

# Description Complexity of Regular Distributions

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

## Abstract

Myerson’s regularity condition of a distribution is a standard assumption in economics. In this paper, we study the complexity of describing a regular distribution within a small statistical distance. Our main result is that  $\tilde{\Theta}(\epsilon^{-0.5})$  bits are necessary and sufficient to describe a regular distribution with support  $[0, 1]$  within  $\epsilon$  Lévy-distance. We prove this by showing that we can learn the regular distribution approximately with  $\tilde{O}(\epsilon^{-0.5})$  queries to the cumulative density function. As a corollary, we show that the pricing query complexity to learn the class of regular distribution with support  $[0, 1]$  within  $\epsilon$  Lévy-distance is  $\tilde{\Theta}(\epsilon^{-2.5})$ . To learn the mixture of two regular distributions,  $\tilde{\Theta}(\epsilon^{-3})$  pricing queries are required.

## 1 Introduction

A Myerson-regular distribution is a distribution with CDF  $F$  such that the revenue curve in quantile space  $R(q) = q \cdot F^{-1}(1 - q)$  is concave. For distributions that can be represented by a PDF  $f$  (i.e., have no point masses) this is equivalent to the (non-decreasing) monotonicity of the virtual value function

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

Myerson-regularity (or simply regularity) is a standard condition in Economics that was originally introduced by Myerson in his seminar paper on optimal auctions [Mye81]. Since then it has played a fundamental role in the design and analysis of various economic setups such as: bilateral trade [MS83], prior-independent mechanism design [BK96, HR09, RTCY12, FHH13, SS13, AB20], auctions from samples [DRY10, CR14, FILS15, HMR15, MR15], approximation in mechanism design and revenue management [CLMN14, PLPV16, AHN<sup>+</sup>19, GLMN21], ...

Besides being widely used in Economics, they also encompass various important distributions:

- distributions with log-concave PDF, i.e. distributions where the PDF is of the form  $f(x) = \exp(-g(x))$  for a convex function  $g$ , such as uniform, exponential and normal.
- distributions satisfying the monotone hazard rate conditions (MHR) condition which is a notion from reliability theory. If a random variable measures the time until a certain failure event happens (e.g. a light bulb goes out) then MHR means that the probability that the failure happens at any given moment conditioned it hasn’t yet happened weakly increases over time.

- equal-revenue distributions where pricing at each point of the support leads to the same revenue. As an example, consider the distribution supported on  $[1, \infty)$  with CDF  $F(x) = 1 - 1/x$ . Those are known to be extremal regular distributions and are often used to derive lower bounds in revenue management.
- certain distributions arising from machine learning, for example: let  $x$  be a random feature vector in  $\mathbb{R}^d$  which is sampled from a uniform or Gaussian distribution restricted to a  $d$ -dimensional convex set  $K$  and let  $w \in \mathbb{R}^d$  be a fixed vector of weights  $w \in \mathbb{R}^d$ . Then the distribution of the dot product  $\langle w, x \rangle$  is regular by the Prekopa-Leidner Theorem [Pré73].

**Describing Regular Distributions** In this paper we ask how many bits of information we need to describe a regular distribution. We obtain a sharp bound and apply to derive new tight guarantees for learning regular distributions.

Since a regular distribution is a continuous object with possibly unbounded support, we need to make a few assumptions to make the problem well-defined.

- **Bounded support:** We will assume that the distribution has support in  $[0, 1]$ . Without this assumption, even representing the subclass of uniform distributions supported on  $[n, n + 1]$  requires infinitely many bits.
- **Lévy-distance approximation** We will allow for  $\epsilon$ -error in the representation measured in Lévy distance, which allows an  $\epsilon$  error both in values and probabilities. Formally:

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

Now we can state our main question formally:

**Definition 1.** *We say that it is possible to describe a class of distributions  $\mathcal{C}$  with  $b$  bits and  $\epsilon$  error, if there is a class of distributions  $\mathcal{C}'$  with  $|\mathcal{C}'| \leq 2^b$  such that for every distribution  $F \in \mathcal{C}$  there exists  $F' \in \mathcal{C}'$  such that  $\text{Lévy}(F, F') \leq \epsilon$ .*

Our main theorem is:

**Theorem 1.1.** *It is possible to describe the class of regular distributions bounded in  $[0, 1]$  within  $\epsilon$  Lévy-distance using  $\tilde{O}(1/\epsilon^{0.5})$  bits. Moreover, this is tight up to polylog factors.*

It is useful to contrast this with the class of *general* distribution supported in  $[0, 1]$  for which  $\Omega(1/\epsilon)$  bits are necessary and  $\tilde{O}(1/\epsilon)$  bits sufficient. For the sufficient part, we can represent a CDF  $F$  by the numbers  $\lfloor F(k\epsilon)/\epsilon \rfloor$  for  $k = 0, 1, \dots, \lceil 1/\epsilon \rceil$ . Given those, we can construct a distribution with CDF:

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

which is  $\epsilon$ -close to  $F$  in Lévy-distance. To see that  $\Omega(1/\epsilon)$  bits are necessary, it is enough to construct a set of  $2^{\Omega(1/\epsilon)}$  distributions such that each pair differ by at least  $\epsilon$  in Lévy-distance. Given bits  $b_0, \dots, b_k \in \{0, 1\}$  for  $k = 1/(2\epsilon)$  construct a distribution such that:

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

For regular distributions, we will be able to construct a more succinct representation, using only square root of the number of bits. However, we will need a more sophisticated sampling procedure.

**Beyond Regular Distributions** While regular distributions are common in auction theory, many distributions encountered in practice are irregular. To generalize our results beyond regularity we define a notion of an irregularity coefficient, which measures how close to regular a distribution is:

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

A 0-irregular distribution is a regular distribution in the usual sense and an  $\infty$ -irregular distribution is a general distribution. The  $\beta$ -irregular class for  $\beta > 0$  contains irregular distributions that are close enough to regularity to afford a low description complexity:

**Theorem 1.2.** *It is possible to describe the class of  $O(1)$ -irregular distributions bounded in  $[0, 1]$  within  $\epsilon$  Lévy-distance using  $\tilde{O}(1/\epsilon^{0.5})$  bits.*

The notion of  $\beta$ -irregularity coincides with the notion of  $\alpha$ -strong regularity of Cole and Roughgarden [CR14] when the sign is negative. The definition was originally intended to interpolate between regular and MHR distributions for positive values of  $\alpha$ . When one considers negative values for  $\alpha$ , we obtain a measure of how close a certain distribution is to regularity. A distribution is  $\alpha$ -strongly-regular in the sense of [CR14] whenever:

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

and hence it is  $\beta$ -irregular when it is  $(-\beta)$ -strongly-regular, i.e.:

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

**Application: Pricing Query Complexity** Our first application is to settle the pricing query complexity of learning a regular distribution with  $\epsilon$ -error in Lévy-distance. The pricing query complexity model was introduced in [PLSTW23] with the goal of learning in economic settings where the only viable mechanism is posted prices. In such settings, we only observe if prices posted for different agents led to a sale or no sale. This notion is also useful when the auction of choice is a first-price auction, since we don't have access to truthful bids, but we still know if the bidder chose to bid above the reserve or not. In such settings, we use the binary sale/no-sale outcomes observed from previous periods to optimize the price in future auctions.

In this model a learning algorithm is able to interact with a distribution  $F$  via pricing queries: in each query, the algorithm chooses a price  $p_t$ , a sample  $v_t$  is drawn from  $F$  and the algorithm only learns the sign  $\{+1, -1\}$  of  $v_t - p_t$ . The goal of the algorithm is to estimate some parameters of the distribution such as the mean, median or monopoly price for a given class of distributions. Paes Leme, Sivan, Teng and Worah [PLSTW23] give matching upper and lower bounds for several parameters of interest, but they leave a gap in estimating the pricing query complexity of learning the CDF of a regular distribution. They give provide a  $\tilde{O}(1/\epsilon^3)$  upper bound and a  $\Omega(1/\epsilon^{2.5})$  lower bound.

We settle this question by providing a  $\tilde{O}(1/\epsilon^{2.5})$ , matching the lower bound up to polylogarithmic factors. This is obtained by computing the  $O(1/\epsilon^{0.5})$  description using  $O(1/\epsilon^2)$  pricing queries to evaluate each point using the Chernoff bound:

**Theorem 1.3.** *There is a  $\tilde{O}(1/\epsilon^{2.5})$  upper bound on the pricing query complexity of learning the CDF of a regular distribution within  $\epsilon$  Lévy-distance error. This result is tight up to polylogarithmic factors.*

**Application: Mixture Distributions** We say that a distribution with CDF  $F$  is a mixture of  $k$  distributions  $F_1, \dots, F_k$  if there are weights  $w_1, \dots, w_k \geq 0$  with  $\sum_i w_i = 1$  such that  $F(v) = \sum_i w_i F_i(v), \forall v$ . In various applications, it is useful to write distributions as mixtures of other distributions. For example, Sivan and Syrgkanis [SS13] design auctions for mixtures of regular distributions whose performance depends on the number of distributions in the mixture. One may ask how many regular distributions are needed to represent a general distribution. The description complexity bounds automatically imply the following corollary:

**Corollary 1.4.** *There exist a irregular distribution that can't be  $\epsilon$ -approximated in Lévy-distance by a mixture of  $o(1/\epsilon^{0.5})$  regular distributions.*

This follows directly from the fact that a mixture of  $k$  regular distributions can be described by  $O(k/\epsilon^{0.5})$  while a general distribution requires  $\Omega(1/\epsilon)$  bits to represent.

It is also useful in ML applications to represent distributions as mixtures of Gaussians. Our result also implies that one may require many Gaussians to represent a regular distribution, since a Gaussian distribution can be represented by  $\tilde{O}(1)$  bits.

