Learning Random Monotone DNF Under the Uniform Distribution

Linda Sellie[*]
University of Chicago, Chicago IL
lmsellie@uchicago.edu

Abstract

We show that randomly generated monotone $c\log(n)$-DNF formula can be learned exactly in probabilistic polynomial time. Our notion of randomly generated is with respect to a uniform distribution. To prove this we identify the class of *well behaved* monotone $c\log(n)$-DNF formulae, and show that almost every monotone DNF formula is well-behaved, and that there exists a probabilistic Turing machine that exactly learns all well behaved monotone $c\log(n)$-DNF formula.

1 Introduction

Intuitively, a monotone $c\log(n)$-DNF, f, is well behaved if it satisfies three smoothness criteria—the "small", "medium," and "large" z properties—that collectively rule out having an unexpectedly large number of terms having a common subset of the variables. Thus by removing terms we maintain our well-behaved criteria and we have:

Theorem (Subset property of the set of well-behaved functions)**.** *If f is well-behaved and f' contains a subset of the terms of f then f' is also well-behaved.*

The question of what is meant by "a randomly generated monotone $c\log(n)$-DNF formula" is somewhat application specific, but because of the subset property of the set of well-behaved functions, our learning algorithm and proof of correctness is quite robust. We imagine a process that randomly selects m terms of size $c\log(n)$; we show that such a function will be well behaved with high probability as long as $m \leq 2\log\log(n)n^c$ (where roughly $\frac{1}{\log(n)}$ of the examples will be false when $m = 2\log\log(n)n^c$.) This subsumes standard notions of randomness that are intended to generate formula which are expected to be true with fixed probability less than one. For functions with the small and medium smoothness properties and for a set of variables, s, of bounded size, we can efficiently determine by sampling whether or not there exists a term $t \in f$ such that $s \subset t$ with high probability. Our algorithm considers all subsets of variables, s, of a given, fixed size. To extend s we make multiple trials of random extension of s, through $|s| = \beta(n) = \log\log\sqrt[3]{n}$. The medium and small subset properties guarantee that with high probability, if $s' \supset s$ has size at most $\beta(n)$ and there exists a term $t \in f$ such that $s' \subset t$, then s' is generated by this process. At this point, the large smoothness property comes into play and guarantees that the previous t is unique, and therefore can be efficiently found. In this way, we find all terms t of f in polynomial time.

1.1 Motivation and Past Work

Mentioning DNF, Valiant [12] states:

> The possible importance of disjunctions of conjunctions as a knowledge representation stems from the observations that on the one hand humans appear to like using it, and, on the other, that there is circumstantial evidence that significantly larger classes may not be learnable in polynomial time.

Many learning theorist have considered learning monotone DNF formula. Angluin [2] completely solved this problem for the case of exact learning using membership queries — all monotone DNF are learnable in polynomial time in this model for all distributions. This problem has proven more difficult if the learner is restricted to sampling, i.e. learning by example. The obstacle seems to be "cluster structure" within the formula, specifically a relatively large set of variables common to a relatively large number of clauses. Existing results in the literature tackle this obstacle in two different ways. (1) allow the running time of the learner to explode in the face of such clusters, e.g. Verbeurgt [13] learns any poly(n)-size DNF in time $n^{O(\log(n))}$ from uniform examples. Or (2) consider classes of formula that do not contain such clusters, specifically by random generation and limited number of terms, e.g. Servedio [9] learns any $2^{\sqrt{\log(n)}}$-term DNF in polynomial time from a product distribution. Other researchers have used similar approaches to other problems, [10], [8], and [6].

The results of this paper belong to group (2). Our result is distinguished from Servedio [9] in that our definition of *well behaved* represents an initial attempt to formalize the obstacle, and to obtain the best possible result based on that formalization. From this, we obtain conditions of greater generality.

Despite the difficulty of learning monotone DNF with random examples drawn from the uniform distribution, the

[*] Computer Science Department, University of Chicago.

naturalness of the class suggests in some restricted form, it must be possible to learn. In their 1994 paper, Aizenstein and Pitt proposed learning most DNF instead of all DNF. They defined "most" as the DNF generated randomly with certain parameters set, one parameter is choosing the variables in a term with probability $\frac{1}{2}$. They left as an open question a more natural setting of those parameters. Jackson and Servedio in 2006 started answering the open question of Aizenstein and Pitt in their paper [7]. They learned "most" monotone DNF where the number of terms is bounded by $O(n^{2-\gamma})$ with fixed term size, $\log m$, where m is the number of terms. We continue this work left open by Aizenstein and Pitt, and Jackson and Servedio.

We expand the approach used by Jackson and Servedio in their paper [7]. To learn random monotone DNF with $O(n^{2-\gamma})$ number of terms, they use a clustering algorithm after using an inclusion/exclusion pair finding algorithm. In our paper, we learn $O(n^c)$ number of terms in polynomial time for any constant c, and fixed term size, $c\log(n)$.

Similar results are independently obtained by Jackson, Lee, Servedio and Wan [5] but are slightly weaker. They use a similar algorithm but significantly different underlying proofs.

Theorem 1. *Given a random monotone DNF, f, Algorithm* **Learn Random Monotone DNF** *finds f in polynomial time with high probability.*

1.2 Our Model and Random Functions

Continuing the work of Aizenstein and Pitt [1] and Jackson and Servedio [7], we explore learning a function chosen randomly from a large class of functions. Jackson and Servedio learn a monotone DNF formula chosen randomly from a subclass of monotone DNF; we do the same except we choose a larger subclass of monotone DNF. As in Jackson and Servedio, we randomly choose the terms for our function from $\binom{n}{k}$ possible terms of size k. We differ from Jackson and Servedio's choice of a class of functions in two ways. The most important is that we learn functions with n^c terms for any c, while they learn only for $c \leq 2 - \gamma$ for $\gamma > 0$. The second way we differ is by loosening Jackson and Servedio's restriction which bounds the function away from 0 and 1 by a constant; we restrict our attention to functions that are bounded away from one by a slow growing function in n, and without restriction on how close the function is to zero. Even in the case of $c \leq 2$, for large n, the set of functions they learn is a subset of the functions we learn. They allow the number of terms, m, to be $\alpha 2^k \leq m \leq 2^{k+1} \ln \frac{2}{\alpha}$ for a constant α, $(0 < \alpha < 0.09)$. Instead, we restrict the number of terms, m, to be $m \leq 2^{k+1} c \log\log(n)$.

As Jackson and Servedio in [7]; we learn in the uniform distribution model; where each example is chosen uniformly at random and labeled according to the unknown function.

Our goal is stronger than theirs, in that we exactly learn with probability $1 - \delta$. (They learn a function which is ϵ close with probability $1 - \delta$.) We run in time polynomial in the probability of an example satisfying a term, (i.e. time polynomial in 2^k.)

The model for our class of random monotone DNF formulas is as follows, let $\mathcal{F}^{n,k,m}$ be the set of monotone DNF over n variables, with terms of size k, and m terms. Or interest is when with $m \leq 2^{k+1} c \log\log(n)$ where $c = \frac{k}{\log(n)}$. Each term is selected independently and uniformly from the set of all k-variable terms.

2 Notation and Definitions

Our function will be defined on n variables; we let $X = \{x_1, x_2, \ldots, x_n\}$ be the set of variables. For $s \subset X$ we define $X \backslash s = \{x \in X \mid x \notin s\}$. Let $t \subset X$ be a term, and $k = |t| = c\log(n)$ be the size of a term. Let m be the number of terms; where $m \leq m_{\max} = 2^{k+1} c \log\log(n) = 2n^c \log\log(n)$. Let $f = \underset{i=1\ldots m}{\cup} t_i$. We define $f \backslash t = \{t' \in f \mid t' \neq t\}$.

Let $E = \{0,1\}^n$ be the set of all examples, and $E^+ = \{e \in E \mid f(e)\}$ be the set of all positive examples. Let $s \subset X$, and let a_s be a partial assignment of the variables in s. For $x \in s$, and a_s a partial assignment, then by an abuse of notation, we define $x(a_s) = 1$ iff the assignment to x is 1 and 0 otherwise. Let $X_{a_s} = \{x \in s \mid x(a_s)\}$ be the set of variables in s that a_s satisfies.

We use a_s to partition the set of examples, E^+, and the set terms in f. We then explore the relationship between these sets in the paper with an inclusion/exclusion algorithm that allows us to find subsets of terms in f. Let $E_{a_s} = \{e \in E \mid \forall x \in s, x(a_s) = x(e)\}$ be the set of assignments who agree with a_s on all the variables s. Let $E^+_{a_s} = E_{a_s} \cap E^+$ be the set of positive assignments who agree with a_s on all the variables s. Let $T_{a_s} = \{t \in f \mid t \cap s = X_{a_s}\}$.

Let $T_e = \{t \in f \mid t(e)\}$ be the set of terms satisfying an example e. Let $\#_0(a_s) = |\{x \in s \mid x(a_s) = 0\}|$, e.g. $\#_0(10110) = 2$.

Let $\beta(n) = \log\log(\sqrt[3]{n})$. We use $\beta(n)$ throughout the paper as a bound that is not constant – and not too large. For technical ease we chose this value of $\beta(n)$, we could have chosen other values for $\beta(n)$, such as $\beta(n) = \log\log(n)$.

For ease of notation, we define $i^{\underline{j}} = \frac{i!}{(i-j)!}$.

For e and a_s, we define a transformation, $e' = e_{s \leftarrow a_s}$ to be $\forall x \notin s, x(e') = x(e)$ and $\forall x \in s, x(e') = x(a_s)$. For e and x' we define $e' = e_{\text{flip}(x')}$ to be $\forall x \in (X \setminus \{x'\}), x(e') = x(e)$ and $x'(e') = 1 - x'(e)$.

Our algorithm discovers f by finding subsets of terms in f. We use the knowledge that a set of variables, s, is a subset of a term iff there is a positive example, e, which becomes false for $e_{\text{flip}(x)}$ for any $x \in s$.

Definition 2. *e is s-minimal for f iff*

- *$e \in E^+$, and*
- *$\forall x \in s, e_{x \leftarrow 0} \notin E^+$.*

We define the set of s-minimal examples:

Definition 3. *Let $\Upsilon_s = \{e \in E^+ \mid e \text{ is } s\text{-minimal}\}$.*

Thus, given any $e \in \Upsilon_s$ and $t \in f$, if $t(e) = 1$ then $s \subset t$.

