Testing Piecewise Functions

Steve Hanneke, Liu Yang

Abstract

This work explores the query complexity of property testing for general piecewise functions on the real line, in the active and passive property testing settings. The results are proven under an abstract zero-measure crossings condition, which has as special cases piecewise constant functions and piecewise polynomial functions. We find that, in the active testing setting, the query complexity of testing general piecewise functions is independent of the number of pieces. We also identify the optimal dependence on the number of pieces in the query complexity of passive testing in the special case of piecewise constant functions.

Keywords: Property testing; Active testing; Learning theory; Real-valued functions

1. Introduction

Property testing is a well-studied class of problems, in which an algorithm must decide (with high success probability) from limited observations of some object whether a given property is satisfied, or whether the object is in fact far from all objects with the property [1, 2]. In many property testing problems, it is natural to describe the object as a function f mapping some instance space $\mathcal{X}$ to a value space $\mathcal{Y}$, and the property as a family of functions $\mathcal{F}$. The property testing problem is then equivalently stated as determining whether $f \in \mathcal{F}$ or whether f is far from all elements of $\mathcal{F}$, for a natural notion of distance. In this setting, the observations are simply $(x, f(x))$ points. The setting has been studied in several variations, depending on how the x observation points are selected (e.g., at random or by the algorithm). In particular, this type of property testing problem is closely related to the PAC *learning* model of [3].[1] One of the main theoretical questions in the study of property testing is how many observations are needed by the optimal tester, a quantity known as the *query complexity* of the testing problem. In the above context, this question is most interesting when one can show that the query complexity of the testing problem is significantly smaller than the query complexity of the corresponding

Email addresses: `steve.hanneke@gmail.com` (Steve Hanneke), `liu.yang0900@outlook.com` (Liu Yang)

PAC learning problem [4, 5, 6, 7, 8, 9].

The property testing literature is by now quite broad, and includes testing algorithms and characterizations of their query complexities for many function classes $\mathcal{F}$; see [10, 11, 12] for introductions to this literature and some of its key techniques. However, nearly all of the work on the above type of property testing problem has focused on the special case of *binary* functions, where $|\mathcal{Y}| = 2$, or in some cases with $\mathcal{Y}$ a general finite field. In this article, we are interested in the study of more general types of functions, including real-valued functions. Previous works on testing real-valued functions include testers for whether f is monotone [13, 14, 15], unate [16], or Lipschitz [15, 17]. In the present work, we study a general problem of testing *piecewise* functions. Specifically, we consider a scenario where $\mathcal{X} = \mathbb{R}$, and where the function class $\mathcal{F}$ can be described as a k-*piecewise* function (for a given $k \in \mathbb{N}$), where each piece is a function in a given base class of functions $\mathcal{H}$. Formally, defining $t_0 = -\infty$ and $t_k = \infty$, for any $t_1, \ldots, t_{k-1} \in \mathbb{R}$ with $t_1 \leq \cdots \leq t_{k-1}$, and any $h_1, \ldots, h_k \in \mathcal{H}$, define (for any $x \in \mathbb{R}$)

$$f\left(x; \{h_i\}_{i=1}^{k}, \{t_i\}_{i=1}^{k-1}\right) = h_i(x) \text{ for the } i \text{ such that } t_{i-1} < x \leq t_i.$$

Then we consider classes $\mathcal{F} = \mathcal{F}_k(\mathcal{H})$ defined as

$$\mathcal{F}_k(\mathcal{H}) = \left\{ f\left(\cdot; \{h_i\}_{i=1}^{k}, \{t_i\}_{i=1}^{k-1}\right) : h_1, \ldots, h_k \in \mathcal{H}, t_1 \leq \cdots \leq t_{k-1} \right\}.$$

In the results below, we will be particularly interested in the dependence on k in the query complexity of the testing problem. To be clear, the values $t_1, \ldots, t_{k-1}$ and functions $h_1, \ldots, h_k$ are all free parameters, with different choices of these yielding different functions, all contained in $\mathcal{F}_k(\mathcal{H})$. Thus, in the testing problem (defined formally below), the algorithm may directly depend on k and $\mathcal{H}$, but in the case of $f \in \mathcal{F}_k(\mathcal{H})$ the specific values of $t_1, \ldots, t_{k-1}$ and functions $h_1, \ldots, h_k \in \mathcal{H}$ specifying f are all considered as *unknown*.

In this work, our primary running example is the scenario where $\mathcal{Y} = \mathbb{R}$ and $\mathcal{H}$ is the set of *degree-p polynomials*: $\mathcal{H} = \{x \mapsto \sum_{i=0}^{p} \alpha_i x^i : \alpha_0, \ldots, \alpha_p \in \mathbb{R}\}$, so that $\mathcal{F}_k(\mathcal{H})$ is the set of k-piecewise degree-p polynomials. A further interesting special case of this is when $p = 0$, in which case $\mathcal{F}_k(\mathcal{H})$ is the set of k-piecewise *constant* functions. However, our general analysis is more abstract, and will also apply to many other interesting function classes $\mathcal{H}$.

Specifically, for the remainder of this article, we consider $\mathcal{Y}$ as an arbitrary nonempty set (equipped with an appropriate σ-algebra), and $\mathcal{H}$ as an arbitrary nonempty set of measurable functions $\mathcal{X} \to \mathcal{Y}$ satisfying the following

[1]In the PAC (*probably approximately correct*) learning model proposed by Valiant [3] (and previously by Vapnik and Chervonenkis [4]), it is assumed that $f \in \mathcal{F}$, and an algorithm is tasked with choosing any $\hat{f}$ that is (with high probability) within distance ϵ of f, given access to a finite number of $(x, f(x))$ pairs drawn at random (or in some variants, selected by the algorithm).

zero-measure crossings property:

$$\forall h, h' \in \mathcal{H}, \quad h \neq h' \implies \lambda(\{x : h(x) = h'(x)\}) = 0, \tag{1}$$

where λ denotes the Lebesgue measure. It is well known that this property is satisfied by polynomial functions: for any two polynomial functions h, h' of degree p, if they agree on a nonzero measure set of points, then we can find in that set distinct points $x_1, \ldots, x_{p+1}$ on which they agree, but since the values on any $p + 1$ distinct points uniquely determine the polynomial function, it must be that $h = h'$. Thus, the analysis below indeed applies to piecewise polynomial functions as a special case. The zero-measure crossings property is also satisfied by many other interesting function classes, such as shifted sine functions $\{x \mapsto \sin(x + t) : t \in \mathbb{R}\}$ or normal pdfs $\{x \mapsto c \cdot e^{-(x-t)^2/2} : t \in \mathbb{R}\}$.

The testing problem is defined as follows. Fix a value $\epsilon \in (0, 1)$ and a probability measure $\mathcal{P}$ over $\mathcal{X}$, and for any measurable $f, g : \mathcal{X} \to \mathcal{Y}$, define $\rho(f, g) = \mathcal{P}(x : f(x) \neq g(x))$, the $L_0(\mathcal{P})$ (pseudo)distance between f and g. Further define $\rho(f, \mathcal{F}) = \inf_{g \in \mathcal{F}} \rho(f, g)$, the distance of f from a set of functions $\mathcal{F}$. In the active testing protocol, the algorithm samples a number s of iid unlabeled examples from $\mathcal{P}$, and then interactively queries for the f values at points of its choosing from among these: that is, it chooses one of the s examples x, queries for its value $f(x)$, then selects another of the s examples x', queries for its value $f(x')$, and so on, until it eventually halts and produces a decision of either Accept or Reject. The following definition is taken from [18].

Definition 1. *An s-sample q-query ϵ-tester for $\mathcal{F}$ under the distribution $\mathcal{P}$ is a randomized algorithm A that draws a sample S of size at most s iid from $\mathcal{P}$, sequentially queries for the value of f on at most q points of S, and satisfies the following properties (regardless of the identity of f):*

- *if $f \in \mathcal{F}$, it decides* Accept *with probability at least $\frac{2}{3}$.*
- *if $\rho(f, \mathcal{F}) \geq \epsilon$, it decides* Reject *with probability at least $\frac{2}{3}$.*

We will be interested in identifying values of q for which there exist s-sample q-query ϵ-testers for $\mathcal{F}$, known as the *query complexity*; we are particularly interested in this when s is polynomial in k, $1/\epsilon$, and the complexity of $\mathcal{H}$ (defined below). In the special case that $q = s$, so that the algorithm queries the values at *all* of the points in S, the algorithm is called a *passive tester* [2], whereas in the general case of $q \leq s$ it is referred to as an *active tester* [18] (in analogy to the setting of *active learning* studied in the machine learning literature [8, 19]).

Remark on general distributions $\mathcal{P}$: For simplicity, we will focus on $\mathcal{P} = \text{Uniform}(0, 1)$ in this work. However, all of our results easily generalize to all distributions $\mathcal{P}$ over $\mathbb{R}$ absolutely continuous with respect to Lebesgue measure. Specifically, [18] discuss a simple technique which produces this generalization, by using the empirical distribution to effectively rescale the real line so that the distribution appears approximately uniform. One can show that this technique is also applicable in the present more-general context as well. The rescaling effectively changes the function class $\mathcal{H}$, but its graph dimension (defined below) remains unchanged, and the zero-measure crossings property is preserved due to $\mathcal{P}$ being absolutely continuous. In the special case of testing piecewise *constant* functions, even the restriction to absolutely continuous distributions can be removed, as then the zero-measure crossings property always holds. The interested reader is referred to [18] for the details of this rescaling technique.

