Information and Computation 307 (2025) 105385

Contents lists available at ScienceDirect

Information and Computation

journal homepage: www.elsevier.com/locate/yinco

One-shot learning for k-SAT ☆

Andreas Galanis, Leslie Ann Goldberg, Xusheng Zhang *

Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford, Oxfordshire, OX1 3QD, United Kingdom

ARTICLE INFO

Article history:
Received 15 July 2025
Received in revised form 21 November 2025
Accepted 24 November 2025
Available online 24 November 2025

Keywords:
Computational learning theory
k-SAT
Maximum likelihood estimation

ABSTRACT

Consider a k-SAT formula Φ where every variable appears at most d times. Let σ be a satisfying assignment, sampled proportionally to $e^{\beta m(\sigma)}$ where $m(\sigma)$ is the number of true variables and β is a real parameter. Given Φ and σ, can we efficiently learn β?

This problem falls into a recent line of work about single-sample ("one-shot") learning of Markov random fields. Our k-SAT setting was recently studied by Galanis, Kalavasis, Kandiros (SODA24). They showed that single-sample learning is possible when roughly $d \leq 2^{k/6.45}$ and impossible when $d \geq (k+1)2^{k-1}$. In addition to the gap in d, their impossibility result left open the question of whether the feasibility threshold for one-shot learning is dictated by the satisfiability threshold for bounded-degree k-SAT formulas.

Our main contribution is to answer this question negatively. We show that one-shot learning for k-SAT is infeasible well below the satisfiability threshold; in fact, we obtain impossibility results for degrees d as low as k^2 when β is sufficiently large, and bootstrap this to small values of β when d scales exponentially with k, via a probabilistic construction. On the positive side, we simplify the analysis of the learning algorithm, obtaining significantly stronger bounds on d in terms of β. For the uniform case $\beta \to 0$, we show that learning is possible under the condition $d \lesssim 2^{k/2}$. This is (up to constant factors) all the way to the sampling threshold – it is known that sampling a uniformly-distributed satisfying assignment is NP-hard for $d \gtrsim 2^{k/2}$.

1. Introduction

A key task that arises in statistical inference is to estimate the underlying parameters of a distribution, frequently based on the assumption that one has access to a sufficiently large number of independent and identically distributed (i.i.d.) samples. However, in many settings it is critical to perform the estimation with substantially fewer samples, driven by constraints in data availability, computational cost, or real-time decision-making requirements. In this paper, we consider the extreme setting where only a single sample is available and investigate the feasibility of parameter estimation in this case. We refer to this setting as "one-shot learning".

Markov random fields (also known as undirected graphical models) are a canonical framework used to model high-dimensional distributions. The seminal work of Chatterjee [11] initiated the study of one-shot learning for the Ising and spin glass models, a significant class of Markov random fields that includes the well-known Sherrington-Kirkpatrick and Hopfield models. This approach was later explored in greater depth for the Ising model by Bhattacharya and Mukherjee [3] and subsequently extended to tensor or weighted variants of the Ising model in [24,34,15]. Beyond the Ising model, Daskalakis

☆ All data is provided in full in the results section of this paper.

* Corresponding author.
E-mail address: xusheng.zhang@cs.ox.ac.uk (X. Zhang).

https://doi.org/10.1016/j.ic.2025.105385

et al. [17,18] examined one-shot learning in more general settings, notably including logistic regression and higher-order spin systems, obtaining various algorithmic results in "soft-constrained" models, i.e., models where the distribution is supported on the entire state space. Bhattacharya and Ramanan [4] showed that efficient parameter estimation using one sample is still possible under the presence of hard constraints which prohibit certain states, relaxing the soft-constrained assumption with "permissiveness"; canonical Markov random fields in this class include various combinatorial models such as the hardcore model (weighted independent sets). Notably, in all these cases, one-shot learning is always feasible with mild average-degree assumptions on the underlying graph (assuming of course access to an appropriate sample).

More recently, Galanis et al. [22] investigated one-shot learning for hard-constrained models that are not permissive, focusing primarily on k-SAT and proper colourings; in contrast to soft-constrained models, they showed that one-shot learning is not always possible and investigated its feasibility under various conditions. Their results left however one important question open for k-SAT, in terms of identifying the "right" feasibility threshold. In particular, their impossibility results were based on the existence of unsatisfiable instances for k-SAT, suggesting that it might be the satisfiability threshold that is most relevant for one-shot learning. Here we refute this in a strong way. We show infeasibility well below the satisfiability threshold, and obtain positive results that align closely with the conjectured threshold for sampling satisfying assignments.

1.1. Definitions and main results

In the k-SAT model, we consider the state space $\Omega_n := \{\mathsf{TRUE}, \mathsf{FALSE}\}^n$, where each element is an assignment to n Boolean variables. The support of the Markov random field is then restricted to the set of assignments that satisfy a given k-CNF formula. More precisely, we define $\Phi_{n,k,d}$ as the set of CNF formulas with n variables such that each clause has exactly k distinct variables and each variable appears in at most d clauses. For an assignment $\sigma \in \Omega_n$ and a formula $\Psi \in \Phi_{n,k,d}$, we denote by $\sigma \models \Psi$ the event that σ satisfies Ψ and we denote by $m(\sigma)$ the number of variables that are assigned to TRUE in σ. (See Section 1.4 for further details.)