The idea behind our proof is that we can determine if Υ_s is non-empty for any s of cardinality greater than $c + 1$ and less than or equal to $\beta(n) + 1$, thus finding a subset of a term.

```
Function Distinguishing Subsets
• $S = \left\{ s \subset X \mid |s| = c+2, \text{ and } I_s > 2^n \cdot \frac{1}{n^{c+\frac{1}{5}}} \right\}$
• For $i = (c+3)$ to $\beta(n)$
    – $S' = \emptyset$
    – For $s \in S$ and $x \in X$
        * If $I_{s\cup\{x\}} > 2^n \cdot \frac{1}{n^{c+\frac{1}{5}}}$ then add $(s \cup \{x\})$ to $S'$
    – $S = S'$
• Return $S$
```

Figure 1: Function **Distinguishing Subsets**

From this knowledge we build the rest of the term. Our goal is to exactly learn f with probability $1-\delta$.

Unfortunately we don't know how to compute the size of Υ_s. Instead we estimate $|\Upsilon_s|$.

Definition 4. *Let* $I_s = \sum_{a_s} (-1)^{\#_0(a_s)} |E^+_{a_s}|$.

Our paper focuses on proving that most monotone k-DNF are well behaved, that for well behaved functions I_s approximates $|\Upsilon_s|$ for $c+2 \le |s| \le \beta(n)+1$, and if $s \subset t \in f$ then Υ_s is sufficiently large.

The organization of our paper is as follows: in Section 3 we present a simple algorithm that exploits the knowledge that we can find a subset of at term. In Subsection 4.1 we partition the set of positive examples and use this partition to define how I_s miscounts the size of Υ_s. In Subsection 4.2 we prove most $f \in \mathcal{F}^{n,k,m}$ are well behaved. In Subsection 4.3 we bound by how much I_s misclassifies $|\Upsilon_s|$ for well behaved functions. In Subsection 4.4 we prove that for well-behaved $f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}$, if $s \subset t \in f$ then Υ_s is sufficiently large enabling us to discover if $s \subset t$.

We put some of the technical details into the appendix. In Appendix 1A, we provide some observations and simplifications of algebraic expressions used in our proofs. In Appendix 1B, we prove that most $f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}$ are well behaved. In Appendix 1C, we bound Υ_t for well behaved DNF. In Appendix 2A, we use standard sampling techniques to prove we can approximate $|E^+_{a_s}|$ sufficiently. In Appendix 2B, for the sake of completeness, we provide the details that show our algorithm finds the unknown monotone DNF in polynomial time with high probability.

3 The Algorithm for Finding f Using I_s

Using I_s as our estimate for $|\Upsilon_s|$, our algorithm builds terms in three stages. First our algorithm tests all subsets of size $c+2$, selecting those that are a subset of a term in f. Next, it builds upon these subsets, variable by variable, till it has found all subsets of terms of f of size $\beta(n)$. Finally, having a subset unique to a single term in f (we prove the uniqueness of terms of size $\beta(n)$ later in the paper in Corollary 46,) we find the rest of variables for this term. The steps of the first two stages are in Figure 1 and the steps for the third stage are in Figure 2.

```
Algorithm Learn Random Monotone DNF
• $S =$ Distinguishing Subsets
• $f = \emptyset$
• For $s \in S$
    – $t = \emptyset$
    – For $x \in X$
        * If $I_{s\cup\{x\}} > 2^n \cdot \frac{1}{n^{c+\frac{1}{5}}}$ then add $x$ to $t$
    – add $t$ to $f$
• Return $f$
```

Figure 2: Algorithm **Learn Random Monotone DNF**

4 Approximating Υ_s by I_s

In this section, we show that with high probability I_s approximates $|\Upsilon_s|$ for $c+2 \le |s| \le \beta(n)+1$ to within $2^n \cdot \frac{4k\log^6(n)n^{2/3}}{n^{c+1}}$ (i.e. $|I_s - |\Upsilon_s|| < 2^n \cdot \frac{4k\log^6(n)n^{2/3}}{n^{c+1}}$.) We use Subsections 4.1 through 4.3 to prove this main theorem. In Subsection 4.4, we prove that $|\Upsilon_s| \ge 2^n \cdot \frac{1}{8\log^{4c}(n)}\frac{1}{n^c}$ if and only if $s \subset t \in f$.

4.1 Observations about I_s

To explore how I_s relates to the size of Υ_s, we partition the set E^+ of positive examples. We partition E^+ by grouping examples that "map" under s to the same example in $E^+_{1_s}$. Observing the behavior of a partition during the calculation of I_s, we bound how I_s misjudges the size of Υ_s. (We bound the size of the miscalculation in Subsection 4.3.)

Definition 5. *For $s \subset X$, and $e \in E^+_{1_s}$, we define a set of partial assignments, $A_{e,s} = \{a_s \mid e_{s\leftarrow a_s} \in E^+\}$, which map e to another positive example under s.*

Next, using this partition of the set of positive examples, we define a criteria for $e \in E^+_{1_s}$ to be correctly counted.

Definition 6. *For $e \in E^+_{1_s}$ we define*

$$\mathcal{I}_e = \sum_{a_s \in A_{e,s}} (-1)^{\#_0(a_s)}.$$

Observation 7. $I_s = \sum_{e\in E^+_{1_s}} \mathcal{I}_e$.

Definition 8. *An $e \in E^+_{1_s}$ is* correctly counted *iff $e \in E^+_{1_s}\setminus\Upsilon_s$ then $\mathcal{I}_e = 0$, and if $e \in \Upsilon_s$ then $\mathcal{I}_e = 1$.*

Note, if all examples in $E^+_{1_s}$ are counted correctly then $I_s = |\Upsilon_s|$. We observe that $e \in \Upsilon_s$ is correctly counted.

Lemma 9. *For all $e \in \Upsilon_s$ then $\mathcal{I}_e = 1$.*

Proof: $A_{e,s} = \{1_s\}$. □

By characterizing examples which are correctly counted, we restrict the number of examples that could be incorrectly counted. We describe two ways examples are correctly counted.

Lemma 10. *An example, $e \in E^+_{1_s}$, is correctly counted if $\exists x \in s$ such that $\forall a_s \in A_{e,s}$, $(a_s)_{\text{flip}(x)} \in A_{e,s}$.*

Proof: Let x be such that $\forall a_s \in A_{e,s}$ and $(a_s)_{\text{flip}(x)} \in A_{e,s}$ then $(-1)^{\#_0(a_s)}$ and $(-1)^{\#_0((a_s)_{\text{flip}(x)})}$ are included in the sum, where the parity of $\#_0(a_s)$ and $\#_0((a_s)_{\text{flip}(x)})$ are opposite. Thus $\mathcal{I}_e = 0$ and e is counted correctly. □

Corollary 11. *An example, $e \in E^+_{1_s}$, is* correctly counted *if $\exists t \in f$ such that $t(e)$ and $t \cap s = \emptyset$.*

Proof: The Corollary follows from Lemma 10 and the definition of $\mathcal{I}_e$ since all partial assignments are contained in $A_{e,s}$. □

If e is not known to be correctly counted by Lemma 10 and Corollary 11, it may or may not be correctly counted, but our proof will not need to consider this option.

Definition 12. *Let $s \subset X$, we define the set of miscounted examples by*

$$M_s = \left\{ e \in E^+_{1_s} \backslash \Upsilon_s \mid \mathcal{I}_e \neq 0 \right\}.$$

Partitioning M_s based on sets of partial assignments, we simplify bounding the number of miscounted examples.

Definition 13. *Let $s \subset X$, and $A \subset \{0,1\}^{|s|}$; we define $M_{s,A} = \{e \in M_s \mid A_{e,s} = A\}$.*

We define a partial order by $a'_s \prec a''_s$ iff $\forall x \in s$ then $x(a'_s) \leq x(a''_s)$ and $a'_s \neq a''_s$. The smallest partial assignments are very important to our proof; they determine if an example $e \in E^+_{1_s}$ is miscounted.

Definition 14. *Let $A \subseteq \{0,1\}^{|s|}$, we define*

$$L(A) = \{a_s \in A \mid \forall a'_s \in A, a'_s \not\prec a_s\}.$$

Lemma 15. *Let $e \in M_{s,A}$, then $\forall a_s \in L(A), \exists t \in T_{a_s}$ where $t(e)$.*

Proof: By definition 5, given any $e \in M_{s,A}$ and $\forall a_s \in L(A)$, $\exists e' \in E^+_{a_s}$ such that e' maps under s to e (i.e. $e = e'_{s \leftarrow 1_s}$) which implies $e' = e_{s \leftarrow a_s}$. Because e' is X_{a_s}-minimal, we know $\exists t \in T_{a_s}$ such that $t(e')$ which implies $t(e)$ since f is monotone. □

We have now proved in Lemmas 10 and 15 that every miscounted example is satisfied by a set of terms whose union contains s. We will use this fact in Lemma 24 where we bound the number of miscounted examples in $M_{s,A}$.

Knowing A is a subset of the partial assignments to s, we calculate by how much an example has been miscounted.

Observation 16. *Let $e \in M_{s,A}$ then $|\mathcal{I}_e| < |A| \leq 2^{|s|}$.*

Definition 17. *Let $\mathbf{A_s} = \{A | A = A_{e,s}$ for an $e \in M_s\}$.*

Observation 18.

$$I_s - |\Upsilon_s| = \sum_{e \in M_s} \mathcal{I}_e = \sum_{A \in \mathbf{A_s}} \sum_{e \in M_{s,A}} \mathcal{I}_e.$$

4.2 Properties of Well Behaved Functions

In this subsection, we describe the properties a function needs for our proof to hold; our algorithm works for functions that are not "clustered" together. We prove that with high probability these properties hold for $f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}$. We will call DNF formulas that have this property "well behaved."

Definition 19. *A monotone DNF function, $f \in \mathcal{F}^{n,k,m}$ is* well behaved *iff for all $s \subset t \in f$ where $|s| \leq \beta(n) + 1$, and $\forall a_s$ where $z = \#_1(a_s)$ then*

- Small z property:
  if $0 < z \leq c$ then $|T_{a_s}| < 3m_{\max}k^{\underline{z}}/n^{\underline{z}}$,
- Medium z property:
  if $c < z < \beta(n)$ then $|T_{a_s}| < \beta(n)$, and
- Large z property:
  if $z \geq \beta(n)$ then $|T_{a_s}| \leq 1$.

Using Chernoff bounds we prove random monotone DNF are well behaved with high probability.