**Corollary 1.5.** *There exist a regular distribution that can't be  $\epsilon$ -approximated in Lévy-distance by a mixture of  $o(1/\epsilon^{0.5})$  Gaussian distributions.*

**Learning Mixtures of Regular Distributions** A mixture of two regular distributions can be described using  $\tilde{O}(1/\epsilon^{0.5})$  bits since we need  $\tilde{O}(1/\epsilon^{0.5})$  bits to describe each distribution and  $\tilde{O}(1)$  bits more to describe the weights. Given that, one would guess that the pricing query complexity of learning a mixture is also  $\tilde{O}(1/\epsilon^{2.5})$ . We conclude the paper with the following rather surprising result:

**Theorem 1.6.** *Let  $\mathcal{C}$  be the class of distributions supported in  $[0, 1]$  that can be written as a mixture of two regular distributions. To estimate the CDF of a distribution in that class within  $\epsilon$  in Lévy-distance we require at least  $\Omega(1/\epsilon^3)$  pricing queries.*

**Why Levy Distance?** There are different ways to measure the distance between two distributions such as Total Variation (TV), Wasserstein, Kolmogorov, and Lévy. We note that bounds on the Lévy-distance automatically imply bounds on the Wasserstein. The Kolmogorov and TV distances are stronger notions but it is impossible to obtain any approximation in either one using pricing queries. This is discussed in detail in [PLSTW23]. The Kolmogorov distance between two distributions  $F$  and  $G$  is given by:

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

If a distribution is a deterministic value in  $[0, 1]$ , to get any meaningful approximation in Kolmogorov distance, one needs to estimate this value exactly, which is impossible using pricing queries.

Another reason to study Lévy-distance is that such a metric has been considered in the literature on the sample complexity of learning revenue-optimal auctions, see Brustle et al [BCD20] and Cherapanamjeri et al [CDIZ22] for examples. Better understanding on the complexity of estimating a distribution within Lévy-distance can lead to improved results on the complexity of auction learning.

**Related Work** Our work is broadly situated in the theme of sample complexity in algorithmic economics, where the goal is to understand what is the minimal amount of information to describe or optimize a certain economic setup. There are several ways one can explore this question. For example, Dhangwatnotai et al [DRY10] and Fu et al [FHH13] ask to what extent one can optimize an auction using a single sample of a distribution. In the other extreme, we can ask how many samples from a certain distribution are required to optimize an auction (Cole and Roughgarden [CR14] and Morgenstern and Roughgarden [MR15]) or compute the optimal reserve price (Huang et al [HMR15]).

Closest to our paper is the paper by Paes Leme et al [PLSTW23] which considers a restricted query model. Instead of having access to samples of a distribution, we are only allowed to post a price and observe for a fresh draw of that distribution whether the price was above or below the posted price. This is motivated by learning in two important scenarios: (i) settings where posted-price mechanisms are used and we only observe purchase/no-purchase decisions; (ii) learning in non-truthful auctions (like first-price auctions) where the bid is not an unbiased sample of the value but we can still observe whether a bidder decided to bid above the reserve price or not. In this setting, Paes Leme et al [PLSTW23] provides tight bounds on how to learn the monopoly price of MHR, regular and general distributions and shows that it requires strictly less samples to learn the reserve than to learn the entire CDF of a regular distribution.

However, [PLSTW23] leaves a gap on the number of pricing queries required to learn the CDF of a regular distribution: they show a lower bound of  $\Omega(1/\epsilon^{2.5})$  and an upper bound of  $\tilde{O}(1/\epsilon^3)$ . We settle this question by providing an algorithm with pricing query complexity  $\tilde{O}(1/\epsilon^{2.5})$ .

Our work is also related to the line of work on query complexity in Computer Science, which asks what is the minimum number of queries to a black box required to perform a certain task. This approach has been applied to learning theory, parallel computing, quantum computing, analysis of Boolean functions, and optimization, among others. Since this literature is too broad to be surveyed here, we refer the reader to the book by Kothari, Lee, Newman, and Szegedy [KL+23].

## 2 Description Complexity of Regular Distributions

An  $\Omega(1/\epsilon^{0.5})$  lower bound on the description complexity of regular distributions is implicit in the  $\Omega(1/\epsilon^{2.5})$  lower bound on the pricing query complexity given in [PLSTW23]. We can even modify their example to strengthen the result to work for the class of distributions with non-decreasing PDF functions (which is a subclass of MHR distributions and regular distributions).

**Theorem 2.1.** *There is a set of  $2^{\Omega(1/\epsilon^{0.5})}$  distributions with non-decreasing PDF supported on  $[0, 1]$  such that for every given pair the Lévy-distance is at least  $\Omega(\epsilon)$ .*

*Proof.* Let  $n = 1/(4\epsilon^{0.5})$  and let  $b = (b_0, \dots, b_{n-1}) \in \{0, 1\}^n$  be a sequence of bits. Now define a PDF  $f_b(x)$  such that for  $v \in [4\sqrt{\epsilon}k, 4\sqrt{\epsilon}(k+1)]$  we have  $f_b(v) = 2v$  if  $b_k = 0$  and

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

if  $b_k = 1$  (see Figure 1). First, we observe that all such functions have non-decreasing PDFs. Now, if two functions differ in a certain bit  $b_t$  then the Lévy-distance must be at least  $\Omega(\epsilon)$  by comparing their CDF around  $v_0 = (4k+2)\sqrt{\epsilon}$ . In particular, let  $b, b' \in \{0, 1\}^n$  be two sequences that only differs at index  $k$ , with  $b_k = 0$  and  $b'_k = 1$ . Let  $F, G$  be the CDFs of distributions constructed by sequence  $b$  and  $b'$  respectively. Then  $F$  and  $G$  have  $\Omega(\epsilon)$  Lévy-distance since  $F(v_0) < G(v_0 - 0.1\epsilon) - 0.1\epsilon$ .  $\square$

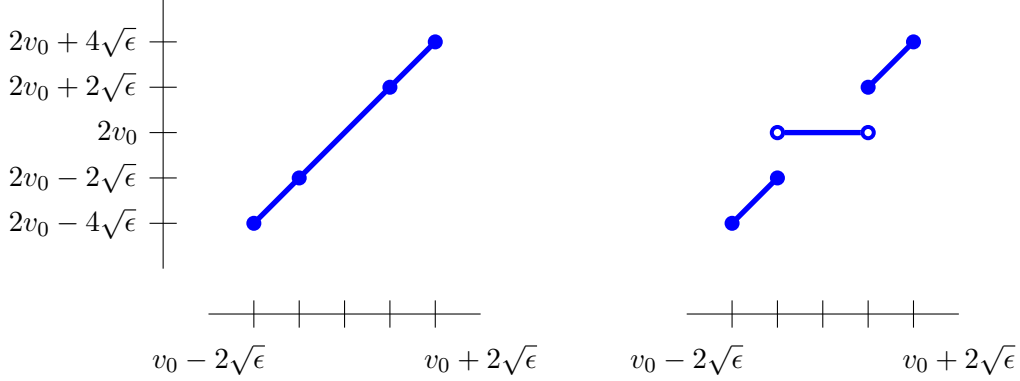


Figure 1: The PDF  $f_b$  in region  $[v_0 - 2\sqrt{\epsilon}, v_0 + 2\sqrt{\epsilon}]$  for bit  $b_k = 0$  (left) and  $b_k = 1$  (right).

The above theorem shows that to describe a regular distribution within  $\epsilon$  Lévy-distance,  $\Omega(\epsilon^{-0.5})$  bits are necessary. The main focus of the remaining section is to provide a matching upper bound  $\tilde{O}(\epsilon^{-0.5})$  on the description complexity.

## 2.1 Learning a smooth distribution

We will start with the assumption that the distribution is smooth, by which we mean that it has no point masses and is described by a PDF of class  $C^1$ , i.e., the derivative  $f'(v)$  exists and is continuous. Under this assumption, regularity can be written as the following condition:

$$f'(v) \cdot (1 - F(v)) \geq -2f^2(v) \quad (6)$$

which corresponds to the derivative of the virtual value function being non-negative. We start proving Lemma 2.2 with sufficient conditions for learning a smooth distribution. In Section 2.2 we describe a simplified version of our algorithm to learn a distribution with a convex PDF ( $f'(v) \geq 0$ ). After we describe our main ideas in this special case, we then turn to learn a smooth regular distribution in Section 2.3. Finally, we drop the smoothness assumption in Section 2.4 by using a regularity-preserving mollification argument.

**Lemma 2.2.** *For any unknown distribution  $F$  and  $m$  points  $x_1 < x_2 < \dots < x_m$ , such that for each  $i$ ,*

- *we can identify  $F_i \in [0, 1]$  such that  $F(x_i) \in [F_i - \epsilon, F_i + \epsilon]$ , and*
- *either  $x_{i+1} - x_i \leq \epsilon$ , or there exists  $\underline{f}_i \leq \bar{f}_i$  such that  $f(x) \in [\underline{f}_i, \bar{f}_i]$  for every  $x \in [x_i, x_{i+1}]$ , and satisfy  $(x_{i+1} - x_i)(\bar{f}_i - \underline{f}_i) \leq \epsilon$ ,*

*we can construct a distribution  $\hat{F}$  within  $O(\epsilon)$  Lévy-distance from  $F$  on  $[x_1, x_m]$ .*

*Proof.* We can assume that  $F_1 \leq F_2 \leq \dots \leq F_m$ . If not, we can sort the values of  $F_i$  in increasing order and the properties in the lemma will continue to hold. To see that, first notice that if  $F_i > F_{i+1}$  then we have  $F(x_i)$  and  $F(x_{i+1})$  are both in range  $[F_i - \epsilon, F_{i+1} + \epsilon]$ . This happens because  $F(x_i) \leq F(x_{i+1})$ , and  $F(x_i) \in [F_i - \epsilon, F_i + \epsilon]$ ,  $F(x_{i+1}) \in [F_{i+1} - \epsilon, F_{i+1} + \epsilon]$ . Then  $F(x_i) \in [F_i - \epsilon, F_{i+1} + \epsilon] \subseteq [F_{i+1} - \epsilon, F_{i+1} + \epsilon]$ , and  $F(x_{i+1}) \in [F_i - \epsilon, F_{i+1} + \epsilon] \subseteq [F_i - \epsilon, F_i + \epsilon]$ . Therefore, if we switch  $F_i$  and  $F_{i+1}$ , the properties in the lemma still hold.

