I^{(\ell)}}f(v)dv\\ &= \epsilon^{-1/2}+2^{70+\ell}\epsilon^{-1/2}\cdot 2^{-\ell}=O(\epsilon^{-1/2}).\end{aligned}$$

因此，如果我们针对该步骤所得区间的所有端点查询 f 和 F，则所需查询总数至多为 $O(\epsilon^{-1/2})$。

步骤 3 对于步骤 2 中划分得到的每个区间 $I_{k,j}=[v_{k,j},v_{k,j}+2^{-k-\ell}\sqrt{\epsilon}]\subseteq I_k$，如果 $f(v_{k,j}+2^{-k-\ell}\sqrt{\epsilon})-f(v_{k,j})=t_{k,j}\cdot 2^{k+\ell}\sqrt{\epsilon}$，我们进一步将该区间划分为 $\tilde{t}_{k,j}$ 个长度为 $\frac{1}{\tilde{t}_{k,j}2^{k+\ell}}\sqrt{\epsilon}$ 的区间 $[v_{k,j,0},v_{k,j,1}]$、$[v_{k,j,1},v_{k,j,2}]$、$\cdots$、$[v_{k,j,\tilde{t}_{k,j}-1},v_{k,j,\tilde{t}_{k,j}}]$，其中 $v_{k,j,0}=v_{k,j}$ 以及 $v_{k,j,\tilde{t}_{k,j}}=v_{k,j+1}$，并对新增端点查询 F。根据正则分布的定义，对于任意 $x\in[v_{k,j},v_{k,j+1}]$，

$$f'(x)\ge-\frac{2f^2(x)}{1-F(x)}>\frac{2\cdot(2^{k+3})^2}{2^{-\ell}}=-2^{\ell+2k+7}$$

由 $f(x)<2^{k+3}$ 和 $F(x)<1-2^{\ell}$。因此，我们可以对 $x\in[v_{k,j},v_{k,j+1}]$ 的 $f(x)$ 作如下界定：

$$\begin{aligned}f(x) &\le f(v_{k,j+1})-(-2^{\ell+2k+7})\cdot(v_{k,j+1}-v_{k,j})\\ &< f(v_{k,j})+t_{k,j}\cdot 2^{k+\ell}\sqrt{\epsilon}+2^{\ell+2k+7}2^{-k-\ell}\sqrt{\epsilon}\\ &= f(v_{k,j})+t_{k,j}\cdot 2^{k+\ell}\sqrt{\epsilon}+2^{k+7}\sqrt{\epsilon}\stackrel{def}{=}\overline{f},\end{aligned}$$

and

$$\begin{aligned} f(x) &\geq f(v_{k,j}) + (-2^{\ell+2k+7}) \cdot (v_{k,j+1} - v_{k,j}) \\ &> f(v_{k,j}) - 2^{\ell+2k+7} 2^{-k-\ell} \sqrt{\epsilon} \\ &= f(v_{k,j}) - 2^{k+7} \sqrt{\epsilon} \stackrel{def}{=} \underline{f}, \end{aligned}$$

因此，对于划分中任意一对相邻点 $v_{k,j,a}, v_{k,j,a+1}$，其距离为 $d = \tilde{t}_{k,j} 2^{k+\ell} \sqrt{\epsilon}$，若 $x \in [v_{k,j,a}, v_{k,j,a+1}]$，则有 $f(x) \in (\underline{f}, \overline{f})$，其中 $\overline{f} - \underline{f} < t_{k,j} \cdot 2^{k+\ell} \sqrt{\epsilon} + 2^{k+8} \sqrt{\epsilon}$，并且

$$\begin{aligned} (v_{k,j,a+1} - v_{k,j,a})(\overline{f} - \underline{f}) &< \frac{1}{\tilde{t}_{k,j} 2^{k+\ell}} \sqrt{\epsilon} (t_{k,j} \cdot 2^{k+\ell} \sqrt{\epsilon} + 2^{k+8} \sqrt{\epsilon}) \\ &= \frac{t_{k,j}}{\tilde{t}_{k,j}} \epsilon + 2^{8-\ell} \epsilon = O(\epsilon). \end{aligned}$$

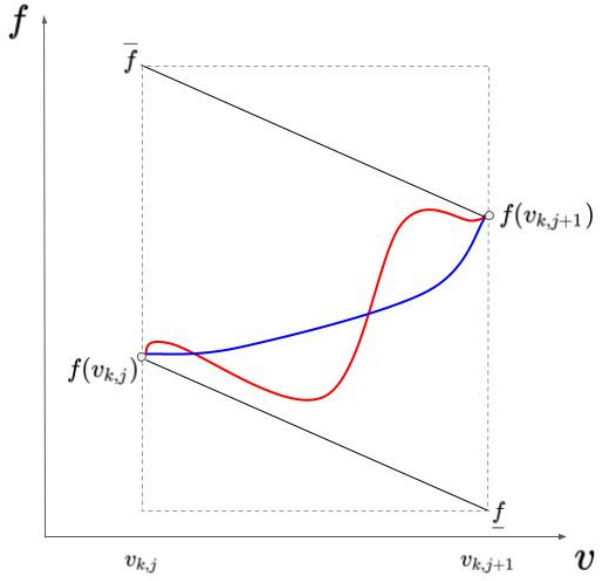

图2：展示在区间 $[v_{k,j}, v_{k,j+1}]$ with $t_{k,\,j} \leq 1$ 上的分析。根据正则分布的定义，在该范围内，两条黑色实线的斜率为 $-2^{\ell+2k+7}$，这是 $f'(v)$ 的下界。给定 $f(v_{k,\,j})$ 和 $f(v_{k,\,j+1})$，正则分布的概率密度函数 f 不会超出这两条黑线。由于任意分布的累积分布函数等于概率密度曲线下方的面积，给定 $v_{k,\,j}$ 和 $v_{k,\,j+1}$ 处的累积分布函数和概率密度函数后，任意两个正则分布的累积分布函数之差（例如红色曲线和蓝色曲线）都小于虚线矩形的面积，即 $(v_{k,\,j+1} - v_{k,\,j})(\overline{f} - \underline{f})$。