Theorem 20. *For a fixed c and sufficiently large n, if $f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}$ for $m \leq 2^{k+1}c \log\log(n)$ then f is well behaved with probability at least $1 - n^{2c}\log(n)\left(\frac{1}{n}\right)^{\beta(n)}$.*

Proof: This follows from Corollaries 42, 44,and 46 (found in the appendix,) and noting that the probability of small, medium and large z properties of being well behaved are not satisfied with probability at most $\frac{1}{3}n^{2c}\log(n)\left(\frac{1}{n}\right)^{\beta(n)-1} + \frac{1}{3}n^{2c}\log(n)/3\left(\frac{1}{n}\right)^{\beta(n)-1} + \frac{1}{3}n^{2c}\log(n)/3\left(\frac{1}{n}\right)^{\beta(n)-1}$. Consequently $f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}$ is well behaved with probability at least $1 - n^{2c}\log(n)\left(\frac{1}{n}\right)^{\beta(n)}$. □

4.3 Observations about well behaved Monotone DNF Formulas

In this subsection, we derive some properties of well behaved functions. First, we bound the number of variables that occur in more than one term from a set of terms, $T \subset f$ for $f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}$. Next we bound the probability an example satisfies every term in T. Third, we bound the size of $M_{s,A}$, using the probability an example satisfies a term in T_{a_s} for every $a_s \in L(A)$. At the end of this subsection we bound $|M_s|$ and $I_s - |M_s|$.

Corollary 21. *Let f be a well behaved monotone k-DNF formula and $T \subset f$, then $|\{x \mid x \in (t \cap t')$ for some $t, t' \in T\}| < |T|^2\beta(n)$.*

Proof: For f, a well behaved monotone k-DNF, we know that a pair of terms $t, t' \in f$ have in common at most $\beta(n)$ variables. Since the number of pairs is $\binom{|T|}{2}$, we bound the total number of variables used by more that one term by $\binom{|T|}{2}\beta(n)$. Note that what we've proved is stronger than what we've claimed. The form of our claim is for our subsequent technical convenience. □

Knowing an upper bound on the number of variables occurring in a set of terms, we bound the probability an example satisfies every term in this set of terms.

Lemma 22. *Let f be a well behaved monotone k-DNF, and $T \subset f$ a subset of terms then*

$$|\{e \in E \mid \forall t \in T, t(e)\}| \leq 2^n \cdot \frac{1}{2^{(|T|k-|T|^2\beta(n))}}.$$

Proof: The T terms share at most $|T|^2\beta(n)$ variables out $|T|k$ variables by Corollary 21. Thus the number of variables that need to be satisfied is at least $|T|k - |T|^2\beta(n)$. □

We note that if we restrict our examples to have the bits in s set to one, we get the following corollary.

Corollary 23. *For $s \subset X$ and $|\{e \in E \mid \forall t \in T, t(e)\}| \leq 2^n \cdot \frac{1}{2^{(|T|k-|T|^2\beta(n))}}$ then*

$$|\{e \in E_{1_s} \mid \forall t \in T, t(e)\}| \leq 2^n \cdot \frac{1}{2^{(|T|k-|T|^2\beta(n))}}.$$

Proof: The size of the set E_{1_s} is $2^{n-|s|}$. Given that $|\{e \in E \mid \forall t \in T, t(e)\}| \leq 2^n \cdot \frac{1}{2^{(|T|k-|T|^2\beta(n))}}$, the restriction of the variables to be from the set E_{1_s} reduces the number of variables that must be satisfied to at least $(|T|k - |T|^2\beta(n) - |s|)$. (i.e. at most $|s|$ bits were forced to one.) Thus $|\{e \in E_{1_s} \mid \forall t \in T, t(e)\}| \leq 2^{n-|s|} \cdot \frac{1}{2^{(|T|k-|T|^2\beta(n)-|s|)}}$ $= 2^n \cdot \frac{1}{2^{(|T|k-|T|^2\beta(n))}}$. □

We now bound the number of examples in $M_{s,A}$.

Lemma 24. *For fixed c and sufficiently large n, let f be a well behaved monotone k-DNF, $s \subset X$ where $c + 2 \leq |s| \leq \beta(n) + 1$, and $A \in \mathbf{A_s}$ then $|M_{s,A}| < 2^n \cdot \frac{k \log^5(n)}{n^{c+1}}$.*

Proof: Let $v = |L(A)|$.

As noted in Lemma 15, $e \in M_{s,A}$ are satisfied by at least one term from every T_{a_s} for every $a_s \in L(A)$. From Corollary 23, we know that the probability an example satisfies a set of v terms in E_{1_s} is at most $2^n \cdot \frac{1}{2^{vk-v^2\beta(n)}}$

Therefore we bound $|M_{s,A}|$ by bounding the number of $e \in M_{s,A}$ which is satisfied by at least one term from every T_{a_s} for every $a_s \in L(A)$. We create this bound by using a Bonferroni type argument.

$$\begin{aligned} & |M_{s,A}| \\ = \;& |\{e \in M_s \mid \forall a_s \in L(A), \exists t \in T_{a_s}, t(e)\}| \quad \text{(Def. 13.)} \\ \leq \;& 2^n \cdot \frac{1}{2^{vk-v^2\beta(n)}} \prod_{a_s \in L(A)} |T_{a_s}| \quad \text{(Lemma 15.)} \end{aligned}$$

In counting the number of possible ways an example $e \in M_{s,A}$ could be satisfied by one term from every T_{a_s}, for every $a_s \in L(A)$, we consider two cases.

In the first case, we assume that for all $a_s \in L(A)$ that $\#_1(a_s) \leq c$. Using the assumption that f is well defined, we know that $|T_{a_s}| < 3m_{\max}\left(\frac{k^{\#_1(a_s)}}{n^{\underline{\#_1(a_s)}}}\right)$, we compute the probability as follows.

$$|M_{s,A}| \leq 2^n \cdot \frac{1}{2^{vk-v^2\beta(n)}} \prod_{a_s \in L(A)} 3m_{\max}\left(\frac{k^{\#_1(a_s)}}{n^{\underline{\#_1(a_s)}}}\right).$$

By Lemma 10 and Corollary 11, $s \subseteq \left(\underset{t \in T_e}{\cup} t\right)$ and $\forall a_s \in A_{e,s}, \#_1(a_s) \geq 1$. Let $w = \sum_{a_s \in L(A)} \#_1(a_s) \geq \max\{|L(A)|, |s|\}$ (and since $v = |L(A)|$.) This implies that

$$\begin{aligned} |M_{s,A}| &\leq 2^n \cdot \frac{3^v m_{\max}^v}{2^{vk-v^2\beta(n)}} \frac{k^w}{n^{\underline{w}}} \\ &\leq 2^n \cdot \frac{2^{v^2\beta(n)}(6c\log\log(n))^v n^{cv}}{2^{vk}} \frac{k^w}{n^{\underline{w}}} \\ &= 2^n \cdot 2^{v^2\beta(n)}(6c\log\log(n))^v \frac{k^w}{n^{\underline{w}}} \quad (n^{cv} = 2^{kv}.) \\ &\leq 2^n \cdot \sqrt[3]{n}(6c\log\log(n))^{c+2} \frac{k^{c+2}}{n^{\underline{c+2}}} \\ &\quad \text{(From Obs. 35, } w \geq v, \text{ and } w \geq |s| \geq c+2.) \\ &\leq 2^n \cdot \frac{1}{n^{c+1}} \quad \text{(From Observation 33.)} \end{aligned}$$

In the second case, there exists an $a'_s \in L(A)$ such that $\#_1(a'_s) > c$; by f being well behaved we know that $|T_{a'_s}| < \beta(n)$. Let $v' = |\{a'_s \in L(A) | \#_1(a'_s) > c\}|$. If $a_s \in L(A)$ where $\#_1(a_s) \leq c$ then by f being well behaved we know that $|T_{a_s}| < 3m_{\max}\left(\frac{k^{\#_1(a_s)}}{n^{\underline{\#_1(a_s)}}}\right)$. Using these bounds, we compute an upper bound by again noting that $e' \in M_{s,A}$ is satisfied by one from each T_{a_s} for all $a_s \in L(A)$.

$$|M_{s,A}| \leq 2^n \cdot \frac{(\beta(n))^{v'}}{2^{vk-v^2\beta(n)}} \prod_{a_s \in L(A), \#_1(a_s) \leq c} 3m_{\max} \cdot \frac{k^{\#_1(a_s)}}{n^{\underline{\#_1(a_s)}}}.$$

By Lemma 11, we know $\#_1(a_s) \geq 1$ for all $a_s \in L(A)$, and $\left(\frac{k^{\#_1(a_s)}}{n^{\underline{\#_1(a_s)}}}\right) \leq \left(\frac{k}{n}\right)$. we reduce the formula so that

$$\begin{aligned} & |M_{s,A}| \\ \leq \;& 2^n \cdot \frac{(\beta(n))^{v'}}{2^{vk-v^2\beta(n)}}(3m_{\max})^{v-v'}\left(\frac{k}{n}\right)^{(v-v')} \\ \leq \;& 2^n \cdot \frac{(\beta(n))^{v'}}{2^{vk-v^2\beta(n)}}(6c\log\log(n)2^k)^{(v-v')}\left(\frac{k}{n}\right)^{(v-v')} \\ & \qquad (\text{since } n^{c(v-v')} = 2^{k(v-v')}.) \\ \leq \;& 2^n \cdot \frac{(\beta(n))^{v'}2^{v^2\beta(n)}}{2^{v'k}}(6c\log\log(n))^{(v-v')}\left(\frac{k}{n}\right)^{(v-v')} \end{aligned}$$

We now break the calculations down into two sub-cases. If $v = 2$ then the equation is largest if $v' = 1$. In this case we bound $|M_{s,A}|$ by $2^n \cdot \frac{\beta(n)2^{4\beta(n)}}{2^k}(6c\log\log(n))\left(\frac{k}{n}\right) \leq 2^n \cdot \frac{\beta(n)\log^4(\sqrt[3]{n})}{2^k}(6c\log\log(n))\left(\frac{k}{n}\right) < 2^n \cdot \frac{\log^5(n)k}{n2^k}$.

If $v \geq 3$, we note [1] this equation is again largest if $v' = 1$, and using Observation 35, we reduce the formula to:

$$2^n \cdot \frac{\beta(n)\sqrt[3]{n}}{2^k}(6c\log\log(n))^{(v-1)}\left(\frac{k}{n}\right)^{(v-1)} \leq 2^n \cdot \frac{1}{n2^k}.$$

Therefore $|M_{s,A}| \leq 2^n \cdot \frac{k\log^5(n)}{n^{c+1}}$. □

Having computed an upper bound on the number of miscounted examples in $M_{s,A}$, we now bound $|M_s|$.

[1] Argument here passes over a minor potential difficulty. i.e. if v' is large, Corollary 23 does not come into play — but the crucial fact is the nevertheless true as we show in Observation 32.