We study the weighted k-SAT model parametrized by β. For a fixed formula $\Psi \in \Phi_{n,k,d}$, the probability for each assignment $\sigma \in \Omega_n$ is given by

$$\Pr_{\Psi,\beta}[\sigma] = \frac{e^{\beta\, m(\sigma)}\, \mathbb{1}[\sigma \models \Psi]}{\sum_{\sigma \in \Omega_n} e^{\beta\, m(\sigma)}\, \mathbb{1}[\sigma \models \Psi]}. \tag{1}$$

Let $\Omega_n(\Psi) := \{\sigma \in \Omega_n : \sigma \models \Psi\}$ be the support of $\Pr_{\Psi,\beta}$. When $\beta = 0$, this distribution reduces to the uniform distribution over all satisfying assignments $\Omega_n(\Psi)$. For general $\beta \neq 0$, it biases the distribution toward assignments with more TRUE if $\beta > 0$ and biases toward those with more FALSE if $\beta < 0$.

We consider the following one-shot learning task for β. The learner knows parameters d, k and a fixed formula $\Psi \in \Phi_{n,k,d}$. Additionally, the learner has access to a single sample $\sigma \in \Omega_n(\Psi)$ drawn from distribution $\Pr_{\Psi,\beta}[\cdot]$. The learner also knows that β lies within a specified range $|\beta| \leq B$, but it does not know the exact value of β. The goal is to estimate β using these inputs.

To quantify the accuracy of our estimate, we say that $\hat{\beta}$ is an ϵ-estimate if $|\beta - \hat{\beta}| \leq \epsilon$. Typically we want ϵ to decrease as n increases so that $\epsilon \to 0$ when $n \to \infty$. In this case we call $\hat{\beta}$ a consistent estimator. On the other hand, if there exists a constant $\epsilon_0 > 0$ such that $\limsup_n |\hat{\beta} - \beta| \geq \epsilon_0$, then $\hat{\beta}$ is not a consistent estimator and we say β is not identifiable by $\hat{\beta}$. Finally, if β is not identifiable by any $\hat{\beta}$, we say it is impossible to estimate β.

Our main algorithmic result is a linear-time one-shot learning algorithm for β in the weighted k-SAT model.

Theorem 1.1. *Let $B > 0$ be a real number. Let $d, k \geq 3$ be integers such that*

$$d \leq \frac{1}{e^3\sqrt{k}} \cdot (1 + e^{-B})^{\frac{k}{2}}. \tag{2}$$

There is an estimation algorithm which, for any β^* with $|\beta^*| \leq B$, given any input $\Phi \in \Phi_{n,k,d}$ and a sample from $\sigma \sim \Pr_{\Phi,\beta^*}$, outputs in $O(n + \log(nB))$ time an $O(n^{-1/2})$-estimate $\hat{\beta}(\sigma)$ such that

$$\Pr_{\Phi,\beta^*}\left[\left|\hat{\beta}(\sigma) - \beta^*\right| = O(n^{-1/2})\right] = 1 - e^{-\Omega(n)}.$$

Our results improve upon the conditions in [22], which ensure a consistent estimate under the requirement when $d \lesssim (1 + e^{-B})^{k/6.45}$. Based on the corresponding threshold for approximate sampling, the conjectured "true" threshold for d is of the order $(1 + e^{-B})^{\frac{k}{2}}$. Consequently, our improved condition in (2) is only off by a polynomial factor in k relative to this conjectured threshold.

For comparison with the approximate sampling threshold—commonly stated for the uniform k-SAT distribution—we specialize to $B \to 0$. In that limit, our algorithmic result for single-sample learning holds roughly when $d \lesssim 2^{k/2}$. The best currently known result for efficient sampling, due to Wang and Yin [37], holds under the condition $d \lesssim 2^{k/4.82}$, see also the series of works [32,19,29,28]. It is conjectured that the sharp condition for efficient sampling is $d \lesssim 2^{k/2}$, supported by

matching hardness results for monotone formulas. It is known in particular that for $d \gtrsim 2^{k/2}$, no efficient sampling algorithm exists (unless $\mathsf{NP} = \mathsf{RP}$).

To complement our algorithmic result, we also present impossibility results, suggesting that conditions like (2) are nearly sharp.

Theorem 1.2. *Let β^* be a real number such that $|\beta^*| > 1$. Let $k \geq 4$ be an even integer, and let n be a multiple of $k/2$ that is large enough. If*

$$d \geq k^3 \left(1 + \frac{e}{e^{|\beta^*|} - e}\right)^{\frac{k}{2}}, \tag{3}$$

then there exists a formula $\Phi \in \Phi_{n,k,d}$ such that it is impossible to estimate β^* from a sample $\sigma \sim \Pr_{\Phi,\beta^*}$ with high probability.