Consider the following distribution  $\hat{F}$ : For each  $x_i$ ,  $\hat{F}(x_i) = F_i$  and

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

In other words,  $\hat{F}$  is defined by the estimation of  $F$  at some points, and filling the curve in between by a linear function.

Now we show that  $\hat{F}$  is within  $O(\epsilon)$  Lévy-distance from  $F$  on  $[x_1, x_m]$ . We prove this by showing that this is true on every interval  $[x_i, x_{i+1}]$ .

If  $x_{i+1} - x_i \leq \epsilon$ , then  $F$  and  $\hat{F}$  are within  $\epsilon$  Lévy-distance on  $[x_i, x_{i+1}]$  since for any  $x \in [x_i, x_{i+1}]$

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

If  $x_{i+1} - x_i > \epsilon$ , we show a stronger statement that  $\hat{F}$  is within  $O(\epsilon)$  Kolmogorov distance from  $F$  on  $[x_i, x_{i+1}]$ , i.e., for any  $x \in [x_i, x_{i+1}]$ ,  $|\hat{F}(x) - F(x)| \leq O(\epsilon)$ . By the lemma statement, this is true for every  $x_i$ . Now we show that this is also true for every  $x \in (x_i, x_{i+1})$ . Let  $\hat{f}_i = \frac{F_{i+1} - F_i}{x_{i+1} - x_i}$  be the PDF of  $\hat{F}$  on  $(x_i, x_{i+1})$ . Then

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

Here the first line is by the definition of the CDF and PDF; the second line is by  $|a + b| \leq |a| + |b|$  for any  $a, b$ ; the third line is by the definition of  $F_i$ . Now we bound  $\hat{f}_i$ . As  $f(x) \leq \bar{f}_i$  for every  $x \in (x_i, x_{i+1})$ , we have

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

The same way, as  $f(x) \geq \underline{f}_i$  for every  $x \in (x_i, x_{i+1})$ , we have

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

Combine the two inequalities above, we can bound  $\hat{f}_i$  by

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

Thus for any  $t \in (x_i, x_{i+1})$ , as  $f(t) \in [\underline{f}_i, \bar{f}_i]$ , we have  $|\hat{f}_i - f(t)| \leq \bar{f}_i - \underline{f}_i + \frac{2\epsilon}{x_{i+1} - x_i}$ . Apply this bound to (7), we get

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

□



## 2.2 Learning a smooth distribution with convex CDF

In this section, we assume we have a smooth distribution  $F$  and an oracle that allows us to sample the value of its CDF  $F(\cdot)$  and its PDF  $f(\cdot)$ . We start with a special case of convex CDF ( $f'(v) > 0$ ) to describe a simplified version of our algorithm.

For a high-level intuition, it is useful to perform the following ‘heuristic’ calculation based on Lemma 2.2: if  $\Delta x$  is the distance between query points and  $\Delta f$  is the variation of the PDF between those points, our goal is to obtain  $\Delta x \cdot \Delta f \leq \epsilon$ . Approximating  $\Delta f \approx f'(x) \cdot \Delta x$  we obtain  $\Delta x \leq \sqrt{\epsilon/f'(x)}$ . Hence, we will try to sample points at a rate  $\sqrt{f'(x)/\epsilon}$ .

We make this formal in the following proof and bound the number of queries we need to guarantee this sample rate. Note that our proof is algorithmic and will be easily converted later to a pricing query complexity bound.

**Theorem 2.3.** *Let  $F$  be a smooth distribution with a convex PDF. Then  $\tilde{O}(\epsilon^{-0.5})$  oracle queries to  $F$  and  $f$  are sufficient to learn the distribution within  $\epsilon$  Lévy-distance.*

*Proof.* We will show it is possible to find a set of  $O(1/\epsilon^{0.5})$  points satisfying the conditions of Lemma 2.2. We start by noticing that for a convex CDF, the PDF  $f$  is monotone.

**Step 1** We set  $K = \log(1/\epsilon)$  and use binary search to identify points  $x_0, x_1, \dots, x_{2K+3}$  such that for each  $k = 0, \dots, K+1$  we have  $|x_{2k+1} - x_{2k+2}| \leq \epsilon$  and for the interval  $I_k = [x_{2k}, x_{2k+1}]$  it holds that:

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

Identifying each point takes  $O(\log(1/\epsilon)) = \tilde{O}(1)$  queries to  $f(\cdot)$ . Given that we have  $\tilde{O}(1)$  such points, we used a total of  $\tilde{O}(1)$  so far. We observe that the last interval  $I_{K+1}$  already satisfies the conditions of Lemma 2.2 since its length  $|I_{K+1}|$  is at most  $\epsilon$ :

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

**Step 2** For each interval  $I_k$  with  $k = 0, \dots, K$ , partition  $I_k$  to intervals  $I_{k,1}, I_{k,2}, \dots$  of length  $\frac{1}{2^k} \sqrt{\epsilon}$ . The number of endpoints added is at most  $\frac{|I_k|}{\sqrt{\epsilon}/2^k} = 2^k \epsilon^{-1/2} |I_k|$  for each interval  $I_k$ , and sums up to at most

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

Thus if we query  $f$  and  $F$  for all the endpoints of the intervals from the first two steps, the total number of queries needed is at most  $O(\epsilon^{-1/2})$ .



**Step 3** For each interval  $I_{k,j} = [v_{k,j}, v_{k,j} + 2^{-k}\sqrt{\epsilon}] \subseteq I_k$  partitioned in Step 2, define  $\underline{f}_{k,j} = f(v_{k,j})$  and  $\bar{f}_{k,j} = f(v_{k,j} + 2^{-k}\sqrt{\epsilon})$ . If  $\bar{f}_{k,j} - \underline{f}_{k,j} = t_{k,j}2^k\sqrt{\epsilon}$ , then partition the interval to  $\tilde{t}_{k,j} := \max(1, \lceil t_{k,j} \rceil)$  intervals of length  $\frac{1}{\tilde{t}_{k,j}2^k}\sqrt{\epsilon}$ , and query  $F$  for the newly added endpoints. As any neighboring points  $\bar{v}, \underline{v}$  have distance  $\bar{v} - \underline{v} \leq \frac{1}{\tilde{t}_{k,j}2^k}\sqrt{\epsilon}$ . Hence this interval satisfies the second condition in Lemma 2.2 :

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

which allows us to learn the distribution up to  $\epsilon$  Lévy-distance.

Finally, we only need to bound the number of queries needed in this step. At most  $\lceil t_{k,j} \rceil$  are needed in  $I_{k,j}$  which has length  $2^{-k}\sqrt{\epsilon}$ . Now if  $I_k = [x_{2k}, x_{2k+1}]$  we have:

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

which implies  $\sum_j t_{k,j} \leq 1/(2\sqrt{\epsilon})$ . Thus in this step, for all  $k$ , the total number of queries on  $F$  is  $\tilde{O}(\epsilon^{-1/2})$ .  $\square$

### 2.3 Learning a smooth regular distribution

Unlike convex functions, the PDF of a regular distribution can increase or decrease, so we can't easily partition the interval as in the previous lemma. However, regularity imposes a rate at which the PDF can decrease (equation (6)). Our first lemma shows that the PDF of a regular distribution can't decrease by a factor of 2 too many times. We make this intuition formal in the next lemma:

**Lemma 2.4.** *Let  $F$  be a smooth regular distribution with PDF  $f$ . Fix an integer  $\ell$ . Now assume  $S$  is a collection of disjoint intervals of the form  $[v_i, v'_i]$  such that:*

- for each interval  $[v_i, v'_i]$  we have  $2^{-\ell} \leq 1 - F(v'_i) \leq 1 - F(v_i) < 2^{-\ell+1}$
- For each  $v \in [v_i, v'_i]$  we have  $f(v'_i) \leq f(v) \leq f(v_i)$  and  $f(v'_i) \leq \frac{1}{2}f(v_i)$

Then the number of intervals is bounded by a constant  $|S| \leq 32$ .

*Proof.* We will count the maximum number of intervals in  $S$ . We can assume w.l.o.g. that  $f(v'_i) = \frac{1}{2}f(v_i)$  as we can always take a subinterval where the last condition holds with equality.

Consider an interval  $[v_i, v'_i]$  in the collection  $S$  with  $f(v_i) \in [2^{k-1}, 2^k)$ . As for any  $v \in [v_i, v'_i]$ ,

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

Here the first inequality is by the definition of regular distributions; the second and the third inequalities are by  $f$  is bounded between  $f(v'_i)$ , thus  $f(v) \leq f(v_i) \leq 2^k$ , and  $1 - F(v) \geq 1 - F(v_i) \geq 2^{-\ell}$ . Therefore

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

Here the first inequality is by  $f(v'_i) = \frac{1}{2}f(v_i)$  and  $f(v_i) \geq 2^{k-1}$ ; the second inequality is by inequality (8). By inequality (9) we can lower bound the length of the interval by  $v'_i - v_i \geq \frac{1}{2}2^{k-1-(2k+\ell+1)} = 2^{-k-\ell-3}$ . Thus the CDF change in the interval can be lower bounded as follows:

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

Here the first inequality is by  $f(v) \geq f(v_i)$  in the interval with a non-increasing PDF function, and the second inequality is by  $f(v'_i) = \frac{1}{2}f(v_i) \geq 2^{k-2}$ , and  $v'_i - v_i \geq 2^{-k-\ell-3}$ . Since the intervals  $[v_i, v'_i)$  are disjoint, the sum of  $F(v'_i) - F(v_i)$  is at most  $2^{-\ell+1} - 2^{-\ell} = 2^{-\ell}$ . Therefore, there are at most  $2^5 = O(1)$  such intervals.  $\square$

Lemma 2.4 shows that in an interval with the quantile bounded by a factor of 2, the PDF can decrease by a factor of  $\frac{1}{2}$  for at most 32 times. A useful corollary is that in this case the PDF can decrease by a constant factor of at most  $1 - 2^{-33}$  between two points.

