Approximate Counting, the Lovász Local Lemma and Inference in Graphical Models

Ankur Moitra*

October 17, 2016

Abstract

In this paper we introduce a new approach for approximately counting in bounded degree systems with higher-order constraints. Our main result is an algorithm to approximately count the number of solutions to a CNF formula Φ with at least k variables per clause and degree at most d when k is logarithmic in d. This closes an exponential gap between the known upper and lower bounds.

Moreover our algorithm extends straightforwardly to approximate sampling, which shows that under Lovász Local Lemma-like conditions it is not only possible to find a satisfying assignment, it is also possible to generate one approximately uniformly at random from the set of all satisfying assignments. Our approach is a significant departure from earlier techniques in approximate counting, and is based on a framework to bootstrap an oracle for computing marginal probabilities on individual variables. Finally, we give an application of our results to show that it is algorithmically possible to sample from the posterior distribution in an interesting class of graphical models.

*Massachusetts Institute of Technology. Department of Mathematics and the Computer Science and Artificial Intelligence Lab. Email: moitra@mit.edu. This work was supported in part by NSF CAREER Award CCF-1453261, NSF Large CCF-1565235, an Alfred P. Sloan Fellowship, an Edmund F. Kelley Research Award, a Google Research Award and the MIT NEC Corporation.

1 Introduction

1.1 Background

In this paper we introduce a new approach for approximately counting in bounded degree systems with higher-order constraints. For example, if we are given a CNF formula Φ with n variables and m clauses with the property that each clause contains at least k variables and each variable belongs to at most d clauses we ask:

Question 1.1. *How does k need to relate to d for there to be algorithms to estimate the number of satisfying assignments to Φ within a $(1 \pm 1/n^c)$ multiplicative factor?*

In the case of a monotone CNF formula where no variable appears negated, the problem is equivalent to the following: Suppose we are given a hypergraph on n nodes and m hyperedges with the property that each hyperedge contains at least k nodes and each node belongs to at most d hyper edges. How does k need to relate to d in order to be able to approximately compute the number of independent sets. Here an independent set is a subset of nodes for which there is no induced hyperedge. Bordewich, Dyer and Karpinski [5] gave an MCMC algorithm for approximating the number of hypergraph independent sets (equivalently, the number of satisfying assignments in a monotone CNF formula) that succeeds whenever $k \geq d+2$. Bezakova et al. [4] gave a deterministic algorithm that succeeds whenever $k \geq d \geq 200$ and proved that when $d \geq 5 \cdot 2^{k/2}$ it is NP-hard to approximate the number of hypergraph independent sets even within an exponential factor.

More broadly, there is a rich literature on approximate counting problems. In a seminal work, Weitz [26] gave an algorithm to approximately count in the hardcore model with parameter λ in graphs of degree at most d whenever

$$\lambda \leq \frac{(d-1)^{d-1}}{(d-2)^d}$$

And in another seminal work, Sly [24] showed a matching hardness result which was later improved in various respects by Sly and Sun [25] and Galanis, Štefankovič and Vigoda [10]. These results show that approximate counting is algorithmically possible if and only if there is spatial mixing. Moreover, Weitz's result can be thought of as a comparison theorem that spatial mixing holds on a bounded degree graph if and only if it holds on an infinite tree with the same degree bound. There have been a number of attempts to generalize these results to hypergraphs, many of which follow the approach of defining analogues of the self-avoiding walk trees used in Weitz's algorithm [26]. However what makes hypergraph versions of these problems more challenging is that spatial mixing fails, even on trees. And we can see that there are *exponential* gaps between the upper and lower bounds, since the algorithms above require k to be linear in d while the lower bounds only rule out $k \leq 2 \log d - O(1)$.

We can take another vantage point to study these problems. Bounded degree CNF formulae are also one of the principal objects of study in the Lovász Local Lemma [9] which is a celebrated result in combinatorics that guarantees when $k \geq \log d + O(1)$ that Φ has at least one satisfying assignment. The original proof of the Lovász Local Lemma was non-constructive and did not yield a polynomial time algorithm for finding such an assignment, even though it was guaranteed to exist. Beck [3] gave an algorithm followed by a parallel

version due to Alon [2] that can find a satisfying assignment whenever $k \geq 8 \log d + O(1)$. And in a celebrated recent result, Moser and Tardos [20] gave an algorithm matching exactly the existential result. This was followed by a number of works giving constructive proofs of various other settings and generalizations of the Lovász Local Lemma [13, 1, 15, 18]. However these works leave open the following question:

Question 1.2. *Under the conditions of the Lovász Local Lemma (i.e. when k is logarithmic in d) is it possible to approximately sample from the uniform distribution on satisfying assignments?*

Approximate counting and approximate sampling problems are well-known to be related. When the problem is self-reducible, they are in fact algorithmically equivalent [16, 22]. However in our setting the problem is not self-reducible because as we fix variables we could violate the assumption that k is at least logarithmic in d. It is natural to hope that under the conditions of the Lovász Local Lemma, that there is an algorithm for approximate sampling that matches the limits of the existential and now algorithmic results. However the hardness results of Bezákova et al. [4] imply that we need at least another factor of two, and that it is NP-hard to approximately count when $k \leq 2 \log d - O(1)$.

In fact, there is another connection between the Lovász Local Lemma and approximate counting. Scott and Sokal [25] showed that given the dependency graph of events in the local lemma, the best lower bound on the probability of an event guaranteed to exist by the Lovász Local Lemma (i.e. the fraction of satisfying assignments) is exactly the solution to some counting problem. Harvey, Srivastava and Vondrák [14] recently adapted techniques of Weitz to complex polydisks and gave an algorithm for approximately computing this lower bound. This yields a lower bound on the fraction of satisfying assignments, however the actual number could be exponentially larger.

1.2 Our Results

Our main result is an algorithm to approximately count the number of solutions when k is at least logarithmic in d. In what follows, let c, k and d be constants. We prove[1]:

Theorem 1.3 (informal)**.** *Suppose Φ is a CNF formula with at least k variables per clause and at most d clauses containing any one variable. For any $k \geq 20 \log d$ there is a deterministic polynomial time algorithm for approximating the number of satisfying assignments to Φ within a multiplicative $(1 \pm 1/n^c)$ factor. Moreover there is a randomized polynomial time algorithm to sample from a distribution that is $1/n^c$-close in total variation distance to the uniform distribution on satisfying assignments.*

This algorithm closes an exponential gap between the known upper [5, 4] and lower [4] bounds. It also shows that under Lovász Local Lemma-like conditions not only is it possible to efficiently find a satisfying assignment, it is possible to find a random one. Moreover our

[1] We have not made an attempt to optimize the constant in this theorem. It remains an interesting question to improve the constant, in hopes of approaching or even finding a sharp phase transition corresponding to where approximate counting is easy and where it is hard. For counting independent sets in hypergraphs, this will necessarily be a *different* threshold than the threshold where the Gibbs measure is unique [4].

approach is a significant departure from earlier techniques based either on path coupling [5] or adapting Weitz's approach to non-binary models and hypergraphs [11, 21, 23, 19, 4]. The results above appear in Theorem 6.3 and Theorem 6.5.

Our approach starts from a thought experiment about what we could do if we had access to a very powerful oracle that could answer questions about the marginal distributions of individual variables under the uniform distribution on satisfying assignments. We use this oracle and properties of the Lovász Local Lemma (namely, bounds it gives on the marginal distribution of individual variables) to construct a coupling between two random satisfying assignments so that both agree outside some logarithmic sized component. If we knew the distribution on what logarithmic sized component this coupling procedure produces, we could brute force and find the ratio of the number of satisfying assignments with $x = T$ to the number with $x = F$ to compute marginals at x. However the distribution of what component the coupling produces depends intimately on the powerful oracle we have assumed that we have access to.

Instead, we abstract the coupling procedure as a random root-to-leaf path in a tree that represents the state of the coupling. We show that at the leaves of this tree, there is a way to fractionally charge assignments where $x = T$ against assignments where $x = F$. Crucially, doing so requires only brute-force search on a logarithmic sized component. Finally, we show that there is a polynomial sized linear program to find a flow through the tree that produces an approximately valid way to fractionally charge assignments with $x = T$ against ones with $x = F$, and that any such solution *certifies* the correct marginal distribution. From these steps, we have thus bootstrapped an oracle for answering queries about the marginal distribution. Our main results then follow from utilizing this oracle. In settings where the problem is self-reducible [22] it is well-known how to go from knowing the marginal to approximate counting and sampling. In our setting, the problem is not self-reducible because setting variables could result in clauses becoming too small in which case k would not be large enough as a function of d. We are able to get around this by using the Lovász Local Lemma once more to find a safe ordering in which to set the variables.

1.3 Further Applications

Our algorithms have an interesting application in graphical models. Directed graphical models are a rich language for describing distributions by the conditional relationships of their variables. However very little is known algorithmically about learning them or performing basic tasks such as inference [7, 8]. In most settings, these problems are computationally hard. However we can study an interesting class of directed graphical models which we call *cause networks*. See Figure 1.

Definition 1.4. In a cause network there is a collection of hidden variables $x_1, x_2, ...x_n$ that are chosen independently to be T or F with equal probability. There is a collection of m observed variables each of which is either an OR or an AND of several variables or their negations.