图 2 给出了该分析的示例。根据引理 2.2，我们可以学习一个与 F 的 ϵ L´evy 距离不超过 $k \in [0, K]$ 的分布，适用于 $\bigcup I_{k,\,j}$。由于 $I_{K+1,\,j}$ 中所有区间的长度都至多为 ϵ，因此无需学习这些区间即可满足 ϵ L´evy 距离要求。

现在我们分析此步骤中还需要多少个额外点（即对 F 的查询）。对于每个区间 $I_{k,j}$，需要 $\tilde{t}_{k,\,j} - 1 = \max(1, \lceil t_{k,j} \rceil) - 1 \leq \max(0, t_{k,j})$ 个额外查询。这意味着，如果 f 在 $I_{i,j}$ 上递增，即 $f(v_{k,j+1}) > f(v_{k,j})$，那么至多还需要 $t_{k,j} = \frac{f(v_{k,j+1}) - f(v_{k,j})}{2^{k+\ell} \sqrt{\epsilon}}$ 个额外查询；否则，~~当 f 在 $I_{i,j}$ 上~~不递增时，则不需要额外查询。

令 $I_{k,\uparrow} = \{x\colon\ f'(x) \geq 0, 2^{k-70} < f(x) < 2^{k+3}\}$ 和 $I_{k,\downarrow} = \{x\colon\ f'(x) < 0, 2^{k-70} < f(x) < 2^{k+3}\}$ 为价值集合，其中 $f(x)$ 的取值范围有界于 $(2^{k-70}, 2^{k+3})$，而 f 分别为非递减和递减。注意，I_k 是 $I_{k,\uparrow} \cup I_{k,\downarrow}$ 的子集。因此，我们可以对 I_k 中额外查询的总数给出上界：

$$\sum_j \max(0, t_{k,j}) \leq \sum_j \frac{f(v_{k,j+1}) - f(v_{k,j})}{2^{k+\ell}\sqrt{\epsilon}} \leq \frac{1}{2^{k+\ell}\sqrt{\epsilon}} \int_{I_{k,\uparrow}} f'(x)dx. \tag{11}$$

由于在 $I_{k,\uparrow} \cup I_{k,\downarrow}$，$f(x)$ 最多可增加到 2^{k+3}，因此我们有 $\int_{I_{k,\uparrow}} f'(x)dx \leq 2^{k+3} - \int_{I_{k,\downarrow}} f'(x)dx$。注意，对于任意满足 $2^{k-70} < f(x) < 2^{k+3}$ 的 x，当 $f'(x) < 0$ 时，$f'(x) > -\frac{2f^2(x)}{1-F(x)} > -2^{2k+\ell+7}$。此外，根据

$$2^{k-70}|I_{k,\downarrow}| < \int_{I_{k,\downarrow}} f(x)dx \leq 2^{-\ell}$$

我们有 $|I_{k,\downarrow}| < 2^{70-k-\ell}$。因此，

$$\begin{aligned}
\int_{I_{k,\uparrow}} f'(x)dx &\leq 2^{k+3} - \int_{I_{k,\downarrow}} f'(x)dx < 2^{k+3} + \int_{I_{k,\downarrow}} 2^{2k+\ell+7}dx \\
&< 2^{k+3} + 2^{2k+\ell+7}|I_{k,\downarrow}| < 2^{k+3} + 2^{2k+\ell+7} \cdot 2^{70-k-\ell} \\
&< 2^{k+3} + 2^{k+77} < 2^{k+78}.
\end{aligned} \tag{12}$$

结合（11）和（12），可知在 I_k 中的额外查询总数至多为

$$\frac{1}{2^{k+\ell}\sqrt{\epsilon}} \int_{I_{k,\uparrow}} f'(x) < \frac{1}{2^{k+\ell}\sqrt{\epsilon}} \cdot 2^{k+78} = 2^{78-\ell}\sqrt{\epsilon}.$$

因此，对于每个 k，步骤 3 中的查询数量为 $O(\epsilon^{-1/2})$，这意味着步骤 3 中的查询总数为 $\tilde{O}(\epsilon^{-1/2})$，因为 k 有 $\tilde{O}(1)$ 个不同的值。

□

关于 β-非正则概率分布的讨论。请注意，方程（6）中的因子 2 并无特殊之处。通过使用不同的非正则系数重复证明，我们可以得到定理1.2的证明。

2.4 一般（非光滑）正则分布的描述复杂度

当正则分布不具有光滑的累积分布函数时，概率密度函数可能不存在，而我们使用概率密度函数的预言机查询来学习累积分布函数分布的算法可能不适用于一般正则分布。然而，我们在以下引理中证明，任意˜正则分布 F 都可以在任意小的 L´evy-距离 δ下由光滑正则分布 F 一致逼近。因此，要在 ϵ L´evy-距离内描述 F，只需在 ϵ L´evy-距离内描述˜F，这可以通过前几小节中的算法实现。

上文将介绍实分析中广为人知的光滑化过程的一种变体：它用一个高度集中的 C^∞-函数与原函数的卷积来替换该函数。唯一需要特别说明的是，我们将论证如何以一种保持正则性的方式进行光滑化。

引理 2.8。 给定任意正则分布 F，对于任意 $\delta > 0$，都存在另一个正则分布 ˜F，使得 $L\acute{}evy(F, \tilde{F}) \leq \delta$，且 $\tilde{F}$没有点质量，并且具有一个累积分布函数。