For the parameter d around the satisfiability threshold, specifically at the uniquely satisfiable threshold $u(k)$ (see, e.g., [31] on the connection between these two thresholds), if

$$d \geq u(k) = \Theta\Big(\frac{2^k}{k}\Big), \tag{4}$$

there exists a formula $\Phi \in \Phi_{n,k,d}$ such that it is impossible to estimate β^* from any number of samples $\sigma \sim \Pr_{\Phi,\beta^*}$ because $\Omega_n(\Phi)$ is a deterministic set consisting of a single satisfying assignment that does not depend on β^*. Galanis et al. [22] explicitly construct such a formula Φ, though it requires an additional $O(k^2)$ factor relative to (4), representing the previous best known condition for the impossibility of estimation. Condition (3) in Theorem 1.2 not only relaxes (4) when $|\beta^*|$ grows large, but it also features the correct $k/2$ exponent, matching that in both (2) and the conjectured threshold. Indeed, when $B \approx |\beta^*| \to \infty$, conditions (2) and (3) both take the form

$$(1 + O(e^{-|\beta^*|}))^{k/2} \cdot k^{O(1)} \tag{5}$$

These findings partially indicate that, at least for the k-SAT model, the sampling threshold is more relevant to one-shot learning than the satisfiability threshold.

In addition, we find that if we allow β^* to be proportional to k, then learning becomes impossible for a significantly larger range of d. Specifically, unlike condition (3), which requires d to be exponential in k, here we only need d to be quadratic in k, leading to a much sparser formula when k is large.

Theorem 1.3. *Let $k \geq 4$ be an even integer. For all $\beta^* \in \mathbb{R}$ such that $|\beta^*| \geq k \ln 2$ the following holds. Let n be a multiple of $k/2$ that is large enough. If $d \geq k^2/2$, then there exists a formula $\Phi \in \Phi_{n,k,d}$ such that it is impossible to estimate β^* from a sample $\sigma \sim \Pr_{\Phi,\beta^*}$ with high probability.*

Remark. In the regimes where Theorem 1.2 or Theorem 1.3 applies, the corresponding formula Φ ensures that there is a single assignment that is the output with all but exponentially-small probability, regardless of the value of β^*. Hence, the proof of this theorem guarantees that it is impossible to learn from exponentially many independent samples. Moreover, for any pair of (β_1, β_2) such that $|\beta_1| > |\beta_2| \geq |\beta^*|$ and $\beta_1\beta_2 \geq 0$, no hypothesis testing for $H_0 : \beta = \beta_1$ versus $H_1 : \beta = \beta_2$ can be done to distinguish $\Pr_{\Phi,\beta_1}$ from $\Pr_{\Phi,\beta_2}$, that is, there exists no sequence of consistent test functions $\phi_n : \Omega_n \to \{0, 1\}$ such that

$$\lim_{n\to\infty} \mathbb{E}_{\sigma\sim\Pr_{\Phi,\beta_1}} \phi_n(\sigma) = 0 \text{ and } \lim_{n\to\infty} \mathbb{E}_{\sigma\sim\Pr_{\Phi,\beta_2}} \phi_n(\sigma) = 1.$$

1.2. Proof overview

We prove Theorem 1.1 by using the maximum pseudo-likelihood estimator. Establishing the consistency of this estimator—as stated in Theorem 2.1—requires demonstrating that the log-pseudo-likelihood function is (strongly) concave; see (11) for the precise formulation. In the k-SAT setting, showing such concavity amounts to showing that with high probability over samples drawn from $\Pr_{\Phi,\beta^*}[\cdot]$, each sample contains a linear number of "flippable variables." We prove this property in Lemma 2.2 by applying the Lovász local lemma (LLL), which enables us to compare $\Pr_{\Phi,\beta^*}[\cdot]$ to a suitable product distribution—under which the number of flippable variables is guaranteed to be linear. Notably, we apply the LLL directly to a non-local set of variables, in contrast to previous analyses that confined the application of the LLL to local neighbourhoods—a restriction that typically imposes stronger constraints on the parameter regime. By circumventing these stronger constraints, our approach achieves its guarantee under the nearly optimal LLL condition.

The main technical novelty of this paper is our negative results. To explain these, we begin by outlining the proof of Theorem 1.3. For simplicity, we focus on the case where $\beta^* \geq 0$. At a high level, we construct a gadget formula Ψ_0 for which the all-true assignment σ^+ carries almost all of the probability mass under $\Pr_{\Psi_0,\beta^*}[\cdot]$ provided that $\beta^* \geq k \ln 2$.

Consequently, a sample $\sigma \sim \Pr_{\Psi_0,\beta^*}[\cdot]$ drawn from this distribution is nearly deterministic, offering virtually no information about β^* and thus rendering learning impossible. The key property of this gadget is established in Lemma 3.1. Specifically, the lemma ensures that σ^+ satisfies Ψ_0 and that any other assignment σ with fewer than $2n/k$ variables set to FALSE is not a satisfying assignment of Ψ_0. We achieve this by incorporating a cyclic structure over the variables that enforces global correlation among the FALSE values in the assignments. In particular, there are no flippable variables in σ^+.