1.1. Related Work: Testing Unions of Intervals

The work of [18] explored the query complexity of testing in both the active and passive models, for a variety of function classes, but all under the restriction $\mathcal{Y} = \{0, 1\}$. Of the results of [18], the most relevant to the present work are results on the query complexity of testing *unions of intervals*: $x \mapsto \mathbb{I}\left[x \in \bigcup_{i=1}^{n}[t_{2i-1}, t_{2i}]\right]$, for a fixed $n \in \mathbb{N}$, defined for all nondecreasing sequences $t_1, \ldots, t_{2n} \in \mathbb{R} \cup \{\pm\infty\}$. They specifically find that the query complexity of testing unions of intervals is $O(1/\epsilon^4)$ — independent of n — in the active testing setting, and is $O(\sqrt{n}/\epsilon^5)$ in the passive testing setting, with an $\Omega(\sqrt{n})$ lower bound; these results strengthened an earlier result of Kearns and Ron [20]. Note that unions of intervals are a special case of piecewise constant functions, and indeed the techniques we employ in constructing the tester below closely parallel the analysis of unions of intervals by [18]. However, to extend that approach to general piecewise functions — even piecewise *constant* functions — requires careful consideration about what the appropriate generalization of the technique should be. In particular, the original proof involved a *self-correction* step, wherein f is replaced by a smoothed function, which is then rounded again to a binary-valued function. This kind of smoothing and rounding interpretation no longer makes sense for general $\mathcal{Y}$-valued functions. We are therefore required to re-interpret these steps in the more general setting, where we find that they can be reformulated as *voting* on the function value at each point (rather than rounding a smoothed version of f). In the end, the active tester below for general $\mathcal{F}_k(\mathcal{H})$ functions has significant differences compared to the original tester for unions of intervals from [18]. Nevertheless, we do recover the dependences on k established by [18] for active and passive testing of unions of intervals, as special cases of our results on testing piecewise constant functions. Our results on testing general piecewise functions further extend this beyond piecewise constant functions, and require the introduction of additional uniform concentration arguments from VC theory. These considerations about general piecewise functions also lead us to an appropriate generalization of the notion of *noise sensitivity*, a quantity commonly appearing in the property testing literature for binary-valued functions [18].

It is also worth mentioning that, for binary-valued functions in higher dimensions $\mathbb{R}^n$, the noise sensitivity has also been used to test for the property that the decision boundary of f has low *surface area* [21]. For simplicity, the present work focuses on the one-dimensional setting, leaving for future work the appropriate generalization of these results to higher-dimensional spaces.

1.2. The Graph Dimension

For any set $\mathcal{Z}$, any collection $\mathcal{C}$ of subsets of $\mathcal{Z}$, and any $m \in \mathbb{N} \cup \{0\}$, following [22], we say $\mathcal{C}$ *shatters* a set $\{z_1, \ldots, z_m\} \subseteq \mathcal{Z}$ if

$$|\{C \cap \{z_1, \ldots, z_m\} : C \in \mathcal{C}\}| = 2^m.$$

The VC dimension of $\mathcal{C}$ is then defined as the largest integer m for which $\exists \{z_1, \ldots, z_m\} \subseteq \mathcal{Z}$ shattered by $\mathcal{C}$, or as infinity if no such largest m exists.

The VC dimension is an important quantity in characterizing the optimal sample complexity of statistical learning for binary classification. It is also useful in our present context, for the purpose of defining a related complexity measure for general functions. Specifically, the *graph dimension* of $\mathcal{H}$, denoted d below, is defined as the VC dimension of the collection $\{\{(x, h(x)) : x \in \mathcal{X}\} : h \in \mathcal{H}\}$ (where $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ in this context). For the remainder of this article, we restrict to the case $0 < d < \infty$, but otherwise we can consider $\mathcal{H}$ as a completely *arbitrary* set of functions (subject to (1)).

To continue the examples from above, we note that in the case of $|\mathcal{Y}| \geq 2$ and $\mathcal{H}$ as the set of constant functions (so that $\mathcal{F}_k(\mathcal{H})$ is the k-piecewise constant functions) one can easily show $d = 1$. Moreover, in the case of $\mathcal{Y} = \mathbb{R}$ and $\mathcal{H}$ as the set of degree-p real-valued polynomial functions (so that $\mathcal{F}_k(\mathcal{H})$ is the k-piecewise degree-p polynomial functions), it follows from basic algebra that $d = p + 1$, since any polynomial f is uniquely determined by any $p + 1$ distinct $(x, f(x))$ pairs (so that $d \leq p + 1$), and any $p + 1$ pairs (x, y) (with distinct x components) can be fit by a polynomial (so that $d \geq p + 1$, by choosing $p + 1$ distinct x points, each with two corresponding y values, and all 2^{p+1} choices of which y value to use for each point can be fit by a degree-p polynomial).

The results below will be expressed in terms of k, ϵ, and d. The dependence of the query complexity on each of these is an important topic to consider. However, in the present work, we primarily focus on identifying the optimal dependence on k. The optimal *joint* dependence on k, ϵ, and the complexity of $\mathcal{H}$ is a problem we leave for future work.

1.3. Main Results

We are now ready to state our main results. We present their proofs in the sections below. The first result is for active testing. Its proof in Section 3 below is based on an analysis of a novel generalization of the notion of the *noise sensitivity* of a function.

Theorem 1. *For any $\epsilon \in (0, 1/2)$, there exists an s-sample q-query ϵ-tester for $\mathcal{F}_k(\mathcal{H})$ under the distribution* Uniform$(0, 1)$*, with $s = O\left(\frac{dk}{\epsilon^6} \ln\left(\frac{1}{\epsilon}\right)\right)$ and $q = O\left(\frac{d}{\epsilon^8} \ln\left(\frac{1}{\epsilon}\right)\right)$.*

In particular, this immediately implies that the optimal dependence on k in the query complexity of active testing is $O(1)$. This independence from k in the query complexity is the main significance of this result.[2]

Our second result is for *both* active and passive testing, and applies specifically to the special case of piecewise *constant* functions (for any $\mathcal{Y}$ space). The upper bound in this result is again based on a generalization of the notion of the *noise sensitivity* of a function, and recovers as a special case the result of [18] for unions of intervals. The lower bound is essentially already known, as it was previously established by [20, 18] for unions of intervals, which are a special case of piecewise constant functions. The proof of the lower bound for general piecewise constant functions follows immediately from this via a simple reduction argument. The complete proof of this theorem is presented in Section 4 below.

Theorem 2. *If $\mathcal{H}$ is the set of* constant *functions, then for any $\epsilon \in (0, 1/2)$, there exists an s-sample q-query ϵ-tester for $\mathcal{F}_k(\mathcal{H})$ under the distribution* Uniform$(0,1)$*, with $s = O\left(\frac{\sqrt{k}}{\epsilon^5}\right)$, and with $q = O\left(\frac{1}{\epsilon^4}\right)$ for active testing, and $q = s$ for passive testing.*

Moreover, in this case, if $\epsilon \in (0, 1/8)$, every *s-sample s-query ϵ-tester for $\mathcal{F}_k(\mathcal{H})$ under the distribution* Uniform$(0,1)$ *has $s = \Omega\left(\sqrt{k}\right)$.*

In particular, this implies that the optimal query complexity of passive testing of k-piecewise constant functions has dependence $\sqrt{k}$ on the number of pieces k. It is also straightforward to extend this lower bound to k-piecewise degree-p polynomial functions. However, our results below do not imply an upper bound with $\sqrt{k}$ dependence on k for passive testing of this larger function class, and as such identifying the optimal query complexity of passive testing for piecewise degree-p polynomials (and for general classes $\mathcal{H}$ satisfying (1)) remains an interesting open problem.

2. A Generalization of Noise Sensitivity

Here we develop a generalization of the definition of the *noise sensitivity* used by [18] in their analysis of testing unions of intervals; this will be the key quantity in the proofs of the above theorems. Throughout this section, we let $\mathcal{P}$ be the Lebesgue measure restricted to $[0, 1]$: i.e., the distribution Uniform$(0, 1)$. For any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, define $\mathcal{H}_{(x,y)} = \{h \in \mathcal{H} : h(x) = y\}$. Let $x \sim$ Uniform$(0, 1)$, and conditioned on x, let $x' \sim$ Uniform$(x - \delta, x + \delta)$.

[2] Since the dependence on ϵ in this result is greater in the bound on q than in s (due to some over-counting in the bound on q when ϵ is small), we should note that this result also implies the existence of an s-sample $\min\{q, s\}$-query ϵ-tester, since we could simply query all s samples and then simulate the interaction with the oracle internally. This becomes relevant for $\epsilon \ll 1/\sqrt{k}$. That said, we describe a simple testing strategy at the end of Section 3 which obtains $s = q = O\left(\frac{dk\ln(k)}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$, which is superior in this range $\epsilon \ll 1/\sqrt{k}$ anyway.

Define the *instantaneous noise sensitivity* at x as[3]

$$\mathrm{NS}_\delta(f,x;\mathcal{H}) = \inf_{h\in\mathcal{H}_{(x,f(x))}} \mathbb{P}(h(x') \neq f(x')|x),$$

or in the event that $\mathcal{H}_{(x,f(x))}$ is empty, define $\mathrm{NS}_\delta(f,x;\mathcal{H}) = 1$. Then define the *noise sensitivity* as

$$\mathbb{NS}_\delta(f;\mathcal{H}) = \mathbb{E}[\mathrm{NS}_\delta(f,x;\mathcal{H})] = \int_0^1 \mathrm{NS}_\delta(f,z;\mathcal{H})\mathrm{d}z.$$

The instantaneous noise sensitivity essentially measures the ability of functions from $\mathcal{H}$ to match the behavior of f in a local neighborhood around a given point x, and the noise sensitivity is simply the average of this over x.

We have the following two key lemmas on this definition of noise sensitivity. Their statements and proofs directly parallel the analysis of unions of intervals by [18], but with a few important changes (particularly in the proof of Lemma 2) to generalize the arguments to general piecewise functions.