证明。令 $R(q) = qF^{-1}(1-q)$ 为与 F 关联的收益曲线。由于 F 是正则的，曲线 R 是凹的，但它可能不是 C^∞。有多种方法可以用 C^∞ 凹函数一致逼近凹函数，其中一种是与非负光滑核函数卷积（例如参见 [AM15] ）。对于我们的目的，只需存在一个函数 C^∞-凹的 $\hat{R}(x)$，满足 $\hat{R}(0) = 0$ 且 $|R(x) - \hat{R}(x)| \le \delta^2/2, \forall x \in [0,1]$。现在，若取 $\tilde{R}(x) = \hat{R}(x) + \delta^2 x(1-x)$，就得到一个强凹函数，并且 $|R(x) - \tilde{R}(x)| \le \delta^2$。

现在，定义 $\tilde{F}$ 使得 $\tilde{R}(q) = q\tilde{F}^{-1}(1-q)$。因为 $\tilde{R}$ 是强凹的，且 C^∞，因此 $f(v)$ 在其定义域内 > 0。因此，$\tilde{F}(v)$ 也 C^∞。

最后注意，对于 $q \in [\delta, 1-\delta]$，我们有 $|\tilde{F}^{-1}(q) - F^{-1}(q)| \le \delta^2/\delta = \delta$。现在，对于给定的 q，令 $v = F^{-1}(q)$。那么我们知道：$v - \delta \le \tilde{F}^{-1}(q) \le v + \delta$。因此 $\tilde{F}(v-\delta) \le q = F(v) \le \tilde{F}(v+\delta)$，这意味着 $\mathrm{L\acute{e}vy}(F, \tilde{F}) \le \delta$。 □

综上所述我们可以将定理2.7 应用于引理2.8 中的光滑逼近，并得到一个 $\tilde{O}(1/\epsilon^{0.5})$ 查询次数上界，用于学习任意正则分布所需的 F 和 f。

为了完成这一结果，我们还需要讨论表示结果所需的比特数，因为被查询的点 x 以及结果 $F(x)$ 和 $f(x)$ 原则上都是实数。不过，我们可以如下看待对 F（或 f）的每次查询：查询 $F(x)$（或 $f(x)$），其中 x 是 ϵ 的整数倍，并获取一个精度在 $F(x) \pm \epsilon$（或 $f(x) \pm \epsilon$）以内的近似答案。所有引理和算法都已考虑到 $\pm\epsilon$ 误差，包括 x、F 和 f 中的误差。这自动给出了一个 $\tilde{O}(1/\epsilon^{0.5})$ 比特复杂度上界。结合定理2.1中的下界，即可证明我们的主要结果（定理1.1）。

3 正则分布学习的定价查询复杂度

本节研究在 ϵ L´evy 距离下学习正则分布所需的定价查询复杂度。具体而言，对于每次定价查询，我们提交一个价格 x，并得到一个实现的价值 $v \sim F$。我们无法直接观察 v，但可以观察 sign$(v-x) \in \{-1, +1\}$。根据 Chernoff界，只需进行 $\tilde{O}(1/\epsilon^2)$ 次定价查询，就足以以 ϵ 加性误差学习 $F(x)$。然而，$\tilde{O}(\epsilon^{-0.5})$ 描述复杂度上界结果不足以证明在 $[0,1]$ 上以 ϵ L´evy 距离学习正则分布的定价查询复杂度。还有两个问题有待解决。首先，证明描述复杂度上界结果的算法不仅依赖对累积分布函数的查询，也依赖对概率密度函数的查询。其次，对于非光滑正则分布，我们通过学习一个与 F 一致接近、但并不完全等于 F 的光滑正则分布 $\tilde{}$，证明了描述复杂度结果。因此，当我们查询 $F(x)$ 时，只能得到 $F(x+\epsilon_1) + \epsilon_2$，其中 $|\epsilon_1|$ 可以任意小（我们使用以下界 $|\epsilon_1| \le \frac{\epsilon^2}{2}$），且 $|\delta_2| < \epsilon$。因此，我们要解决的问题等价于以下定理：

定理 3.1。对于任意具有光滑概率密度函数的正则分布 F，假设我们可以访问扰动预言机 $F\tilde{x}$ 每次查询 $F\ x + \epsilon_1\ + \epsilon_2$ $\epsilon_1 \in [-\frac{\epsilon^2}{2}, \frac{\epsilon^2}{2}]$ $\epsilon_2 \in [-\epsilon, \epsilon]$ $\tilde{O}\ \epsilon^{-0.5}$ 时，预言机 F^* 都会返回 $F^*()()$，其中 和 均未知。那么，对扰动预言机进行() 次查询 F^*，足以在 ϵ L´évy 距离内学习 F。

3.1 凸 F分布的查询复杂度

在证明3.1 一般正则分布的定理之前，我们先证明具有凸 F 的分布情形，以便提供更多直观理解。一般情形的思路几乎完全相同。

定理 3.2。 对于任意具有光滑概率密度函数的正则分布 F，假设我们可以访问 F 扰动预言机 $\mathcal{O}$，即每次查询 x 时，预言机 F^* 返回 $F^*(x) = F(x + \epsilon_1) + \epsilon_2$，其中某些参数未知，且 $\epsilon_1 \in [-\frac{\epsilon^2}{2}, \frac{\epsilon^2}{2}]$ 和 $\epsilon_2 \in [-\epsilon, \epsilon]$。那么，对扰动预言机进行 $\tilde{O}(\epsilon^{-0.5})$ 次查询 F^*，足以在 ϵ 以 Lévy-距离下学习 F。