Towards the proof of Theorem 1.2, we first leverage the gadget Ψ_0 to show the existence of a stronger gadget Ψ_2, parametrized by $b > 1$ in Lemma 3.4, which guarantees that the all-true assignment σ^+ satisfies Ψ_2 and any other assignment with fewer than n/b variables set to FALSE fails to satisfy Ψ_2. Then we choose b appropriately in terms of β^* to make sure that σ^+ carries nearly all of the probability mass, using some more technical estimates for the corresponding partition function. To build Ψ_2, we take a finite number of replicas of Ψ_0 on randomly permuted sets of variables. The existence of the desired formula is established using the probabilistic method; specifically, we demonstrate an upper bound on the expectation of $m(\sigma)$ for any satisfying assignment $\sigma \neq \sigma^+$, over the choice of the permutations.

1.3. Related work

Parameter Estimation in Markov Random Fields. A large body of work has focused on parameter estimation under the one-shot learning paradigm (see, e.g., [11,3,17,18,24,15,34]), particularly for Ising-like models in statistical physics and for dependent regression models in statistics. In this work, we follow a similar approach by establishing the consistency of the maximum pseudo-likelihood estimator. Earlier studies (e.g., [25,14,13,23]) have also explored parameter estimation in Markov random fields using the maximum likelihood estimator.

Before our work, the papers Bhattacharya and Ramanan [4], Galanis et al. [22] were the first to study one-shot learning in hard-constrained models. In particular, the hardcore model analysed in [4] can be viewed as a weighted monotone 2-SAT model, and one natural extension of the hardcore model to k-uniform hypergraphs corresponds to the class of weighted monotone k-SAT models—a special case of the weighted k-SAT models that we consider. Because a typical assignment in these monotone formulas possesses $\Omega(n)$ flippable variables, the pseudo-likelihood estimator remains consistent across all parameter regimes, and no phase transition is expected. The weighted k-SAT problem was analysed in [22], where the authors derived both a consistency condition and an impossibility condition, though a substantial gap remained between them. By tightening the bounds on both ends, our work considerably narrows this gap, nearly closing it entirely.

Related Works in Structural Learning/Testing. An alternative direction in learning Markov random fields involves estimating the interaction matrix between variables—a question originally posed by Chow and Liu [12]. For the Ising model, this problem has been extensively studied (see, e.g., [6,36,9] and the references therein), and subsequent work has extended the results to higher-order models [30,27,20]. Recent work [38,15,21] has also considered the joint learning of both structure and parameters. Moreover, Santhanam and Wainwright [35] establishes the information-theoretic limits on what any learner can achieve, and similar analyses have been conducted for hardcore models [8,7]. While some approaches in this line of work require multiple independent samples, as noted in [15], it is also possible to reduce learning with $O(1)$ samples to a class of special cases within one-shot learning. Related problems in one-shot testing for Markov random fields have also been studied in [10,16,33,2,5].

1.4. k-SAT notation

We use standard notation for the k-SAT problem. A formula $\Phi = (V, C)$ denotes a *CNF (Conjunctive Normal Form) formula* with variables $V = \{x_1, \ldots, x_n\}$ and clauses C. We use $\sigma(x_i)$ and σ_i to denote the truth value of the variable x_i under an assignment $\sigma : V \to \{\mathsf{TRUE}, \mathsf{FALSE}\}$. For any clause $c \in C$, $var(c)$ denotes the set of variables appearing in c (negated or not). The *degree* of a variable x_j in Φ is the number of clauses in which x_j or $\neg x_j$ appears, namely $|\{c \in C : x_j \in var(c)\}|$. The degree of Φ is the maximum, over $x_j \in V$, of the degree of x_j in Φ. As noted in the introduction, $m(\sigma)$ denotes the number of variables that are assigned to TRUE by an assignment σ. Namely, $m(\sigma) := |\{i \in [n] : \sigma_i = \mathsf{TRUE}\}|$.

2. Maximum pseudo-likelihood estimator: the proof of Theorem 1.1

Section 2.1 introduces the fundamentals of the maximum pseudo-likelihood estimator and analyses its running time for solving the weighted k-SAT problem. In Sections 2.2 and 2.3, we establish the estimator's consistency.

2.1. Overview of maximum (pseudo)-likelihood estimation

For a k-SAT formula Ψ and a satisfying assignment σ, we will use $f(\beta; \sigma)$ to denote the quantity $\Pr_{\Psi,\beta}(\sigma)$ from (1), i.e., $f(\beta; \sigma) = e^{\beta m(\sigma)}/Z(\beta; \Psi)$, where $Z(\beta; \Psi)$ is the normalising constant of the distribution (the partition function).