Lemma 1. *For any $\delta > 0$, $\forall f \in \mathcal{F}_k(\mathcal{H})$, $\mathbb{NS}_\delta(f;\mathcal{H}) \leq (k-1)\frac{\delta}{2}$.*

Proof. For $f \in \mathcal{F}_k(\mathcal{H})$, let $h_1,\ldots,h_k \in \mathcal{H}$ and $t_1,\ldots,t_{k-1} \in \mathbb{R}$ be such that $f(\cdot) = f\big(\cdot;\{h_i\}_{i=1}^k,\{t_i\}_{i=1}^{k-1}\big)$. Thus, for any $i \leq k$ and $x \in (t_{i-1},t_i]$, and any $y \in \mathbb{R}$, we have $h_i(x) = f(x)$, and we would have $h_i(y) \neq f(y)$ only if x and y are separated by one of the boundaries t_{i-1} or t_i; in particular, $x \leq t_i \leq y$ or $y \leq t_{i-1} \leq x$.

Let $x \sim \mathrm{Uniform}(0,1)$ and (conditioned on x) $y \sim \mathrm{Uniform}(x-\delta,x+\delta)$. Denoting by $i(x)$ the i with $t_{i-1} < x \leq t_i$, we have

$$\mathrm{NS}_\delta(f,x;\mathcal{H}) \leq \mathbb{P}(h_{i(x)}(y) \neq f(y)|x) \leq \mathbb{P}(x \leq t_{i(x)} \leq y|x) + \mathbb{P}(y \leq t_{i(x)-1} \leq x|x),$$

so that

$$\begin{aligned}\mathbb{NS}_\delta(f;\mathcal{H}) &\leq \mathbb{P}(x \leq t_{i(x)} \leq y) + \mathbb{P}(y \leq t_{i(x)-1} \leq x)\\ &\leq \sum_{i=1}^{k-1}\left(\mathbb{P}(x \leq t_i \leq y) + \mathbb{P}(y \leq t_i \leq x)\right),\end{aligned}$$

where the last inequality uses the facts that $\mathbb{P}(t_k \leq y) = 0$ and $\mathbb{P}(y \leq t_0) = 0$.

For any fixed $t \in \mathbb{R}$,

$$\mathbb{P}(x \leq t \leq y) \leq \int_0^\delta \mathbb{P}_{y'\sim\mathrm{Uniform}(t-z-\delta,t-z+\delta)}[y' \geq t]\mathrm{d}z = \int_0^\delta \frac{\delta - z}{2\delta}\mathrm{d}z = \frac{\delta}{4},$$

noting that, if t is outside $[\delta,1]$, then the probability can only become smaller.

[3]The original definition of [18] essentially defined $\mathrm{NS}_\delta(f,x) = \mathbb{P}(f(x') \neq f(x)|x)$. The involvement of $\mathcal{H}$ in our generalization of the definition will be crucial to the results below.

Similarly, any $t \in \mathbb{R}$ has $\mathbb{P}(y \leq t \leq x) \leq \frac{\delta}{4}$, again noting that the probability only becomes smaller if t is outside $[0, 1-\delta]$.

Combining these inequalities with the above bound on $\mathbb{NS}_\delta(f;\mathcal{H})$ yields $\mathbb{NS}_\delta(f;\mathcal{H}) \leq (k-1)\frac{\delta}{2}$, as claimed. □

Lemma 2. *Fix any $\epsilon \in (0, 1/2)$ and let $\delta = \frac{\epsilon^2}{32k}$. Let $f : \mathcal{X} \to \mathcal{Y}$ be any function with $\mathbb{NS}_\delta(f;\mathcal{H}) \leq (k-1)\frac{\delta}{2}(1+\frac{\epsilon}{4})$. Then $\rho(f, \mathcal{F}_k(\mathcal{H})) < \epsilon$.*

Proof. Let $k' = \lfloor 1 + (k-1)(1+\frac{\epsilon}{2}) \rfloor$. We first argue that f is $\frac{\epsilon}{2}$-close to a function in $\mathcal{F}_{k'}(\mathcal{H})$, and then we argue that every function in $\mathcal{F}_{k'}(\mathcal{H})$ is $\frac{\epsilon}{2}$-close to $\mathcal{F}_k(\mathcal{H})$.

For each $h \in \mathcal{H}$, consider the function $f^h_\delta : [0,1] \to [0,1]$ defined by

$$f^h_\delta(x) = \frac{1}{2\delta}\int_{x-\delta}^{x+\delta} \mathbb{I}[f(t) = h(t)]\mathrm{d}t.$$

The function f^h_δ is the convolution of $t \mapsto \mathbb{I}[f(t) = h(t)]$ and the uniform kernel $\phi : \mathbb{R} \to [0,1]$ defined by $\phi(x) = \frac{1}{2\delta}\mathbb{I}[|x| \leq \delta]$. Note that, since any distinct $h, h' \in \mathcal{H}$ have $\int_0^1 \mathbb{I}[h(t) = h'(t)]\mathrm{d}t = 0$ (by the zero-measure crossings assumption (1)), the sum (over $h \in \mathcal{H}$) of all $f^h_\delta(x)$ values is at most 1. In particular, at most one $h \in \mathcal{H}$ has $f^h_\delta(x) > 1/2$ for any x.

Fix $\tau = \frac{4}{\epsilon}\mathbb{NS}_\delta(f;\mathcal{H})$. Since $\mathbb{NS}_\delta(f;\mathcal{H}) \leq (k-1)\frac{\delta}{2}(1+\frac{\epsilon}{4}) < \frac{\epsilon^2}{32}$, we have $\tau < 1/8$. For each x, let $h_x = \mathrm{argmax}_{h\in\mathcal{H}} f^h_\delta(x)$ (breaking ties arbitrarily); since the sum (over $h \in \mathcal{H}$) of $f^h_\delta(x)$ values is finite (bounded by 1), it follows that the value $\sup_{h\in\mathcal{H}} f^h_\delta(x)$ is actually realized by some $f^h_\delta(x)$ with $h \in \mathcal{H}$, so that h_x is well-defined. Define a function $g^* : [0,1] \to \mathcal{Y} \cup \{*\}$ by $g^*(x) = h_x(x)$ if $f^{h_x}_\delta(x) \geq 1-\tau$, and $g^*(x) = *$ otherwise. Next, define a function $g : \mathbb{R} \to \mathcal{Y}$ by setting, for any $x \in [0,1]$, $g(x) = h_z(x)$ where z is the largest value in $[0,x]$ for which $g^*(z) \neq *$, and let $g_x = h_z$; if no such z exists, take z minimal in $[x,1]$ with $g^*(z) \neq *$ instead; if that also does not exist, we can define $g(x)$ and g_x arbitrarily, as this case will not come up in our present context. As we discuss below, $f^{h_x}_\delta(x) = \sup_{h\in\mathcal{H}} f^h_\delta(x)$ is continuous in x, which entails that at least one of these two possible z values will exist if g^* is not everywhere equal $*$ in $[0,1]$. For completeness, also define, for any $x < 0$, $g(x) = g_0(x)$, and for any $x > 1$, $g(x) = g_1(x)$.

Now note that

$$\begin{aligned}
\rho(f,g) = \mathcal{P}(x : f(x) \neq g(x)) &\leq \mathcal{P}(x : g^*(x) = *) + \mathcal{P}(x : * \neq g^*(x) \neq f(x)) \\
&= \mathcal{P}\left(x : \sup_{h\in\mathcal{H}} f^h_\delta(x) < 1-\tau\right) \\
&\quad + \mathcal{P}\left(x : \mathcal{H}\setminus\mathcal{H}_{(x,f(x))} \neq \emptyset, \sup_{h\in\mathcal{H}\setminus\mathcal{H}_{(x,f(x))}} f^h_\delta(x) \geq 1-\tau\right). \qquad (2)
\end{aligned}$$

Because $\tau < 1/2$, at most one h can have $f^h_\delta(x) \geq 1-\tau$ (as discussed above), so that if the event $\sup_{h\in\mathcal{H}} f^h_\delta(x) < 1-\tau$ holds or the event $\mathcal{H}\setminus\mathcal{H}_{(x,f(x))} \neq \emptyset$

and $\sup_{h\in\mathcal{H}\setminus\mathcal{H}_{(x,f(x))}} f_\delta^h(x) \geq 1-\tau$ holds, then either way we would have $\sup_{h\in\mathcal{H}_{(x,f(x))}} f_\delta^h(s) < 1-\tau$ (or $\mathcal{H}_{(x,f(x))}=\emptyset$); thus, since the two events implying this are disjoint, the sum of probabilities in (2) is at most

$$\mathcal{P}\left(x : \mathcal{H}_{(x,f(x))} = \emptyset \text{ or } \sup_{h\in\mathcal{H}_{(x,f(x))}} f_\delta^h(x) < 1-\tau\right).$$

Now observe that $\mathrm{NS}_\delta(f,x;\mathcal{H}) = 1 - \sup_{h\in\mathcal{H}_{(x,f(x))}} f_\delta^h(x)$ if $\mathcal{H}_{(x,f(x))} \neq \emptyset$, and $\mathrm{NS}_\delta(f,x;\mathcal{H}) = 1$ if $\mathcal{H}_{(x,f(x))} = \emptyset$. Together with Markov's inequality, this implies that

$$\begin{aligned}
&\mathcal{P}\left(x : \mathcal{H}_{(x,f(x))} = \emptyset \text{ or } \sup_{h\in\mathcal{H}_{(x,f(x))}} f_\delta^h(x) < 1-\tau\right) \\
&\qquad\qquad = \mathcal{P}(x : \mathrm{NS}_\delta(f,x;\mathcal{H}) > \tau) < \frac{\mathbb{NS}_\delta(f;\mathcal{H})}{\tau} = \frac{\epsilon}{4}.
\end{aligned}$$

Thus, we have established that $\rho(f,g) \leq \frac{\epsilon}{4}$.