Our goal is: Given a random sample $x_1, x_2, ...x_n$ from the model where we observe the truth value of each of the m clauses, to sample from the posterior distribution on the hidden variables. This generalizes graphical models such as the symptom-disease network where

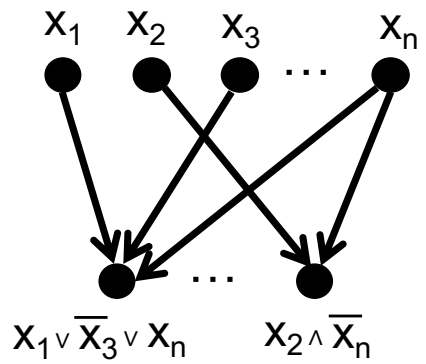

Figure 1: An example cause network with n hidden variables. A sample is generated by choosing each hidden variable to be T/F independently with equal probability, and observing the truth values of each clause.

the hidden variables represent diseases that a patient may have, and the clauses represent observed symptoms. We will require the following regularity condition on our observations:

Definition 1.5. A collection of observations is regular if for every observed variable, the corresponding clause is adjacent to (i.e. shares a variable with) at most $k/6$ OR clauses that are false and at most $k/6$ AND clauses that are true.

Now, as an immediate corollary we have:

Corollary 1.6. *Given a cause network where each observed variable depends on at least k hidden variables, each hidden variable affects at most d observed variables and $k \geq 30 \log d$, there is a polynomial time algorithm for sampling from the posterior distribution for any regular collection of observations.*

This is a rare setting where there is an algorithm to solve an inference problem in graphical models but (i) the underlying graph does not have bounded treewidth and (ii) correlation decay fails. We believe that our techniques may eventually be applicable to settings where the observed variables are noisy functions of the hidden variables and where the hidden variables are not distributed uniformly.

2 Preliminaries

In this paper, we will be interested in approximately counting the number of satisfying assignments to a CNF formula. For example, we could be given:

$$\Phi = (\overline{x}_1 \vee x_3 \vee x_5) \wedge (x_2 \vee x_3 \vee \overline{x}_8) \wedge ... \wedge (x_4 \vee \overline{x}_5 \vee x_9)$$

Let's fix some parameters. We will assume that there are n variables and there are m clauses each of which is an OR of at least k distinct variables. Finally, we will require a degree bound that each variable appears in at most d clauses. We will be interested in the relationships between k and d that allow us to approximately count the number of satisfying assignments in polynomial time.

The celebrated Lovász Local Lemma tells us conditions on k and d where we are guaranteed that there is at least one satisfying assignment:

Theorem 2.1. *[9] If $e(d+1)2^{-k} \leq 1$ then Φ has at least one satisfying assignment.*

Moser and Tardos [20] gave an algorithm to find a satisfying assignment under these same conditions . However the assignment that their randomized algorithm finds is not uniform from the set of all satisfying assignments. Our goal is to solve to be able to both approximately count and uniformly sample when k is logarithmic in d.

There are many more related results, but we will not review them all here. Instead we state a version of the asymmetric local lemma given in [12] which gives us some control on the uniform distribution on assignments. Let $\mathcal{C}$ be the collection of clauses in Φ. Let $\Pr[\cdot]$ denote the uniform distribution all assignments – i.e. uniform on $\{T, F\}^n$. Finally, for a clause b let $\Gamma(b)$ denote all the clauses that intersect b. We can abuse notation and for any event a that depends on some set of the variables, let $\Gamma(a)$ denote all the clauses that contain any of the variables on which a depends.

Theorem 2.2. *Suppose there is an assignment $x : \mathcal{C} \to (0, 1)$ such that for all $c \in \mathcal{C}$ we have*

$$Pr[c \text{ is unsatisfied}] \leq x(c) \prod_{b \in \Gamma(c)} \Big(1 - x(b)\Big)$$

then there is at least one satisfying assignment. Moreover the uniform distribution $\mathcal{D}$ on satisfying assignments satisfies that for any event a

$$Pr_{\mathcal{D}}[a] \leq Pr[a] \prod_{b \in \Gamma(a)} \Big(1 - x(b)\Big)^{-1}$$

Notice that this inequality is one-sided, as it ought to be. After all if we take b to be some clause, and a to be the event that b is not satisfied then we know that $\Pr_{\mathcal{D}}[a] = 0$ even though $\Pr[a]$ is nonzero. However what this theorem does tell us is that the marginal distribution of $\mathcal{D}$ on any variable is close to uniform. We will establish a quantitative version of this statement in the following corollary:

Corollary 2.3. *Suppose that $ed^6 2^{-k} \leq 1$. Then for every variable x_i, we have*

$$\frac{1}{2} - \frac{2}{d^5} \leq Pr_{\mathcal{D}}[x_i = T] \leq \frac{1}{2} + \frac{2}{d^5}$$

Proof. Set $x(c) = \frac{1}{d^6}$ for each clause c, and consider the event a that $x_i = T$. Now invoking Theorem 2.2 we calculate:

$$\begin{aligned} \Pr_{\mathcal{D}}[x_i = T] &\leq \Pr[x_i = T] \prod_{b \in \Gamma(a)} \Big(1 - x(b)\Big)^{-1} \\ &\leq \Big(\frac{1}{2}\Big)\Big(1 - \frac{1}{d^6}\Big)^{-1} \leq \frac{1}{2} + \frac{2}{d^5} \end{aligned}$$

where the last inequality follows because $(1 - \frac{1}{d^6})^{-d} \leq e^{\frac{2}{d^5}} \leq 1 + \frac{4}{d^5}$. An identical calculation works for the event $x_i = F$. All that remains is to check that the condition in Theorem 2.2 holds, which is a standard calculation: If c is a clause then

$$\Pr[c \text{ is unsatisfied}] \leq \Big(\frac{1}{d^6}\Big)\Big(1 - \frac{1}{d^6}\Big)^{d}$$

The left hand side is at most 2^{-k} because each clause has at least k distinct variables, and the right hand side is at least $(\frac{1}{d^6})(\frac{1}{e})$. Rearranging completes the proof. □

Notice that k is still only logarithmic in d but with a larger constant, and by increasing this constant we get some useful facts about the marginals of the uniform distribution on satisfying assignments.

3 A Coupling Procedure

3.1 Marked Variables

Now we are almost ready to define a coupling procedure. The basic strategy that we will employ is to start from either $x = T$ and $x = F$, and then sample from the corresponding marginal distribution on satisfying assignments. If we sample a variable y next, then Corollary 2.3 tells us that regardless of whether $x = T$ or $x = F$, each clause has at least $k - 1$ variables remaining and so the marginal distribution on y is still close to uniform.

Thus we will try to couple the conditional distributions, when starting from $x = T$ or $x = F$ as well as we can, to show that the marginal distribution on variables that are all at least some distance Δ away must converge in total variation distance. There is, however, an important catch that motivates the need for a fix. Imagine that we continue in this fashion, sampling variables from the appropriate conditional distribution. We can reach a situation where a clause c has all of its variables except y set and yet the clause is still unsatisfied. The marginal distribution on y is no longer close to uniform. Hence, reaching small clauses is problematic because then we cannot say much about the marginal distribution on the remaining variables and it would be difficult to construct a good coupling.

Instead, our strategy is to use the Lovász Local Lemma once more, but to decide on a set of variables in advance which we call *marked*.

Lemma 3.1. *Set $c_0 = e^{(\frac{1}{2})(\frac{1}{6})^2}$. Suppose that $2e(d+1)c_0^{-k} \leq 1$. Then there is an assignment*

$$\mathcal{M} : \{x_i\}_{i=1}^n \to \{\textit{marked}, \textit{unmarked}\}$$

such that for every clause c, it has at least $\frac{k}{3}$ marked and at least $\frac{k}{3}$ unmarked variables.

Proof. We will choose each variable to be marked or unmarked with equal probability, and independently. Consider the m bad events, one for each clause c, that c does not have enough marked or enough unmarked variables. Then we have

$$\Pr[c \text{ is bad}] \leq 2e^{-(\frac{1}{2})(\frac{1}{6})^2 k} = 2c_0^{-k}$$

which follows from the Chernoff bound. Now we can appeal to the Lovász Local Lemma to get the desired conclusion. □

Only the variables that are marked will be allowed to be set to either T or F by the coupling procedure. The above lemma guarantees that every clause c always has enough remaining variables that can make it true that the marginal distribution on any marked variable always is close to uniform.

3.2 Factorizing Formulas

Now fix a variable x. We will build up two partial assignments, and will use the notation

$$\mathcal{A}_1(x) = T \text{ and } \mathcal{A}_2(x) = F$$

to indicate that the first partial assignment sets x to T, and the second one sets x to F. Furthermore we will refer to the conditional distribution that is uniform on all satisfying assignments consistent with the decisions made so far in $\mathcal{A}_1$ and $\mathcal{D}_1$. Similarly we will refer to the other conditional distribution as $\mathcal{D}_2$. Note that these distributions are updated as more variables are set.