A standard approach to parameter estimation is to find $\hat{\beta}_{\mathrm{MLE}}(\sigma) := \arg\max_\beta f(\beta; \sigma)$, which is commonly referred to as the maximum likelihood estimate (MLE). However, two main obstacles arise when applying MLE directly to the weighted k-SAT problem. First, (approximately) computing $Z(\beta; \Psi)$ is generally intractable because it is an NP-hard computation. Second, even if an approximation algorithm exists for computing $\hat{\beta}_{\mathrm{MLE}}(\sigma)$, there is no guarantee of its consistency, i.e., there

is no guarantee that with high probability it is close to β^*. Hence, we take a computationally more tractable variant of MLE from [1] which is called the maximum pseudo-likelihood estimation. Let $f_i(\beta;\sigma)$ be the conditional probability of σ_i, in a distribution with parameter β, conditioned on the value of $\sigma_{-i} := (\sigma_j)_{j\neq i}$. The maximum pseudo-likelihood estimate (MPLE) is defined as

$$\hat{\beta}_{\mathrm{MPLE}}(\sigma) := \arg\max_{\beta} \prod_{i\in V} f_i(\beta;\sigma) = \arg\max_{\beta} \sum_{i\in V} \ln f_i(\beta;\sigma).$$

Here, the objective function $F(\beta;\sigma) := \sum_{i\in V} \ln f_i(\beta;\sigma)$ is the so-called *log-pseudo-likelihood function*. For the weighted k-SAT problem, it is not hard to compute $f_i(\beta;\sigma)$ as

$$f_i(\beta;\sigma) = \frac{e^{\beta\sigma_i}}{e^{\beta}\mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{TRUE})] + \mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{FALSE})]},$$

where $\mathbb{1}[\omega]$ is shorthand for $\mathbb{1}[\omega\models\Psi]$. So we can write the log-pseudo-likelihood function for k-SAT as

$$\begin{aligned} F(\beta;\sigma) &= \sum_{i\in V} \ln\left(\frac{e^{\beta\sigma_i}}{e^{\beta}\mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{TRUE})] + \mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{FALSE})]}\right), \\ &= \beta m(\sigma) - \sum_{i\in V} \ln\left(e^{\beta}\mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{TRUE})] + \mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{FALSE})]\right). \end{aligned}$$

For a fixed σ, $F(\cdot;\sigma):\mathbb{R}\to\mathbb{R}$ is a function of β. By taking derivative with respect to β, we obtain

$$\frac{\partial F(\beta;\sigma)}{\partial\beta} = m(\sigma) - \sum_{i\in V} \frac{e^{\beta}\mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{TRUE})]}{e^{\beta}\mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{TRUE})] + \mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{FALSE})]}. \tag{6}$$

Clearly, $\frac{\partial F(\beta;\sigma)}{\partial\beta}$ is a decreasing function of β, which implies that $F(\beta;\sigma)$ has a unique global maximum that is achieved when $\frac{\partial F(\beta;\sigma)}{\partial\beta} = 0$. Therefore, $\hat{\beta}_{\mathrm{MPLE}}(\sigma)$ can be uniquely defined to be the maximum of $F(\beta;\sigma)$.

Moreover, provided $|\hat{\beta}_{\mathrm{MPLE}}(\sigma)| \le 2B$, an ϵ-close estimate of $\hat{\beta}_{\mathrm{MPLE}}(\sigma)$ can be computed, using $O(\ln(B/\epsilon))$ steps of binary search for the solution to $\frac{\partial F(\beta;\sigma)}{\partial\beta} = 0$. At each step of the binary search, we evaluate $\frac{\partial F(\beta;\sigma)}{\partial\beta}$ and adjust the binary search interval based on its sign. A naive evaluation of $\frac{\partial F(\beta;\sigma)}{\partial\beta}$ as in (6) would require $\Theta(n)$ operations per step. We can reduce this by exploiting the fact that the summand

$$S_i(\beta) := \frac{e^{\beta}\mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{TRUE})]}{e^{\beta}\mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{TRUE})] + \mathbb{1}[\sigma_{-i}\wedge(\sigma_i\leftarrow\mathsf{FALSE})]}$$

can take only three values $\{0, 1, e^{\beta}/(1+e^{\beta})\}$. Hence, by grouping the summands according to their values, we obtain

$$m(\sigma) - \sum_{i\in V} S_i(\beta) = m(\sigma) - |\{i\in V : S_i(\beta) = 1\}| - \frac{e^{\beta}}{1+e^{\beta}} \cdot \left|\left\{i\in V : S_i(\beta) = \frac{e^{\beta}}{1+e^{\beta}}\right\}\right|. \tag{7}$$

Crucially, the sets $\{i\in V : S_i(\beta) = 1\}$ and $\{i\in V : S_i(\beta) = \frac{e^{\beta}}{1+e^{\beta}}\}$ do not depend on β, and thus can be computed only once before the binary search. After this preprocessing, each evaluation of $\frac{\partial F(\beta;\sigma)}{\partial\beta}$ can be done in $O(1)$ time using the decomposition in (7). Overall, to achieve an $O(n^{-1/2})$-close estimate, the total running time is $O(n+\log(nB))$.

2.2. Consistency of MPLE

In Section 2.1 we gave an algorithm with running time $O(n+\log(nB))$, which takes as input a formula Ψ and an assignment $\sigma\in\Omega_n(\Psi)$ and outputs an estimate $\hat{\beta}(\sigma)$ which satisfies

$$\left|\hat{\beta}(\sigma) - \hat{\beta}_{\mathrm{MPLE}}(\sigma)\right| = O(n^{-1/2}),$$

provided $|\hat{\beta}_{\mathrm{MPLE}}(\sigma)| \le 2B$. Theorem 1.1 follows immediately from the Theorem 2.1, which demonstrates $O(\sqrt{n})$-consistency.