Next we show that $g \in \mathcal{F}_{m+1}(\mathcal{H})$ for some nonnegative integer $m \leq (k-1)(1+\frac{\epsilon}{2})$. Since each f_δ^h is the convolution of $\mathbb{I}[f(\cdot)=h(\cdot)]$ with a uniform kernel of width 2δ, it is $\frac{1}{2\delta}$-Lipschitz smooth. Also recall that $\tau < 1/2$, and the sum of all $f_\delta^h(x)$ values for a given x is at most 1. Thus, if we consider any two points $x,z \in [0,1]$ with $g^*(x) \neq *$, $g^*(z) \neq *$, $x<z$, and $h_x \neq h_z$, then it must be that $|x-z| \geq 2\delta(1-2\tau)$, and that there is at least one point $t \in (x,z)$ with $\sup_{h\in\mathcal{H}} f_\delta^h(t) = 1/2$. Since each f_δ^h is $\frac{1}{2\delta}$-Lipschitz, so is $\sup_{h\in\mathcal{H}} f_\delta^h$, so that we have

$$\int_{t-2\delta(\frac{1}{2}-\tau)}^{t+2\delta(\frac{1}{2}-\tau)} \sup_{h\in\mathcal{H}} f_\delta^h(s)\mathrm{d}s \leq 2\int_0^{2\delta(\frac{1}{2}-\tau)} \left(\frac{1}{2}+\frac{s}{2\delta}\right)\mathrm{d}s = 2\delta\left(\frac{1}{2}-\tau\right)\left(\frac{3}{2}-\tau\right).$$

Therefore,

$$\begin{aligned}
&\int_x^z \mathrm{NS}_\delta(f,s;\mathcal{H})\mathrm{d}s \geq \int_x^z \left(1-\sup_{h\in\mathcal{H}} f_\delta^h(s)\right)\mathrm{d}s \\
&\geq (z-x) - 2\delta\left(\frac{1}{2}-\tau\right)\left(\frac{3}{2}-\tau\right) \geq 2\delta\left(1-2\tau\right) - 2\delta\left(\frac{1}{2}-\tau\right)\left(\frac{3}{2}-\tau\right) \\
&= 2\delta\left(\frac{1}{2}-\tau\right)\left(\frac{1}{2}+\tau\right) = 2\delta\left(\frac{1}{4}-\tau^2\right).
\end{aligned}$$

Since any x with $g^*(x) \neq *$ has $g(x) = g^*(x)$, and since g_t is extrapolated from the left in $*$ regions of g^* (aside from the case of an interval of $*$ values including 0, where it is extrapolated from the right), for every point $x>0$ for which there exist arbitrarily close points y having $g_y \neq g_x$, we must have that $g^*(x) \neq *$, and that there is a point $z<x$ such that $g^*(z) \neq *$ and such that every $t \in (z,x)$ has $g_t = g_z \neq g_x$. Combined with the above, we have that

$\int_z^x \mathrm{NS}_\delta(f,s;\mathcal{H})\mathrm{d}s \geq 2\delta(\frac{1}{4}-\tau^2)$. Altogether, if g has m such "transition" points, then

$$\mathbb{NS}_\delta(f;\mathcal{H}) = \int_0^1 \mathrm{NS}_\delta(f,s;\mathcal{H})\mathrm{d}s \geq m2\delta\left(\frac{1}{4}-\tau^2\right).$$

By assumption, $\mathbb{NS}_\delta(f;\mathcal{H}) \leq (k-1)\frac{\delta}{2}(1+\frac{\epsilon}{4})$. Therefore, we must have

$$m \leq \frac{(k-1)\delta(1+\frac{\epsilon}{4})}{4\delta(\frac{1}{4}-\tau^2)} \leq (k-1)\frac{1+\frac{\epsilon}{4}}{1-4\tau^2} \leq (k-1)\frac{1+\frac{\epsilon}{4}}{(1-2\tau)^2} \leq (k-1)\left(1+\frac{\epsilon}{2}\right),$$

since $\tau < 1/8$. In particular, this means $g \in \mathcal{F}_{m+1}(\mathcal{H})$ for an $m \leq (k-1)(1+\frac{\epsilon}{2})$, as claimed.

As a second step in the proof, we show that for any nonnegative integer $m \leq (k-1)(1+\frac{\epsilon}{2})$, any function $g' \in \mathcal{F}_{m+1}(\mathcal{H})$ is $\frac{\epsilon}{2}$-close to a function in $\mathcal{F}_k(\mathcal{H})$. Let $t_1,\ldots,t_m \in \mathbb{R}$ with $t_1 \leq \cdots \leq t_m$, and $h_1,\ldots,h_{m+1} \in \mathcal{H}$, be such that $g'(\cdot) = f(\cdot;\{h_i\}_{i=1}^{m+1},\{t_i\}_{i=1}^m)$. For each $i \in \{1,\ldots,m+1\}$, let $\ell_i = \mathcal{P}((t_{i-1},t_i])$ denote the probability mass in the i^{th} region. In particular, $\ell_1+\cdots+\ell_{m+1} = 1$, so there must be a set $S \subseteq \{1,\ldots,m+1\}$ with $|S| = (m+1)-k \leq (k-1)\frac{\epsilon}{2}$ such that

$$\sum_{i\in S}\ell_i \leq \frac{(m+1)-k}{(m+1)} \leq \frac{(k-1)\epsilon/2}{1+(k-1)(1+\epsilon/2)} < \frac{\epsilon}{2}.$$

Define a function $f' : \mathcal{X} \to \mathcal{Y}$ such that, for each $i \in \{1,\ldots,m+1\}$ and $x \in (t_{i-1},t_i]$, we set $f'(x) = h_j(x)$ for the $j \in \{1,\ldots,m+1\}\setminus S$ of smallest $|i-j|$ (breaking ties to favor smaller j). The function f' is then contained in $\mathcal{F}_k(\mathcal{H})$, and has $f'(x) = g'(x)$ for every $x \in (t_{i-1},t_i]$ with $i \notin S$, and hence $\rho(g',f') < \frac{\epsilon}{2}$. This completes the proof, since taking $g' = g$ yields $\rho(f,f') \leq \rho(f,g)+\rho(g,f') < \epsilon$. $\square$

3. Active Testing

We can use the above lemmas to construct an active tester for $\mathcal{F}_k(\mathcal{H})$ as follows. Fix any $\epsilon \in (0,1/2)$. Let $m = \lceil\frac{c}{\epsilon^4}\rceil$, $\ell = \left\lceil\frac{c'd}{\epsilon^4}\ln\left(\frac{c''}{\epsilon}\right)\right\rceil$, $\delta = \frac{\epsilon^2}{32k}$, and $s = m + \left\lceil\max\left\{\frac{2\ell}{\delta},\frac{8}{\delta}\ln(12m)\right\}\right\rceil$, for appropriate choices of numerical constants $c,c',c'' \geq 1$ from the analysis below. Sample s points $x_1'',\ldots,x_s''$ independent Uniform$(0,1)$. Define $x_i = x_i''$ for each $i \leq m$. For each $i \leq m$, denoting $t_{i0} = m$, for each $j \in \{1,\ldots,\ell\}$, let $t_{ij} = \min\{t \in \{t_{i(j-1)}+1,\ldots,s\} : x_t'' \in (x_i-\delta,x_i+\delta)\}$ if such a value exists (if it does not exist, the tester may return any response, as this is a failure case), and define $x_{ij}' = x_{t_{ij}}''$. Thus, the random variables $x_1,\ldots,x_m$ are iid Uniform$(0,1)$ and, given x_i, the random variables $x_{i1}',\ldots,x_{i\ell}'$ are conditionally iid Uniform$((x_i-\delta,x_i+\delta)\cap[0,1])$ (given x_i and the event that they exist). The tester requests the f values for all $m(\ell+1)$ of these points x_i,

x'_{ij}, $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, \ell\}$. It then calculates, for each $i \le m$,

$$\widehat{\mathrm{NS}}_\delta(f, x_i; \mathcal{H}) = \min_{h \in \mathcal{H}_{(x_i, f(x_i))}} \frac{1}{\ell} \sum_{j=1}^{\ell} \mathbb{I}[h(x'_{ij}) \neq f(x'_{ij})],$$

or $\widehat{\mathrm{NS}}_\delta(f, x_i; \mathcal{H}) = 1$ in the event that $\mathcal{H}_{(x_i, f(x_i))}$ is empty. Then define

$$\widehat{\mathbb{NS}}_\delta(f; \mathcal{H}) = \frac{1}{m} \sum_{i=1}^{m} \widehat{\mathrm{NS}}_\delta(f, x_i; \mathcal{H}).$$

Lemma 3. *If $k \ge 80/\epsilon$, then for appropriate choices of numerical constants c, c', c'', for any measurable function $f : \mathcal{X} \to \mathcal{Y}$, with probability at least $2/3$, all of the above x'_{ij} points exist, and the following two claims hold:*

$$\mathbb{NS}_\delta(f; \mathcal{H}) \le (k-1)\frac{\delta}{2} \implies \widehat{\mathbb{NS}}_\delta(f; \mathcal{H}) \le (k-1)\frac{\delta}{2}\left(1 + \frac{\epsilon}{8}\right)$$
$$\mathbb{NS}_\delta(f; \mathcal{H}) > (k-1)\frac{\delta}{2}\left(1 + \frac{\epsilon}{4}\right) \implies \widehat{\mathbb{NS}}_\delta(f; \mathcal{H}) > (k-1)\frac{\delta}{2}\left(1 + \frac{\epsilon}{8}\right).$$

Proof. For each $i \le m$, any $t \in \{m+1, \ldots, s\}$ has conditional probability (given x_i) at least δ of having $x''_t \in (x_i - \delta, x_i + \delta)$. Therefore, by a Chernoff bound (applied under the conditional distribution given x_i) and the law of total probability, with probability at least $1 - \exp\{-\delta(s-m)/8\}$, the number of $t \in \{m+1, \ldots, s\}$ with $x''_t \in (x_i - \delta, x_i + \delta)$ is at least $(1/2)\delta(s-m) \ge \ell$. By the union bound, this holds simultaneously for all $i \le m$ with probability at least $1 - m\exp\{-\delta(s-m)/8\} \ge 11/12$, and on this event all of the x'_{ij} points exist.