证明。 我们首先说明如何通过扰动预言机定义一个对 f 的（扰动）查询。假设我们要查询 f。当我们向扰动预言机 F^* 查询 $x+\gamma$ 和 $x-\gamma$ 时，将得到 $F^*(x+\gamma) = F(x+\gamma+\epsilon_1)+\epsilon_2$，以及 $F^*(x-\gamma) = F(x-\gamma+\epsilon_3)+\epsilon_4$，其中某些未知量为 $\epsilon_1, \epsilon_3 \in [-\frac{\epsilon^2}{2}, \frac{\epsilon^2}{2}]$ 和 $\epsilon_2, \epsilon_4 \in [-\epsilon, \epsilon]$。令

$$f^*_\gamma(x) = \frac{F^*(x+\gamma) - F^*(x-\gamma)}{2\gamma}.$$

根据拉格朗日中值定理，对于光滑函数 F，存在 $x^* \in [x-\gamma+\epsilon_3, x+\gamma+\epsilon_1]$ 使得

$$\begin{aligned} f(x^*) &= \frac{F(x+\gamma+\epsilon_1) - F(x-\gamma+\epsilon_3)}{(x+\gamma+\epsilon_1) - (x-\gamma+\epsilon_3)} \\ &= \frac{(F^*(x+\gamma) - \epsilon_2) - (F^*(x-\gamma) - \epsilon_4)}{2\gamma + \epsilon_1 - \epsilon_3} \\ &= \frac{2\gamma f^*_\gamma(x) - \epsilon_2 + \epsilon_4}{2\gamma + \epsilon_1 - \epsilon_3}. \end{aligned}$$

那么我们有

$$f^*_\gamma(x) = \frac{2\gamma + \epsilon_1 - \epsilon_3}{2\gamma} f(x^*) + \frac{\epsilon_2 - \epsilon_4}{2\gamma} = (1+\delta_1) f(x+\gamma_1) + \delta_2, \tag{13}$$

对于某些 $\delta_1, \delta_2, \gamma_1$，其中 $|\delta_1| \le \frac{\epsilon^2}{2\gamma}$, $|\delta_2| \le \frac{\epsilon}{\gamma}$, $|\gamma_1| < 2\gamma$。随后，我们将通过 $f^*_\gamma(x)$ 对 f 进行扰动查询，并指定学习算法每一步所需的精度参数 γ。

引理 2.2 的改动。假设我们只能识别 $F_i \in [0,1]$，使得 $F(x_i+\epsilon_i) \in [F_i - \epsilon, F_i + \epsilon]$，其中某个未知的小量 $\epsilon_i \in [-\epsilon^2/2, \epsilon^2/2]$。我们证明引理 2.2 仍然成立。注意，在引理 2.2 的证明中，由输入 $(x_1, \cdots, x_m)$ 和 $(F_1, \cdots, F_m)$ 构造的分布，与由输入 $(x'_1, x'_2, \cdots, x'_m) = (x_1+\epsilon_1, x_2+\epsilon_2, \cdots, x_m+\epsilon_m)$ 和 $(F_1, \cdots, F_m)$ 构造的分布之间的 Lévy 距离仅为 $O(\epsilon)$。此外，由于 $x'_{i+1} - x'_i$ 要么是 $O(\epsilon)$ 要么是 $2(x_{i+1} - x_i)$，要么 $x'_{i+1} - x'_i \le x_{i+1} - x_i + \epsilon^2 < 2(x_{i+1} - x_i)$，要么 $f_i = f_i$ 要么 $f_i - f_i = O(\epsilon)$（因为 $F_i, F_i \in [0,1]$），由输入 $(x'_1, \cdots, x'_m)$ 和 $(F_1, \cdots, F_m)$ 构造的分布与 F 之间的 Lévy 距离至多为 ϵ。因此，由输入 $(x_1, \cdots, x_m)$ 和 $(F_1, \cdots, F_m)$ 构造的分布在 $[x_1, x_m]$ 上与 F 之间的 Lévy 距离至多为 $O(\epsilon)$。总之，应用引理 2.2 时，我们不需要 F_i 是 $F(x)$ 的一个估计值，且误差为 $O(\epsilon)$；令 $F_i = F^*(x)$ 就足够了。

定理 *2.3* 的变更。 现在我们详细说明在什么情况下 F 是凸的（即具有单调的 f），算法会如何变化。

步骤 1 我们设置 $K = \log(1/\epsilon)$，并使用二分搜索来确定点 x_0，x_1，……，x_{2K+3}，使得对于每个 $k = 0$，……，$K+1$，我们都有 $|x_{2k+1} - x_{2k+2}| = O(\epsilon)$，并且对于区间 $I_k = [x_{2k}, x_{2k+1}]$，有：

- $f(v) \le 2$ 对于 $v \in I_0$
- $2^{k-2} \le f(v) \le 2^{k+1}$ 适用于 $v \in I_k$, $k = 1, ..., K$

- $f(v) \geq 2^{K-1}$ 适用于 $v \in I_{K+1}$

注意，现在我们无法访问 f。因此，为了执行二分搜索，我们查询f^*_γ，其中 $\gamma = 4\epsilon$。根据（13），对于任意查询 x，$f^*_\gamma(x) = (1+\delta_1) f(x+\gamma_1) + \delta_2$，存在某个 δ_1，δ_2，γ_1满足 $|\delta_1| \leq \frac{\epsilon^2}{2\gamma} \leq \frac{1}{8}$，$|\delta_2| \leq \frac{\epsilon}{\gamma} = \frac{1}{4}$，$|\gamma_1| \leq 2\gamma = 8\epsilon$。因此

$$\frac{7}{8} f(x-8\epsilon) - \frac{1}{4} \leq f^*_\gamma(x) \leq \frac{9}{8} f(x+8\epsilon) + \frac{1}{4}.$$