**Corollary 2.5.** *Let  $F$  be a smooth regular distribution and  $\ell$  an integer. For any two points  $x < x'$  with  $2^{-\ell} \leq 1 - F(x') < 1 - F(x) \leq 2^{-\ell+1}$  it holds that  $f(x') > 2^{-33}f(x)$ .*

*Proof.* If  $f(x) \geq 2^{33}f(x')$ , as  $f$  is continuous,  $[x, x')$  can be partitioned to 33 intervals  $[x, v_1) = [v_0, v_1)$ ,  $[v_1, v_2)$ ,  $[v_2, v_3)$ ,  $\dots$ ,  $[v_{31}, v_{32})$ ,  $[v_{32}, x') = [v_{32}, v_{33})$  such that for any interval  $[v_i, v_{i+1})$ ,  $f(v_{i+1}) \leq \frac{1}{2}f(v_i)$  which contradicts Lemma 2.4.  $\square$

With Lemma 2.4 we can also obtain the following key component of the learning algorithm.

**Lemma 2.6.** *Let  $F$  be a smooth regular distribution and  $\ell$  an integer and  $\epsilon > 0$  a parameter. For any interval  $[\underline{x}, \bar{x}]$  with  $2^{-\ell} \leq 1 - F(\bar{x}) < 1 - F(\underline{x}) \leq 2^{-\ell+1}$ , using  $\tilde{O}(1)$  queries to  $f$  we can find a sequence of points  $\underline{x} = x_1 \leq x_2 \leq \dots \leq x_m = \bar{x}$  in the interval, such that: for any sub-interval  $[x_i, x_{i+1})$ , either*

- $x_{i+1} - x_i \leq \epsilon$ ;
- $f(x_{i+1}) < 2^{-33}$ , which implies for any point  $x \in [x_i, x_{i+1})$ ,  $f(x) \leq 1$ ;
- for any two points  $x, x' \in [x_i, x_{i+1})$ ,  $2^{-36}f(x) < f(x') < 2^{36}f(x)$ .

*Proof.* Consider the following binary search algorithm: For any interval  $[a, b]$  (begin with  $[a, b] = [\underline{x}, \bar{x} - \epsilon]$  and only  $\underline{x}, \bar{x} - \epsilon, \bar{x}$  in the output sequence),

1. If  $b - a \leq \epsilon$ , stop searching  $[a, b]$ ;
2. If  $f(b) < 4f(a)$ , stop searching  $[a, b]$ ;
3. If  $f(b) < 2^{-33}$ , stop searching  $[a, b]$ ;
4. Otherwise, add  $\frac{a+b}{2}$  to the output sequence, then search both  $[a, \frac{a+b}{2}]$  and  $[\frac{a+b}{2}, b]$ .

We show that all the first three cases in the algorithm are valid stopping conditions.

In Case 1, we don't need to search the interval  $[a, b]$  as the endpoints are close enough to satisfy the required property of the lemma.

In Case 2, consider interval  $[a, b]$  with  $f(b) \leq 4f(a)$ , and any  $x < x'$  in  $[a, b]$ . By Corollary 2.5  $f(x') < 2^{36}f(x)$ . If  $f(x') \geq 2^{36}f(x)$ , as  $f(b) < 4f(a)$ , there exist at least 33 intervals  $[a, v_1)$ ,  $[v_1, v_2)$ ,  $\dots$ ,  $[v_{j-2}, v_{j-1})$ ,  $[v_{j-1}, x)$ ,  $[x', v_{j+1})$ ,  $[v_{j+1}, v_{j+2})$ ,  $\dots$ ,  $[v_{31}, v_{32})$ ,  $[v_{32}, b)$  such that for any interval listed above, the PDF between the two endpoints decreases by at least a factor of  $\frac{1}{2}$ . This also contradicts Lemma 2.4 that in  $[\underline{x}, \bar{x}]$  the PDF can only decrease by half for at most 32 times. Similarly,  $f(x') > 2^{36}f(x)$  otherwise there are at least 33 intervals between  $x$  and  $x'$  where the PDF decreases by half.

In Case 3, if  $f(b) < 2^{-33}$ , by Corollary 2.5 there does not exist any  $x \in [a, b)$  such that  $f(x) > 1 > 2^{33}f(b)$ . Thus for any  $x \in [a, b)$ ,  $f(x) \leq 1$ .

Thus, all first 3 cases of the algorithm are valid stopping conditions. Now we analyze how many points are added to the sequence. First notice that for any  $v \leq \bar{x} - \epsilon$ ,  $f(v) < \frac{2^{33}}{\epsilon}$ : otherwise if  $f(v) \geq \frac{2^{33}}{\epsilon}$ , as  $f(x) > 2^{-33}f(v) \geq \frac{1}{\epsilon}$  for every  $x \in [v, v+\epsilon]$  by Corollary 2.5, we have  $F(v+\epsilon) - F(v) = \int_v^{v+\epsilon} f(t)dt > \frac{1}{\epsilon}((v+\epsilon) - v) = 1$ , which is impossible.

Consider the following set  $S$  containing all searched intervals  $[a_i, b_i]$  that are one level from the leaves: in other words, the algorithm searches  $[a_i, \frac{a_i+b_i}{2}]$  and  $[\frac{a_i+b_i}{2}, b_i]$ , but does not search deeper into either interval. If we list all intervals in  $S$  in increasing order (of either endpoint)  $[a_1, b_1], [a_2, b_2], \dots, [a_s, b_s]$ , then  $f$  increases by a factor of 4 from  $a_i$  to  $b_i$ , but decreases by at least half from  $b_i$  to  $a_{i+1}$  for at most 32 times due to Lemma 2.4. Therefore,  $f(b_{i+1}) > 2f(b_i)$  for all but at most 32 times. Partition  $S$  to at most 32 subsequence of intervals  $S_d = ([a_{d+1}, b_{d+1}], [a_{d+2}, b_{d+2}], \dots, [a_{d+j}, b_{d+j}])$  with  $f(b_{d+i+1}) > 2f(b_{d+i})$  for every  $i \in [j-1]$ . Since for every  $[a_i, b_i] \in S$ ,  $2^{-33} \leq f(b_i) < \frac{2^{33}}{\epsilon}$ ,  $S_d$  contains at most  $O(\log \frac{1}{\epsilon}) = \tilde{O}(1)$  intervals, thus  $S$  also contains  $\tilde{O}(1)$  intervals.

Notice that each interval in the search tree satisfies one of the 4 conditions:

1. The interval is in  $S$ ;
2. The interval is a child of an interval in  $S$ ;
3. In the search tree, the interval is an ancestor of an interval in  $S$ ;
4. In the search tree, the interval is a child of an ancestor of an interval in  $S$ .

Since the length of each searched interval is at least  $\epsilon$ , the depth of the search tree is  $O(\log \frac{1}{\epsilon}) = \tilde{O}(1)$ . Therefore, each interval  $[a_i, b_i] \in S$  defined in the previous paragraph has  $\tilde{O}(1)$  ancestors in the search tree. Thus each interval in  $S$  can define at most  $\tilde{O}(1)$  intervals in the above 4 conditions, which means there are  $\tilde{O}(|S|) = \tilde{O}(1)$  intervals in the search tree. We conclude that the proposed algorithm uses  $\tilde{O}(1)$  queries to  $f$  to output a sequence of points satisfying the conditions in the lemma statement. □

Now we are ready to generalize the algorithm for distributions with monotone PDF to arbitrary regular distributions. The algorithm has almost the same structure, but we need to be more careful in the analysis of query complexity.

**Theorem 2.7.** *Let  $F$  be a smooth regular distribution. Then  $\tilde{O}(\epsilon^{-0.5})$  oracle queries to  $F$  and  $f$  are sufficient to learn the distribution within  $\epsilon$  Lévy-distance.*

*Proof.* Let  $L = \log \frac{1}{\epsilon}$ . We describe the steps of the algorithm as follows. We start with “Step 0” just to keep the correspondence of Steps 1, 2, and 3.

**Step 0** Use binary search to identify  $2(L+1)$  points  $z_i$  such that for the intervals  $I^{(\ell)} = [z_{2\ell-1}, z_{2\ell}]$  the CDF satisfies  $1-2^{-\ell+1} \leq F(v) < 1-2^{-\ell}$  and  $I^{(L+1)}$  contains all values with  $1-F(v) \geq 1-2^{-L} = 1-\epsilon$ . And  $z_{2\ell+1} - z_{2\ell} \leq \epsilon$ . Binary search requires  $O(\log \frac{1}{\epsilon}) = \tilde{O}(1)$  queries for each interval, thus  $\tilde{O}(1)$  queries in total for all intervals as there are only  $\tilde{O}(1)$  intervals. We are going to run an algorithm similar to the case with monotone PDF for each interval and show that  $\tilde{O}(\epsilon^{-0.5})$  queries to  $F$  and  $f$  are enough to learn the CDF for each  $I^{(\ell)}$ , thus  $\tilde{O}(\epsilon^{-0.5})$  queries are enough in total.

**Step 1** All operations below are applied to interval  $I^{(\ell)}$  with fixed  $\ell$ . We only focus on  $I^{(\ell)}$  with  $\ell \leq L$ , as there is no need to learn  $I^{(L+1)}$  with  $\epsilon$  Levy distance. When there is no ambiguity, we omit the superscript  $\ell$ .