The VC dimension of the collection of sets $\{\{(x, h(x)) : x \in \mathcal{X}\} : h \in \mathcal{H}\}$ is d (by definition of d). Therefore, denoting

$$A_{\ell,m} = 4\frac{d\ln(2e\ell/d) + \ln(96m)}{\ell},$$

applying standard VC "relative deviation" bounds [4] (see Theorem 5.1 of [23]), to obtain a concentration inequality for the frequency of $(x'_{ij}, f(x'_{ij})) \in \{(x, h(x)) : x \in \mathcal{X}\}$, holding for all $h \in \mathcal{H}$, we obtain that, for each $i \le m$, with probability at least $1 - 1/(12m)$, if the points x'_{ij} exist, then every $h \in \mathcal{H}$ has

$$\frac{1}{\ell}\sum_{j=1}^{\ell} \mathbb{I}[h(x'_{ij}) \neq f(x'_{ij})] \le \mathbb{P}(h(x'_{i1}) \neq f(x'_{i1}) | x_i) + \sqrt{\mathbb{P}(h(x'_{i1}) \neq f(x'_{i1}) | x_i) A_{\ell,m}} + A_{\ell,m} \quad (3)$$

and

$$\frac{1}{\ell}\sum_{j=1}^{\ell}\mathbb{I}[h(x'_{ij}) \neq f(x'_{ij})] \geq \mathbb{P}(h(x'_{i1}) \neq f(x'_{i1})|x_i) - \sqrt{\mathbb{P}(h(x'_{i1}) \neq f(x'_{i1})|x_i)A_{\ell,m}}. \quad (4)$$

The union bound implies this is true simultaneously for all $i \leq m$ with probability at least $11/12$.

Furthermore, for any $x_i \in (\delta, 1-\delta)$, the conditional distribution of x'_{i1} given x_i is Uniform$(x_i - \delta, x_i + \delta)$, so that

$$\mathrm{NS}_\delta(f, x_i; \mathcal{H}) = \inf_{h \in \mathcal{H}_{(x_i, f(x_i))}} \mathbb{P}(h(x'_{i1}) \neq f(x'_{i1})|x_i)$$

in the case $\mathcal{H}_{(x_i,f(x_i))} \neq \emptyset$. Thus, on the above events, for each $i \leq m$ with $x_i \in (\delta, 1-\delta)$ and $\mathcal{H}_{(x_i,f(x_i))} \neq \emptyset$, taking the infimum over $h \in \mathcal{H}_{(x_i,f(x_i))}$ on both sides of (3) yields

$$\widehat{\mathrm{NS}}_\delta(f, x_i; \mathcal{H}) \leq \mathrm{NS}_\delta(f, x_i; \mathcal{H}) + \sqrt{\mathrm{NS}_\delta(f, x_i; \mathcal{H})A_{\ell,m}} + A_{\ell,m}. \quad (5)$$

For the other inequality, note that the left hand side of (4) is nonnegative, so that the inequality remains valid if we include a maximum with 0 on the right hand side. Then noting that $x \mapsto x - \max\left\{\sqrt{xA_{\ell,m}}, 0\right\}$ is nondecreasing on $[0,1]$, we obtain, on the above events, for each $i \leq m$ with $x_i \in (\delta, 1-\delta)$ and $\mathcal{H}_{(x_i,f(x_i))} \neq \emptyset$,

$$\widehat{\mathrm{NS}}_\delta(f, x_i; \mathcal{H}) \geq \mathrm{NS}_\delta(f, x_i; \mathcal{H}) - \sqrt{\mathrm{NS}_\delta(f, x_i; \mathcal{H})A_{\ell,m}}. \quad (6)$$

Both of these inequalities are trivially also satisfied in the case $\mathcal{H}_{(x_i,f(x_i))} = \emptyset$.

Furthermore, since $k \geq 80/\epsilon$, we have $2\delta \leq \frac{\epsilon^3}{16\cdot 80}$, so that a Chernoff bound implies that, for an appropriately large choice of the numerical constant c, with probability at least $11/12$,

$$\frac{1}{m}\sum_{i=1}^{m}\mathbb{I}[x_i \notin (\delta, 1-\delta)] \leq \frac{\epsilon^3}{16\cdot 65}.$$

Furthermore, note that since $k \geq 80/\epsilon$, we have $\frac{\epsilon^3}{16\cdot 65} = (k-1)\frac{\delta}{2}\frac{k}{k-1}\frac{64\epsilon}{16\cdot 65} < (k-1)\frac{\delta}{2}\frac{\epsilon}{16}$, so that on the above event,

$$\frac{1}{m}\sum_{i=1}^{m}\mathbb{I}[x_i \notin (\delta, 1-\delta)] < (k-1)\frac{\delta}{2}\frac{\epsilon}{16}. \quad (7)$$

Additionally, since $k \geq 80/\epsilon$, we have $(k-1)\frac{\delta}{2} > \frac{\epsilon^2}{65}$, so that $m > \frac{c/65}{(k-1)(\delta/2)\epsilon^2}$, which clearly also means $m > \frac{c/65}{(k-1)(\delta/2)(1+\epsilon/4)\epsilon^2}$. Therefore, another application of a Chernoff bound implies that, for an appropriately large choice of the numerical constant c, with probability at least 11/12,

$$\mathbb{NS}_\delta(f;\mathcal{H}) \le (k-1)\frac{\delta}{2} \implies \frac{1}{m}\sum_{i=1}^{m} \mathrm{NS}_\delta(f, x_i; \mathcal{H}) \le (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{33}\right) \tag{8}$$

and

$$\begin{aligned} &\mathbb{NS}_\delta(f;\mathcal{H}) > (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{4}\right) \\ &\implies \frac{1}{m}\sum_{i=1}^{m} \mathrm{NS}_\delta(f, x_i; \mathcal{H}) > (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{4}\right)\left(1-\frac{\epsilon}{33}\right) \ge (k-1)\frac{\delta}{2}\left(1+\frac{7}{33}\epsilon\right). \end{aligned} \tag{9}$$

The union bound implies that all four of the above events hold simultaneously with probability at least 2/3. Let us suppose all of these events indeed hold. In this case, if $\mathbb{NS}_\delta(f;\mathcal{H}) \le (k-1)\frac{\delta}{2}$, then (5) and Jensen's inequality imply

$$\begin{aligned} &\widehat{\mathbb{NS}}_\delta(f;\mathcal{H}) \\ &\le \frac{1}{m}\sum_{i=1}^{m}\left(\mathrm{NS}_\delta(f,x_i;\mathcal{H}) + \sqrt{\mathrm{NS}_\delta(f,x_i;\mathcal{H})A_{\ell,m}} + A_{\ell,m}\right) + \frac{1}{m}\sum_{i=1}^{m}\mathbb{I}[x_i \notin (\delta, 1-\delta)] \\ &\le \left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{NS}_\delta(f,x_i;\mathcal{H})\right) + \sqrt{\left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{NS}_\delta(f,x_i;\mathcal{H})\right)A_{\ell,m}} \\ &\quad + A_{\ell,m} + \frac{1}{m}\sum_{i=1}^{m}\mathbb{I}[x_i \notin (\delta, 1-\delta)], \end{aligned}$$

and (7) and (8) imply this is at most

$$(k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{33}\right) + \sqrt{(k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{33}\right)A_{\ell,m}} + A_{\ell,m} + (k-1)\frac{\delta}{2}\frac{\epsilon}{16}.$$

For appropriately large choices of the numerical constants c', c'', we can obtain $A_{\ell,m} \le \frac{\epsilon^4}{65\cdot 68\cdot 33} \le (k-1)\frac{\delta}{2}\frac{\epsilon^2}{68\cdot 33}$, so that the above expression is at most

$$(k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{33}+\sqrt{\left(1+\frac{\epsilon}{33}\right)\frac{\epsilon^2}{68\cdot 33}} + \frac{\epsilon^2}{68\cdot 33} + \frac{\epsilon}{16}\right) \le (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{8}\right),$$

which verifies the first claimed implication from the lemma. On the other hand, for the second implication, if $\mathbb{NS}_\delta(f;\mathcal{H}) > (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{4}\right)$, then (6) and Jensen's inequality imply

$$
\begin{aligned}
&\widehat{\mathbb{NS}}_\delta(f;\mathcal{H})\\
&\geq \frac{1}{m}\sum_{i=1}^{m}\left(\mathrm{NS}_\delta(f,x_i;\mathcal{H}) - \sqrt{\mathrm{NS}_\delta(f,x_i;\mathcal{H})A_{\ell,m}}\right) - \frac{1}{m}\sum_{i=1}^{m}\mathbb{I}[x_i \notin (\delta,1-\delta)]\\
&\geq \left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{NS}_\delta(f,x_i;\mathcal{H})\right) - \sqrt{\left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{NS}_\delta(f,x_i;\mathcal{H})\right)A_{\ell,m}}\\
&\qquad - \frac{1}{m}\sum_{i=1}^{m}\mathbb{I}[x_i \notin (\delta,1-\delta)],
\end{aligned}
$$

and (7) implies this is greater than

$$
\left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{NS}_\delta(f,x_i;\mathcal{H})\right) - \sqrt{\left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{NS}_\delta(f,x_i;\mathcal{H})\right)A_{\ell,m}} - (k-1)\frac{\delta}{2}\frac{\epsilon}{16}.
$$