Corollary 25. *Let f be a well behaved monotone k-DNF, and let $s \subset X$ where $c + 2 \leq |s| \leq \beta(n) + 1$ then $|M_s| < 2^n \cdot \frac{4k \log^5(n) n^{2/3}}{n^{c+1}}$.*

Proof: This follows from $|M_s| = \sum_{A \in \mathbf{A_s}} |M_{s,A}| < |\mathbf{A_s}| \left(2^n \cdot \frac{k \log^5(n)}{n^{c+1}}\right)$. We note that $|\mathbf{A_s}|$ is bounded by the number of subsets of the subsets of s, i.e. $2^{2^{|s|}} \leq 2^{2^{\beta(n)+1}} = 2^{2^{\log\log(\sqrt[3]{n})+1}} \leq 4n^{2/3}$.

Thus $|M_s| < 2^n \cdot \frac{4k \log^5(n) n^{2/3}}{n^{c+1}}$. □

Knowing $|M_s|$, we now compute the difference between I_s and $|\Upsilon_s|$. This bound is computed by multiplying $|M_s|$ and a bound of how large the misclassification is for an example.

Theorem 26. *Let f be a well behaved monotone k-DNF formula, and $s \subset X$ where $c + 2 \leq |s| \leq \beta(n) + 1$ then $|I_s - |\Upsilon_s|| < 2^n \cdot \frac{4k \log^6(n) n^{2/3}}{n^{c+1}}$.*

Proof: As noted earlier, $I_s - |\Upsilon_s| = \sum_{e \in M_s} \mathcal{I}_e$.

Using Corollary 25, we know $|M_s| < 2^n \cdot \frac{4k \log^5(n) n^{2/3}}{n^{c+1}}$. From Observation 16, we know that that for all e, $|\mathcal{I}_e| \leq \log(\sqrt[3]{n})$.

Consequently,

$$|I_s - |\Upsilon_s|| \leq |M_s| \log(\sqrt[3]{n}) < 2^n \cdot \frac{4k \log^6(n) n^{2/3}}{n^{c+1}}.$$

□

4.4 Bounding $|\Upsilon_s|$

Definition 27. *Let $E_{f \backslash t} = \{e \in E \mid \exists t' \in f \backslash t, t'(e)\}$.*

Next we prove that every term has a high probability of being uniquely satisfied. Jackson and Servedio have a similar lemma, Lemma (3.6).

Lemma 28. *Let $f \in \mathcal{F}^{n,k,m}$ be a well behaved monotone k-DNF function, $t \in f$ then $|E^+_{1_t} - E_{f \backslash \{t\}}| \geq 2^n \cdot \frac{1}{8 \log^{4c}(n)} \frac{1}{n^c}$*

The proof of this lemma is found in Appendix 1 in Subsection C.

We note that if f is a monotone DNF and $e \in (E_t - E_{f \backslash \{t\}})$, then $e \in \Upsilon_t$.

Corollary 29. *Let $f \in \mathcal{F}^{n,k,m}$ be a well behaved monotone k-DNF, and $s \subset t \in f$ then $|\Upsilon_s| > 2^n \cdot \frac{1}{8 \log^{4c}(n)} \frac{1}{n^c}$.*

The following theorem is crucial; it is the key computation we use in our algorithm **Learn Random Monotone DNF**.

Theorem 30. *Let $f \in \mathcal{F}^{n,k,m}$ be well behaved, and let $c + 2 \leq |s| \leq \beta(n) + 1$:*

- *if $s \subset t \in f$ then $I_s \geq 2^n \cdot \frac{1}{8 \log^{4c}(n)} \frac{1}{n^c} - 2^n \cdot \frac{4k \log^6(n) n^{2/3}}{n^{c+1}}$,*
- *if $s \not\subset t \in f$ then $I_s \leq 2^n \cdot \frac{4k \log^6(n) n^{2/3}}{n^{c+1}}$.*

The previous theorem shows there exists a large gap that reliably determines if $s \subset t \in f$ for $c+2 \leq |s| \leq \beta(n)+1$ by computing I_s. This means that given a small set $s \subset t \in f$ and $x \in X \backslash s$ we can determine whether or not $s \cup x \subset t \in f$, and this is the key to our algorithm.

In this section, we proved we could determine if a set, s, is a subset of a term if $c + 2 \leq |s| \leq \beta(n) + 1$ by computing I_s. Unfortunately, we cannot efficiently compute I_s since we cannot compute $|E^+_{a_s}|$ in polynomial time. Instead we approximate I_s using standard sampling techniques. We estimate this value by sampling $g_s = n^{2c+3} 2^{k+|s|}$ uniformly chosen labeled examples from E. Thus we can effectively estimate I_s with high probability. Details can be seen in Appendix 2 in Subsection A. Our fairly straightforward algorithm is easily adapted to use our sampled values of I_s, and thus runs in polynomial time in n and 2^k. Details can be found in Appendix 2 in Subsection B.

5 Future Work

Extensions of the ideas presented here can also handle the non-monotone case. We are currently writing up this case and checking the proofs. We are also working on relaxing the requirement that k is fixed.

Acknowledgement

I would like to thank Stuart Kurtz for many conversations and help with the presentation, and Carsten Lund for help with polishing the paper.

References

[1] H. Aizenstein and L. Pitt. *On the Learnability of Disjunctive Normal Form Formulas.* Machine Learning, 19:183, 1995.

[2] D. Angluin. *Queries and concept learning.* Machine Learning, 2(4):319–342, 1988.

[3] Canny. *Lecture 10 CS 174*
www.cs.berkeley.edu/ jfc/cs174/lecs/lec10/lec10.pdf.

[4] J. Jackson. *An efficient membership-query algorithm for learning DNF with respect to the uniform distributon.* Journal of Computer and System Sciences, 55(3):414–440, 1997.

[5] J. Jackson, H. Lee, R. Servedio and A. Wan. *Learning random monotone DNF.* Electronic Colloquium on Computational Complexity, Report No. 129, 2007.

[6] J. Jackson and R. Servedio. *Learning Random Log-Depth Decision Trees under Uniform Distribution.* SIAM J. on Computing, 34(5), 2005.

[7] J. Jackson and R. Servedio. *On Learning Random DNF Formulas Under the Uniform Distribution.* Theory of Computing, 2(8):147–172, 2006.

[8] E. Mossel, R. O'Donnell,R. Servedio. *Learning juntas.* STOC 206–212, 2003.

[9] R. Servedio. *On learning monotone DNF under product distributions.* Information and Computation, 193(1):57–74, 2004.

[10] S. Smale. *On the average number of steps of the simplex method of linear programming.* Math. Programming 27: 241–267, 1983.

[11] Valiant, L.G. (1984). *A theory of the learnable.* Communications of the ACM, 27(11):1134–1142.

[12] Valiant, L.G. *Learning disjunctions of conjunctions.* In Proceedings of the 9th n International Joint Conference on Artificial Intelligence, Vol. 1, pages 560–566, 1985.

[13] K. Verbeurgt. *Learning DNF under the uniform distribution in quasi-polynomial time.* In Proceedings of the Third Annual Workshop on Computational Learning Theory, pp. 314326, 1990. [ACM:92571.92657]. 1.1, 2.2

APPENDIX 1

A Useful Observations

We use the following observations and simplifications of algebraic expressions in our proofs.

Observation 31. *For f a well behaved monotone k-DNF, $T \subset f$ where $|T| \geq 2\log\log(n)$ then $|\{e \in E_{1_s} \mid t(e)\forall t \in T\}| < 2^n \cdot \frac{1}{n^{\log\log(n)}}$ for sufficiently large n.*

Proof: First, we observe that if $T' \subset T$ then $\{e \in E_{1_s} \mid t(e), \forall t \in T'\} \supset \{e \in E_{1_s} \mid t(e), \forall t \in T\}$.

Therefore, using Corollary 23 we know given any $T' \subset T$ where $|T'| = 2\log\log(n)$, then $|\{e \in E_{1_s} \mid t(e), \forall t \in T'\} \leq 2^n \cdot \frac{1}{2^{2\log\log(n)k - (2\log\log(n))^2\beta(n)}} < 2^n \cdot \frac{1}{n^{\log\log(n)}}$. □

Observation 32. *For a well-behaved monotone k-DNF, given $A \subset \{0,1\}^{|s|}$ where $|\{a_s \in A \mid \#_1(a_s) > c\}| \geq 2\log\log(n)$ then $|M_{s,A}| < \frac{\beta(n)^{2\log\log(n)}}{n^{\log\log(n)}}$.*

Proof: Let $A' = \{a_s \in A \mid \#_1(a_s) > c\}$. Using Observation 31, if $|A'| \geq 2\log\log(n)$ then $|M_{s,A}| \leq |\{e \in E_{1_s} \mid \forall a_s \in A', \exists t \in T_{a_s}$ such that $t(e)\}|$
$\leq 2^n \cdot \frac{1}{n^{\log\log(n)}} \prod_{a_s \in A'} |T_{a_s}| \leq 2^n \cdot \frac{\beta(n)^{2\log\log(n)}}{n^{\log\log(n)}}$. The last inequality follows from noting that $\forall \#_1(a_s) > c, |T_{a_s}| \leq \beta(n)$ by the large z property, and from noting that the product is maximized for $|A'| = 2\log\log(n)$. □

Observation 33. *For c a constant, then $n^{\underline{c+1}} = n(n-1)\cdots(n-c) > n^{c+1} - \left(\frac{(c(c+1))}{2}\right) n^c$ and $n^{\underline{c+1}} < n^{c+1} - \left(\frac{(c(c+1))}{2}\right) n^c + \left(\frac{(c(c+1))}{2}\right)^2 n^{c-1}$.*

Observation 34. *Let $s \subset X$, $e \in E^+$, and $a_s = e_{|_s}$ if $\#_1(a_s) \leq c, \forall a_s \in L(A_{e,s})$ then $|L(A_{e,s})| < |s|^{\frac{(c+1)c}{2}}$.*

Proof: This follows from observing that $\binom{|s|}{c} \leq |s|^c$. □

Observation 35. *Let $s \subset X$ where $|s| \leq \beta(n)+1$, and $e \in E^+$ then if $\#_1(a_s) \leq c, \forall a_s \in L(A_{e,s})$ then $2^{|L(A_{e,s})|^2\beta(n)} < \sqrt[3]{n}$ for large enough n.*

Proof: Let $s \subset X$ where $|s| \leq \beta(n)+1$, and $A_{e,s}$ be such that $\#_1(a_s) \leq c, \forall a_s \in L(A_{e,s})$. Then using Observation 34 we know $|L(A_{e,s})| < |s|^{\frac{(c+1)c}{2}}$.

$$\begin{aligned}
2^{|L(A_{e,s})|^2\beta(n)} &\leq 2^{\left((\beta(n)+1)^{\frac{(c+1)c}{2}}\right)^2\beta(n)} \\
&= 2^{(\beta(n)+1)^{(c+1)c}\beta(n)} \\
&< 2^{(\log\log\sqrt[3]{n}+1)^{(c+1)c}\log\log\sqrt[3]{n}} \\
&< (\log\sqrt[3]{n})^{(\log\log\sqrt[3]{n}+1)^{(c+1)c}} \\
&< \sqrt[3]{n}.
\end{aligned}$$