因此，当 $f^*_\gamma(x) \geq 2^{k-1}$时，$f(x+8\epsilon) > 2^{k-2}$；当 $f^*_\gamma(x) \leq 2^k$时，$f(x-8\epsilon) < 2^{k+1}$。所以，假设我们对 f^*_γ进行二分搜索，以找到一个点序列 x'_0，x'_1，$\cdots$，x'_{2K+3}，使得对于每个 $k = 0$，$\cdots$，K，$f^*_\gamma(x'_{2k+1}) \leq 2^k$，$f^*_\gamma(x'_{2k+2}) \geq 2^k$。然后，若设 $x_{2k+1} = x'_{2k+1} - 8\epsilon$，$x_{2k+2} = x'_{2k+2} + 8\epsilon$，序列 x_0，$\cdots$，x_{2K+3} 便满足此步骤所需的所有性质。

~识别每个点仍需要 $O(\log(1/\epsilon)) = O(1)$ 次查询来 $f^*_\gamma(\cdot)$（因此 $F(\cdot)$）。鉴于 ~ ~我们有 $O(1)$ 个这样的点，总共使用了 $O(1)$ 次查询。最后一个区间 I_{K+1} 同样满足引理2.2的条件。

步骤 2 对于每个区间 I_k，其中 $k = 0, ..., K$，将 I_k 划分为长度为 $\frac{1}{2^k}\sqrt{\epsilon}$ 的区间 $I_{k,1}, I_{k,}, \cdots$。每个区间 I_k至多增加 $\frac{|I_k|}{\sqrt{\epsilon}/2^k} = 2^k \epsilon^{-1/2} |I_k|$ 个端点，总数至多为

$$\begin{aligned}
\sum_{k=0}^{K} 2^k \epsilon^{-1/2} |I_k| &= \epsilon^{-1/2} |I_0| + \sum_{k=1}^{K} 2^k \epsilon^{-1/2} |I_k| \\
&\leq \epsilon^{-1/2} + \sum_{k=1}^{K} \epsilon^{-1/2} \int_{v \in I_k} 4 f(v) dv \\
&\leq \epsilon^{-1/2} + \epsilon^{-1/2} \int_{v \in [0,1]} 4 f(v) dv \\
&= \epsilon^{-1/2} + 4\epsilon^{-1/2} = 5\epsilon^{-1/2}.
\end{aligned}$$

与上一节相比，唯一的变化是第二行中的 $2^k \leq 4f(v)$ 替代了 $2f(v)$。因此，如果我们查询 F^* 以获取前两个步骤中各区间的所有端点，所需查询总数至多为 $O(\epsilon^{-1/2})$。

步骤 3 对于步骤 2 中划分的每个区间 $I_{k,\ j} = [v_{k,j}, v_{k,j+1}] = [v_{k,j}, v_{k,j} \mp 2^{-k}\sqrt{\epsilon}] \subseteq I_k$ 假设 $2^{-k}\sqrt{\epsilon} > 10\epsilon$，否则假设 $|I_{k,\ j}| = O(\epsilon)$，它已经满足引理2.2的条件。

定义 $\underline{f}_{k,\ j}$和 $\overline{f}_{k,\ j}$如下。设 $\gamma^* = 2^{-k-2}\sqrt{\epsilon}$。如果$I_{k,\ j}$是 I_k 中最左边的区间，那么$\underline{f}_{k,\ j} = 2^{k-2}$，否则 $\underline{f}_{k,\ j} = f^*_{\gamma^*}(v_{k,\ j} - 2^{-k-1}\sqrt{\epsilon}) - 2^{k+3}\sqrt{\epsilon}$；如果 $I_{k,\ j}$是 I_k 中最右边的区间，那么 $\overline{f}_{k,\ j} = 2^{k+1}$，否则 $\overline{f}_{k,\ j} = f^*_{\gamma^*}(v_{k,\ j} + 2^{-k}\sqrt{\epsilon} + 2^{-k-1}\sqrt{\epsilon}) + 2^{k+3}\sqrt{\epsilon}$。注意，根据 (13)，如果$\underline{f}_{k,\ j} \neq 2^{k-2}$，那么

$$f^*_{\gamma^*}(v_{k,j} - 2^{-k-1}\sqrt{\epsilon}) = (1+\delta_1) f(v_{k,j} - 2^{-k-1}\sqrt{\epsilon} + \gamma_1) + \delta_2$$

对于某些 $\delta_1, \delta_2, \gamma_1$，其中 $|\delta_1| \leq \frac{\epsilon}{2\gamma^*} = 2^{k+3}\epsilon^{1.5}$、$\delta_2 \leq \frac{\epsilon}{\gamma^*} = 2^{k+2}\sqrt{\epsilon}$、$|\gamma_1| < 2\gamma = 2^{-k-1}\sqrt{\epsilon}$。这意味着 $v_{k\,j} - 2^{-k-1}\sqrt{\epsilon} + \gamma_1 \in [v_{k,j} - 2^{-k}\sqrt{\epsilon}, v_{k,j}]$ $f\ v_{k\,j} \leq 2^{K+1} = \frac{2}{\epsilon}$ $f\ v_{k\,j}$ $|\delta_1| \leq$。由于 ()，我们有 ()。然后 $\frac{2}{\epsilon} \cdot 2^{k+1}\epsilon^{1.5} = 2^{k+2}\sqrt{\epsilon}$

$$\underline{f}_{k,j} = f^*_{\gamma^*}(v_{k,j} - 2^{-k-1}\sqrt{\epsilon}) - 2^{k+3}\sqrt{\epsilon}$$

$$\begin{aligned} &= (1+\delta_1)f(v_{k,j} - 2^{-k-1}\sqrt{\epsilon} + \gamma_1) + \delta_2 - 2^{k+3}\sqrt{\epsilon} \\ &\leq (1+\delta_1)f(v_{k,j}) + 2^{k+2}\sqrt{\epsilon} - 2^{k+3} < f(v_{k,j}). \end{aligned}$$