We can now state our goal. Suppose we have partial assignments $\mathcal{A}_1$ and $\mathcal{A}_2$. Then we will want to write

$$\Phi_{\mathcal{A}_1} = \Phi_{I_1} \wedge \Phi_{O_1}$$

where $\Phi_{\mathcal{A}_1}$ is the subformula we get after making the assignments in $\mathcal{A}_1$ and simplifying – i.e. removing variables that are F, and deleting clauses that already have a variable set to T. Similarly we will want to write

$$\Phi_{\mathcal{A}_2} = \Phi_{I_2} \wedge \Phi_{O_2}$$

Finally, we want the following conditions to be met:

(1) $\Phi_{O_1} = \Phi_{O_2} (:= \Phi_O)$

(2) Φ_{I_1} and Φ_O share no variables, and similarly for Φ_{I_2} and Φ_O

The crucial point is that if we can find partial assignments $\mathcal{A}_1$ and $\mathcal{A}_2$ where $\Phi_{\mathcal{A}_1}$ and $\Phi_{\mathcal{A}_2}$ meet the above conditions, then the conditional distribution on all variables in Φ_O is exactly the same. We will use the notation

$$\mathcal{D}_1\Big|_{\text{vars}(\Phi_O)}$$

to denote the conditional distribution of $\mathcal{D}_1$ projected onto just the variables in Φ_O. Then we have:

Lemma 3.2. *If the above factorization conditions are met, then*

$$\mathcal{D}_1\Big|_{vars(\Phi_O)} = \mathcal{D}_2\Big|_{vars(\Phi_O)}$$

Proof. From the assumption that $\Phi_{\mathcal{A}_1} = \Phi_{I_1} \wedge \Phi_O$ and because Φ_{I_1} and Φ_O share no variables, it means that there are no clauses that contain variables from both the subformulas Φ_{I_1} and Φ_O. Any such clause would prevent us from writing the formula $\Phi_{\mathcal{A}_1}$ in such a factorized form. Thus the distribution $\mathcal{D}_1$ is simply the cross product of the uniform distributions on satisfying assignments to Φ_{I_1} and Φ_O. An identical statement holds for $\mathcal{D}_1$ which completes the proof. □

Note that meeting the factorization conditions does *not* mean that the *number* of satisfying assignments to $\Phi_{\mathcal{A}_1}$ and $\Phi_{\mathcal{A}_2}$ are the same.

3.3 Factorization via Coupling

Our goal in this subsection is to give a coupling procedure to generate partial assignments $\mathcal{A}_1$ and $\mathcal{A}_2$ starting from $x = T$ and $x = F$ respectively, that result in a factorized formula. In fact, we will set exactly the same set S of variables in both, although not all variables will be set to the same value in the two partial assignments and this set S will also be random.

There are two important constraints that we will impose on how we construct the partial assignments, that will make it somewhat more tricky. First, suppose we have only set the variable x and next we choose to set the variable y in both $\mathcal{A}_1$ and $\mathcal{A}_2$. We will want that the distribution on how we set y in the coupling procedure in $\mathcal{A}_1$ to match the conditional distribution $\mathcal{D}_1$ and similarly for $\mathcal{A}_2$. Now suppose we terminate with some set S having been set. We can continue sampling the variables in $\bar{S}$ from $\mathcal{D}_1$, and we are now guaranteed that the full assignment we generate is uniform from the set of assignments with $x = T$. An identical statement holds when starting with $x = F$. Second, we will want that with very high probability, the coupling procedure terminates with not too many variables in the formula Φ_{I_1} or Φ_{I_2}. Finally, we will assume that we are given access to a powerful oracle:

Definition 3.3. We will call the following a *conditional distribution* oracle: Given a CNF formula Φ, a partial assignment $\mathcal{A}$ and a variable y it can answer with the probability that $y = T$ in a uniformly random satisfying assignment that is also consistent with $\mathcal{A}$

Such an oracle is obviously very powerful, and it is well known that if we had access to it we could compute the number of satisfying assignments to Φ exactly with a polynomial number of queries. However one should think of the coupling procedure as a though experiment, which will be useful in an indirect way to build up towards our algorithm for approximate counting.

Notice that a clause c can only trigger the WHILE loop at most once. If it ends up in Case # 1 then it is deleted from the formula. If it ends up in Case # 2 then all its variables are included in V_I and once a variable is included in V_I it is never removed. Thus the procedure clearly terminates. Our first step is to show that when it does, the formula factorizes. Let $\mathcal{C}_I$ be the set of remaining clauses which have all of their variables in V_I. Similarly let $\mathcal{C}_O$ be the set of remaining clauses which have all of their variables in V_O. Then set

$$\Phi'_I = \wedge_{c \in \mathcal{C}_I} c$$

and let Φ_{I_1} and Φ_{I_2} be the simplification of Φ'_I with respect to the partial assignments $\mathcal{A}_1$ and $\mathcal{A}_2$. Similarly set

$$\Phi'_O = \wedge_{c \in \mathcal{C}_O} c$$

and let Φ_{O_1} and Φ_{O_2} be the simplification of Φ'_O with respect to the partial assignments $\mathcal{A}_1$ and $\mathcal{A}_2$.

Claim 3.4. *All variables with different truth assignments in $\mathcal{A}_1$ and $\mathcal{A}_2$ are in V_I.*

Proof. A variable is set in response to it being contained in some clause c that triggers the WHILE loop. Any such variable is moved into V_I in both Case # 1 and Case # 2. □

Now we have an immediate corollary that helps us towards proving that we have found partial assignments for which Φ factorizes:

Algorithm 1 COUPLING PROCEDURE,
Input: Monotone CNF Φ, variable x and conditional distribution oracle F

1. Using Lemma 3.1, label variables as marked or unmarked
2. Initialize $\mathcal{A}_1(x) = T$ and $\mathcal{A}_2(x) = F$
3. Initialize $V_I = \{x\}$ and $V_O = \{x_i\}_{i=1}^n \setminus \{x\}$
4. While there is a clause c with variables in both V_I and V_O
5. Sequentially sample its marked variables (if any) from $\mathcal{D}_1$ and $\mathcal{D}_2$, using F to construct best coupling
6. Case # 1: c is satisfied by variables already set in both $\mathcal{A}_1$ and $\mathcal{A}_2$
7. Let S be the variables in c that have different truth values in $\mathcal{A}_1$ and $\mathcal{A}_2$.
8. Update $V_I \leftarrow V_I \cup S$, $V_O \leftarrow V_O \setminus S$
9. Delete c
10. Case # 2: c is not satisfied by variables already set in either $\mathcal{A}_1$ or $\mathcal{A}_2$
11. Let S be all variables in c (marked or unmarked)
12. Update $V_I \leftarrow V_I \cup S$, $V_O \leftarrow V_O \setminus S$
13. End

Corollary 3.5. $\Phi_{O_1} = \Phi_{O_2}$

Proof. Recall that Φ_{O_1} and Φ_{O_2} come from simplifying Φ'_O (which contains only variables in V_O) according to $\mathcal{A}_1$ and $\mathcal{A}_2$. From Claim 3.4, we know that $\mathcal{A}_1$ and $\mathcal{A}_2$ are the same restricted to V_O and thus we get the same formula in both cases. □

Now that we know they are equal, we can define $\Phi_O = \Phi_{O_1} = \Phi_{O_2}$. What remains is to show that the subformulas we have are actually factorizations of the original formula Φ:

Lemma 3.6. $\Phi_{\mathcal{A}_1} = \Phi_{I_1} \wedge \Phi_O$ *and* $\Phi_{\mathcal{A}_2} = \Phi_{I_2} \wedge \Phi_O$

Proof. When the WHILE loop terminates, every clause c in the original formula Φ either has all of its variables in V_I or in V_O, or was deleted because it already contains at least one variable in both $\mathcal{A}_1$ and $\mathcal{A}_2$ that satisfies it (although it need not be the same variable). Hence every clause in Φ that is not already satisfied in both $\mathcal{A}_1$ and $\mathcal{A}_2$ shows up in $\Phi'_I \wedge \Phi'_O$. Some clauses that are already satisfied in both may show up as well. In any case, this completes the proof because the remaining operation just simplifies the formulas according to the partial assignments. □

3.4 How Quickly Does the Coupling Procedure Terminate?

What remains is to bound the probability that the number of variables included in V_I is at most t. First we need an elementary definition:

Definition 3.7. When a variable x_i is given different truth assignments in $\mathcal{A}_1$ and $\mathcal{A}_2$, we call it a *type* 1 error. When a clause c has all of its marked variables set in both $\mathcal{A}_1$ and $\mathcal{A}_2$, but in at least one of them is not yet satisfied, we call it a *type* 2 error.

Note that it is possible for a variable to participate in both a type 1 and type 2 error. In any case, these are the *only* reasons that a variable is included in V_I in an execution of the coupling procedure:

Observation 1. *All variables in V_I are included either due to a type* 1 *error or a type* 2 *error, or both.*

Now our approach to showing that V_I contains not too many variables with high probability is to show that if it did, there would be a large collection of disjoint errors. First we construct a useful graph underlying the process:

Definition 3.8. Let G be the graph on vertices V_I where we connect variables if and only if they appear in the same clause together (*any* clause from the original formula Φ).

The crucial property is that it is connected:

Observation 2. *G is connected*

Proof. This property holds by induction. Assume that at the start of the WHILE loop, the property holds. Then at the end of the loop, any variable x_i added to V_I must have been contained in a clause c that at the outset had one of its variables in V_I. This completes the proof. □

Now by Observation 1, for every variable in V_I we can blame it on either a type 1 or a type 2 error. Both of these types of errors are unlikely. But for each variable, charging it to an error is problematic because of overlaps in the events. In particular, suppose we have two variables x_i and x_j that are both included in V_I. It could be that both variables are in the same clause c which resulted in a type 2 error, in which case we could only charge one of the variables to it. This turns out not to be a major issue.