For interval  $I^{(\ell)} = [\underline{x}, \bar{x}]$ , by Lemma 2.6 we can find a sequence of points  $x_1 = \underline{x} \leq x_2 \leq \dots \leq x_m = \bar{x}$  in the interval, such that: for any subinterval  $[x_i, x_{i+1})$ , either (a).  $x_{i+1} - x_i \leq \epsilon$ ; or (b).  $f(x_{i+1}) < 2^{-33}$ , which implies for any point  $x \in [x_i, x_{i+1})$ ,  $f(x) \leq 1$ ; or (c). for any two points  $x, x' \in [x_i, x_{i+1})$ ,  $2^{-36}f(x) < f(x') < 2^{36}f(x)$ .

Construct interval collections  $I_0, I_1, \dots, I_K, I_{K+1}$  for some  $K = \tilde{O}(1)$  as follows. For all subintervals  $[x_i, x_{i+1})$  satisfying (a), group them to an interval collection  $I_{K+1}$ ; for all subintervals satisfying (b), group them to an interval collection  $I_0$ ; for all subintervals satisfying (c), add  $[x_i, x_{i+1})$  to interval collection  $I_k$ , if  $2^{k-1} \leq 2^{33}f(x_{i+1}) < 2^k$ . Different from the analysis for monotone non-decreasing  $f$ , for any interval  $[x_i, x_{i+1})$  in  $I_k$ , point  $x \in [x_i, x_{i+1})$  satisfies  $f(x) < 2^{36}f(x_{i+1}) < 2^{k+3}$ , and  $f(x) > 2^{-36}f(x_{i+1}) \geq 2^{k-70}$ . As the intervals in the collections are constructed by Lemma 2.6 which only uses  $\tilde{O}(1)$  queries, the total number of queries to  $f$  used in this step is  $\tilde{O}(1)$ .

**Step 2** For each interval collection  $I_k$  with  $0 \leq k \leq K$ , partition all points in  $I_k$  to intervals  $I_{k,1}, I_{k,2}, \dots$  of length  $\frac{1}{2^{k+\ell}}\sqrt{\epsilon}$ . Denote by  $|I_k|$  the total length of all intervals in  $I_k$ . The number of endpoints added is at most  $\frac{|I_k|}{\sqrt{\epsilon/2^{k+\ell}}} = 2^{k+\ell}\epsilon^{-1/2}|I_k|$  for each interval collection  $I_k$ , and sums up to at most

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

Thus if we query  $f$  and  $F$  for all the endpoints of the intervals from this step, the total number of queries needed is at most  $O(\epsilon^{-1/2})$ .

**Step 3** For each interval  $I_{k,j} = [v_{k,j}, v_{k,j} + 2^{-k-\ell}\sqrt{\epsilon}] \subseteq I_k$  partitioned in Step 2, if  $f(v_{k,j} + 2^{-k-\ell}\sqrt{\epsilon}) - f(v_{k,j}) = t_{k,j} \cdot 2^{k+\ell}\sqrt{\epsilon}$ , we further partition the interval to  $\tilde{t}_{k,j} := \max(1, \lceil t_{k,j} \rceil)$  intervals  $[v_{k,j,0}, v_{k,j,1}]$ ,  $[v_{k,j,1}, v_{k,j,2}]$ ,  $\dots$ ,  $[v_{k,j,\tilde{t}_{k,j}-1}, v_{k,j,\tilde{t}_{k,j}}]$  of length  $\frac{1}{\tilde{t}_{k,j}2^{k+\ell}}\sqrt{\epsilon}$  (with  $v_{k,j,0} = v_{k,j}$  and  $v_{k,j,\tilde{t}_{k,j}} = v_{k,j+1}$ ), and query  $F$  for the newly added endpoints. By the definition of regular distribution, for any  $x \in [v_{k,j}, v_{k,j+1}]$ ,

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

by  $f(x) < 2^{k+3}$  and  $F(x) < 1 - 2^\ell$ . Thus we can bound  $f(x)$  for  $x \in [v_{k,j}, v_{k,j+1}]$  as follows:

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

and

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

Thus for any neighboring points  $v_{k,j,a}, v_{k,j,a+1}$  in the partition with distance  $d = \frac{1}{\tilde{t}_{k,j} 2^{k+\ell}} \sqrt{\epsilon}$ , if  $x \in [v_{k,j,a}, v_{k,j,a+1}]$ , we have  $f(x) \in (\underline{f}, \bar{f})$  with  $\bar{f} - \underline{f} < t_{k,j} \cdot 2^{k+\ell} \sqrt{\epsilon} + 2^{k+8} \sqrt{\epsilon}$ , and

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

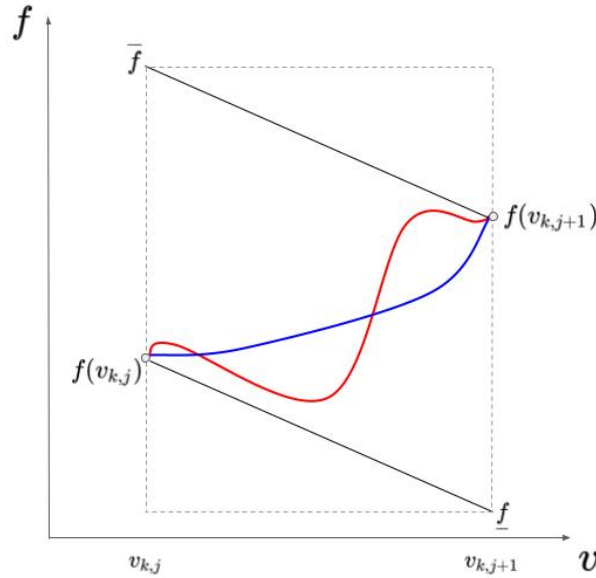


Figure 2: Illustration of the analysis on interval  $[v_{k,j}, v_{k,j+1}]$  with  $t_{k,j} \leq 1$ . The two black solid lines have slope  $-2^{\ell+2k+7}$ , which is a lower bound of  $f'(v)$  in this range by the definition of regular distribution. Given  $f(v_{k,j})$  and  $f(v_{k,j+1})$ , the PDF function  $f$  of a regular distribution cannot go beyond the two black lines. As the CDF of any distribution is the area below the PDF curve, the CDF difference of any two regular distributions given the CDF and the PDF at the  $v_{k,j}$  and  $v_{k,j+1}$  (for example, the red curve and the blue curve) is smaller than the area of the dashed rectangle, which is  $(v_{k,j+1} - v_{k,j})(\bar{f} - \underline{f})$ .

Figure 2 provides an illustrative example of the analysis. By Lemma 2.2 we can learn a distribution within  $\epsilon$  Lévy-distance from  $F$  on  $\bigcup I_{k,j}$  for  $k \in [0, K]$ . As all intervals in  $I_{K+1,j}$  have length at most  $\epsilon$ , we do not need to learn those intervals to satisfy  $\epsilon$  Lévy-distance.

Now we analyze how many additional points (i.e. queries to  $F$ ) are needed in this step.  $\tilde{t}_{k,j} - 1 = \max(1, \lceil t_{k,j} \rceil) - 1 \leq \max(0, t_{k,j})$  additional queries are needed for each interval  $I_{k,j}$ . This means that if  $f$  increases on  $I_{i,j}$ , i.e.  $f(v_{k,j+1}) > f(v_{k,j})$ , then at most  $t_{k,j} = \frac{f(v_{k,j+1}) - f(v_{k,j})}{2^{k+\ell} \sqrt{\epsilon}}$  additional queries are needed; otherwise when  $f$  does not increase on  $I_{i,j}$ , no additional query is required.

Let  $I_{k,\uparrow} = \{x : f'(x) \geq 0, 2^{k-70} < f(x) < 2^{k+3}\}$  and  $I_{k,\downarrow} = \{x : f'(x) < 0, 2^{k-70} < f(x) < 2^{k+3}\}$  be the sets of values such that  $f(x)$  are bounded in range  $(2^{k-70}, 2^{k+3})$ , while  $f$  is non-decreasing and decreasing respectively. Notice that  $I_k$  is a subset of  $I_{k,\uparrow} \cup I_{k,\downarrow}$ . Therefore, we can upper-bound the total number of additional queries in  $I_k$  by

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

As on  $I_{k,\uparrow} \cup I_{k,\downarrow}$ ,  $f(x)$  can increase to at most  $2^{k+3}$ , we have  $\int_{I_{k,\uparrow}} f'(x) dx \leq 2^{k+3} - \int_{I_{k,\downarrow}} f'(x) dx$ . Notice that for any  $x$  such that  $2^{k-70} < f(x) < 2^{k+3}$ , when  $f'(x) < 0$ ,  $f'(x) > -\frac{2f^2(x)}{1-F(x)} > -2^{2k+\ell+7}$ . Also by

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

we have  $|I_{k,\downarrow}| < 2^{70-k-\ell}$ . Therefore,

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

Combine (11) and (12), we have the total number of additional query in  $I_k$  is at most

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

Thus for every  $k$ , the number of queries in Step 3 is  $O(\epsilon^{-1/2})$ , which means that the total number of queries in Step 3 is  $\tilde{O}(\epsilon^{-1/2})$  since there are  $\tilde{O}(1)$  different values of  $k$ . □

**Discussion on  $\beta$ -irregular distributions.** Observe that there is nothing special about the factor of 2 in equation (6). By repeating the proof with a different irregularity coefficient we obtain a proof of Theorem 1.2.

## 2.4 Description complexity of a general (non-smooth) regular distribution

When the regular distribution does not have a smooth CDF function, the PDF function may not exist, and our algorithm for learning the CDF distribution using oracle queries to the PDF function may not apply to general regular distributions. However, we show in the following lemma, that any regular distribution  $F$  can be uniformly approximated by a smooth regular distribution  $\tilde{F}$  within arbitrarily small Lévy-distance  $\delta$ . Thus, to describe  $F$  within  $\epsilon$  Lévy-distance, it suffices to describe  $\tilde{F}$  within  $\epsilon$  Lévy-distance, which is doable via the algorithm in previous subsections.