这意味着 $\underline{f}_{k,j}$ 确实是 $f(v_{k,j})$ 的下界。对称地，$\overline{f}_{k,j}$ 确实是 $f(v_{k,j+1})$ 的上界。

如果 $f_{k,j} - f_{k,j} = t_{k,j}2^k\sqrt{\epsilon}$，则将区间划分为 $\tilde{t}_{k,j} := \max(1, \lceil t_{k,j} \rceil - 2^5)$ 个长度为 $\tilde{t}_{k,j}2^k\sqrt{\epsilon}$ 的区间，并对新添加的端点进行查询。由于任意相邻点 v，v 之间的距离为 $v - v \leq \tilde{t}_{k,j}2^k\sqrt{\epsilon}$。因此，该区间满足引理 2.2 中的第二个条件：

$$(\overline{v} - \underline{v})(\overline{f}_{k,j} - \underline{f}_{k,j}) \leq \frac{1}{\tilde{t}_{k,j}2^k}\sqrt{\epsilon} \cdot (\tilde{t}_{k,j} + 2^5)2^k\sqrt{\epsilon} = O(\epsilon)$$

这使我们能够将分布学习到最多相差 ϵ L´evy距离。

最后，我们只需要界定这一步所需的查询次数。注意，如果 $I_{k,j}$ 不是 I_k 中最左侧的区间，

$$\begin{aligned} \underline{f}_{k,j} &= (1+\delta_1)f(v_{k,j} - 2^{-k-1}\sqrt{\epsilon} + \gamma_1) + \delta_2 - 2^{k+3}\sqrt{\epsilon} \\ &\geq (1-|\delta_1|)f(v_{k,j} - 2^{-k}\sqrt{\epsilon}) - 2^{k+2}\sqrt{\epsilon} - 2^{k+3}\sqrt{\epsilon} \\ &\geq f(v_{k,j-1}) - 2^{k+4}\sqrt{\epsilon}. \end{aligned}$$

对称地，如果 $I_{k,j}$ 不是 I_k，$f_{k,j} \leq f(v_{k,j+2}) + 2^{k+4}\sqrt{\epsilon}$ 中最右侧的区间。因此，

$$\overline{f}_{k,j} - \underline{f}_{k,j} \leq f(v_{k,j+2}) - f(v_{k,j-1}) + 2^{k+5}\sqrt{\epsilon}.$$

由于区间 $I_{k,j}$ 被划分为 $\max(1, \lceil t_{k,j} \rceil - 2^5)$ 个区间，因此对 $f^*_{\gamma^*}$ 的查询次数至多为 $\frac{f(v_{k,j+2}) - f(v_{k,j-1})}{2^k\sqrt{\epsilon}}$。将每个 j 对应的 $\frac{f(v_{k,j+2}) - f(v_{k,j-1})}{2^k\sqrt{\epsilon}}$ 相加，可得对 $f^*_{\gamma^*}$ 的查询总次数至多为[1]

$$\sum_j \frac{f(v_{k,j+2}) - f(v_{k,j-1})}{2^k\sqrt{\epsilon}} \leq \frac{3}{2^k\sqrt{\epsilon}}(2^{k+1} - 2^{k-2}) = O(\epsilon^{-0.5})$$

通过化简望远镜求和。因此，在这一步中，对于每个 k，为新增点而对 F^* 进行的查询次数为 $\tilde{O}(\epsilon^{-1/2})$。此外，对 $f^*_{\gamma^*}$ 进行的 I_k 查询次数是 $I_{k,j}$中端点总数的两倍，根据步骤 2，该数量为 $O(\epsilon^{-1/2})$。由于 k只有 $\tilde{O}(1)$ 个不同的取值，因此对 F^* 的查询总次数为 $\tilde{O}(\epsilon^{-1/2})$。

总而言之，我们介绍了如何修改使用针对 F 和 f 的预言机查询来学习具有光滑凸累积分布函数的分布的算法，使其只需针对F^*进行预言机查询。其查询复杂度渐近不变，为 $\tilde{O}(\epsilon^{-1/2})$。 □

3.2 一般正则分布的查询复杂度

基本上，针对凸累积分布函数所描述的相同修改，也可应用于一般正则分布，从而得到定理 3.1 的证明。由于这基本上只是对定理 2.7 证明的重写，我们省略细节，转而描述主要修改。

[1]更准确地说，对于 I_k 中最右侧的区间，$f(v_{k,j+2})$ 被 2^{k+1} 取代；对于 I_k 中最左侧的区间，$f(v_{k,j-1})$ 被 2^{k-2} 取代。

在步骤 0 中，将 $[0,1]$ 划分为分位数相差不超过 2 倍的区间，也可以通过向 F^*发起查询来完成。在定理 2.7 的证明的步骤 1 中，将查询f 替换为对 $f^*_{O(\epsilon)}$的查询即可完成划分，因为引理 2.6 只要求在一个常数因子范围内估计 f，这与定理 3.2的步骤 1 相同。唯一的区别是，每个区间内 f 的价值范围可能会扩大一个常数因子。因此，步骤 2 中的查询次数会增加一个常数因子。在步骤 3 的分析中，与定理 3.2 中修改后算法的步骤 3 类似，估计 f 所需的精度为 $O(2^{k+\ell}\sqrt{\epsilon})$，这可以通过向 $f^*_{O\,(2^{-k-\ell}\sqrt{\epsilon})}$发起查询来达到。