The more challenging type of overlap is when two clauses c and c' both experience type 2 errors and overlap. In isolation, each clause would be unlikely to experience a type 2 error. But it could be that c and c' share all but one of their marked variables, in which case once we know that c experiences a type 2 error, then c' has a reasonable chance of experiencing one as well. We will get around this issue by building what we call a 3-tree. This approach is inspired by Noga Alon's parallel algorithmic local lemma [2] where he uses a $2,3$-tree.

Definition 3.9. We call a graph T on subset of V_I a 3-tree if each vertex is distance at least 3 from all the others, and when we add edges between vertices at distance exactly 3 the tree is connected.

Next we show that G contains a large 3-tree:

Lemma 3.10. *Any maximal* 3*-tree contains at least $\frac{|V_I|}{(d+1)dk}$ vertices.*

Proof. Consider a maximal 3-tree T. We claim that every vertex $x_i \in V_I$ must be distance at most 2 from some x_j in T. If not, then we could take the shortest path from x_i to T and move along it, and at some point we would encounter a vertex that is also not in T whose distance from T is exactly 3, at which point we could add it, contradicting T's maximality. Now for every x_i in T, we remove from consideration at most $(d+1)dk$ other variables (all those at distance at most 2 from x_i in G). This completes the proof. □

Now we can indeed charge every variable in T to a disjoint error:

Claim 3.11. *If two variables x_i and x_j in T are the result of type 2 errors for c and c', then*

$$vars(c_i) \cap vars(c_j) = \emptyset$$

Proof. For the sake of contradiction, suppose that $\text{vars}(c_i) \cap \text{vars}(c_j) \neq \emptyset$. Then since c and c' experience type 2 errors, all of their variables are included in V_I. This gives a length 2 path from x_i to x_j in G, which if they were both included in T, would contradict the assumption that T is a 3-tree. □

We are now ready to prove the main theorem of this section:

Theorem 3.12. *Suppose that $d \geq k \geq 20 \log d$. Then*

$$Pr[|V_I| \geq (d+1)dkt] \leq \left(\frac{3}{d}\right)^t$$

Proof. Suppose that $|V_I| \geq (d+1)dkt$. Then by Lemma 3.10 we can find a 3-tree T with at least t vertices. The probability of any particular 3-tree on t vertices can be bounded by:

$$\left(\frac{2}{d^5} + d\left(\frac{3}{7}\right)^{k/3}\right)^t$$

This is because by Observation 1 each vertex is caused by either a type 1 or type 2 error (or both). Moreover by Claim 3.11 the clauses that cause the type 2 errors for each vertex in T are disjoint. Now the first term in the expression above is the probability of a type 1 error, which follows from Corollary 2.3. The second term follows from the fact that each variable is contained in at most d clauses each of which could cause a type 2 error, from Lemma 3.1 which implies that each clause has at least $k/3$ marked variables, and again from Corollary 2.3 which implies that for any variable x_i, the minimum of the probability it is set to T or to T in *either* $\mathcal{A}_1$ or $\mathcal{A}_2$ is conservatively at least $3/7$ (because we chose the best coupling).

Now it is well-known (see [17, 2]) that the number of trees of size t in a graph of degree at most D is at most $(eD)^t$. Moreover if we connect pairs of vertices in G that are distance exactly 3 from each other, then we get a new graph H whose maximum degree is at most $D = d^3k$. Thus putting it all together we have that the probability that $|V_I| > (d+1)dkt$ can be bounded by

$$\left(\frac{2k}{d^2} + d^4k\left(\frac{3}{7}\right)^{k/3}\right)^t \leq \left(\frac{3}{d}\right)^t$$

where the last inequality follows from the constraint $d \geq k \geq 20 \log d$. □

Thus we can conclude that with high probability, the number of variables in V_I is at most logarithmic. We can now brute-force search over all assignments to count the number of satisfying assignments to either Φ_{I_1} or Φ_{I_2}. The trouble is that we do not have access to the marginal probabilities, so we cannot actually execute the coupling procedure. We will need to circumvent this issue next.

4 Implications of the Coupling Procedure

In this section, we give an abstraction that allows us to think about the coupling procedure as a randomly chosen root-to-leaf path in a certain tree whose nodes represent states. First, we make an elementary observation that will be useful in discussing how this tree is constructed. Recall that the coupling procedure chooses *any* clause that contains variables in both V_I and V_O and then samples *all* marked variables in it. We will assume without loss of generality that the choices it makes are done in lexicographic order. So if the clauses in Φ are ordered arbitrarily as $c_1, c_2, ..., c_m$ and the variables are ordered as $x_1, x_2, ..., x_n$ when executing the WHILE loop, if it has a choice of more than one clause it chooses among them the clause c_i with the lowest subscript i. Similarly, given a choice of which marked variable to sample next, it chooses among them the x_j with the lowest subscript j.

The important point is that now we can think of a state associated with the coupling procedure, which we will denote by σ.

Definition 4.1. The state σ of the coupling procedure specifies the following:

1. The set of remaining clauses $\mathcal{C}'$ – i.e. that have not yet been deleted
2. The partition of the variables into V_I and V_O
3. The set S of variables whose values have been set, along with their values in both $\mathcal{A}_1$ and $\mathcal{A}_2$
4. The current clause c^* being operated on in the while loop, if any

We will assume that the set $\mathcal{M}$ of marked variables is fixed once and for all. Now the transition rules are that if c^* has any marked variables that are unset, it chooses the lexicographically first and sets it. And when c^* has no remaining marked variables to set, it updates $\mathcal{C}'$, V_I and V_O according to whether it falls into Case # 1 or Case #2 and sets the current clause to empty. Finally, if the current clause is empty then it chooses the lexicographically first clause from $\mathcal{C}'$ which has at least one variable in each of V_I and V_O to be c^*.

Finally, we can define the next variable operation:

Definition 4.2. Let $\mathcal{R} : \Sigma \to \{x_i\}_{i=1}^n \cup \{\emptyset\} \times \Sigma$ be the function that takes in a state σ, transitions to the next state σ' that sets some variable y and outputs (y, σ').

Note that some states σ do not immediately set a variable – e.g. if the next operation is to choose the next clause, or update $\mathcal{C}'$, V_I and V_O. These latter transitions are deterministic, so we let σ' be the end resulting state and y be the variable that it sets. Now we can define the stochastic decision tree underlying the coupling procedure:

Algorithm 2 Decision Tree Sampling,
Input: Monotone CNF Φ, stochastic decision tree S

1. Choose a random root-to-leaf path in S
2. Choose a uniformly random assignment A_1 consistent with $\mathcal{A}_1$
3. Choose a uniformly random assignment A_2 consistent with $\mathcal{A}_2$
4. Output A_1 with probability $q = \Pr_{\mathcal{D}}[x = T]$, and otherwise output A_2

Definition 4.3. Given a conditional distribution oracle F, the function $\mathcal{R}$ and a stopping threshold s, the associated *stochastic decision tree* is the following:

(1) The root node corresponds to the state where only x is set, $\mathcal{A}_1(x) = T$, $\mathcal{A}_2(x) = F$, $V_I = \{x\}$ and $V_O = \{x_i\}_{i=1}^n \setminus \{x\}$.

(2) Each node has either zero or four descendants. If the current node corresponds to state σ, let $(y, \sigma') = \mathcal{R}(\sigma)$. Then if $y = \emptyset$ or if $|V_I| = s$ there are no descendants and the current node is a leaf corresponding to the termination of the coupling procedure or $|V_I|$ being to large. Otherwise the four descendants correspond to the four choices for how to set y in $\mathcal{A}_1$ and $\mathcal{A}_2$, and are marked with the state σ'' which incorporates their respective choices into σ'.

(3) Moreover the probability on an edge from a state σ' to a state σ'' where y has been set as $\mathcal{A}_1(y) = T$ and $\mathcal{A}_2(y) = T$ is equal to

$$\min(\mathcal{D}_1(y), \mathcal{D}_2(y))$$

and the transition to the state where $\mathcal{A}_1(y) = F$ and $\mathcal{A}_2(y) = F$ has probability

$$\min(1 - \mathcal{D}_1(y), 1 - \mathcal{D}_2(y))$$

Finally if $\mathcal{D}_1(y) > \mathcal{D}_2(y)$ then the transition to $\mathcal{A}_1(y) = T$ and $\mathcal{A}_2(y) = F$ is non-zero and is assigned all the remaining probability. Otherwise the transition to $\mathcal{A}_1(y) = F$ and $\mathcal{A}_2(y) = T$ is non-zero and is assigned all the remaining probability.

Now we can use the stochastic decision tree to give an alternative procedure to sample a uniformly random satisfying assignment of Φ. We will refer to the process of starting from the root, and choosing a descendant with the corresponding transition probability, until a leaf node is reached as "choosing a random root-to-leaf path".

Claim 4.4. *The decision tree sampling procedure outputs a uniformly random satisfying assignment of* Φ.