What we will describe above is a variation of the well-known mollification procedure in real analysis, which replaces a function with its convolution with a very concentrated  $C^\infty$ -function. The only specific detail below is that we argue we can do mollification in a way that is regularity-preserving.

**Lemma 2.8.** *Given any regular distribution  $F$ , for any  $\delta > 0$  there is another regular distribution  $\tilde{F}$  such that  $\text{Lévy}(F, \tilde{F}) \leq \delta$  and  $\tilde{F}$  has no point-masses and has a  $C^\infty$  CDF.*

*Proof.* Let  $R(q) = qF^{-1}(1-q)$  be the revenue curve associated  $F$ . Since  $F$  is regular, the curve  $R$  is concave, but it may not be  $C^\infty$ . There are different techniques to obtain a uniform approximation of concave functions by  $C^\infty$  concave functions, one of which is taking a convolution with a non-negative mollifier (see [AM15] for example). For our purpose, it is enough that it exists a function  $C^\infty$ -concave  $\hat{R}(x)$  such that  $\hat{R}(0) = 0$  and  $|R(x) - \hat{R}(x)| \leq \delta^2/2, \forall x \in [0, 1]$ . Now, if we take  $\tilde{R}(x) = \hat{R}(x) + \delta^2x(1-x)$  we obtain a function that is  $C^\infty$  strongly concave and  $|R(x) - \tilde{R}(x)| \leq \delta^2$ .

Now, define  $\tilde{F}$  such that  $\tilde{R}(q) = q\tilde{F}^{-1}(1-q)$ . Because  $\tilde{R}$  is strongly concave,  $\tilde{F}$  has no point masses. The function  $\tilde{F}$  is strictly increasing and  $C^\infty$ , so  $\tilde{f}(v) > 0$  in its domain. As a consequence,  $\tilde{F}(v)$  is also  $C^\infty$ .

Finally observe that for  $q \in [\delta, 1-\delta]$  we have  $|\tilde{F}^{-1}(q) - F^{-1}(q)| \leq \delta^2/\delta = \delta$ . Now, for a given  $q$ , let  $v = F^{-1}(q)$ . Then we know that:  $v - \delta \leq \tilde{F}^{-1}(q) \leq v + \delta$ . Hence  $\tilde{F}(v - \delta) \leq q = F(v) \leq \tilde{F}(v + \delta)$  which implies that  $\text{Lévy}(F, \tilde{F}) \leq \delta$ .  $\square$

**Putting it all together** We can apply Theorem 2.7 to the smooth approximation from Lemma 2.8 and obtain a  $\tilde{O}(1/\epsilon^{0.5})$  bound on the number of queries to  $F$  and  $f$  to learn any regular distribution.

To complete the result, we still need to argue about the number of bits necessary to represent the result since both the points  $x$  queried and the results  $F(x)$  and  $f(x)$  are in principle real numbers. However, we can see each query to  $F$  (or  $f$ ) as follows: query  $F(x)$  (or  $f(x)$ ) where  $x$  is an integer times  $\epsilon$ , and an approximate answer within  $F(x) \pm \epsilon$  (or  $f(x) \pm \epsilon$ ). All of the lemmas and algorithms already accommodate  $\pm\epsilon$  errors both in  $x$ ,  $F$ , and  $f$ . This automatically leads to a  $\tilde{O}(1/\epsilon^{0.5})$  bit complexity bound. Combined with the lower bound in Theorem 2.1, this proves our main result (Theorem 1.1).

### 3 Pricing Query Complexity of Learning a Regular Distribution

In this section, we study the pricing query complexity of learning a regular distribution within  $\epsilon$  Lévy-distance. In particular, for each pricing query, we submit a price  $x$ , and a value  $v \sim F$  is realized. We cannot directly observe  $v$ , but can observe  $\text{sign}(v-x) \in \{-1, +1\}$ . By Chernoff bound,  $\tilde{O}(1/\epsilon^2)$  pricing queries are sufficient to learn  $F(x)$  with  $\epsilon$  additive error. However, the  $\tilde{O}(\epsilon^{-0.5})$  description complexity upper bound result is not sufficient to show that the pricing query complexity of learning a regular distribution on  $[0, 1]$  within  $\epsilon$  Lévy-distance. There are two problems remaining. Firstly, the algorithms for showing description complexity upperbound results rely on queries to not only the CDF, but also the PDF. Secondly, for non-smooth regular distributions, we prove the description complexity result by learning a smooth regular distribution  $\tilde{F}$  that is uniformly close to  $F$ , but not exactly  $F$ . Thus when we query  $F(x)$ , we are only able to get  $\tilde{F}(x + \epsilon_1) + \epsilon_2$  where  $|\epsilon_1|$  can be arbitrarily small (and we use the following bound  $|\epsilon_1| \leq \frac{\epsilon^2}{2}$ ), and  $|\epsilon_2| < \epsilon$ . Thus, the problem we want to solve is equivalent to the following theorem:

**Theorem 3.1.** *For any regular distribution  $F$  with smooth PDF function, suppose that we have perturbed oracle access to  $F$  that every time when querying  $x$ , the oracle  $F^*$  returns  $F^*(x) = F(x + \epsilon_1) + \epsilon_2$  for some unknown  $\epsilon_1 \in [-\frac{\epsilon^2}{2}, \frac{\epsilon^2}{2}]$ , and  $\epsilon_2 \in [-\epsilon, \epsilon]$ . Then  $\tilde{O}(\epsilon^{-0.5})$  queries to the perturbed oracle  $F^*$  are sufficient to learn  $F$  within  $\epsilon$  Lévy-distance.*

#### 3.1 Query complexity of distributions with convex $F$

Before proving Theorem 3.1 for general regular distributions, we first prove the theorem for a distribution with convex  $F$  to give more intuition. The spirit of the general case is almost identical.



**Theorem 3.2.** For any regular distribution  $F$  with smooth PDF function, suppose that we have perturbed oracle access to  $F$  that every time when querying  $x$ , the oracle  $F^*$  returns  $F^*(x) = F(x + \epsilon_1) + \epsilon_2$  for some unknown  $\epsilon_1 \in [-\frac{\epsilon^2}{2}, \frac{\epsilon^2}{2}]$ , and  $\epsilon_2 \in [-\epsilon, \epsilon]$ . Then  $\tilde{O}(\epsilon^{-0.5})$  queries to the perturbed oracle  $F^*$  are sufficient to learn  $F$  within  $\epsilon$  Lévy-distance.

*Proof.* We first show how to define a (perturbed) query to  $f$  via the perturbed oracle. Suppose that we want to query  $f$ . When we query  $x + \gamma$  and  $x - \gamma$  to the perturbed oracle  $F^*$ , we will get  $F^*(x + \gamma) = F(x + \gamma + \epsilon_1) + \epsilon_2$ , and  $F^*(x - \gamma) = F(x - \gamma + \epsilon_3) + \epsilon_4$  for some unknown  $\epsilon_1, \epsilon_3 \in [-\frac{\epsilon^2}{2}, \frac{\epsilon^2}{2}]$ , and  $\epsilon_2, \epsilon_4 \in [-\epsilon, \epsilon]$ . Let

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

By Lagrange mean value theorem, for smooth function  $F$ , there exists  $x^* \in [x - \gamma + \epsilon_3, x + \gamma + \epsilon_1]$  such that

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

Then we have

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

for some  $\delta_1, \delta_2, \gamma_1$  where  $|\delta_1| \leq \frac{\epsilon^2}{2\gamma}$ ,  $|\delta_2| \leq \frac{\epsilon}{\gamma}$ ,  $|\gamma_1| < 2\gamma$ . Later, we will make perturbed queries to  $f$  via  $f_\gamma^*(x)$ , and specify the accuracy parameter  $\gamma$  needed in each step of the learning algorithm.

**Changes to Lemma 2.2.** Suppose that we are only able to identify  $F_i \in [0, 1]$  such that  $F(x_i + \epsilon_i) \in [F_i - \epsilon, F_i + \epsilon]$  for some unknown small  $\epsilon_i \in [-\epsilon^2/2, \epsilon^2/2]$ . We show that Lemma 2.2 still holds. Observe that the distribution constructed from input  $(x_1, \dots, x_m)$  and  $(F_1, \dots, F_m)$  in the proof of Lemma 2.2 is only  $O(\epsilon)$  Lévy-distance from the distribution constructed from input  $(x'_1, x'_2, \dots, x'_m) = (x_1 + \epsilon_1, x_2 + \epsilon_2, \dots, x_m + \epsilon_m)$  and  $(F_1, \dots, F_m)$ . Furthermore, as either  $x'_{i+1} - x'_i$  is  $O(\epsilon)$ , or  $(x'_{i+1} - x'_i)(\bar{f}_i - \underline{f}_i) < 2(x_{i+1} - x_i)(\bar{f}_i - \underline{f}_i) = O(\epsilon)$  (since  $x'_{i+1} - x'_i \leq x_{i+1} - x_i + \epsilon^2 < 2(x_{i+1} - x_i)$ ), the distribution constructed from input  $(x'_1, \dots, x'_m)$  and  $(F_1, \dots, F_m)$  is within  $(\epsilon)$  Lévy-distance from  $F$ . Thus the distribution constructed from input  $(x_1, \dots, x_m)$  and  $(F_1, \dots, F_m)$  is within  $O(\epsilon)$  Lévy-distance from  $F$  on  $[x_1, x_m]$ . To summarize, to apply Lemma 2.2, we do not need  $F_i$  to be an estimate of  $F(x)$  with  $O(\epsilon)$  error; letting  $F_i = F^*(x)$  is good enough.