Lemma 36. *For a positive integer $a \geq \beta(n)$ and $c \leq \log(n)/(3\log\log(n))$ then*

$$\binom{m}{a}\left(\frac{k^{\underline{c+1}}}{n^{\underline{c+1}}}\right)^a > \binom{m}{a+1}\left(\frac{k^{\underline{c+1}}}{n^{\underline{c+1}}}\right)^{a+1}.$$

Proof: The proof follows from expanding the formulas:

$$\begin{aligned}
\binom{m}{a}\left(\frac{k^{\underline{c+1}}}{n^{\underline{c+1}}}\right)^a &> \binom{m}{a+1}\left(\frac{k^{\underline{c+1}}}{n^{\underline{c+1}}}\right)^{a+1} &\Leftrightarrow \\
\frac{m^{\underline{a}}}{a!}\left(\frac{1}{\left(n^{\underline{c+1}}\right)^a}\right) &> \frac{m^{\underline{a+1}}}{(a+1)!}\left(\frac{k^{\underline{c+1}}}{\left(n^{\underline{c+1}}\right)^{a+1}}\right) &\Leftrightarrow \\
1 &> \frac{m-a}{a+1}\left(\frac{k^{\underline{c+1}}}{n^{\underline{c+1}}}\right).
\end{aligned}$$

□

Observation 37. *For positive integer k, $\left(1-\frac{1}{2^k}\right)^{2^k} \geq \frac{1}{4}$*

Proof: The proof follows from expanding the formula and noting that

$$\binom{2^k}{2i}\left(-\frac{1}{2^k}\right)^{2i} + \binom{2^k}{2i+1}\left(-\frac{1}{2^k}\right)^{2i+1} \geq 0,$$

and

$$\binom{2^k}{2}\left(-\frac{1}{2^k}\right)^2 + \binom{2^k}{3}\left(-\frac{1}{2^k}\right)^3 \geq \frac{1}{4}.$$

Thus

$$\begin{aligned}
\left(1-\frac{1}{2^k}\right)^{2^k} &= \sum_{i=0\ldots 2^k}\binom{2^k}{i}\left(-\frac{1}{2^k}\right)^i \\
&\geq \frac{1}{4}.
\end{aligned}$$

□

Observation 38. *For constant c and sufficiently large n, $\log(n)\binom{m}{\beta(n)}\left(\frac{k^{\underline{c+1}}}{n^{\underline{c+1}}}\right)^{\beta(n)} + m\binom{m}{\log(n)}\left(\frac{k^{\underline{c+1}}}{n^{\underline{c+1}}}\right)^{\log(n)} < \frac{1}{n^{\beta(n)-1}}$.*

Proof: The proof follows from the following calculations:

$$\begin{aligned}
&\log(n)\binom{m}{\beta(n)}\left(\frac{k^{\underline{c+1}}}{n^{\underline{c+1}}}\right)^{\beta(n)} \\
&+m\binom{m}{\log(n)}\left(\frac{k^{\underline{c+1}}}{n^{\underline{c+1}}}\right)^{\log(n)} \\
<\ & \frac{\log(n)(2c\log\log(n))^{\beta(n)} n^{c\beta(n)} k^{(c+1)\beta(n)}}{(n^{c+1} - \frac{(c+1)(c+2)}{2}n^c)^{\beta(n)}} \\
&+\frac{(2c\log\log(n))^{\log(n)+1} n^{c(\log(n)+1)} k^{(c+1)\log(n)}}{\left(n^{c+1} - \frac{(c+1)(c+2)}{2}n^c\right)^{\log(n)}} \\
\leq\ & \frac{\log(n)(2c\log\log(n))^{\beta(n)} k^{(c+1)\beta(n)}}{\left(n - \frac{(c+1)(c+2)}{2}\right)^{\beta(n)}} \\
&+\frac{(2c\log\log(n))^{\log(n)+1} n^c k^{(c+1)\log(n)}}{\left(n - \frac{(c+1)(c+2)}{2}\right)^{\log(n)}} \\
<\ & \frac{1}{n^{\beta(n)-1}}.
\end{aligned}$$

The first inequality follows from Observation 33 and substitution using $m \leq 2c \log\log(n) n^c$ and $\binom{m}{i} \leq m^i$. The second inequality can be seen by multiplying the first summand by $\frac{1/n^{c\beta(n)}}{1/n^{c\beta(n)}}$ and the second summand by $\frac{1/n^{c\log(n)}}{1/n^{c\log(n)}}$. The last inequality can be seen by: $\left(n - \frac{(c+1)(c+2)}{2}\right)^{\beta(n)} > n^{\beta(n)} - \frac{\beta(n)(c+1)(c+2)}{2} n^{\beta(n)-1}$ and $\left(n - \frac{(c+1)(c+2)}{2}\right)^{\log(n)} > n^{\log(n)} - \frac{\log(n)(c+1)(c+2)}{2} n^{\log(n)-1}$ and $\frac{\log(n)(2c\log\log(n))^{\beta(n)} k^{(c+1)\beta(n)}}{\left(1 - \frac{\beta(n)(c+1)(c+2)}{2n}\right)} + \frac{(2c\log\log(n))^{\log(n)+1} n^c k^{(c+1)\log(n)}}{\left(n^{\log(n)-\beta(n)} - \frac{\beta(n)(c+1)(c+2) n^{\log(n)-\beta(n)-1}}{2}\right)} < n.$ □

B Proving Random Monotone Functions are Well Behaved with High Probability

In this sections we prove that most monotone DNF in $\mathcal{F}^{n,k,m}$ are well behaved.

For the small z property of being well behaved, bounding $|T_{a_s}|$ for $\#_1(a_s) \leq c$, we first find the expected value of $|T_{a_s}|$. Next we use Chernoff bounds to give an upper bound on how far $|T_{a_s}|$ is away from the expected value with high probability.

Observation 39. *For $s \subset X$, a_s where $z = \#_1(a_s)$, and $|s| \leq k \leq \log^2(n)$ then*

$$m\frac{k^{\underline{z}}}{2n^{\underline{z}}} < \mathbf{E}_{f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}} \{|T_{a_s}|\} < m\frac{k^{\underline{z}}}{n^{\underline{z}}}.$$

Proof: We first observe that

$$\mathbf{E}_{f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}} \{|T_{a_s}|\} = m\frac{\binom{n-|s|}{k-z}}{\binom{n}{k}} = m\frac{k^{\underline{z}}(n-k)^{\underline{|s|-z}}}{n^{\underline{|s|}}}.$$

The upper bound follows by observing that

$$\frac{(n-k)^{\underline{|s|-z}}}{n^{\underline{|s|}}} = \frac{(n-k)^{\underline{|s|-z}}}{n^{\underline{z}}(n-z)^{\underline{|s|-z}}} \leq \frac{1}{n^{\underline{z}}}$$

(remember $z \leq k$,) and thus

$$\mathbf{E}_{f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}} \{|T_{a_s}|\} \leq m\frac{k^{\underline{z}}}{n^{\underline{z}}}.$$

The lower bound follows from

$$\begin{array}{rcl}
\frac{(n-k)^{\underline{|s|-z}}}{n^{\underline{|s|}}} & = & \frac{1}{n^{\underline{z}}} \prod_{i=0\ldots|s|-z-1} \frac{n-k-i}{n-z-i} \\
 & = & \frac{1}{n^{\underline{z}}} \prod_{i=0\ldots|s|-z-1} \left(1 - \frac{k-z}{n-z-i}\right) \\
 & \geq & \frac{1}{n^{\underline{z}}} \left(1 - \frac{k-z}{n-|s|}\right)^{|s|} \\
 & & \text{By} \left(1 - \frac{k-z}{n-|s|}\right)^{|s|} > 1 - |s|\frac{k-z}{n-|s|}. \\
 & > & 1/2\frac{1}{n^{\underline{z}}},
\end{array}$$

and thus $m\frac{k^{\underline{z}}}{2n^{\underline{z}}} < m\frac{k^{\underline{z}}(n-k)^{\underline{|s|-z}}}{n^{\underline{|s|}}}$. □

Next, we state the simplified Chernoff upper bound from Canny's lecture notes [3]; we use this Chernoff bound to bound the expected value.

Lemma 40 (Chernoff)**.** *Let $\delta < 2e - 1$, μ be the expected value, and χ be a series of independent Poisson trials, then* $\mathbf{Pr}\{\chi > (1+\delta)\mu\} < e^{(-\mu\delta^2/4)}$.

We now bound the number of terms in T_{a_s} for $\#_1(a_s) \leq c$, with high probability.

Lemma 41. *Fix $s \subset X$, a_s where $z = \#_1(a_s) \leq c$, and $k \leq \log^2(n)$, then*

$$\mathbf{Pr}_{f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}} \left\{|T_{a_s}| > 3m_{\max}\frac{k^{\underline{z}}}{n^{\underline{z}}}\right\} < e^{-c\log\log(n)k^{\underline{c}}}.$$

Proof: We note that $z \leq c < c\log(n) = k$, thus we know the bounds of Observation 39 hold. Let f' be a random extension of f to $m_{\max}$ terms.

$$\begin{array}{rl}
 & \mathbf{Pr}_{f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}} \{|T_{a_s}| \geq 3m_{\max}k^{\underline{z}}/n^{\underline{z}}\} \\
\leq & \mathbf{Pr}_{f' \in_{\mathcal{R}} \mathcal{F}^{n,k,m_{\max}}} \left\{|T'_{a_s}| \geq 3m_{\max}k^{\underline{z}}/n^{\underline{z}}\right\} \\
\leq & \mathbf{Pr}_{f' \in_{\mathcal{R}} \mathcal{F}^{n,k,m_{\max}}} \left\{|T'_{a_s}| \geq (1+\delta)\mathbf{E}\{T'_{a_s}\}\right\} \\
\leq & e^{-u'\delta^2/4},
\end{array}$$

where $\mu' = \mathbf{E}\{T'_{a_s}\}$ by Chernoff.