Proof. We could alternatively think of the decision tree sampling procedure as deciding on A_1 or A_2 with probability q vs. $1 - q$ at the outset. Then if we choose A_1, and we only keep track of the choices made for $\mathcal{A}_1$, marginally these correspond to sequentially sampling the assignment of variables from $\mathcal{D}_1$. And when we reach a leaf node in S we can interpret the

remaining choices to A_1 as sampling all unset variables from $\mathcal{D}_1$. Thus the output in this case is a uniformly random satisfying assignment with $x = T$. An identical statement holds for when we choose A_2, and because we decided between them at the outset with the correct probability, this completes the proof of the claim. □

Now let σ be the state of a leaf node u and let $\mathcal{A}_1$ and $\mathcal{A}_2$ be the resulting partial assignments. Let p_1 be the product of certain probabilities along the root-to-leaf path. In particular, suppose along the path there is a transition with y being set. Let q_1 be the probability of the transition to $(\mathcal{A}_1(y), \mathcal{A}_2(y))$ – i.e. along the branch that it actually went down. And let q_2 be the probability of the transition to $(\mathcal{A}_1(y), \overline{\mathcal{A}_2(y)})$ – i.e. where y is set the same in $\mathcal{A}_1$ but is set to the opposite value as it was in $\mathcal{A}_2$. We let p_1 be the product of all $\frac{q_1}{q_1+q_2}$ over all such decision on the root-to-leaf path.

Lemma 4.5. *Let A be an assignment that agrees with $\mathcal{A}_1$. Then for the* Decision Tree Sampling *procedure*

$$Pr\Big[\textit{terminates at leaf } u \,\Big|\, \textit{outputs assignment } A\Big] = p_1$$

Proof. The idea behind this proof is to think of the random choice of which of the four descendants to transition to as being broken down into two separate random choices where we first choose $\mathcal{A}_1(y)$ and then we choose $\mathcal{A}_2(y)$. See Figure 2. Now we can make the random choices in the Decision Tree Sampling procedure in an entirely different order. Instead of choosing the transition in the first layer, then the second layer and so on, we instead make *all* of the choices in the odd layers. Moreover at each leaf, we choose which assignment consistent with $\mathcal{A}_1$ we would output. This is the first phase. Next we choose whether to output the assignment consistent with $\mathcal{A}_1$ or with $\mathcal{A}_2$. Finally, we make all the choices in the even layers which fixes the root-to-leaf path and then we choose an assignment consistent with $\mathcal{A}_2$. This is the second phase.

The key point is that once the output A is fixed, all of the choices in the first phase are determined, because every time a variable y is set it must agree with its setting in A. Moreover each leaf node must choose A for its assignment consistent with $\mathcal{A}_1$. And finally, we know that the sampling procedure must output the assignment consistent with $\mathcal{A}_1$ because A agrees with $\mathcal{A}_1$ and not $\mathcal{A}_2$ (because they differ on how they set x). Thus conditioned on outputting A the only random choices left are those in the second phase. Now the lemma follows because the probability of reaching leaf node u is exactly the probability along the path of all of the even layer choices, which is how we defined p_1. □

We can define p_2 in an analogous way to how we defined p_1 (i.e. as the product of certain probabilities along the root-to-leaf path), and the lemma above shows that p_2 is exactly the probability of all the decisions made along the root-to-leaf path conditioned on the output being A where A agrees with $\mathcal{A}_2$.

The key lemma is the following:

Lemma 4.6. *Let N_1 be the number of satisfying assignments consistent with $\mathcal{A}_1$ and let N_2 be the number of satisfying assignments consistent with $\mathcal{A}_2$. Then*

$$\frac{p_1 N_1}{p_2 N_2} = \frac{q}{1-q}$$

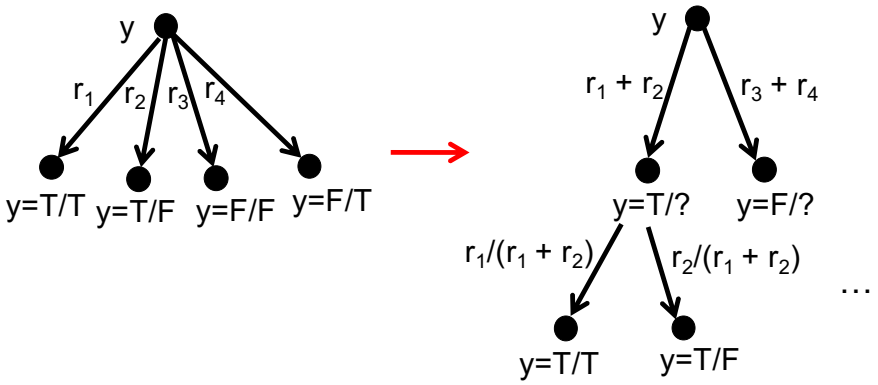

Figure 2: A transformation on the stochastic decision tree that makes it easier to understand what happens when we condition on the assignment A that is output by Decision Tree Sampling.

Proof. Let u be a leaf node. Consider a random variable Z_u that when we run the decision tree sampling procedure is non-zero if and only if we end at u. Moreover let $Z_u = (1-q)$ if an assignment with $x = T$ is output, and $Z_u = -q$ if an assignment with $x = F$ is output. Then clearly $\mathbf{E}[Z_u] = 0$. Now alternatively we can write:

$$\mathbf{E}[Z_u] = \mathop{\mathbf{E}}_{A}[\mathbf{E}[Z_u | A \text{ is output}]]$$

where A is a uniformly random satisfying assignment of Φ, precisely because of Lemma 4.4. Let N be the total number of such assignments. Then

$$\mathop{\mathbf{E}}_{A}[\mathbf{E}[Z_u|A]] = \left(\frac{N_1}{N}\right)(p_1)(1-q) + \left(\frac{N_2}{N}\right)(p_2)(-q)$$

This follows because the only assignments A that can be output at u must be consistent with either $\mathcal{A}_1$ or $\mathcal{A}_2$. Note that these are disjoint events because in one of them $x = T$ while in the other $x = F$. Then once we know that A is consistent with $\mathcal{A}_1$ (which happens with probability $\frac{N_1}{N}$) the probability for the decisions made in $\mathcal{A}_2$ being such that we reach u is exactly p_1, as this was how it was defined. The final term in the product of three terms is just the value of Z_u. An identical argument justifies the second term. Now using the fact that the above expression evaluates to zero and rearranging completes the proof. □

5 Certifying the Marginal Distribution

5.1 One-Sided Stochastic Decision Trees

The stochastic decision tree that we defined in the previous section is a natural representation of the trajectory of the coupling procedure. However it has an important drawback that we will remedy here. Its crucial property is captured in Lemma 4.6 which gives a relation between

(1) p_i – the conditional probability of an assignment consistent with $\mathcal{A}_i$ reaching u and

(2) N_i – the number of assignments consistent with $\mathcal{A}_i$

for $i = 1, 2$. However p_1 is the product of various ratios of probabilities along the root-to-leaf path. This means that if we think of the transition probabilities as variables, the constraint imposed by Lemma 4.6 is far from linear[2].

In this section, we will transform a stochastic decision tree into two separate trees, that we call *one-sided stochastic decision trees*. These will have the property that the constraint imposed by Lemma 4.6 will be linear in the unknown probabilities that we think of as variables. Ultimately we will show that any such pair can (1) certify that a given value q is within an additive inverse polynomial factor of $\Pr_{\mathcal{D}}[x = T]$ and (2) can be constructed in polynomial time through linear programming. First we explain the transformation from a stochastic decision tree to a one-sided stochastic decision tree. We will then formally define its properties and what we require of it.

Now suppose we are given a stochastic decision tree S. Let's construct the one-sided stochastic decision tree S_1 that represents the trajectory of the partial assignment $\mathcal{A}_1$. When we start from the starting state σ (see Definition 4.1), the four descendants of it in S will now be four grand-children. It's immediate descendants will be two nodes u and u', one representing the choice $\mathcal{A}_1(y) = T$ and one representing $\mathcal{A}_1(y) = F$, where y is the next variable set (see Definition 4.2). The two children of σ in S that correspond to $\mathcal{A}_1(y) = T$ will now be the children of u and the other two children will now be the children of u'. We will continue in this way so that alternate layers represent nodes present in S and new nodes.

This alone does not change much the semantics of the trajectory. All we are doing is breaking up the decision of which of the four children to proceed to, into two separate decisions. The first decision is based on just $\mathcal{A}_1$ and the second is based on $\mathcal{A}_2$. However we will change the semantics of what probabilities we associate with different transitions. For starters, we will work with total probabilities. So the total probability incoming into the starting node is 1. Let's see how this works inductively. Let's now suppose that σ represents the state of some node in S (not necessarily the starting state) and u and u' are its descendants in S_1. Then if the total probability into σ in S_1 is z, we place z along both the edges to u and to u'. This is because the decision tree is now from the perspective of $\mathcal{A}_1$, who perhaps has already chosen his assignment uniformly at random from those with $x = T$ but has not set all of those values in $\mathcal{A}_1$. Hence his decision is not a random variable, since given the option of transition to u or u' he must go to whichever one is consistent with his hidden values.