简单的 Chernoff界表明，对 F^*（定理 3.1中定义的对象）的一次查询可以通过 $O(1/\epsilon^2)$ 次定价查询来模拟。结合定理3.1 与 Chernoff界，我们得到了定理 1.3的证明，与 [PLSTW23]中确立的下界相匹配。

4 混合分布

我们的第二个应用是分析正则分布的混合分布。推论 1.4 和 1.5可直接由定理 1.1得出。

我们以证明一个更令人惊讶的结果（定理 1.6）作为本文的结尾：尽管 O个正则分布的混合分布可以用 $\tilde{O}(1/\epsilon^{0.5})$ 比特描述，但要将其学习到误差不超过 $1/\epsilon^3$，仍需要 $\Omega(\epsilon)$ 次定价查询，其渐近数量与一般分布相同。

定理（定理1.6的重述。令 $\mathcal{C}$ 为支撑于 $[0,1]$ 且可表示为两个正则分布的混合分布的概率分布类。若要以 ϵ 的误差在 *L´evy* 距离下估计该类中某个分布的累积分布函数，至少需要$\Omega(1/\epsilon^3)$ 次定价查询。

定理*1.6*的证明。我们将使用 [PLSTW23] 中的引理 3.5，该引理指出，要区分累积分布函数分别为 F_1 和 F_2 的任意两个满足以下条件的分布，需要 $\Omega(1/\epsilon^2)$ 次定价查询

$$\frac{1}{1+\epsilon} \le \frac{F_1(v)}{F_2(v)}, \frac{1-F_1(v)}{1-F_2(v)} \le 1+\epsilon, \ \forall v. \tag{14}$$

现在，我们将构造一个集合 S，其中包含 $\Omega(1/\epsilon)$ 个分布，每个分布都满足：（1）它是两个正则分布的混合分布；（2）它满足公式14。此外，任意两个分布之间的 L´evy 距离均为 $\Omega(\epsilon)$。

构造一类由 $a < 2, \delta > 0$ 参数化的正则分布如下。对于概率密度函数为 $f_{a,\delta}$ 的分布，

$$f_{a,\delta}(v) = \begin{cases} a, & \text{if } v \in [0, \frac{1-\delta}{a}], \\ \frac{a^2}{2\delta}(\frac{1+\delta}{a} - v), & \text{if } v \in (\frac{1-\delta}{a}, \frac{1+\delta}{a}), \\ 0, & \text{if } v \in [\frac{1+\delta}{a}, 1]. \end{cases}$$

换言之，f_a 是一个密度为 a 的均匀分布，直到累积分布函数达到 $1-\delta$ 的点 $v = \frac{1-\delta}{a}$。然后 $f_{a,\delta}$ 线性递减至 0，此后保持为 0。我们首先证明此类中的任何分布都是正则的。我们要证明，对于任意 $x \in [\frac{1-\delta}{a}, \frac{1+\delta}{a}]$，$f'_{a,\delta}(x) \ge -\frac{2f^2_{a,\delta}(x)}{1-F(x)}$。首先，在这个范围内，$f'_{a,\delta}(x) = -\frac{a^2}{2\delta}$。其次，由于在这个范围内 $1 - F_{a,\delta}(x) = \frac{1}{2}(\frac{1+\delta}{a} - x) f_{a,\delta}(x)$，我们有

$$-\frac{2f^2_{a,\delta}(x)}{1-F_{a,\delta}(x)} = -\frac{2f_{a,\delta}(x)}{\frac{1}{2}(\frac{1+\delta}{a}-x)} = -\frac{2\cdot\frac{a^2}{2\delta}(\frac{1+\delta}{a}-x)}{\frac{1}{2}(\frac{1+\delta}{a}-x)} = -\frac{2a^2}{\delta} < -\frac{a^2}{2\delta} = f'_{a,\delta}(x).$$

因此，对于任意 a, δ, $f_{a,\delta}$ ，它都是正则分布的概率密度函数。考虑任意概率密度函数为$g_{a,\delta}(v) = 2 - f_{a,\delta}(v)$ 的分布，其中每个 $v \in [0,1]$ 都满足该条件。那么这种分布也是正则分布，因为它的概率密度函数是单调非递减的。此外，$f_{a,\delta}$ 与$g_{a,\delta}$ 的均匀混合分布是在 $[0,1]$ 上的均匀分布。

令 F 为定义在 $[0,1]$ 上的均匀分布。定义如下由 $a \in [1.5, 1.6]$ 参数化的一类分布，其累积分布函数为 F_a，概率密度函数为 f_a。对于任意 $v \in [0,1]$,

$$f_a(v) = f_{a,\epsilon/2}(v) + g_{a,\epsilon}.$$

那么，除了 $v \in [\frac{1-\epsilon}{a}, \frac{1+\epsilon}{a}]$ 外， f_a 是均匀分布。对于任意 $v \in [\frac{1-\epsilon}{a}, \frac{1+\epsilon}{a}]$， F_a，$_\epsilon(v)$ 与 F_a，$_{\epsilon/2}(v)$ 的差至多为 ϵ；因此， $F_a(v) = {}_{F_a}$，$_{\epsilon/2}(v)+G_a$，$_\epsilon(v)$ 2 与 $F(v) = {}_{F_a}$，$_\epsilon(v)+G_a$，$_\epsilon(v)$ 2 的差至多为 ϵ。此外，由于 $F_a(v)$ 和 $F(v)$ 在 1 以外均为常数，因此满足 (14)，且需要 $\Omega(1/\epsilon^2)$ 次定价查询才能区分 F_a 和 F。