We can obtain an upper bound for this expression from a lower bound for its unnegated exponent. Let $\delta = 2$.

$$\begin{array}{rcl}
\mu'\delta^2/4 & = & \mu' \\
 & > & m_{\max}k^{\underline{z}}/2n^{\underline{z}} \\
 & > & 2n^c c \cdot \log\log(n)k^{\underline{z}}/2n^{\underline{z}} \\
 & > & c \cdot \log\log(n)k^{\underline{c}} \qquad \text{Since } k^{\underline{z}}/2n^{\underline{z}} \geq k^c/2n^c.
\end{array}$$

Therefore

$$\mathbf{Pr}_{f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}} \left\{|T_{a_s}| > 3m_{\max}\frac{k^{\underline{z}}}{n^{\underline{z}}}\right\} \leq e^{-c\log\log(n)k^{\underline{c}}}.$$

□

Corollary 42 (The Small z Property Holds with High Probability)**.** *Therefore for $s \subset t \in f$ where $|s| \leq \beta(n) + 1$ and $z = \#_1(a_s)$ then* $\mathbf{Pr}_{f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}}\{\exists a_s, 0 < \#_1(a_s) \leq c, |T_{a_s}| > 3m_{\max}\frac{k^{\underline{z}}}{n^{\underline{z}}}\} < n^{2c}\log(n)\left(\frac{1}{n}\right)^{\beta(n)-1}$.

Proof: We assume $c \geq 1$; if $c < 1$ then there does not exist a z since $0 < z \leq c < 1$ and z is an integer.

If $c \geq 1$ then the number of $s \subset t$ and a_s where $z = \#_1(a_s) \leq c$ is bounded by

$$\begin{array}{rl}
 & m \sum_{|s|=1,\ldots,\beta(n)+1} \binom{k}{|s|} \sum_{z=1,\ldots,c} \binom{|s|}{z} \\
< & m_{\max} \cdot \beta(n)k^{\beta(n)+1} \cdot 2^{\beta(n)} \\
< & \frac{1}{3}n^{c+1}.
\end{array}$$

(i.e. For a given term, the number of different sets of size s is $\binom{k}{|s|}$. The number of terms is m. For a given set, s, the number of ways to choose z items from the set is $\binom{|s|}{z}$.)

Thus, by Lemma 41:

$$\begin{array}{rcl}\mathbf{Pr}_{f\in_{\mathcal{R}}\mathcal{F}}\{\exists a_s, \#_1(a_s)\le c, |T_{a_s}| & > & 3m_{\max}\dfrac{k^{\underline{z}}}{n^{\underline{z}}}\} \\ & < & \dfrac{1}{3}n^{c+1}e^{-c\log\log(n)k^{\underline{c}}} \\ & < & \dfrac{1}{3}n^{2c}\log(n)\left(\dfrac{1}{n}\right)^{\beta(n)-1}\end{array}.$$

Therefore, with high probability, the small z property for $f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}$ is proved. □

Next we prove the medium z property: that for $\#_1(a_s)$ where $c < \#_1(a_s) < \beta(n)$ then $|T_{a_s}| < \beta(n)$ with high probability.

Lemma 43. *For fixed c, and sufficiently large n, let $s \subset X$, and a_s with $\#_1(a_s) > c$ then*

$$\mathbf{Pr}_{f\in_{\mathcal{R}}\mathcal{F}^{n,k,m}}\left\{|T_{a_s}| \ge \beta(n)\right\} < \left(\frac{1}{n}\right)^{\beta(n)-1}.$$

Proof: Let $z = \#_1(a_s)$.

If $z > k$ then $|T_{a_s}| = 0$, since there does not exist a term with more that k variables.

If $z \le k$ then the probability a random term $t \in T_{a_s}$ is $\frac{\binom{n-|s|}{k-z}}{\binom{n}{k}} < \frac{k^{\underline{z}}}{n^{\underline{z}}}$. Consequently,

$$\begin{array}{rl} & \mathbf{Pr}_{f\in_{\mathcal{R}}\mathcal{F}^{n,k,m}}\left\{|T_{a_s}| \ge \beta(n)\right\} \\ < & \displaystyle\sum_{j=\beta(n)\ldots m}\binom{m}{j}\left(\frac{k^{\underline{z}}}{n^{\underline{z}}}\right)^j\left(1-\frac{k^{\underline{z}}}{n^{\underline{z}}}\right)^{m-j} \\ < & \displaystyle\sum_{j=\beta(n)\ldots\log(n)-1}\binom{m}{j}\left(\frac{k^{\underline{z}}}{n^{\underline{z}}}\right)^j\left(1-\frac{k^{\underline{z}}}{n^{\underline{z}}}\right)^{m-j} \\ & \displaystyle+\sum_{j=\log(n)\ldots m}\binom{m}{j}\left(\frac{k^{\underline{z}}}{n^{\underline{z}}}\right)^j\left(1-\frac{k^{\underline{z}}}{n^{\underline{z}}}\right)^{m-j} \\ < & \log(n)\dbinom{m}{\beta(n)}\left(\dfrac{k^{\underline{c+1}}}{n^{\underline{c+1}}}\right)^{\beta(n)} \\ & +m\dbinom{m}{\log(n)}\left(\dfrac{k^{\underline{c+1}}}{n^{\underline{c+1}}}\right)^{\log(n)} \\ < & \left(\dfrac{1}{n}\right)^{\beta(n)-1}.\end{array}$$

The third inequality follows from the observation that the sum is maximized for $z = c+1$, and from Lemma 36. The fourth inequality follows from Observation 38. □

Corollary 44 (The Medium z Property Holds with High Probability)**.** *Therefore for $s \subset t \in f$ where $|s| \le \beta(n)+1$, $\mathbf{Pr}_{f\in_{\mathcal{R}}}\mathcal{F}^{n,k,m}\{\exists a_s, c < \#_1(a_s) < \beta(n), |T_{a_s}| \ge \beta(n)\} < \frac{1}{3}n^{2c}\log(n)\left(\frac{1}{n}\right)^{\beta(n)-1}$ for sufficiently large n.*

Proof: The number of $s \subset t$ and a_s where $c < \#_1(a_s) < \beta(n)$ is bounded by

$$\begin{array}{rl} & m\displaystyle\sum_{|s|=\lceil c\rceil,\ldots,\beta(n)+1}\binom{k}{|s|}\sum_{z=c+1,\ldots,\lfloor\beta(n)\rfloor}\binom{|s|}{z} \\ < & m_{\max}\cdot\beta(n)k^{\beta(n)+1}\cdot 2^{\beta(n)} \\ \le & n^{c+1} \\ < & \dfrac{1}{3}n^{2c}\log(n).\end{array}$$

(i.e. $\sum_{|s|=\lceil c\rceil,\ldots,\beta(n)+1}\binom{k}{|s|}$ is the number of ways to find a subset of $t \in f$ of size greater than c and less than or equal to $\beta(n)+1$. The sum $\sum_{z=c+1,\ldots,\lfloor\beta(n)\rfloor}\binom{|s|}{z}$ is the number of ways to choose a set of size z from $|s|$ elements.)

Therefore using Lemma 43, we know $\mathbf{Pr}_{f\in_{\mathcal{R}}\mathcal{F}^{n,k,m}}\{\exists a_s, c < \#_1(a_s) < \beta(n), |T_{a_s}| \ge \beta(n)\} \le \frac{1}{3}n^{2c}\log(n)\left(\frac{1}{n}\right)^{\beta(n)-1}$.

Consequently, the medium z property is also satisfied by a random $f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}$ with high probability. □

The large z property is that two terms in f overlap by at most $\beta(n)$; we prove this with a counting argument. Jackson and Servedio's paper [7] has a similar lemma, Lemma (3.5).

Lemma 45. *Let $s, s' \subseteq X$ be sets of $k \le \sqrt[3]{n}$ variables chosen independently at random, then the* $\mathbf{Pr}\{|s\cap s'| \ge \beta(n)\} < \left(\frac{1}{n}\right)^{\underline{\beta(n)-1}}$.

Proof

$$\begin{array}{rl} & \mathbf{Pr}\left\{|s\cap s'| \ge \beta(n)\right\} \\ = & \displaystyle\sum_{j=\beta(n)}^{k}\frac{\binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j}}{\binom{n}{k}^2} \\ = & \displaystyle\sum_{j=\beta(n)}^{k}\frac{(k^{\underline{j}})^2(n-k)^{\underline{k-j}}}{j!n^{\underline{k}}} \\ & \qquad\text{(The sum is maximized for } j=\beta(n).\text{)} \\ < & k\dfrac{(k^{\underline{\beta(n)}})^2}{\beta(n)!n^{\underline{\beta(n)}}} \\ < & \dfrac{1}{n^{\underline{\beta(n)-1}}}.\end{array}$$

□

Corollary 46 (The Large z Property Holds with High Probability)**.** *Therefore,* $\mathbf{Pr}_{f\in_{\mathcal{R}}\mathcal{F}^{n,k,m}}\{\exists t, t' \in f, |t\cap t'| \ge \beta(n)\} < \frac{1}{3}n^{2c}\log(n)\left(\frac{1}{n}\right)^{\beta(n)-1}$.

Proof: The proof follows from noting that $\mathbf{Pr}_{f\in_{\mathcal{R}}\mathcal{F}^{n,k,m}}\{\exists t, t' \in f, |t\cap t'| \ge \beta(n)\} \le \binom{m}{2}\frac{1}{n^{\underline{\beta(n)-1}}} \le \binom{m_{\max}}{2}\frac{1}{n^{\underline{\beta(n)-1}}} < \binom{m_{\max}}{2}\frac{1}{n^{\beta(n)-1}}\frac{1}{1-\frac{\beta(n)(\beta(n)-1)}{2n}}$ $< \frac{1}{3}n^{2c}\log(n)\left(\frac{1}{n}\right)^{\beta(n)-1}$.

The third inequality follows from Observation 33 and $n^{\beta(n)-1} - \frac{(\beta(n)-1)\beta(n)}{2}n^{\beta(n)-2} = n^{\beta(n)-1}\left(1-\frac{(\beta(n)-1)\beta(n)}{2n}\right)$.

Since two terms in $f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}$ share less than $\beta(n)$ variables with high probability, a random $f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}$ satisfies the large z property in being well behaved with high probability. □

Recalling Corollaries 42, 44, and 46 and the definition of "well behaved," we note that $f \in_{\mathcal{R}} \mathcal{F}^{n,k,m}$ is well behaved with high probability.

C Bounding Υ_s

Next we present the proof of Lemma 28 that for f a well behaved monotone k-DNF function, $m \leq 2^{k+1} c \log\log n$ and $t \in f$ then $|E^+_{1_t} - E^+_{f\backslash\{t\}}| \geq 2^n \cdot \frac{1}{8\log^{4c}(n)} \frac{1}{n^c}$.

Proof: Divide $f \backslash \{t\}$ into three disjoint sets,

- $T_{\text{disjoint}} = \{t' \in f \backslash \{t\} \mid t \cap t' = \emptyset\}$,
- $T_{\text{small}} = \{t' \in f \backslash \{t\} \mid 1 \leq |t' \cap t| \leq c\}$ and
- $T_{\text{not small}} = f \backslash (T_{\text{disjoint}} \cup T_{\text{small}})$.