However from this perspective, the choices corresponding to $\mathcal{A}_2$ are random because he has no knowledge of the assignment that the other player is working with. If we have z total probability coming into u, then the total probability into its two descendants will be $(\frac{q_1}{q_1+q_2})z$ and $(\frac{q_2}{q_1+q_2})z$ respectively, where q_1 and q_2 were the probabilities on the transitions in S into the two corresponding descendants. In particular, if q_1 is the probability of setting $\mathcal{A}_1(y) = T$ and $\mathcal{A}_2(y) = T$ and q_2 is the probability of setting $\mathcal{A}_1(y) = T$ and $\mathcal{A}_2(y) = F$ then $(\frac{q_1}{q_1+q_2})z$ is the total probability on the transition from u to the descendant where $\mathcal{A}_2(y) = T$ and $(\frac{q_2}{q_1+q_2})z$ is the total probability on the transition from u to the descendant where $\mathcal{A}_2(y) = F$.

[2]What's worse is that the contribution of a particular decision to p_1 and p_2 is a multiplication by one of two ratios of probabilities, *which have different denominators*. For reasons that we will not digress into, this makes it challenging to encode the total probability p_1 as a flow in a tree.

Note that from Corollary 2.3 we have that

$$\Big(\frac{q_2}{q_1+q_2}\Big)z \le \Big(\frac{6}{d^5}\Big)z$$

This is an important property that we will make crucial use of later. Notice that it is a linear constraint in the total probability. Now we are ready to define a one-sided stochastic decision tree, which closely mirrors Definition 4.3.

Definition 5.1. Given the function $\mathcal{R}$ and a stopping threshold s, the associated *one-sided stochastic decision tree* for $\mathcal{A}_1$ is the following:

(1) The root node corresponds to the state where only x is set, $\mathcal{A}_1(x) = T$, $\mathcal{A}_2(x) = F$, $V_I = \{x\}$ and $V_O = \{x_i\}_{i=1}^n \setminus \{x\}$.

(2) Each node has either two descendants and four *grand*-descendants or zero descendants. If the current node a corresponds to state σ, let $(y, \sigma') = \mathcal{R}(\sigma)$. Then if $y = \emptyset$ or if $|V_I| = s$ there are no descendants and the current node is a leaf corresponding to the termination of the coupling procedure or $|V_I|$ being to large. Otherwise the two descendants correspond to the two choices for how to set y in $\mathcal{A}_1$. Each of their two descendants correspond to the two choices for how to set y in $\mathcal{A}_2$. Each grand-descendant is marked with the state σ' which incorporates their respective choices.

(3) Let z be the total probability into a. Then the total probability into each descendant is z. Moreover let the total probability into the grand-descendants with states $\mathcal{A}_1(y) = T$ and $\mathcal{A}_2(y) = T$ and $\mathcal{A}_1(y) = T$ and $\mathcal{A}_2(y) = F$ be z_1 and z_2 respectively. Then z_1 and z_2 are nonnegative, sum to z and satisfy $z_2 \le (\frac{6}{d^5})z$. Similarly, let the total probability into the grand-descendants with states $\mathcal{A}_1(y) = F$ and $\mathcal{A}_2(y) = F$ and $\mathcal{A}_1(y) = F$ and $\mathcal{A}_2(y) = T$ be z_3 and z_4 respectively. Then z_3 and z_4 are nonnegative, sum to z and satisfy $z_4 \le (\frac{6}{d^5})z$.

The one-sided stochastic decision tree for $\mathcal{A}_2$ is defined analogously, in the obvious way. Finally we record an elementary fact:

Claim 5.2. *There is a perfect matching between the root-to-leaf paths in S_1 and S_2, so that any pair of assignments A_1 and A_2 that takes a root-to-leaf path p in S_1, must also take the root-to-leaf path in S_2 to which p is matched.*

Proof. Recall that the odd levels in S_1 and S_2 correspond to the nodes in S. Therefore from a root-to-leaf path p in S_1 we can construct the root-to-leaf path in S, which in turn uniquely defines a root-to-leaf path in S_2 (because it specifies which nodes are visited in odd layers, and all paths end on a node in an odd layer). □

5.2 An Algorithm for Finding a Valid S_1 and S_2

We are now ready to prove one of the two main theorems of this section:

Theorem 5.3. *Let $q = Pr_{\mathcal{D}}[x = T]$ and $q' \leq q \leq q''$. Then there are two one-sided stochastic decision trees S_1 and S_2 that for any pair of matched root-to-leaf paths terminating in u and u' respectively satisfy*

$$\Big(\frac{q'}{1-q'}\Big)p_2N_2 \leq p_1N_1 \leq \Big(\frac{q''}{1-q''}\Big)p_2N_2$$

where N_1 and N_2 are number of satisfying assignments consistent with $\mathcal{A}_1$ and $\mathcal{A}_2$ respectively, and p_1 and p_2 are the total probability into u and u' respectively.

Moreover given q' and q'' that satisfy $q' \leq q \leq q''$ there is an algorithm to construct two one-sided stochastic decision trees S_1 and S_2 that satisfy the above condition on all matched leaf nodes corresponding to a termination of the coupling procedure, which runs in time polynomial in m and 4^s where s is the stopping size.

Proof. The first part of the theorem follows from the transformation we gave from a stochastic decision tree to two one-sided stochastic decision trees. Then Claim 5.2 combined with Lemma 4.6 implies $\frac{q}{1-q} = \frac{p_1N_1}{p_2N_2}$, which then necessarily satisfies $\frac{q'}{1-q'} \leq \frac{p_1N_1}{p_2N_2} \leq \frac{q''}{1-q''}$. Rearranging completes the proof of the first part.

To prove the second part of the theorem, notice that if s is the stopping size, then the number of leaf nodes in S_1 and in S_2 is bounded by 4^s. At each leaf node that corresponds to a termination of the coupling procedure, from Lemma 3.6 we can compute the ratio of N_1 to N_2 as the ratio of the number of satisfying assignments to Φ_{I_1} to the number of satisfying assignments to Φ_{I_2}. This can be done in polynomial in m and 2^s time by brute-force. Finally, the constraints in Definition 5.1 are all linear in the variables that represent total probability (if we treat $\frac{6}{d^5}$, $\frac{q'}{1-q'}$, $\frac{q''}{1-q''}$ and all ratios $\frac{N_1}{N_2}$ as given constants). Thus we can find a valid choice of the total probability variables by linear programming. This completes the proof of the second part. □

Recall that we will be able to choose $s = O(d^2k \log n)$ and Theorem 3.12 will imply that at most an inverse polynomial fraction of the distribution fails to couple. Thus the algorithm above runs in polynomial time for any constants d and k. What remains is to show that any valid choice of total probabilities *certifies* that $q' \leq \Pr_{\mathcal{D}}[x = T] \leq q''$.

5.3 A Fractional Matching to Certify q

We are now ready to prove the second main theorem of this section. We will show that having any two one-sided stochastic decision trees that meet the constraints on the leaves imposed by Theorem 5.3 is enough to certify that $\Pr_{\mathcal{D}}[x = T]$ is approximately between q' and q''. This result will rest on two facts. Fix any assignment A. Then either

(1) The assignment has too many clauses that restricted to marked variables are all F or

(2) The total probability of A reaching a leaf node u where the coupling procedure failed to terminate before reaching size s is at most $O(\frac{1}{n^c})$.

Theorem 5.4. *Suppose that $d \geq k \geq 20 \log d$. Then any two one-sided stochastic decision trees S_1 and S_2 that meet the constraints on the leaves imposed by Theorem 5.3 and satisfy $s = 10cd^2k \log n$ imply that*

$$q' - O\Big(\frac{1}{n^c}\Big) \leq Pr_{\mathcal{D}}[x = T] \leq q'' + O\Big(\frac{1}{n^c}\Big)$$

The proof of this theorem will use many of the same tools that appeared in the proof of Theorem 3.12, since in essence we are performing a one-sided charging argument.

Proof. The proof will proceed by constructing a complete bipartite graph $H = (U, V, E)$ and finding a fractional approximate matching as follows. The nodes in U represent the satisfying assignments of Φ with $x = T$. The nodes in V represent the satisfying assignments of Φ with $x = F$. Moreover all but a $O(\frac{1}{n^c})$ fraction of the nodes on the left will send between $1 - q'' - O(\frac{1}{n^c})$ and $1 - q' + O(\frac{1}{n^c})$ flow along their outgoing edges. Finally all but a $O(\frac{1}{n^c})$ fraction of the nodes on the right will receive between $q' - O(\frac{1}{n^c})$ and $q'' + O(\frac{1}{n^c})$ flow along their incoming edges.

First notice that any assignment A (say with $x = T$) is mapped by S_1 to a distribution over leaf nodes, some of which correspond to a coupling and some of which correspond to a failure to couple before reaching size s. Now consider matched pairs of leaf nodes (according to Claim 5.2) that correspond to a coupling. Let p_1 and p_2 be the total probability of the leaf nodes in S_1 and S_2 respectively. Let N_1 and N_2 be the total number of assignments that are consistent with $\mathcal{A}_1$ and $\mathcal{A}_2$, and let $\mathcal{N}_1$ and $\mathcal{N}_2$ be the corresponding sets of assignments. From the assumption that

$$\Big(\frac{q'}{1-q'}\Big)p_2N_2 \leq p_1N_1 \leq \Big(\frac{q''}{1-q''}\Big)p_2N_2$$

and the intermediate value theorem it follows that there is a $q' \leq q^* \leq q''$ which satisfies

$$\Big(\frac{q'^*}{1-q^*}\Big)p_2N_2 = p_1N_1$$

Hence there is a flow that sends exactly $(1 - q^*)p_1$ units of flow out of each node in $\mathcal{N}_1$ and which each node in $\mathcal{N}_2$ receives exactly q^*p_2 units of flow.