Now since our choices of constants c, c', c'' above imply $A_{\ell,m} \leq (k-1)\frac{\delta}{2}\frac{\epsilon^2}{68\cdot 33} \leq (k-1)\frac{\delta}{2}\left(1+\frac{7}{33}\epsilon\right)$, and since $x \mapsto x - \sqrt{xA_{\ell,m}}$ is increasing for $x \geq A_{\ell,m}$, (9) implies the above expression is greater than

$$
\begin{aligned}
&(k-1)\frac{\delta}{2}\left(1+\frac{7}{33}\epsilon\right) - \sqrt{(k-1)\frac{\delta}{2}\left(1+\frac{7}{33}\epsilon\right)A_{\ell,m}} - (k-1)\frac{\delta}{2}\frac{\epsilon}{16}\\
&\geq (k-1)\frac{\delta}{2}\left(1+\frac{7}{33}\epsilon\right) - \sqrt{(k-1)\frac{\delta}{2}\left(1+\frac{7}{33}\epsilon\right)(k-1)\frac{\delta}{2}\frac{\epsilon^2}{68\cdot 33}} - (k-1)\frac{\delta}{2}\frac{\epsilon}{16}\\
&= (k-1)\frac{\delta}{2}\left(1+\frac{7}{33}\epsilon - \sqrt{\left(1+\frac{7}{33}\epsilon\right)\frac{\epsilon^2}{68\cdot 33}} - \frac{\epsilon}{16}\right) > (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{8}\right).
\end{aligned}
$$

This verifies the second claimed implication from the lemma, and thus completes the proof. ☐

We are now ready to finish describing the tester and prove its correctness.

Theorem 3. *If $k \geq 80/\epsilon$, then the procedure that outputs* Accept *if $\widehat{\mathbb{NS}}_\delta(f;\mathcal{H}) \leq (k-1)\frac{\delta}{2}(1+\frac{\epsilon}{8})$, and otherwise outputs* Reject, *is an s-sample q-query ϵ-tester for the class $\mathcal{F}_k(\mathcal{H})$ of k-piecewise $\mathcal{H}$ functions under the distribution* Uniform$(0,1)$*, for s as defined above, and for $q = m(\ell+1)$ (where m and ℓ are as defined above).*

Proof. If $f \in \mathcal{F}_k(\mathcal{H})$, then Lemma 1 implies it has $\mathbb{NS}_\delta(f;\mathcal{H}) \leq (k-1)\frac{\delta}{2}$, so that Lemma 3 implies that with probability at least 2/3, $\widehat{\mathbb{NS}}_\delta(f;\mathcal{H}) \leq (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{8}\right)$, and hence the tester will output Accept.

On the other hand, if f is ϵ-far from every function in $\mathcal{F}_k(\mathcal{H})$, then Lemma 2 implies that $\mathbb{NS}_\delta(f;\mathcal{H}) > (k-1)\frac{\delta}{2}(1+\frac{\epsilon}{4})$, so that Lemma 3 implies that with probability at least $2/3$, $\widehat{\mathbb{NS}}_\delta(f;\mathcal{H}) > (k-1)\frac{\delta}{2}(1+\frac{\epsilon}{8})$, and hence the tester will output Reject.

The claim about the number of samples and number of queries is immediate from the definition of the tester. □

Theorem 1 immediately follows from this result for any $k \geq 80/\epsilon$, since $m(\ell+1) = O\big(\frac{d}{\epsilon^8}\ln\big(\frac{1}{\epsilon}\big)\big)$ and $s = O\big(\frac{dk}{\epsilon^6}\ln\big(\frac{1}{\epsilon}\big)\big)$.

For $k < 80/\epsilon$, there is a trivial tester satisfying Theorem 1, based on the learn-then-validate technique of [20], which in fact works for any distribution $\mathcal{P}$. Specifically, in this case, we can take $s = \left\lceil \frac{c_1 dk}{\epsilon}\ln(2ek)\ln\big(\frac{1}{\epsilon}\big)\right\rceil + \left\lceil\frac{c_2}{\epsilon}\right\rceil = O\big(\frac{d}{\epsilon^2}\ln^2\big(\frac{1}{\epsilon}\big)\big)$ iid $\mathcal{P}$ samples, for appropriate numerical constants $c_1, c_2 \geq 1$, and query for the f values for these s samples (so $q = s$ here). We then find a function $\hat{f} \in \mathcal{F}_k(\mathcal{H})$ consistent with f on the first $\left\lceil \frac{c_1 dk}{\epsilon}\ln(2ek)\ln\big(\frac{1}{\epsilon}\big)\right\rceil$ of these ($\hat{f}$ chosen independently from the rest of the samples), if such a function $\hat{f}$ exists; we then check whether $\hat{f}$ agrees with f on at least $(1-\epsilon/2)\left\lceil\frac{c_2}{\epsilon}\right\rceil$ of the remaining $\left\lceil\frac{c_2}{\epsilon}\right\rceil$ samples. If this $\hat{f}$ exists and satisfies this condition, then we output Accept, and otherwise we output Reject. We can bound the graph dimension of $\mathcal{F}_k(\mathcal{H})$ as follows. For any n distinct points $x_1,\ldots,x_n \in \mathbb{R}$ and n values $y_1,\ldots,y_n \in \mathcal{Y}$, the number of distinct $\{(x,g(x)) : x \in \mathcal{X}\} \cap \{(x_1,y_1),\ldots,(x_n,y_n)\}$ sets that can be realized by functions $g \in \mathcal{F}_k(\mathcal{H})$ is at most $\left(\frac{en}{d}\right)^{dk}\left(\frac{en}{k}\right)^k$, obtained by applying Sauer's lemma within each subset $(t_{j-1},t_j] \cap \{x_1,\ldots,x_n\}$ and multiplying them to get at most $\left(\frac{en}{d}\right)^{dk}$ possible classifications for any fixed t_j values, and then multiplying by a bound $\left(\frac{en}{k}\right)^k$ on the number of ways to partition $\{x_1,\ldots,x_n\}$ into at most k intervals. Since $\left(\frac{en}{d}\right)^{dk}\left(\frac{en}{k}\right)^k$ is strictly less than 2^n for any $n > 4dk\log_2(2ek)$, the graph dimension of $\mathcal{F}_k(\mathcal{H})$ is at most $4dk\log_2(2ek)$. Thus, if $f \in \mathcal{F}_k(\mathcal{H})$, then standard VC bounds for the realizable case [4, 5] imply that, for an appropriate choice of the numerical constant c_1, with probability at least $5/6$, the function $\hat{f}$ will have $\mathcal{P}(x : \hat{f}(x) \neq f(x)) < \epsilon/4$. Also, for an appropriately large numerical constant c_2, a Chernoff bound implies that, with probability at least $5/6$, if $\mathcal{P}(x : \hat{f}(x) \neq f(x)) < \epsilon/4$ then $\hat{f}$ will agree with f on at least $(1-\epsilon/2)\left\lceil\frac{c_2}{\epsilon}\right\rceil$ of the last $\left\lceil\frac{c_2}{\epsilon}\right\rceil$ samples. By the union bound, both of these events occur simultaneously with probability at least $2/3$, and the tester will output Accept when they occur. On the other hand, if f is ϵ-far from every function in $\mathcal{F}_k(\mathcal{H})$, then either $\hat{f}$ will not exist (in which case the tester outputs Reject), or else $\hat{f}$ is some function in $\mathcal{F}_k(\mathcal{H})$, so that $\mathcal{P}(x : \hat{f}(x) \neq f(x)) > \epsilon$. Therefore, for an appropriately large choice of the numerical constant c_2, a Chernoff bound implies that with probability at least $2/3$, if $\hat{f}$ exists, then it disagrees with f on strictly more than $\frac{\epsilon}{2}\left\lceil\frac{c_2}{\epsilon}\right\rceil$ of the last $\left\lceil\frac{c_2}{\epsilon}\right\rceil$ samples, so that the tester will output Reject.

4. Piecewise Constant Functions

Next, we restrict focus to the special case of piecewise constant functions: that is, throughout this subsection, take $\mathcal{H}$ as the set of all *constant* functions $\mathcal{X} \to \mathcal{Y}$. In this case, we study both active and passive testing. We construct a passive tester achieving the bound in Theorem 2, as well as an active tester whose number of queries has an improved dependence on ϵ compared to Theorem 1. Unlike the above general active tester, this construction follows more-closely the construction of testers for unions of intervals from [18], and indeed recovers the same dependences on k and ϵ from that work, now for this more-general problem of testing piecewise-constant functions.