Theorem 2.1. *Let β^* be a real number, and let $d, k \ge 3$ be integers such that*

$$d \le \frac{1}{e^3\sqrt{k}} \cdot (1+e^{-|\beta^*|})^{\frac{k}{2}}. \tag{8}$$

For any integer n and $\Phi\in\Phi_{n,k,d}$,

$$\Pr_{\Phi,\beta^*}\left[\left|\hat{\beta}_{\mathrm{MPLE}}(\sigma) - \beta^*\right| = O(n^{-1/2})\right] = 1 - e^{-\Omega(n)}. \tag{9}$$

As in the standard literature [11], the consistency result can be proved using bounds on the derivatives of the log-pseudo-likelihood function F. In particular, following the analysis in [22], (9) is a consequence of these two bounds:

1. A uniform linear upper bound of the second moment of the first derivative of F: for all $\beta \in \mathbb{R}$,

$$\mathbb{E}_{\Phi,\beta}\left[\left(\frac{\partial F(\beta;\sigma)}{\partial \beta}\right)^2\right] \leq kdn. \tag{10}$$

2. With high probability over $\sigma \sim \Pr_{\Phi,\beta^*}$, there is a uniform lower bound of the second derivative of F:

$$\Pr_{\Phi,\beta^*}\left[\inf_{\beta\in\mathbb{R}} \frac{\partial^2 F(\beta;\sigma)}{\partial \beta^2} = \Omega(n)\right] = 1 - e^{-\Omega(n)}. \tag{11}$$

In brief, (10) and (11) control the first and second-derivative of the log–pseudo-likelihood function, for most σ. A second-order Taylor approximation around β^*, combined with these bounds, yields (9). We refer interested readers to the proof of Theorem 1.1 in [22], where the complete argument is carried out in detail.

Moreover, Lemma 3.1 in [22] proves the bound (10) for all $\beta \in \mathbb{R}$ and all $d, k \geq 3$, and they give a combinatorial expression (equations (3.9) and (3.10) in [22]) for $\frac{\partial^2 F(\beta;\sigma)}{\partial \beta^2}$ that will be the useful for proving (11). To state this expression, we introduce the notion of flippable variables. We say a variable v_i is *flippable* in σ if the assignment, obtained by flipping the value of variable v_i while keeping the values of other variables in σ, is still a satisfying assignment, that is, $(\sigma_{-i} \wedge (\neg\sigma_i)) \models \Psi$. We use $e_{v_i}(\sigma)$ to denote the indicator of the event that variable v_i is flippable in a satisfying assignment σ. By differentiating (6) using the expression in (7), we obtain the following expression for the second derivative of F (shown in [22]):

$$\frac{\partial^2 F(\beta;\sigma)}{\partial \beta^2} = \frac{e^\beta}{(e^\beta+1)^2} \sum_{v\in V} e_v(\sigma). \tag{12}$$

Hence, proving (11) reduces to establishing a linear lower bound on the number of flippable variables. The main ingredient in the proof of our positive result is Lemma 2.2, which provides such a lower bound under the condition (8).

Lemma 2.2. *Let β^* be a real number, and let $d, k \geq 3$ be integers such that*

$$d \leq \frac{1}{e^3\sqrt{k}} \cdot (1 + e^{-|\beta^*|})^{\frac{k}{2}}.$$

Then for a fixed $\Phi = (V, C) \in \Phi_{n,k,d}$,

$$\Pr_{\Phi,\beta^*}\left[\sum_{v\in V} e_v(\sigma) = \Omega(n)\right] = 1 - e^{-\Omega(n)}. \tag{13}$$

Using Lemma 2.2 and the identity (12), we derive (11) under the condition (8), thereby completing the proof of Theorem 2.1. The proof of Lemma 2.2 will be presented in the next section.

2.3. Proof of Lemma 2.2: applying the Lovász local lemma in a batch

The following version of the Lovász Local Lemma (LLL) from [26] will be useful for our proof of Lemma 2.2.

Lemma 2.3 *(Lemma 26, [26]). Let $\Phi = (V, C)$ be a CNF formula. Let μ be the product distribution on $\{\mathsf{TRUE}, \mathsf{FALSE}\}^{|V|}$, where each variable is set to TRUE independently with probability $e^\beta/(1+e^\beta)$. For each $c \in C$, let A_c be the event that clause c is unsatisfied. If there exists a sequence $\{x_c\}_{c\in C}$ such that for each $c \in C$,*

$$\Pr_{\mu}[A_c] \leq x_c \cdot \prod_{j\in\Gamma(c)} (1 - x_j), \tag{14}$$

where $\Gamma(c) \subseteq C$ is the set of clauses that contain a variable in $var(c)$, then Φ has a satisfying assignment. Moreover, the distributions $\Pr_{\Phi,\beta}$ and $\Pr_{\mu}$ can be related as follows: for any event E that can be completely determined by the assignment of a set S of variables,

$$\Pr_{\Phi,\beta}[E] \leq \Pr_{\mu}[E] \cdot \prod_{j\in\Gamma(E)} \frac{1}{1 - x_j}, \tag{15}$$

where $\Gamma(E)$ denotes the set of all clauses that contain a variable in S.