If every leaf node corresponding to a coupling, we would indeed have the fractional matching we are looking for, just by summing these flows over all leaf nodes. What remains is to handle the leaf nodes that do not correspond to the coupling terminating before size s. Consider any such leaf node u in S_1 and the corresponding leaf node v in S_2. From Lemma 3.10 we have that there is a 3-tree T of size at least $10c \log n$. For each node in T, from Claim 3.11 we have there are at least $10c \log n$ disjoint type 1 or type 2 errors.

Case # 1: Suppose that there are at least $5c \log n$ disjoint type 1 errors. Fix the 3-tree T, and look at all root-to-leaf paths that are consistent with just the type 1 errors. Then the sum of their total probabilities is at most

$$\Big(\frac{6}{d^{12}}\Big)^{5c \log n}$$

This follows because the constraint that $z_2 \leq (\frac{6}{d^{12}})z$ (and similarly for z_4) in Definition 5.1 implies that for each path we can factor out the above term corresponding to just the decisions where there are type 1 errors. The remaining probabilities are conditional distributions on the paths (after having taken into account the type 1 errors) and sum to at most one. Finally the total number of 3-trees of size $10c \log n$ is at most $(ed^3k)^{10c \log n}$. Thus for any assignment

A, if we ignore what happens to it when it ends up at a leaf node which did not couple and which has at least $5c\log n$ disjoint type 1 errors, in total we have ignored at most

$$\Big(\frac{6e}{d}\Big)^{5c\log n} \leq 1/n^c$$

of its probability.

Case # 2: Suppose that there are at least $5c\log n$ disjoint type 2 errors. Each type 2 error can be blamed on either $\mathcal{A}_1$ or $\mathcal{A}_2$ or both (e.g. it could be that the clause c might only have all of its marked variables set to F in $\mathcal{A}_1$). Let's suppose that the assignment A contributes at least $5/2c\log n$ disjoint type 2 errors. In this case we will completely ignore A in the constraints imposed by our flow. How many such assignments can there be? The probability of getting any such assignment is bounded by

$$\Big(ed^3k\Big)^{10c\log n}\Big(\Big(\frac{3}{7}\Big)^{k/3}\Big)^{5/2c\log n} \leq 1/n^c$$

Thus if we ignore the flow constraints for all such assignments, we will be ignoring at most a $1/n^c$ fraction of the nodes in U and the nodes in V. The only remaining case is when the assignment A ends up at a leaf node u that has at least $5c\log n$ disjoint type 2 errors, but it contributes less than $5/2c\log n$ itself. For each type 2 error that it does not contribute to, it contributes to another type 1 error. The only minor complication is that the node responsible might not be in the 3-tree T. However it is distance at most 1 from the 3-tree because it is contained in a clause that results in type 2 error that does contain a node in T. Now by an analogous reasoning as in Case #1 above, if we fix the pattern of these type 1 errors – i.e. we fix the 3-tree and the extra nodes at distance 1 from it that contribute the missing type 1 errors – the sum of the total probability of all consistent root-to-leaf paths is at most

$$\Big(\frac{6}{d^{12}}\Big)^{5c\log n}$$

Now the number of patterns can be bounded by $(ed^4k)^{10c\log n}$, which accounts for the inclusion of extra nodes that are not in T. Once again, for such an assignment A if we ignore what happens to it when it ends up at a leaf node which did not couple and which has at least $5c\log n$ disjoint type 2 but it contributes less than $5/2c\log n$ itself, in total we have ignored at most

$$\Big(\frac{6e}{d}\Big)^{5c\log n} \leq 1/n^c$$

of its probability.

Now returning to the beginning of the proof and letting N_1 and N_2 be the total number of satisfying assignments with $x = T$ and $x = F$ respectively. We have that the flow in the bipartite graph implies

$$(1-q'')N_1 - O\Big(\frac{1}{n^c}\Big) \leq \text{flowout}_U = \text{flowin}_V \leq q''N_2 + O\Big(\frac{1}{n^c}\Big)$$

and the further condition

$$q'N_2 - O\Big(\frac{1}{n^c}\Big) \leq \text{flowin}_V = \text{flowout}_U \leq (1-q')N_1 + O\Big(\frac{1}{n^c}\Big)$$

which gives $\frac{q'}{1-q'} - O(\frac{1}{n^c}) \leq \frac{q}{1-q} \leq \frac{q''}{1-q''} + O(\frac{1}{n^c})$ which completes the proof of the theorem. $\square$

6 Applications

Here we show how to use our algorithm for computing marginal probabilities when k is logarithmic in d for approximate counting and sampling from the uniform distribution on satisfying assignments.

6.1 Approximate Counting

First, we show how to use an algorithm for computing marginal probabilities to do approximate counting in a monotone CNF, where no variable is negated. This approach is standard, and appears in [4].

Corollary 6.1. *Suppose we are given a monotone CNF formula Φ on n variables with at least k variables per clause and at most d clauses containing any one variable with $d \geq k \geq 20 \log d$. Let OPT be the number of satisfying assignments. Then there is an algorithm that outputs a quantity count that satisfies*

$$\Big(1 - \frac{1}{n^c}\Big) OPT \leq count \leq (1 + \frac{1}{n^c}\Big) OPT$$

and runs in time polynomial in m and n^{cd^2k}.

Proof. First, we fix an ordering of the variables $x_1, x_2, ...x_n$ and a sequence of formulas $\Phi_1, \Phi_2, ...\Phi_n$. Let $\Phi_1 = \Phi$ and let Φ_i be the subformula we get when substituting $x_1 = T, x_2 = T, ...x_{i-1} = T$ into Φ and simplifying. Notice that each such formula is a monotone CNF and inherits the properties we need from Φ. In particular, each clause has at least k variables because the only clauses left in Φ_i (i.e. not already satisfied) are the ones which have all of their variables unset. Also, each variable belongs to at most d clauses because we have only removed variables and clauses.

Thus we can appeal to Theorem 5.3 and Theorem 5.4 to compute for each variable x_i the quantity $p_i \triangleq \Pr_{\mathcal{D}_i}[x_i = T]$ to within an additive $1/2n^{c+1}$ where $\mathcal{D}_i$ is the uniform distribution on satisfying assignments to Φ_i. Let our estimate be q_i. Since $p_i \geq 1/2$ we have that q_i and p_i are also multiplicatively close, with $(1 \pm 1/n^{c+1})$. Now if we take the product of the p_i's we get a telescoping product which computes the ratio of the number of satisfying assignments with all variables set to T divided by the number of satisfying assignments to Φ. Moreover each $p_i \geq 1/2$. Thus we conclude

$$\text{count} \triangleq \prod_{i=1}^{n} \Big(\frac{1}{q_i}\Big) = \Big(1 \pm \frac{1}{n^c}\Big) \text{OPT}$$

which completes the proof. □

The above approach heavily used monotonicity to ensure that no clause becomes too small (i.e. contains few variables, but is still unsatisfied). This is a similarly issue to what happened with the coupling procedure, which necessitating using *marked* and *unmarked* variables, the latter being variables that are never set and are used to make sure no clause becomes too small. We can take a similar approach here. In what follows we will no longer assume Φ is monotone.

Lemma 6.2. *Set $c_0 = \max(e^{(\frac{1}{2})(\frac{1}{6})^2}, 3/4)$. Suppose that $e(d+1)c_0^{-k} \leq 1$. Then there is a partial assignment $\mathcal{A}$ so that every clause is satisfied and each clause has at least $k/3$ unset variables. Moreover there is a randomized algorithm to find such a partial assignment that runs in time polynomial in m, n, k and d. Alternatively there is a deterministic algorithm that runs in time polynomial in m and $n^{O(d^2)}$.*

Proof. We will choose independent for each variable to set it to T with probability $1/4$, to set it to F with probability $1/4$ and to leave it unset with probability $1/2$. Now consider the m bad events, one for each clause c, that c is either unsatisfied or has not enough unset variables (or both). Then we have

$$\Pr[c \text{ is bad}] \leq e^{-(\frac{1}{2})(\frac{1}{6})^2 k} + \left(\frac{3}{4}\right)^k \leq 2c_0^{-k}$$

where the first term follows from the Chernoff bound and represents the probability that there are not enough unset variables, and the second term is the probability that the clause is unsatisfied. Once again we can appeal to the Lovász Local Lemma to show the existence. Finally we can use the algorithm of Moser and Tardos [20] to find such a partial assignment in randomized polynomial time. Moreover Moser and Tardos [20] also give a deterministic algorithm that runs in time polynomial in m and $n^{O(d^2)}$. □

Theorem 6.3. *Suppose we are given a CNF formula Φ on n variables with at least k variables per clause and at most d clauses containing any one variable with $d \geq k \geq 20 \log d$. Let OPT be the number of satisfying assignments. Then there is a deterministic algorithm that outputs a quantity count that satisfies*

$$\left(1 - \frac{1}{n^c}\right) OPT \leq count \leq (1 + \frac{1}{n^c}) OPT$$

and runs in time polynomial in m and n^{cd^2k}.