Since $\mathcal{H}$ is fixed as the set of constant functions $\mathcal{X} \to \mathcal{Y}$ in this section, we simply write $\mathbb{NS}_\delta(f)$ to abbreviate $\mathbb{NS}_\delta(f;\mathcal{H})$, which (as we argue below) is consistent with the notion of noise sensitivity used in the prior literature [18]. Fix any $\epsilon \in (0, 1/2)$ and consider the case $k \geq 80/\epsilon$. Let $\delta = \frac{\epsilon^2}{32k}$, $m' = \left\lceil \frac{c}{\epsilon^4} \right\rceil$, $n = 1 + \left\lceil 2\sqrt{\lceil 1/\delta \rceil} \right\rceil$, and $s' = 4nm'$, for an appropriate choice of numerical constant $c \geq 1$ from the analysis below. Now the active and passive testers both sample s' points $z'_1, \ldots, z'_{s'}$ independent Uniform$(0, 1)$. Let $t_1, \ldots, t_{m'}$ be the first m' distinct values t in $\{1, \ldots s'/n\}$ for which $\exists i, j \in \{(t-1)n+1, \ldots, tn\}$ with $i < j$ and $|z'_i - z'_j| < \delta$ while $z'_i \in (\delta, 1-\delta)$, and for each $r \in \{1, \ldots, m'\}$, denote by i_r the smallest integer $i > (t_r - 1)n$ with $\min_{j \in \{i+1,\ldots,t_r n\}} |z'_i - z'_j| < \delta$ while $z'_i \in (\delta, 1-\delta)$, and denote by j_r the smallest integer $j > i_r$ with $|z'_{i_r} - z'_j| < \delta$; if there do not exist m' such values t_r, then the tester may output any response, and this is considered a failure event. If these values do exist, then for each $r \leq m'$, denote $z_r = z'_{i_r}$ and $y_r = z'_{j_r}$. The active tester queries the f values for z_r and y_r, for each $r \leq m'$, whereas the passive tester (necessarily) queries the f values for all s' points z'_i. Both testers then calculate the following quantity

$$\widehat{\mathbb{NS}}'_\delta(f) = \frac{1-2\delta}{m'} \sum_{r=1}^{m'} \mathbb{I}[f(z_r) \neq f(y_r)]$$

and outputs Accept if $\widehat{\mathbb{NS}}'_\delta(f) \leq (k-1)\frac{\delta}{2}(1 + \frac{\epsilon}{8})$, and otherwise outputs Reject.

Comparing $\widehat{\mathbb{NS}}'_\delta(f)$ to the quantity $\widehat{\mathbb{NS}}_\delta(f;\mathcal{H})$ defined above, the main difference is that, for each of the points z_r, we use only a *single* point y_r sampled from $(z_r - \delta, z_r + \delta)$, rather than ℓ points. For this reason, the total number of examples (both labeled and unlabeled) required to calculate $\widehat{\mathbb{NS}}'_\delta(f)$ is significantly smaller than the number required to calculate $\widehat{\mathbb{NS}}_\delta(f;\mathcal{H})$. Nevertheless, in this special case of piecewise constant functions, we find that the guarantees we had for $\widehat{\mathbb{NS}}_\delta(f;\mathcal{H})$ from Lemma 3 above remain valid for the quantity $\widehat{\mathbb{NS}}'_\delta(f)$. Specifically, we have the following lemma.

Lemma 4. *If $k \geq 80/\epsilon$, then for an appropriate choice of the numerical constants c, for any measurable function $f : \mathcal{X} \to \mathcal{Y}$, with probability at least $2/3$,* *the values $t_1, \ldots, t_{m'}$ exist, and*

$$\mathbb{NS}_\delta(f) \leq (k-1)\frac{\delta}{2} \implies \widehat{\mathbb{NS}}'_\delta(f) \leq (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{8}\right)$$
$$\mathbb{NS}_\delta(f) > (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{4}\right) \implies \widehat{\mathbb{NS}}'_\delta(f) > (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{8}\right).$$

Proof. The existence of the values $t_1, \ldots, t_{m'}$ (with high probability) follows from the so-called *birthday problem* as follows. Let $\beta = 1/\lceil 1/\delta \rceil$ and partition $[0,1)$ into disjoint intervals $[(i-1)\beta, i\beta)$, $i \in \{1, \ldots, 1/\beta\}$. For any n iid Uniform$(0,1)$ samples $w_1, \ldots, w_n$ (for n as defined above), the probability none of these intervals contains more than one w_j value is $\prod_{j=1}^{n-1}(1-j\beta) \leq \exp\{-\beta n(n-1)/2\}$, and noting that $n \geq 1 + 2\sqrt{1/\beta}$, this is at most e^{-2}. Furthermore, on the event that there exists at least one interval containing more than one w_j value, for $\hat{j}$ defined as the first j such that $\exists j' < j$ with w_j and $w_{j'}$ in the same interval, we note that the (unconditional) distribution of $w_{\hat{j}}$ is Uniform$(0,1)$. Therefore, with probability at least $1 - e^{-2} - 4\beta$, there exists some $i \in \{3, \ldots, (1/\beta) - 2\}$ such that at least two w_j values are in $[(i-1)\beta, i\beta)$. Since $2\beta > 2\delta/(1+\delta) > \delta$, these intervals $[(i-1)\beta, i\beta)$ are strictly contained within $(\delta, 1-\delta)$. Furthermore, since $\beta < \delta < 1/32$, we have $1-e^{-2}-4\beta > 1-e^{-2}-1/8 > 1/2$. Thus, for each $t \in \{1, \ldots, s'/n\}$, the sequence $\{z'_i : i \in \{(t-1)n+1, \ldots, tn\}\}$ has probability at least $1/2$ of containing a pair z'_i, z'_j $(i < j)$ with $|z'_i - z'_j| < \delta$ and $z'_i \in (\delta, 1-\delta)$. In particular, the expected number of indices t for which such a pair exists is at least $(1/2)s'/n \geq 2m'$. Since these sequences are independent (over t), a Chernoff bound implies that with probability at least $1 - \exp\{-(1/2)(s'/n)/8\} \geq 5/6$ (for any choice of $c \geq 4\ln(6)$), at least m' of these sequences contain such a pair, so that on this event the values $t_1, \ldots, t_{m'}$ indeed exist.

Next, note that since $\mathcal{H}$ is the set of *constant* functions, for any x, x' and any $h \in \mathcal{H}_{(x,f(x))}$, we have $h(x') = h(x)$, which implies that for $x \sim$ Uniform$(0,1)$ and for $x' \sim$ Uniform$(x-\delta, x+\delta)$ given x, we have $\mathbb{NS}_\delta(f) = \mathbb{P}(f(x) \neq f(x'))$. Furthermore, z_1 has distribution Uniform$(\delta, 1-\delta)$ and the conditional distribution of y_1 given z_1 is Uniform$(z_1 - \delta, z_1 + \delta)$. Therefore,

$$\begin{aligned}\mathbb{P}(f(z_1) \neq f(y_1)) &= \mathbb{P}(f(x) \neq f(x') | x \in (\delta, 1-\delta)) \\ &= \frac{1}{1-2\delta}\mathbb{P}(f(x) \neq f(x') \wedge x \in (\delta, 1-\delta)),\end{aligned}$$

and this rightmost quantity is at least as large as

$$\begin{aligned}\frac{1}{1-2\delta}\left(\mathbb{P}(f(x) \neq f(x')) - 2\delta\right) &= \frac{1}{1-2\delta}\left(\mathbb{NS}_\delta(f) - 2\delta\right) \\ &\geq \frac{1}{1-2\delta}\left(\mathbb{NS}_\delta(f) - (k-1)\frac{\delta}{2}\frac{\epsilon}{19}\right)\end{aligned}$$

and at most as large as

$$\frac{1}{1-2\delta}\mathbb{P}(f(x) \neq f(x')) = \frac{1}{1-2\delta}\mathbb{NS}_\delta(f).$$

Thus,

$$\mathbb{NS}_\delta(f) \leq (k-1)\frac{\delta}{2} \implies \mathbb{P}(f(z_1) \neq f(y_1)) \leq \frac{1}{1-2\delta}(k-1)\frac{\delta}{2}$$

and since $\frac{\epsilon}{4} - \frac{\epsilon}{19} = \frac{15}{76}\epsilon$,

$$\mathbb{NS}_\delta(f) > (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{4}\right) \implies \mathbb{P}(f(z_1) \neq f(y_1)) > \frac{1}{1-2\delta}(k-1)\frac{\delta}{2}\left(1+\frac{15}{76}\epsilon\right).$$

Now recall that $\frac{1}{1-2\delta}(k-1)\frac{\delta}{2} > \frac{\epsilon^2}{65}$, and note that, if they exist, the pairs (z_r, y_r) are iid over $r \leq m'$. Therefore (recalling the definition of $\widehat{\mathbb{NS}}'_\delta(f)$ from above), a Chernoff bound implies that, for an appropriately large choice of the numerical constant c, with probability at least 5/6,

$$\mathbb{P}(f(z_1) \neq f(y_1)) \leq \frac{1}{1-2\delta}(k-1)\frac{\delta}{2} \implies \widehat{\mathbb{NS}}'_\delta(f) \leq (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{8}\right)$$

and

$$\begin{aligned}&\mathbb{P}(f(z_1) \neq f(y_1)) > \frac{1}{1-2\delta}(k-1)\frac{\delta}{2}\left(1+\frac{15}{76}\epsilon\right)\\ &\implies \widehat{\mathbb{NS}}'_\delta(f) > (k-1)\frac{\delta}{2}\left(1+\frac{15}{76}\epsilon\right)\left(1-\frac{\epsilon}{16}\right) > (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{8}\right).\end{aligned}$$

Altogether, on the above two events, we have

$$\mathbb{NS}_\delta(f) \leq (k-1)\frac{\delta}{2} \implies \widehat{\mathbb{NS}}'_\delta(f) \leq (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{8}\right)$$

and

$$\mathbb{NS}_\delta(f) > (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{4}\right) \implies \widehat{\mathbb{NS}}'_\delta(f) > (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{8}\right).$$

To complete the proof, we note that both of these events occur simultaneously with probability at least 2/3 by the union bound. □

Finally, we have the following result on testing of piecewise constant functions.

Theorem 4. *If $k \geq 80/\epsilon$, then the procedure that outputs* Accept *if $\widehat{\mathbb{NS}}'_\delta(f) \leq (k-1)\frac{\delta}{2}(1+\frac{\epsilon}{8})$, and otherwise outputs* Reject*, is an s'-sample q-query ϵ-tester for the class of k-piecewise constant functions under the distribution* Uniform$(0,1)$*, where $q = 2m'$ in the active testing variant and $q = s'$ for the passive testing variant.*

Proof. This proof is nearly identical to that of Theorem 3. If $f \in \mathcal{F}_k(\mathcal{H})$, then Lemma 1 implies it has $\mathbb{NS}_\delta(f) \leq (k-1)\frac{\delta}{2}$, so that Lemma 4 implies that with probability at least 2/3, $\widehat{\mathbb{NS}}'_\delta(f) \leq (k-1)\frac{\delta}{2}\left(1+\frac{\epsilon}{8}\right)$, and hence the tester will output Accept.