In many previous works utilising the LLL for sampling purposes, Lemma 2.3, or a variant of it, is typically applied to local events, including, for instance, the event that a specific variable is flippable or, more generally, to events happening in the neighbourhood of a vertex. This is present in the approach of [22] which, in the end, imposed stricter conditions on d (relative to k) since it requires "marking" variables appropriately (see [32]). Here, we prove Lemma 2.2 by applying Lemma 2.3 directly to a *batch* of random variables scattered around the graph in one go, removing the need for marking, and relaxing significantly the conditions on d. This simple idea enables our stronger result.

Proof of Lemma 2.2. Let $\beta = |\beta^*|$ and let $V = \{v_1, \ldots, v_n\}$. For any $c \in C$, we set $x_c = 1/(d^2k+1)$. Note $\Gamma(c) \le dk$. By (8) and the trivial bound $1 \le d^2k$, we have

$$d^2k + 1 \le \frac{1}{2e}(1+e^{-\beta})^k = \frac{1}{2e}\left(\frac{1+e^{\beta}}{e^{\beta}}\right)^k.$$

Also, since $d, k \ge 3$, we have

$$\frac{1}{2e} \le \left(1 - \frac{1}{dk}\right)^{dk}.$$

Thus, for each $c \in C$,

$$\begin{aligned}\Pr_{\mu}[A_c] &\le \left(\frac{e^{\beta}}{1+e^{\beta}}\right)^k \le \frac{1}{2e}\cdot\frac{1}{d^2k+1} \le \left(1-\frac{1}{dk}\right)^{dk}\frac{1}{d^2k+1}\\ &\le \left(1-\frac{1}{d^2k+1}\right)^{dk}\frac{1}{d^2k+1} \le x_c\cdot\prod_{j\in\Gamma(c)}(1-x_j),\end{aligned}$$

establishing condition (14) in Lemma 2.3.

Next we will show (13) under condition (14). Recall for an assignment σ of $\Omega_n(\Phi)$ and a variable $v_i \in V$, $e_{v_i}(\sigma) = 1$ if in every clause c containing v_i, there is a variable $v_j \ne v_i$ that satisfies c. Since d and k are bounded, for sufficiently large n, there exists a set U of $R = \Omega(n)$ variables $v_1, \ldots, v_R$ such that $d(v_i, v_j) \ge 100$ for all $1 \le i < j \le R$, where distance $d(\cdot,\cdot)$ is defined as the graphical distance in the hypergraph corresponding to Φ. Let E denote the event $\{\sum_{i=1}^{R} e_{v_i}(\sigma) < \frac{R}{3}\}$.

First, we compute the probability of E under the product distribution μ, where each variable is set to TRUE independently with probability $e^{\beta}/(1+e^{\beta})$. We apply a union bound by noting that if E occurs then there are at least $\frac{2R}{3}$ variables in U that are not flippable.

$$\Pr_{\mu}[E] = \Pr_{\mu}\left[\sum_{i=1}^{R} e_{v_i}(\sigma) < \frac{R}{3}\right] \le \binom{R}{2R/3}\left(\max_{v_i\in U}\Pr_{\mu}[e_{v_i}(\sigma)=0]\right)^{2R/3}. \tag{16}$$

If $e_{v_i}(\sigma) = 0$, then there exists a clause c_j such that $v_i \in var(c_j)$ and c_j is not satisfied by $var(c_j)\setminus\{v_i\}$ in σ. We apply another union bound over all c_j in which v_i appears, and obtain

$$\Pr_{\mu}[e_{v_i}(\sigma)=0] \le d\cdot\left(\frac{e^{\beta}}{1+e^{\beta}}\right)^{k-1}. \tag{17}$$

From (16) and (17), we have

$$\Pr_{\mu}[E] \le \binom{R}{2R/3}\left[d\cdot\left(\frac{e^{\beta}}{1+e^{\beta}}\right)^{k-1}\right]^{2R/3}. \tag{18}$$

We now apply Lemma 2.3 to relate the distribution $\Pr_{\Phi,\beta}[\cdot]$ to $\Pr_{\mu}[\cdot]$. Let S be the set of variables that are either in U or share a clause with variables in U. Then E is determined by the variables S. So $\Gamma(E)$ is the set of all clauses containing variables in S, and thus $|\Gamma(E)| \le Rd^2k$. It follows from (15), (18), and the standard bound $\binom{n}{m} \le \left(\frac{ne}{m}\right)^m$ that

$$\begin{aligned}\Pr_{\Phi,\beta}[E] &\le \Pr_{\mu}[E]\cdot\left(1-\frac{1}{d^2k+1}\right)^{-|\Gamma(E)|}\\ &\le \binom{R}{2R/3}\left[d\cdot\left(\frac{e^{\beta}}{1+e^{\beta}}\right)^{k-1}\right]^{2R/3}\left(1-\frac{1}{d^2k+1}\right)^{-Rd^2k}\\ &\le (3e/2)^{2R/3}\left[d\cdot\left(\frac{e^{\beta}}{1+e^{\beta}}\right)^{k-1}\right]^{2R/3}e^{R} \le \left[e^3\cdot d\cdot\left(\frac{e^{\beta}}{1+e^{\beta}}\right)^{k-1}\right]^{2R/3}.\end{aligned}$$

This completes the proof since we have $e^3\cdot d\cdot\left(\frac{e^{\beta}}{1+e^{\beta}}\right)^{k-1} < 1$ by our assumption (8). □

3. Impossibility of learning: proofs of Theorems 1.2 and 1.3

For the construction of the impossibility instances, we begin with a "gadget" that will serve as a building block.