**Changes to Theorem 2.3.** We now describe in detail when  $F$  is convex (i.e. with monotone  $f$ ), how does the algorithm change.

**Step 1** We set  $K = \log(1/\epsilon)$  and use binary search to identify points  $x_0, x_1, \dots, x_{2K+3}$  such that for each  $k = 0, \dots, K + 1$  we have  $|x_{2k+1} - x_{2k+2}| = O(\epsilon)$  and for the interval  $I_k = [x_{2k}, x_{2k+1}]$  it holds that:

- $f(v) \leq 2$  for  $v \in I_0$
- $2^{k-2} \leq f(v) \leq 2^{k+1}$  for  $v \in I_k, k = 1, \dots, K$

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

Notice that now we do not have access to  $f$ . Thus in order to perform the binary search, we query  $f_\gamma^*$  with  $\gamma = 4\epsilon$ . By (13), for any query  $x$ ,  $f_\gamma^*(x) = (1 + \delta_1)f(x + \gamma_1) + \delta_2$  for some  $\delta_1, \delta_2, \gamma_1$  with  $|\delta_1| \leq \frac{\epsilon^2}{2\gamma} \leq \frac{1}{8}$ ,  $|\delta_2| \leq \frac{\epsilon}{\gamma} = \frac{1}{4}$ ,  $|\gamma_1| \leq 2\gamma = 8\epsilon$ . Thus

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

Therefore, when  $f_\gamma^*(x) \geq 2^{k-1}$ ,  $f(x + 8\epsilon) > 2^{k-2}$ ; when  $f_\gamma^*(x) \leq 2^k$ ,  $f(x - 8\epsilon) < 2^{k+1}$ . Thus, suppose we do binary search on  $f_\gamma^*$  to find a sequence of points  $x'_0, x'_1, \dots, x'_{2K+3}$  such that for every  $k = 0, \dots, K$ ,  $f_\gamma^*(x'_{2k+1}) \leq 2^k$ ,  $f_\gamma^*(x'_{2k+2}) \geq 2^k$ . Then if we set  $x_{2k+1} = x'_{2k+1} - 8\epsilon$ ,  $x_{2k+2} = x'_{2k+2} + 8\epsilon$ , the sequence  $x_0, \dots, x_{2K+3}$  satisfies all the properties required for the step.

Identifying each point still takes  $O(\log(1/\epsilon)) = \tilde{O}(1)$  queries to  $f_\gamma^*(\cdot)$  (thus  $F(\cdot)$ ). Given that we have  $\tilde{O}(1)$  such points, we used a total of  $\tilde{O}(1)$  queries. The last interval  $I_{K+1}$  again satisfies the conditions of Lemma 2.2.

**Step 2** For each interval  $I_k$  with  $k = 0, \dots, K$ , partition  $I_k$  to intervals  $I_{k,1}, I_{k,2}, \dots$  of length  $\frac{1}{2^k}\sqrt{\epsilon}$ . The number of endpoints added is at most  $\frac{|I_k|}{\sqrt{\epsilon}/2^k} = 2^k\epsilon^{-1/2}|I_k|$  for each interval  $I_k$ , and sums up to at most

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

The only change compared to the previous section is that  $2^k \leq 4f(v)$  instead of  $2f(v)$  in the second line. Thus if we query  $F^*$  for all the endpoints of the intervals from the first two steps, the total number of queries needed is at most  $O(\epsilon^{-1/2})$ .

**Step 3** For each interval  $I_{k,j} = [v_{k,j}, v_{k,j+1}] = [v_{k,j}, v_{k,j} + 2^{-k}\sqrt{\epsilon}] \subseteq I_k$  partitioned in Step 2, assume that  $2^{-k}\sqrt{\epsilon} > 10\epsilon$ , otherwise as  $|I_{k,j}| = O(\epsilon)$ , it already satisfies the condition of Lemma 2.2.

Define  $\underline{f}_{k,j}$  and  $\bar{f}_{k,j}$  as follows. Let  $\gamma^* = 2^{-k-2}\sqrt{\epsilon}$ . If  $I_{k,j}$  is the leftmost interval in  $I_k$ , then  $\underline{f}_{k,j} = 2^{k-2}$ , otherwise  $\underline{f}_{k,j} = f_{\gamma^*}^*(v_{k,j} - 2^{-k-1}\sqrt{\epsilon}) - 2^{k+3}\sqrt{\epsilon}$ ; if  $I_{k,j}$  is the rightmost interval in  $I_k$ , then  $\bar{f}_{k,j} = 2^{k+1}$ , otherwise  $\bar{f}_{k,j} = f_{\gamma^*}^*(v_{k,j} + 2^{-k}\sqrt{\epsilon} + 2^{-k-1}\sqrt{\epsilon}) + 2^{k+3}\sqrt{\epsilon}$ . Notice that by (13), if  $\underline{f}_{k,j} \neq 2^{k-2}$ , then

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

for some  $\delta_1, \delta_2, \gamma_1$  where  $|\delta_1| \leq \frac{\epsilon}{2\gamma^*} = 2^{k+3}\epsilon^{1.5}$ ,  $\delta_2 \leq \frac{\epsilon}{\gamma^*} = 2^{k+2}\sqrt{\epsilon}$ ,  $|\gamma_1| < 2\gamma = 2^{-k-1}\sqrt{\epsilon}$ . This means that  $v_{k,j} - 2^{-k-1}\sqrt{\epsilon} + \gamma_1 \in [v_{k,j} - 2^{-k}\sqrt{\epsilon}, v_{k,j}]$ . Since  $f(v_{k,j}) \leq 2^{K+1} = \frac{2}{\epsilon}$ , we have  $f(v_{k,j})|\delta_1| \leq \frac{2}{\epsilon} \cdot 2^{k+1}\epsilon^{1.5} = 2^{k+2}\sqrt{\epsilon}$ . Then

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

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

This means that  $\underline{f}_{k,j}$  is indeed a lower bound of  $f(v_{k,j})$ . Symmetrically,  $\bar{f}_{k,j}$  is indeed an upper bound of  $f(v_{k,j+1})$ .

If  $\bar{f}_{k,j} - \underline{f}_{k,j} = t_{k,j}2^k\sqrt{\epsilon}$ , then partition the interval to  $\tilde{t}_{k,j} := \max(1, \lceil t_{k,j} \rceil - 2^5)$  intervals of length  $\frac{1}{\tilde{t}_{k,j}2^k}\sqrt{\epsilon}$ , and query  $F$  for the newly added endpoints. As any neighboring points  $\bar{v}, \underline{v}$  have distance  $\bar{v} - \underline{v} \leq \frac{1}{\tilde{t}_{k,j}2^k}\sqrt{\epsilon}$ . Hence this interval satisfies the second condition in Lemma 2.2 :

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

which allows us to learn the distribution up to  $\epsilon$  Lévy-distance.

Finally, we only need to bound the number of queries needed in this step. Observe that if  $I_{k,j}$  is not the leftmost interval in  $I_k$ ,

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

Symmetrically, if  $I_{k,j}$  is not the rightmost interval in  $I_k$ ,  $\bar{f}_{k,j} \leq f(v_{k,j+2}) + 2^{k+4}\sqrt{\epsilon}$ . Thus,

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

Since the interval  $I_{k,j}$  is partitioned to  $\max(1, \lceil t_{k,j} \rceil - 2^5)$  intervals, the number of queries to  $f_{\gamma^*}$  is at most  $\frac{f(v_{k,j+2}) - f(v_{k,j-1})}{2^k\sqrt{\epsilon}}$ . If we sum up  $\frac{f(v_{k,j+2}) - f(v_{k,j-1})}{2^k\sqrt{\epsilon}}$  for every  $j$ , we get the total number of queries to  $f_{\gamma^*}$  is at most<sup>1</sup>

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

by simplifying the telescoping sum. Thus in this step, for each  $k$ , the number of queries on  $F^*$  for added points is  $\tilde{O}(\epsilon^{-1/2})$ . Also, the number of queries on  $f_{\gamma^*}$  for  $I_k$  is twice the total number of endpoints in  $I_{k,j}$ , which is  $O(\epsilon^{-1/2})$  by Step 2. As there are only  $\tilde{O}(1)$  different values for  $k$ , the total number of queries on  $F^*$  is  $\tilde{O}(\epsilon^{-1/2})$ .

To summarize, we have described how to modify the algorithm for learning a distribution with smooth convex CDF with oracle queries to  $F$  and  $f$ , to an algorithm with only oracle queries to  $F^*$ . The query complexity is asymptotically the same, which is  $\tilde{O}(\epsilon^{-1/2})$ .  $\square$

### 3.2 Query complexity of general regular distributions

Essentially the same modifications described for the convex CDF can be applied to the general regular distributions, obtaining a proof of Theorem 3.1. We omit the details since it is essentially a re-writing of the proof of Theorem 2.7 and instead describe the main modifications.

<sup>1</sup>To be more precise, for the rightmost interval in  $I_k$ ,  $f(v_{k,j+2})$  is replaced by  $2^{k+1}$ ; for the leftmost interval in  $I_k$ ,  $f(v_{k,j-1})$  is replaced by  $2^{k-2}$ .

In Step 0, the partition of  $[0, 1]$  to intervals with quantile within a factor of 2 can also be done via queries to  $F^*$ . In Step 1 of the proof of Theorem 2.7, the partition is doable with queries to  $f$  replaced by queries to  $f_{O(\epsilon)}^*$ , as Lemma 2.6 only requires estimating  $f$  within a constant factor, which is the same as Step 1 for Theorem 3.2. The only difference is that the value range of  $f$  in each interval may get expanded by a constant factor. This leads to the number of queries in Step 2 increasing by a constant factor. In the analysis of Step 3, similar to Step 3 of the modified algorithm in Theorem 3.2 the accuracy needed for estimating  $f$  is  $O(2^{k+\ell}\sqrt{\epsilon})$ , which is achievable via queries to  $f_{O(2^{-k-\ell}\sqrt{\epsilon})}^*$ .