注意，对于任意 a, F_a ，均匀分布 F 的 L´evy 距离至少为 $\Omega(\epsilon)$，因为当$F_{a,\epsilon/2}(\frac{1-\epsilon/2}{a}) - F_{a,\epsilon}(\frac{1-\epsilon/2}{a}) = \Omega(\epsilon)$。取 $S = \{F_a | a = 1.5 + 4k\epsilon$为 $k = 1, 2, \cdots, \frac{1}{40\epsilon}\}$。由于任意$F_a \in S$与均匀分布 F 仅在 $[\frac{1-\epsilon}{a}, \frac{1+\epsilon}{a}]$上不同，我们知道，要区分 F_a和 F，在 $[\frac{1-\epsilon}{a}, \frac{1+\epsilon}{a}]$ 上需要 $\Omega(1/\epsilon^2)$ 次定价查询。此外，由于对于每个 $a' < a, \frac{1+\epsilon}{a} < \frac{1-\epsilon}{a'}$， 当$a, a' \in [1.5, 1.6]$ 且 ϵ足够小时，我们知道，对于不同的 a， F_a与 F不同的区间互不重叠。因此，给定任意分布 F_a，其中 $a = 1.5 + 4k\epsilon$对某个 $k = 1, 2, \cdots, \frac{1}{40\epsilon}$成立，学习一个与 F_a的 L´evy 距离在 $o(\epsilon)$ 以内的分布需要 $\Omega(1/\epsilon^3)$ 次查询。

□

参考文献

[AB20] Amine Allouah 和 Omar Besbes。先验无关的最优拍卖。《管理科学》，66(10)：4417–4432，2020年。[AHN+19] Saeed Alaei、Jason Hartline、Rad Niazadeh、Emmanouil Pountourakis 和 Yang Yuan。最优拍卖与匿名定价。《博弈与经济行为》，118：494–510，2019年。[AM15] Daniel Azagra 和 Carlos Mudarra。在巴拿赫空间上用可微凸函数对凸函数进行整体逼近。《凸分析杂志》，22(4)：1197–1205，2015年。[BCD20] Johannes Brustle、Yang Cai 和 Constantinos Daskalakis。没有物品独立性的多物品机制：通过鲁棒性实现可学习性。载于第*21*届*ACM*经济与计算会议论文集，第715–761页，2020年。[BK96] Jeremy Bulow 和 Paul Klemperer。拍卖与谈判。《美国经济评论》，86(1)：180–194，1996年3月。[CDIZ22] Yeshwanth Cherapanamjeri、Constantinos Daskalakis、Andrew Ilyas 和 Manolis Zampetakis。标准拍卖模型的估计。载于第*23*届*ACM*经济与计算会议论文集，第602–603页，2022年。[CLMN14] L Elisa Celis、Gregory Lewis、Markus Mobius 和 Hamid Nazerzadeh。立即购买或碰碰运气：通过随机拍卖实现价格歧视。《管理科学》，60(12)：2927–2948，2014年。

[CR14] Richard Cole 和 Tim Roughgarden。收益最大化的样本复杂度。载于《计算理论》第46届年度ACM研讨会论文集，第243–252页，2014年。[DRY10] Peerapong Dhangwatnotai、Tim Roughgarden 和 Qiqi Yan。使用单个样本的收益最大化。载于第11届ACM电子商务会议论文集，第129–138页，2010年。[FHH13] Hu Fu、Jason Hartline 和 Darrell Hoy。面向风险厌恶型代理人的先验无关拍卖。载于第14届ACM电子商务会议论文集，第471–488页，2013年。[FILS15] Hu Fu、Nicole Immorlica、Brendan Lucier 和 Philipp Strack。随机化胜过二价拍卖，成为先验无关拍卖。载于第16届ACM经济与计算会议论文集，第323–323页，2015年。[GLMN21] Negin Golrezaei、Max Lin、Vahab Mirrokni 和 Hamid Nazerzadeh。增强型二价拍卖：异质竞标者的收益优化。载于第27届ACM SIGKDD知识发现与数据挖掘会议论文集，第447–457页，2021年。[HMR15] Zhiyi Huang、Yishay Mansour 和 Tim Roughgarden。充分利用你的样本。载于第16届ACM经济与计算会议论文集，第45–60页，2015年。[HR09] Jason D Hartline 和 Tim Roughgarden。简单机制与最优机制。载于第10届ACM电子商务会议论文集，第225–234页，2009年。[KL+23] Robin Kothari、Troy Lee、Ilan Newman 和 Mario Szegedy。查询复杂度。世界科学出版公司Pte Limited，2023年。[MR15] Jamie H Morgenstern 和 Tim Roughgarden。关于近似最优拍卖的伪维数。神经信息处理系统进展，28，2015年。[MS83] Roger B Myerson 和 Mark A Satterthwaite。双边交易的有效机制。《经济理论杂志》，29(2)：265–281，1983年。[Mye81] Roger B Myerson。最优拍卖设计。《运筹学数学》，6(1)：58–73，1981年。[PLPV16] Renato Paes Leme、Martin Pal 和 Sergei Vassilvitskii。个性化保留价指南。载于第25届万维网国际会议论文集，第1093–1102页，2016年。[PLSTW23] Renato Paes Leme、Balasubramanian Sivan、Yifeng Teng 和 Pratik Worah。收益最大化的定价查询复杂度。ACM-SIAM离散算法研讨会（SODA23），2023年。[Pr´e73] Andr´as Pr´ekopa。关于对数凹测度与函数。《数学科学学报》，34：335–343，1973年。[RTCY12] Tim Roughgarden、Inbal Talgam-Cohen 和 Qiqi Yan。供应限制机制。载于第13届ACM电子商务会议论文集，第844–861页，2012年。

[SS13] Balasubramanian Sivan 和 Vasilis Syrgkanis。针对非正则分布的维克里拍卖。载于国际网络与互联网经济学会议，第 422–435 页。施普林格，2013 年。