Looking only at examples in $E^+_{1_t}$, we now calculate the probability that each of these sets is not satisfied. Remembering that f is monotone, we note that if one set is not satisfied, it increases the chance another set is not satisfied. (i.e. $\mathbf{Pr}_{e\in E}\{\neg t \mid \neg t'\} \geq \mathbf{Pr}_{e\in E}\{\neg t\}$ since if we know at least one variable is set to zero that increases the odds of another term to be set to zero if they share a variable.)

In the first case for T_{disjoint}, $\mathbf{Pr}_{e\in E_{1_t}}\{\forall t' \in T_{\text{disjoint}}, \neg t'(e)\} > (1 - \frac{1}{2^k})^m \geq (1 - \frac{1}{2^k})^{2^{k+1} c \log\log(n)} = \left((1 - \frac{1}{2^k})^{2^k}\right)^{2c\log\log(n)} \geq \frac{1}{4^{2c\log\log(n)}} = \frac{1}{\log^{4c}(n)}$, by Observation 37.

In the second case, if $t' \in T_{\text{small}}$ and $r = t \cap t'$ with a_t such that $X_{a_t} = r$ then by f being well behaved we know that $|T_{a_t}| \leq 3m_{\max}\left(\frac{k^{\underline{|r|}}}{n^{\underline{|r|}}}\right)$. Therefore

$$\begin{aligned}
&\mathbf{Pr}_{e\in E^+_{1_t}}\{\forall t' \in T_{\text{small}}, \neg t'(e)\}\\
> &\prod_{r\subset t, 1\leq |r|\leq c}\left(1 - \frac{2^{|r|}}{2^k}\right)^{3m_{\max}\left(\frac{k^{\underline{|r|}}}{n^{\underline{|r|}}}\right)}\\
\geq &\prod_{1\leq |r|\leq c}\left(1 - \frac{2^{|r|}}{2^k}\right)^{2^{k+3} c\log\log(n)\left(\frac{k^{\underline{|r|}}}{n^{\underline{|r|}}}\right)\binom{k}{|r|}}\\
\geq &\prod_{1\leq |r|\leq c}\left(1 - \frac{2^{|r|}}{2^k}\right)^{2^{(k-|r|)} 2^{|r|+3} c\log\log(n)\left(\frac{k^{\underline{|r|}}}{n^{\underline{|r|}}}\right)\binom{k}{|r|}}\\
\geq &\prod_{1\leq |r|\leq c}\left(\frac{1}{4}\right)^{2^{|r|+3} c\log\log(n)\left(\frac{k^{\underline{|r|}}}{n^{\underline{|r|}}}\right)\binom{k}{|r|}} \quad \text{By Obs. 37.}\\
\geq &\left(\frac{1}{4}\right)^{\frac{16c^2\log\log(n)k^2}{n}}.
\end{aligned}$$

The last inequality follows from noticing the product is maximized for $|r| = 1$, thus

$$\mathbf{Pr}_{e\in E^+_{1_t}}\left\{e \in_{\mathcal{R}} E^+_{1_t} \mid \forall t' \in T_{\text{small}}, \neg t'(e)\right\} > \frac{1}{4}.$$

We now bound the third case. Since f is well behaved, we know that a term in f overlaps another term by at most $\beta(n)$ variables, and the number of terms overlapping by a set $r \subset t$ in $T_{\text{not small}}$ is at most $\beta(n)$. Therefore $\mathbf{Pr}_{e\in E^+_{1_t}}\{\forall t' \in T_{\text{not small}}, \neg t'(e)\}$ $> \prod_{r\subset t, c<|r|\leq\beta(n)}\left(1 - \frac{2^{|r|}}{2^k}\right)^{\beta(n)}$ $> \left(1 - \frac{2^{\beta(n)}}{2^k}\right)^{\beta^2(n)\binom{k}{\beta(n)}} \geq \frac{1}{2}$, (since $\binom{k}{|r|} \leq \binom{k}{\beta(n)}$.)

Therefore (remembering $2^k = n^c$)

$$\begin{aligned}
&|\left\{e \in_{\mathcal{R}} E^+_{1_t} \mid \forall t' \in f \backslash \{t\}, \neg t'(e)\right\}|\\
> \quad &2^{n-k} \cdot \left(\frac{1}{4}\right)\left(\frac{1}{\log^4 c(n)}\right)\left(\frac{1}{2}\right) = 2^n \cdot \frac{1}{8\log^{4c}(n)} \frac{1}{n^c}.
\end{aligned}$$

□

APPENDIX 2

In the next two sections we present the standard arguments for the sake of completeness. In Section A we prove that we can sample to find a sufficient approximation to $I_s(E^+)$. In Section B we prove that our very straightforward algorithm runs in polynomial time and produces f.

A Sampling and Approximating I_s

In Section 4, we proved we could determine if a set, s, is a subset of a term if $c + 2 \leq |s| \leq \beta(n) + 1$ by computing I_s. Unfortunately, we cannot efficiently compute I_s since we cannot efficiently compute $|E^+_{a_s}|$. Instead, we show how to approximate I_s. We estimate this value by sampling g_s uniformly chosen labeled examples from E.

Definition 47. *For $s \subset X$, let $E_{\text{Sample}(g_s)} \subset E$ be a random sample of g_s labeled examples drawn uniformly from E.*

Definition 48. *Given $E_{\text{Sample}(g_s)} \subset E$, let $E^+_{\text{Sample}(g_s)} = E_{\text{Sample}}(g_s) \cap E^+$ be the set of positive examples in $E_{\text{Sample}(g_s)}$. Similary, let $\Upsilon_{\text{Sample}_s(g_s)} = E_{\text{Sample}(g_s)} \cap \Upsilon_s$ be the set of positive examples from $E_{\text{Sample}(g_s)}$ which satisfy only terms in T_{1_s}.*

Observation 49. *Let $s \subset X$, we note that $\mathbf{E}\left(\left|\Upsilon_{\text{Sample}_s(g_s)}\right|\right) = g_s \cdot \frac{|\Upsilon_s|}{|E|} = \frac{g_s}{2^n}|\Upsilon_s|$.*

Using sampled labeled examples, we compute the following function to approximate I_s.

Definition 50. *Let $s \subset X$, we define $I_{s,\text{Sample}(g_s)} = \sum_{e\in E^+_{\text{Sample}(g_s)}} (-1)^{\#_0(a_s(e))}$ to be our approximation of I_s, where $a_s(e)$ is $e_{|_s}$.*

Observation 51. *We note that the $\mathbf{E}\left(I_{s,\text{Sample}(g_s)}\right) = \frac{g_s}{2^n} I_s$.*

This observation follows since the expected value of $|E_{\text{Sample}(g_s)} \cap E^+_{a_s}|$ is $g_s \frac{|E^+_{a_s}|}{|E|}$ and $\mathbf{E}\left(I_{s,\text{Sample}(g_s)}\right) = \sum_{a_s} (-1)^{(\#_0(a_s))} g_s \frac{|E^+_{a_s}|}{|E|}$.

Next, we bound how different our sampled $I_{s,\text{Sample}(g_s)}$ is from the expected value, As we have not yet provided a lower tail bound, we state it next, as it is described in Canny [3].

Lemma 52 (Chernoff). *Let $\delta \in (0,1]$, μ be the expected value, and χ be a series of independent Poisson trials then* $\mathbf{Pr}\left\{\chi < (1-\delta)\mu\right\} < e^{-\mu\delta^2/2}$.

Applying the lower and upper Chernoff bounds from Lemmas 40 and 52, we prove that $I_{s,\mathrm{Sample}(g_s)}$ is within $\frac{2g_s}{n^{c+1}}$ fraction of $\mathbf{E}\left(I_{s,\mathrm{Sample}(g_s)}\right)$.

Lemma 53. *For $g_s = n^{2c+3}2^{k+|s|}$ and given access to examples drawn from a well behaved monotone k-DNF then* $\left|I_{s,\mathrm{Sample}(g_s)} - \mathbf{E}\left(I_{s,\mathrm{Sample}(g_s)}\right)\right| < g_s \cdot \frac{2}{n^{c+1}}$ *with probability* $1-4e^{-n/4}$.

Proof: To apply the Chernoff bounds, our main difficulty is our sum has both positive and negative values, we overcome this difficulty by bounding the positive and negative values separately. We define two indicator functions. Let $r_{\mathrm{even}}(e) = 1$ iff $f(e) = 1$ and $\#_0(e_{|_s})$ is even, and let $r_{\mathrm{odd}}(e) = 1$ iff $f(e) = 1$ and $\#_0(e_{|_s})$ is odd.

Let $E_{\mathrm{Sample}(g)}$ be a randomly generated set of g_s examples from E. Let $X_{\mathrm{even}} = \sum_{e \in E_{\mathrm{Sample}(g_s)}} r_{\mathrm{even}}(e)$. (Similarly for X_{odd}.)

We observe that $I_{s,\mathrm{Sample}(g_s)} = X_{\mathrm{even}} - X_{\mathrm{odd}}$.

If $\exists a_s$ such that $E^+_{a_s} \neq \emptyset$, then there is a term consistent with at least one a_s. This term satisfies the examples in E_{a_s} with probability at least $\frac{1}{2^k}$. There are $2^{|s|}$ different a_s, thus if $\#_0(a_s)$ is even, we expect at least $\frac{1}{2^{|s|}2^k}$ fraction of total examples are set to one by r_{even}. Therefore in $g_s = n^{2c+3}2^{k+|s|}$ examples, the expected value of the indicator function is either zero, or the expected value is at least n^{2c+3}. (Similarly for the case where $\#_0(a_s)$ is odd.)

Using the Chernoff bounds with $\delta = \frac{1}{n^{c+1}}$, we bound $\mathbf{E}(X_{\mathrm{even}})$, in the cases where the expected value is not zero. $\mathbf{Pr}\left\{|X_{\mathrm{even}} - (1 \pm \delta)\mathbf{E}(X_{\mathrm{even}})\right\} \leq 2e^{-\frac{n^{2c+3}}{4n^{2c+2}}} = 2e^{-n/4}$. (Similarly for $\mathbf{E}(X_{\mathrm{odd}})$.) Consequently, the indicator functions will be $\frac{1}{n^{c+1}}$ close to their respective expected value functions.

Therefore we know $|(X_{\mathrm{even}} - X_{\mathrm{odd}}) - \mathbf{E}(I_{s,\mathrm{Sample}(g_s)})| \leq \frac{1}{n^{c+1}}(\mathbf{E}(X_{\mathrm{even}}) + \mathbf{E}(X_{\mathrm{odd}})) \leq g_s \frac{2}{n^{c+1}}$. Thus $I_{s,\mathrm{Sample}(g_s)}$ differs from $\mathbf{E}\left(I_{s,\mathrm{Sample}(g_s)}\right)$ by at most $g_s \cdot \frac{2}{n^{c+1}}$ with high probability. □

Using the previous Lemma 53, Theorem 26, and Observation 49, we note that we can determine if $s \subset X$ is a subset of a term in a well behaved monotone k-DNF function by sampling labeled examples from the uniform distribution.