A simple Chernoff bound shows that a query to  $F^*$  (defined in Theorem 3.1) can be simulated by  $O(1/\epsilon^2)$  pricing queries. Combining Theorem 3.1 with Chernoff bound we obtain a proof of Theorem 1.3, matching the lower bound established in [PLSTW23].

## 4 Mixture Distributions

Our second application is the analysis of mixtures of regular distributions. Corollaries 1.4 and 1.5 are straightforward from Theorem 1.1.

We end the paper with the proof of the more surprising result (Theorem 1.6) that even though mixtures of  $O(1)$  regular distributions can be described with  $\tilde{O}(1/\epsilon^{0.5})$  bits, they require  $\Omega(1/\epsilon^3)$  pricing queries to be learned within  $\epsilon$  errors, the same asymptotic amount as a general distribution.

**Theorem** (Restatement of Theorem 1.6). *Let  $\mathcal{C}$  be the class of distributions supported in  $[0, 1]$  that can be written as a mixture of two regular distributions. To estimate the CDF of a distribution in that class within  $\epsilon$  in Lévy-distance we require at least  $\Omega(1/\epsilon^3)$  pricing queries.*

*Proof of Theorem 1.6.* We will use Lemma 3.5 in [PLSTW23] which says that it takes  $\Omega(1/\epsilon^2)$  pricing queries to distinguish between any two distributions with CDFs  $F_1$  and  $F_2$  such that

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

Now, we will construct a set  $S$  of  $\Omega(1/\epsilon)$  distributions such that each of which is: (1) a mixture of two regular distributions and (2) each follows Equation 14. Moreover, the Lévy-distance between any two pairs is  $\Omega(\epsilon)$ .

Construct a class of regular distributions parameterized by  $a < 2, \delta > 0$  as follows. For a distribution with PDF  $f_{a,\delta}$ ,

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

In other words,  $f_a$  is a distribution that is uniform with density  $a$ , until the point  $v = \frac{1-\delta}{a}$  where the CDF reaches  $1 - \delta$ . Then  $f_{a,\delta}$  decreases linearly to 0, and remains 0 afterwards. We first show that any distribution in this class is regular. We want to show that for any  $x \in [\frac{1-\delta}{a}, \frac{1+\delta}{a}]$ ,  $f'_{a,\delta}(x) \geq -\frac{2f_{a,\delta}^2(x)}{1-F_{a,\delta}(x)}$ . Firstly, in this range,  $f'_{a,\delta}(x) = -\frac{a^2}{2\delta}$ . Secondly, as in this range  $1 - F_{a,\delta}(x) = \frac{1}{2}(\frac{1+\delta}{a} - x)f_{a,\delta}(x)$ , we have

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

Thus for any  $a, \delta$ ,  $f_{a,\delta}$  is the PDF of a regular distribution. Consider any distribution with PDF  $g_{a,\delta}(v) = 2 - f_{a,\delta}(v)$  for every  $v \in [0, 1]$ . Then such a distribution is also regular, since its PDF is monotone non-decreasing. Furthermore, the uniform mixture of  $f_{a,\delta}$  and  $g_{a,\delta}$  is a uniform distribution on  $[0, 1]$ .

Let  $F$  be the uniform distribution on  $[0, 1]$ . Define the following class of distributions with CDF  $F_a$  and PDF  $f_a$  parameterized by  $a \in [1.5, 1.6]$  as follows. For any  $v \in [0, 1]$ ,

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

Then  $f_a$  is a uniform distribution, except for  $v \in [\frac{1-\epsilon}{a}, \frac{1+\epsilon}{a}]$ . For any  $v \in [\frac{1-\epsilon}{a}, \frac{1+\epsilon}{a}]$ ,  $F_{a,\epsilon}(v)$  and  $F_{a,\epsilon/2}(v)$  differs by at most  $\epsilon$ , thus  $F_a(v) = \frac{F_{a,\epsilon/2}(v) + G_{a,\epsilon}(v)}{2}$  and  $F(v) = \frac{F_{a,\epsilon}(v) + G_{a,\epsilon}(v)}{2}$  differs by at most  $\epsilon$ . Furthermore, since  $F_a(v)$  and  $F(v)$  are both constant away from 1, thus (14) is satisfied and  $\Omega(1/\epsilon^2)$  pricing queries are needed to distinguish  $F_a$  and  $F$ .

Observe that for any  $a$ ,  $F_a$  and the uniform distribution  $F$  has Lévy-distance at least  $\Omega(\epsilon)$ , as  $F_{a,\epsilon/2}(\frac{1-\epsilon/2}{a}) - F_{a,\epsilon}(\frac{1-\epsilon/2}{a}) = \Omega(\epsilon)$ . Take  $S = \{F_a | a = 1.5 + 4k\epsilon \text{ for } k = 1, 2, \dots, \frac{1}{40\epsilon}\}$ . As any  $F_a \in S$  and the uniform distribution  $F$  differs only on  $[\frac{1-\epsilon}{a}, \frac{1+\epsilon}{a}]$ , we know that  $\Omega(1/\epsilon^2)$  pricing queries are needed on  $[\frac{1-\epsilon}{a}, \frac{1+\epsilon}{a}]$  to distinguish  $F_a$  and  $F$ . Also since for every  $a' < a$ ,  $\frac{1+\epsilon}{a} < \frac{1-\epsilon}{a'}$  for  $a, a' \in [1.5, 1.6]$  and small enough  $\epsilon$ , we know that for different  $a$ , the interval where  $F_a$  and  $F$  differs are non-overlapping. Thus given any distribution  $F_a$  where  $a = 1.5 + 4k\epsilon$  for some  $k = 1, 2, \dots, \frac{1}{40\epsilon}$ ,  $\Omega(1/\epsilon^3)$  queries are required to learn a distribution within  $o(\epsilon)$  Lévy-distance from  $F_a$ . □

## References

- [AB20] Amine Allouah and Omar Besbes. Prior-independent optimal auctions. *Management Science*, 66(10):4417–4432, 2020.
- [AHN<sup>+</sup>19] Saeed Alaei, Jason Hartline, Rad Niazadeh, Emmanouil Pountourakis, and Yang Yuan. Optimal auctions vs. anonymous pricing. *Games and Economic Behavior*, 118:494–510, 2019.
- [AM15] Daniel Azagra and Carlos Mudarra. Global approximation of convex functions by differentiable convex functions on banach spaces. *Journal of Convex Analysis*, 22(4):1197–1205, 2015.
- [BCD20] Johannes Brustle, Yang Cai, and Constantinos Daskalakis. Multi-item mechanisms without item-independence: Learnability via robustness. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 715–761, 2020.
- [BK96] Jeremy Bulow and Paul Klemperer. Auctions versus negotiations. *American Economic Review*, 86(1):180–194, March 1996.
- [CDIZ22] Yeshwanth Cherapanamjeri, Constantinos Daskalakis, Andrew Ilyas, and Manolis Zampetakis. Estimation of standard auction models. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 602–603, 2022.
- [CLMN14] L Elisa Celis, Gregory Lewis, Markus Mobius, and Hamid Nazerzadeh. Buy-it-now or take-a-chance: Price discrimination through randomized auctions. *Management Science*, 60(12):2927–2948, 2014.

- [CR14] Richard Cole and Tim Roughgarden. The sample complexity of revenue maximization. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 243–252, 2014.
- [DRY10] Peerapong Dhangwatnotai, Tim Roughgarden, and Qiqi Yan. Revenue maximization with a single sample. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 129–138, 2010.
- [FHH13] Hu Fu, Jason Hartline, and Darrell Hoy. Prior-independent auctions for risk-averse agents. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, pages 471–488, 2013.
- [FILS15] Hu Fu, Nicole Immorlica, Brendan Lucier, and Philipp Strack. Randomization beats second price as a prior-independent auction. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 323–323, 2015.
- [GLMN21] Negin Golrezaei, Max Lin, Vahab Mirrokni, and Hamid Nazerzadeh. Boosted second price auctions: Revenue optimization for heterogeneous bidders. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 447–457, 2021.
- [HMR15] Zhiyi Huang, Yishay Mansour, and Tim Roughgarden. Making the most of your samples. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 45–60, 2015.
- [HR09] Jason D Hartline and Tim Roughgarden. Simple versus optimal mechanisms. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 225–234, 2009.
- [KL<sup>+</sup>23] Robin Kothari, Troy Lee, , Ilan Newman, and Mario Szegedy. *Query Complexity*. World Scientific Publishing Company Pte Limited, 2023.
- [MR15] Jamie H Morgenstern and Tim Roughgarden. On the pseudo-dimension of nearly optimal auctions. *Advances in Neural Information Processing Systems*, 28, 2015.
- [MS83] Roger B Myerson and Mark A Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of economic theory*, 29(2):265–281, 1983.
- [Mye81] Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.
- [PLPV16] Renato Paes Leme, Martin Pal, and Sergei Vassilvitskii. A field guide to personalized reserve prices. In *Proceedings of the 25th international conference on world wide web*, pages 1093–1102, 2016.
- [PLSTW23] Renato Paes Leme, Balasubramanian Sivan, Yifeng Teng, and Pratik Worah. Pricing query complexity of revenue maximization. *ACM-SIAM Symposium on Discrete Algorithms (SODA23)*, 2023.
- [Pré73] András Prékopa. On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum*, 34:335–343, 1973.
- [RTCY12] Tim Roughgarden, Inbal Talgam-Cohen, and Qiqi Yan. Supply-limiting mechanisms. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 844–861, 2012.

- [SS13] Balasubramanian Sivan and Vasilis Syrgkanis. Vickrey auctions for irregular distributions. In *International Conference on Web and Internet Economics*, pages 422–435. Springer, 2013.