On the other hand, if f is ϵ-far from every function in $\mathcal{F}_k(\mathcal{H})$, then Lemma 2 implies that $\mathbb{NS}_\delta(f) > (k-1)\frac{\delta}{2}(1+\frac{\epsilon}{4})$, so that Lemma 4 implies that with probability at least 2/3, $\widehat{\mathbb{NS}}'_\delta(f) > (k-1)\frac{\delta}{2}(1+\frac{\epsilon}{8})$, and hence the tester will output Reject.

The number of samples and number of queries in the claim are immediate from the definition of the two testers. □

The upper bounds claimed in Theorem 2 immediately follow from this theorem, noting that $s' = O\left(\frac{\sqrt{k}}{\epsilon^5}\right)$, and also noting that for any $k < 80/\epsilon$ we can obtain the result with the tester based on the learn-then-validate technique of [20], as described above at the end of Section 3. In this latter case, the tester uses $\propto \frac{k}{\epsilon}\ln(2ek)\ln\left(\frac{1}{\epsilon}\right) = O\left(\frac{1}{\epsilon^2}\ln^2\left(\frac{1}{\epsilon}\right)\right)$ samples (since $d=1$ when $\mathcal{H}$ is the set of constant functions).

For the lower bound claimed in Theorem 2, first note that our assumption of $d > 0$ implies, in the case of $\mathcal{H}$ the set of constant functions, that $|\mathcal{Y}| \geq 2$. Therefore, we can *reduce* testing unions of $\lfloor (k-1)/2 \rfloor$ intervals to testing k-piecewise constant functions by associating the binary labels 0 and 1 with any two distinct labels $y_0, y_1 \in \mathcal{Y}$. Then any $\{0,1\}$-valued function f_{01} can be mapped to a corresponding $\{y_0, y_1\}$-valued function f, where f is a k-piecewise constant function if f_{01} is a union of $\lfloor (k-1)/2 \rfloor$ intervals, while f is ϵ-far from any k-piecewise constant function if f_{01} is ϵ-far from any union of $\lfloor (k-1)/2 \rfloor + 1$ intervals. This increase by one in the latter case is because the complement of a union of $\lfloor (k-1)/2 \rfloor$ intervals is also k-piecewise constant, but is possibly only representable as a union of $\lfloor (k-1)/2 \rfloor + 1$ intervals. To account for this increase by one, noting that the claim of an $\Omega\left(\sqrt{k}\right)$ lower bound only regards large values of k, if we suppose $k > 1/\epsilon$, then any union of $\lfloor (k-1)/2 \rfloor + 1$ intervals is within distance $1/k < \epsilon$ of a union of $\lfloor (k-1)/2 \rfloor$ intervals. Thus, f_{01} is ϵ-far from any union of $\lfloor (k-1)/2 \rfloor + 1$ intervals if it is 2ϵ-far from any union of $\lfloor (k-1)/2 \rfloor$ intervals. Altogether, we have that f is ϵ-far from any k-piecewise constant function if f_{01} is 2ϵ-far from any union of $\lfloor (k-1)/2 \rfloor$ intervals. So, by this reduction, the $\Omega\left(\sqrt{\lfloor (k-1)/2 \rfloor}\right) = \Omega\left(\sqrt{k}\right)$ lower bound of [20, 18] for passive testing of unions of $\lfloor (k-1)/2 \rfloor$ intervals implies a corresponding $\Omega\left(\sqrt{k}\right)$ lower bound for testing k-piecewise constant functions.[4] This completes the proof of

[4]Technically, the result of [20] establishes this lower bound for the problem of testing whether f_{01} is a union of n intervals or returns uniform random $\{0,1\}$ labels. However, essentially the same argument would apply if, in the latter case, instead of random labels, we took f_{01} to be a randomly-chosen binary *function* based on a partition of $[0,1]$ into $n' \gg n$ equal-sized regions, which would be at least $1/4 > 2\epsilon$ distance from any union of n intervals

Theorem 2.

5. Open Problem on the Query Complexity of Testing Polynomials

While the result above for active testing, when specialized to testing k-piecewise degree-p polynomials, obtains the optimal dependence on the number of pieces k, we were not able to show optimality in the degree p. This leads to an even more-basic question:

Open Problem: What is the optimal dependence on p in the query complexity of testing degree-p real-valued polynomials under $\mathcal{P} = \text{Uniform}(0, 1)$?

This question is open at this time, for *both* the active and passive property testing settings, and indeed also for the stronger value-query setting (where the tester can query for $f(x)$ at any $x \in \mathcal{X}$).

There is a trivial $p + 1 + \frac{1}{\epsilon}\ln(3)$ upper bound for both active and passive testing, based on the learn-then-validate technique, since $p+1$ random samples uniquely specify any degree-p polynomial (with probability one): that is, if we fit a degree-p polynomial $\hat{f}$ to the first $p + 1$ random points, then if f is a degree-p polynomial we would have $\hat{f} = f$ and thus the two would agree on the remaining $\frac{1}{\epsilon}\ln(3)$ points, in which case we may decide ACCEPT; on the other hand, if f is ϵ-far from all degree-p polynomials, then it is ϵ-far from $\hat{f}$, and hence with probability at least 2/3 at least one of $\frac{1}{\epsilon}\ln(3)$ random samples x will have $\hat{f}(x) \neq f(x)$, in which case we may decide REJECT.

One might naïvely think that, since it is possible to fit any $p + 1$ values (at distinct x's) with a degree-p polynomial, a *lower bound* of $\Omega(p)$ should also hold. However, we note that it is also possible to fit any k values (at distinct x's) with a k-piecewise constant function, and yet above we proved it is possible to test k-piecewise constant functions using a number of queries with $\sqrt{k}$ dependence on k by passive testing or *independent* of k by active testing. So the mere ability to fit $p+1$ arbitrary values with a degree-p polynomial is not in-itself sufficient as a basis for proving a lower bound on the query complexity of testing degree-p polynomials.

Bibliography

[1] R. Rubinfeld, M. Sudan, Robust characterizations of polynomials with applications to program testing, SIAM Journal on Computing 25 (1996) 287–294.

[2] O. Goldreich, S. Goldwasser, D. Ron, Property testing and its connection to learning and approximation, Journal of the ACM 45 (4) (1998) 653–750.

with very high probability.

[3] L. G. Valiant, A theory of the learnable, Communications of the Association for Computing Machinery 27 (11) (1984) 1134–1142.

[4] V. Vapnik, A. Chervonenkis, Theory of Pattern Recognition, Nauka, Moscow, 1974.

[5] A. Blumer, A. Ehrenfeucht, D. Haussler, M. Warmuth, Learnability and the Vapnik-Chervonenkis dimension, Journal of the Association for Computing Machinery 36 (4) (1989) 929–965.

[6] A. Ehrenfeucht, D. Haussler, M. Kearns, L. Valiant, A general lower bound on the number of examples needed for learning, Information and Computation 82 (3) (1989) 247–261.

[7] S. Hanneke, The optimal sample complexity of PAC learning, Journal of Machine Learning Research 17 (38) (2016) 1–15.

[8] S. Hanneke, Theory of disagreement-based active learning, Foundations and Trends in Machine Learning 7 (2–3) (2014) 131–309.

[9] S. Dasgupta, Coarse sample complexity bounds for active learning, in: Advances in Neural Information Processing Systems 18, 2005.

[10] D. Ron, Property testing: A learning theory perspective, Foundations and Trends in Machine Learning 1 (3) (2008) 307–402.

[11] D. Ron, Algorithmic and analysis techniques in property testing, Foundations and Trends in Theoretical Computer Science 5 (2) (2010) 73–205.

[12] O. Goldreich, Introduction to Property Testing, Cambridge University Press, 2017.

[13] O. Goldreich, S. Goldwasser, E. Lehman, D. Ron, A. Samorodnitsky, Testing monotonicity, Combinatorica 20 (3) (2000) 301–337.

[14] F. Ergun, S. Kannan, S. R. Kumar, R. Rubinfeld, M. Viswanathan, Spotcheckers, Journal of Computer and System Sciences 60 (3) (2000) 717–751.

[15] P. Berman, S. Raskhodnikova, G. Yaroslavtsev, l_p-testing, in: Proceedings of the 46$^{\text{th}}$ Annual ACM Symposium on Theory of Computing, 2014.

[16] R. Baleshzar, D. Chakrabarty, R. K. S. Pallavoor, S. Raskhodnikova, C. Seshadhri, Optimal unateness testers for real-valued functions: Adaptivity helps, in: 44$^{\text{th}}$ International Colloquium on Automata, Languages, and Programming, 2017.

[17] M. Jha, S. Raskhodnikova, Testing and reconstruction of Lipschitz functions with applications to data privacy, SIAM Journal on Computing 42 (2) (2013) 700–731.

[18] M.-F. Balcan, E. Blais, A. Blum, L. Yang, Active property testing, in: Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science, 2012.

[19] B. Settles, Active Learning, Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Publishers, 2012.

[20] M. Kearns, D. Ron, Testing problems with sub-learning sample complexity, Journal of Computer and System Sciences 61 (3) (2000) 428–456.

[21] P. Kothari, A. Nayyeri, R. O'Donnell, C. Wu, Testing surface area, in: Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms, 2014.

[22] V. Vapnik, A. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, Theory of Probability and its Applications 16 (2) (1971) 264–280.

[23] S. Boucheron, O. Bousquet, G. Lugosi, Theory of classification: A survey of some recent advances, ESAIM: Probability and Statistics 9 (2005) 323–375.