Proof. Our proof follows the same basic outline as in Corollary 6.1. First we (deterministically) find a partial assignment that meets Lemma 6.2 and let $x_1, x_2, ...x_t$ be an ordering of the set variables. We define $\Phi_1, \Phi_2, ...\Phi_t$ in the same way as the subformula we get by substituting in the assignments for $x_1, x_2, ...x_{i-1}$ and simplifying to get Φ_i. Again let q_i be our estimate for the marginal probabilities.

The key point is that Φ_{t+1} would be empty, because all clauses are satisfied. Moreover each clause that appears in any formula Φ_i for $1 \leq i \leq t$ has at least $k/3$ variables because it has at least that many unset variables in the partial assignment. Moreover we can now output

$$\text{count} \triangleq 2^{n-t} \prod_{i=1}^{n} \left(\frac{1}{q_i}\right) = \left(1 \pm \frac{1}{n^c}\right) \text{OPT}$$

because Φ_{t+1} has exactly 2^{n-t} satisfying assignments (every choice of the unset variables) and we have used the same telescoping product, but now to compute the ratio of the number of satisfying assignments to Φ_{t+1} divided by the number of satisfying assignments to Φ. □

Algorithm 3 SAMPLING PROCEDURE,
Input: CNF Φ, oracle F for approximating marginals of variables

1. Using Lemma 3.1, label variables as marked or unmarked
2. While there is a marked variable x that is unset
3. Sample x using F
4. Initialize $V_I = \{x\}$ and V_O to be all unset variables (x is already set)
5. While there is a clause c with variables in both V_I and V_O
6. Sequentially sample its marked variables (if any) using F
7. Case # 1: c is satisfied
8. Delete c
9. Case # 2: c is unsatisfied
10. Let S be all variables in c (marked or unmarked)
11. Update $V_I \leftarrow V_I \cup S$, $V_O \leftarrow V_O \setminus S$
12. End
13. End
14. For each connected component of the remaining clauses
15. Enumerate and uniformly choose a satisfying assignment of the unset variables
16. End

6.2 Approximate Sampling

Here we give an algorithm to generate an assignment approximately uniformly from the set of all satisfying assignments. Again, the complication is that our oracle for approximating the marginals works only if k is at least logarithmic in d so we need some care in the order we choose to sample variables. First we give the algorithm:

First, we prove that the output is close to uniform.

Lemma 6.4. *If the oracle F outputs a marginal probability that is $1/n^{c+1}$ close to the true marginal distribution for each variable queried, then the output of the* SAMPLING PROCEDURE *is a random assignment whose distribution is $1/n^c$-close in total variation distance to the uniform distribution on all satisfying assignments.*

Proof. The proof of this lemma is in two parts. First, imagine we were instead given access to an oracle G that answered each query for a marginal distribution with the exact value. Then each variable set using the oracle is chosen from the correct marginal distribution. And in the last step, the set of satisfying assignments is a cross-product of the satisfying assignments for each component. Thus the procedure would output a uniformly random assignment from the set of all satisfying assignments. Second, since at most n variables are queried, we have that with probability at least $1 - 1/n^c$ all of the random decision of the

procedure would be the same if we had given it answers from G instead of from F. This now completes the proof. □

The key step in the analysis of this algorithm rests on showing that with high probability each connected component is of logarithmic size.

Theorem 6.5. *Suppose we are given a CNF formula Φ on n variables with at least k variables per clause and at most d clauses containing any one variable with $d \leq k \leq 20 \log d$. There is an algorithm that outputs a random assignment whose distribution is $1/n^c$-close in total variation distance to the uniform distribution on all satisfying assignments. Moreover the algorithm runs in time polynomial in m and n^{cd^2k}.*

Proof. The proof of this theorem uses many ideas from the coupling procedure as analyzed in Section 3. Let Φ' be the formula at the start of some iteration of the inner WHILE loop. Then at the end of the inner WHILE loop, using Lemma 3.6 we can write:

$$\Phi' = \Phi'_I \wedge \Phi'_O$$

where Φ'_I is a formula on the variables in V_I and Φ'_O is a formula on the variables in V_O. In particular, no clause has variables in both because the inner WHILE loop terminated. Now we can appeal to the analysis in Theorem 3.12 which gives a with high probability bound on the size of V_I. The analysis presented in its proof is nominally for a different procedure, the Coupling Procedure, but the inner WHILE loop of the Sampling Procedure is identical except for the fact that there are no type 1 errors because we are building up just one assignment. Thus

$$\Pr[|V_I| \geq (d+1)dkt] \leq \left(\frac{3}{d}\right)^t$$

The inner WHILE loop is run at most n times and so if we choose $t \geq c \log n$ we get that with probability at least $1 - 1/n^c$ no component has size larger than $(d+1)dkc \log n$. Now the brute force search in the last step can be implemented in time polynomial in m and n^{cd^2k}, which combined with Lemma 6.4 completes the proof. □

We can also now prove Corollary 1.6.

Proof. Recall, we are given a cause network and the truth assignment of each observed variable. First we do some preprocessing. If an observed variable is an OR of several hidden variables or their negation, and the observed variable is set to F we know the assignment of each hidden variable on which it depends. Similarly, if an observed variable is an AND and it is set to T again we know the assignment of each of its variables. For all the remaining observed variables, we know there is exactly one configuration of its variables that is prohibited so each yields a clause in a CNF formula Φ. Moreover each clause depends on at least $2k/3$ variables whose truth value has not been set because the collection of observations is regular. Finally each variable is contained in at most d clauses. The posterior distribution on the remaining hidden variables (whose value has not already been set) is uniform on the set of satisfying assignments to Φ and thus we can appeal to Theorem 6.5 to complete the proof. □

Acknowledgements

We are indebted to Elchanan Mossel and David Rolnick for many helpful discussions at an earlier stage of this work.

References

[1] D. Achlioptas and F. Iliopoulos. Random walks that find perfect objects and the Lovász local lemma. In *FOCS*, page 494–503, 2014.

[2] N. Alon. A parallel algorithmic version of the local lemma. In *Random Structures and Algorithms*, 2(4):367–378, 1991.

[3] J. Beck. An algorithmic approach to the Lovász local lemma. *Random Structure and Algorithms*, 2(4):343–365, 1991.

[4] I. Bezákova, A. Galanis, L. A. Goldberg, H. Guo and D. Štefankovič. Approximation via correlation decay when strong spatial mixing fails. In *ArXiv*:1510.09193, 2016.

[5] M. Bordewich, M. Dyer and M. Karpinski. Stopping times, metrics and approximate counting. In *ICALP*, pages 108–119, 2006.

[6] G. Bresler, E. Mossel and A. Sly. Reconstruction of Markov random fields from samples: Some observations and algorithms. *SIAM Journal on Computing*, 42(2):563–578, 2013.

[7] P. Dagum and M. Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60:141–153, 1993.

[8] P. Dagum and M. Luby. An optimal approximation algorithm for Bayesian inference. *Artificial Intelligence*, 93(1-2):1–27, 1997.

[9] P. Erdös and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In *Infinite and Finite Sets*, pages 609–627, 1975.

[10] A. Galanis, D. Štefankovič and E. Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combinatorics, Probability and Computing*, 25(4):500–559, 2016.

[11] D. Gamarnik and D. Katz. Correlation decay and deterministic FPTAS for counting colorings of a graph. *Journal of Discrete Algorithms*, 12:29–47, 2012.

[12] B. Haeupler, B. Saha and A. Srinivasan. New constructive aspects of the Lovász Local Lemma. In *Journal of the ACM*, 58(6): 28, 2011.

[13] D. Harris and A. Srinivasan. A constructive algorithm for the Lovász local lemma on permutations. In *SODA*, pages 907–925, 2014.

[14] N. Harvey, P. Srivastava and J. Vondrák. Computing the independence polynomial in Shearer's region for the LLL. *ArXiv*:1608.02282, 2016.

[15] N. Harvey and J. Vondrák. An algorithmic proof of the Lovász local lemma via resampling oracles. In *FOCS*, pages 1327–1346, 2015.

[16] M. Jerrum, L. Valiant and V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

[17] D. Knuth. *The Art of Computer Programming*, Vol I. Addison Wesley, London, page 396 (exercise 11), 1969.

[18] V. Kolmogorov. Commutativity in the algorithmic Lovász local lemma. In *FOCS* 2016, to appear.

[19] J. Liu and P. Lu. FPTAS for counting monotone CNF. In *SODA*, pages 1531–1548, 2015.

[20] R. Moser and G. Tardos. A constructive proof of the Lovász Local Lemma. In *Journal of the ACM*, 57(2):1–15, 2010.

[21] C. Nair and P. Tetali. The correlation decay (CD) tree and strong spatial mixing in multi-spin systems. *ArXiv:*0701494, 2007.

[22] A. Sinclair and M. Jerrum. Approximately counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82:93–133, 1989.

[23] A. Sinclair, P. Srivastava and M. Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems. *Journal of Statistical Physics*, 155(4):666–686, 2014.

[24] A. Sly. Computational transition at the uniqueness threshold. In *FOCS*, pages 287–296, 2010.

[25] A. Sly and N. Sun. Counting in two-spin models on d-regular graphs. *Annals of Probability*, 42(6):2383–2416, 2014.

[26] D. Weitz. Counting independent sets up to the tree threshold. In *STOC*, pages 140–149, 2006.