Lemma 54. *Let f be a well behaved monotone k-DNF formula, $s \subset X$ where $c+2 \leq |s| \leq \beta(n)+1$, and $g_s = n^{2c+3}2^{k+|s|}$;*

- *if $s \subset t \in f$ then $I_{s,\mathrm{Sample}(g_s)} > g_s \cdot \frac{1}{n^{c+\frac{1}{5}}}$ with probability* $1-4e^{-n/4}$.
- *If $s \not\subset t \in f$ then $I_{s,\mathrm{Sample}(g_s)} < g_s \cdot \frac{1}{n^{c+\frac{1}{5}}}$ with probability* $1-4e^{-n/4}$.

Proof: From Lemma 53 and Observation 51, we know

$$-\frac{2g_s}{n^{c+1}} + \frac{g_s}{2^n}I_s \leq I_{s,\mathrm{Sample}(g_s)} \leq \frac{2g_s}{n^{c+1}} + \frac{g_s}{2^n}I_s$$

with probability greater than $1-4e^{-n/4}$.

```
Algorithm Learn Random Monotone DNF
  1. $S =$ Distinguishing Subsets
  2. $f = \emptyset$
  3. For $s \in S$
     (a) $t = \emptyset$
     (b) For $x \in X$
         • If $I_{s \cup \{x\},\mathrm{Sample}(g_s)} > g_s \cdot \frac{1}{n^{c+\frac{1}{5}}}$ then add $x$
           to $t$
     (c) add $t$ to $f$
  4. Return $f$
```

Figure 3:

```
Function Distinguishing Subsets
  1. $S = \{s \subset X \mid |s| = c+2,$
     : and $I_{s,\mathrm{Sample}(g_s)} > g_s \cdot \frac{1}{n^{c+\frac{1}{5}}}\}$
  2. For $i = (c+3)$ to $\beta(n)$
     (a) $S' = \emptyset$
     (b) For $s \in S$ and $x \in X$
         • If $I_{s \cup \{x\},\mathrm{Sample}(g_s)} > g_s \cdot \frac{1}{n^{c+\frac{1}{5}}}$ then add $(s \cup$
           $\{x\})$ to $S'$
     (c) $S = S'$
  3. Return $S$
```

Figure 4:

By Theorem 30, Observation 51, and Lemma 53 we know:

- if $s \subset t \in f$ then $I_s \geq 2^n \cdot \frac{1}{8\log^{4c}(n)}\frac{1}{n^c} - 2^n \cdot \frac{4k\log^6(n)n^{2/3}}{n^{c+1}}$. Thus, $I_{s,\mathrm{Sample}(g_s)} \geq -g_s \cdot \frac{2}{n^{c+1}} + g_s \cdot I_s \geq -g_s \cdot \frac{2}{n^{c+1}} + g_s \cdot \frac{1}{8\log^{4c}(n)}\frac{1}{n^c} - g_s \cdot \frac{4k\log^6(n)n^{2/3}}{n^{c+1}} > g_s \cdot \frac{1}{n^{c+\frac{1}{5}}}$ with probability greater than $1-4e^{-n/4}$.

- If $s \not\subset t \in f$ then $I_s \leq 2^n \cdot \frac{4k\log^6(n)n^{2/3}}{n^{c+1}}$. Thus, $I_{s,\mathrm{Sample}(g_s)} \leq g_s \cdot \frac{2}{n^{c+1}} + g_s \cdot I_s \leq g_s \cdot \frac{2}{n^{c+1}} + g_s \cdot \frac{4k\log^6(n)n^{2/3}}{n^{c+1}} < g_s \cdot \frac{1}{n^{c+\frac{1}{5}}}$ with probability greater than $1-4e^{-n/4}$.

□

B Learning Random Monotone DNF by Finding Terms in Polynomial Time

Next, we restate our algorithm to use $I_{\mathrm{Sample}(g_s)}$.

Referring to our algorithm in Figures 3 and 4, the lemmas, and theorems in the previous sections, we prove our algorithm discovers the unknown well behaved monotone k-DNF from random examples drawn from the uniform distribution with high probability in polynomial time. We show this by following the steps our algorithm takes; first our algorithm finds all $(c+2)$-sized subsets of s in time $O(g_{c+2}n^{c+2})$ with probability greater than $1-4n^{c+2}e^{-n/4}$. Next, given all $(c+2)$-sized subsets of terms in f, our algorithm grows those subsets till they are of size $\beta(n)$ with probability greater than $1-4nmk^{\beta(n)}e^{-n/4}$ in time $O(nmg_{\beta(n)}k^{\beta(n)})$. Finally, given a subset of a term of size $\beta(n)$, our algorithm discovers all the variables in that term in time $mng_{\beta(n)+1}k^{\beta(n)}$ with probability at least $1 - mnk^{\beta(n)}(4e^{n/4})$.

Observation 55. *For $s \subset X$, computing $I_{s,\text{Sample}(g_s)}$ takes time $O(g_s)$.*

In step 1, our algorithm finds all the $(c+2)$-sized subsets of terms in f.

Lemma 56. *Given a well behaved $f \in \mathcal{F}^{n,k,m}$, our function* **Distinguishing Subsets** *finds $\{s \mid s \subset t \in f, |s| = c+2\}$ in time $O(g_{c+2}n^{c+2})$ with probability greater than $1 - 4n^{c+2}e^{-n/4}$ in step 1.*

Proof: Let $s \subset X$ where $|s| = c+2$. By Lemma 54, iff $s \subset t \in f$ then $I_{s,\text{Sample}(g_s)} \geq g_s \cdot \frac{1}{n^{c+\frac{1}{5}}}$ with probability greater than $1 - 4e^{-n/4}$. Function **Distinguishing Subsets** tests all subsets of size $c+2$, thus our function has correctly selected the sets which are subset of terms in f with probability greater than $1-4n^{c+2}e^{-n/4}$ in time $O(g_{c+2}n^{c+2})$. □

Having found all subsets of $t \in f$ of size $c+2$ with high probability, our algorithm builds these sets till all the subsets of terms has size $\beta(n)$.

Lemma 57. *Given a well behaved $f \in \mathcal{F}^{n,k,m}$, and $T = \{s \mid s \subset t \in f, |s| = c+2\}$, function* **Distinguishing Subsets** *in step 2 returns $\{s \mid s \subset t \in f, |s| = \beta(n)\}$ with probability greater than $1 - 4mnk^{\beta(n)}e^{-n/4}$ in time bounded by $O(nmg_{\beta(n)}k^{\beta(n)})$.*

Proof: Using the result of Lemma 54, each iteration of our loop is given a set $S = \{s \mid s \subset t \in f, |s| = i\}$ and produces $S' = \{s \mid s \subset t \in f, |s| = i+1\}$ with probability more than $1-4nm\binom{k}{i}e^{-n/4}$ for $i = c+2 \ldots \beta(n)-1$ in time bounded by $O(g_i nmk^i)$. Thus in $\beta(n)-1-(c+2)$ iterations our algorithm produces $S = \{s \mid s \subset t \in f, |s| = \beta(n)\}$ with probability greater than $1 - 4\beta(n)nmk^{\beta(n)-1}e^{-n/4}$ in time bounded by $O(nmg_{\beta(n)}k^{\beta(n)-1})$. □

Given all $\beta(n)$-sized subsets of $t \in f$, algorithm **Learn Random Monotone DNF** finds all the terms of f.

Lemma 58. *Given a well behaved $f \in \mathcal{F}^{n,k,m}$, and $S = \{s \mid s \subset t \in f, |s| = \beta(n)\}$ our algorithm,* **Learn Random Monotone DNF**, *finds f in time bounded by $O(mng_{\beta(n)+1}k^{\beta(n)})$ with probability greater than $1 - nmk^{\beta(n)}(4e^{-n/4})$ in step 3.*

Proof: Algorithm **Learn Random Monotone DNF** uses Corollary 46 and Lemma 54.

Corollary 46 states that, for a well behaved monotone k-DNF, $\forall s \in S$ where $|s| \geq \beta(n)$ then $|\{t \mid s \subset t \in f\}| \leq 1$. Thus every $s \in S$ is associated with at most one term $t \in f$.

Lemma 54 states that for a given $s \subset X$ and $x \in X$ where $|s \cup \{x\}| = \beta(n)+1$ iff $I_{s\cup\{x\},\text{Sample}(g_s)} \geq g_s \cdot \frac{1}{n^{c+\frac{1}{5}}}$ then $s \cup \{x\} \subset t \in f$ with probability at least $1 - 4e^{-n/4}$. Thus for $|s| = \beta(n)$, $\exists! t$ such that $s \subset t$, we can determine if $x \in t$ with high probability.

Combining these ideas, given $s \in S$ we can find a term in the inside loop of step 3 by testing every $x \in X$ to determine if $\{x\} \cup s \subset t \in f$, and thus find $\{x \mid I_{s\cup\{x\},\text{Sample}(g_s)} \geq g_s \cdot \frac{1}{n^{c+\frac{1}{5}}}\} = t \in f$ in time $O(g_{\beta(n)+1}n)$ with probability greater than $1 - 4ne^{-n/4}$.

Together, the outside loop in step 3 selects every $s \in S$ and the inside loop finds t where $s \subset t$. Since $\forall t \in f$, there exists $s \in S$ such that $s \subset t$, Algorithm **Learn Random Monotone DNF** produces f.

The time it takes to do this is the time is bounded by $O(g_{\beta(n)+1}nmk^{\beta(n)})$ with probability bounded by

$$1 - 4nmk^{\beta(n)}e^{-n/4}.$$

□

Theorem 59. *Given a well behaved $f \in \mathcal{F}^{n,k,m}$, Algorithm* **Learn Random Monotone DNF** *finds f in time bounded by $O(mng_{\beta(n)+1}k^{\beta(n)})$ with probability greater than $1 - 9mnk^{\beta(n)}e^{-n/4}$.*

Proof: Using Lemmas 56, 57, 58 we have proven that our algorithm finds all subsets of size $c+2$ of terms in f in Lemma 56, and having found these subsets it builds upon till our algorithm has found all subsets of terms of f of size $\beta(n)$ in Lemma 57; it then uses the uniqueness of terms of size $\beta(n)$ to find all the variables of a term in f; thus finding the entire function.

The algorithm runs in time bounded by

$$O(mng_{\beta(n)+1}k^{\beta(n)})$$

with probability greater than $1 - 9mnk^{\beta(n)}e^{-n/4}$. □