Lemma 3.1. *Let $k \geq 4$ be an even integer and let $n \geq k$ be a multiple of $k/2$. If $d \geq k^2/2$, then there is a formula $\Psi_0 \in \Phi_{n,k,d}$ such that if an assignment σ of Ψ_0 is satisfying then either*

1. *σ has n TRUEs, or*
2. *σ has at least $2n/k$ FALSEs.*

Similarly (by symmetry), there is a formula $\Psi_1 \in \Phi_{n,k,d}$ such that, for any satisfying assignment σ of Ψ_1, either σ has n FALSEs, or σ has at least $2n/k$ TRUEs.

The instance Ψ_0 and Ψ_1 will lead to a proof of Theorem 1.3. For clarity, we denote the assignment to n variables with n TRUEs as σ^+.

Proof of Theorem 1.3. Assume $\beta^* \geq k \ln 2$ without loss of generality. Let Ψ_0 be the formula given by Lemma 3.1 (when $\beta^* \leq -k \ln 2$ we use Ψ_1 instead). Suppose σ is drawn from $\Pr_{\Psi_0,\beta^*}$, and we directly estimate the probability of $\{\sigma = \sigma^+\}$: since σ has at most 2^n possibilities and on event $\{\sigma \neq \sigma^+\}$ the number of TRUEs is at most $n - 2n/k$, we have

$$\Pr_{\Psi_0,\beta^*}[\sigma = \sigma^+] \geq \frac{e^{\beta^* n}}{e^{\beta^* n} + 2^n \cdot e^{\beta^* \cdot (n-2n/k)}} \geq \frac{2^{kn}}{2^{kn} + 2^{n+k\cdot(n-2n/k)}} = \frac{1}{1+2^{-n}}.$$

As the samples from $\Pr_{\Psi_0,\beta^*}$ are insensitive to β^* with high probability, learning β^* from σ is impossible. □

We now present a detailed description of the formula that defines Ψ_0 in Lemma 3.1. Let $k \geq 3$ be an even integer and let $n \geq k$ be a multiple of $k/2$. Let $N = \{0, \ldots, n-1\}$. The variables of Ψ_0 are $\{x_i \mid i \in N\}$.

For the construction, it will be helpful to group variables in batches of $k/2$ variables in a cyclic manner. Specifically, for $i \in \mathbb{N}$, consider the i-th batch of indices

$$\Xi_i = \{i + j \pmod n \mid j \in \{0, \ldots, k/2 - 1\}\}$$

and let $C_i = \{x_\ell \mid \ell \in \Xi_i\}$ be the corresponding variables in the i-th batch. We now introduce two types of length-$k/2$ clauses $W_{i,\ell}$ and Π_i that will be used to form the final length-k clauses. Specifically, for each ℓ in the batch Ξ_i, $W_{i,\ell}$ is the length-$(k/2)$ clause with variable set C_i in which x_ℓ appears positively and all other variables are negated. Let Π_i be the length-$(k/2)$ clause with variable set C_i in which all variables are negated. Finally,

$$\Psi_0 := \bigwedge_{i \in N, \ell \in \Xi_i} (W_{i,\ell} \vee \Pi_{i+k/2}).$$

(so Ψ_0 is the formula with variable set $\{x_i \mid i \in N\}$ and clause set $\{W_{i,\ell} \vee \Pi_{i+k/2} \mid i \in N, \ell \in \Xi_i\}$). The formula Ψ_1 is obtained from Ψ_0 by negating all of the literals.

Note that $\Psi_0, \Psi_1 \in \Phi_{n,d,k}$ for every integer $d \geq k^2/2$, since for each $j \in N$, the literals x_j and $\neg x_j$ occur (together) $k^2/2$ times. To see this, note that for each $j \in N$, x_j or $\neg x_j$ comes up in $W_{i,\ell} \vee \Pi_{i+k/2}$ for all $i \in \{j - k - 1 \pmod n, \ldots, j\}$ and all $\ell \in \Xi_i$ so this is k different i's and $k/2$ different ℓ's. In the proof of Lemma 3.2 we will demonstrate that Ψ_0 (and analogously, Ψ_1) satisfies the requirements of Lemma 3.1.

Lemma 3.2. *Let $k \geq 4$ be an even integer and let $n \geq k$ be a multiple of $k/2$. If $\sigma \neq \sigma^+$ satisfies Ψ_0 then, for all $\ell \in \{0, 1, \ldots, 2n/k - 1\}$, σ assigns at least one variable in $C_{\ell k/2} \cup C_{(\ell+1)k/2}$ to FALSE and σ has at least $2n/k$ FALSEs in total.*

Proof of Lemma 3.2. We will use the following claim as a key step of the proof.

Claim. *If $\sigma \neq \sigma^+$ satisfies Ψ_0, and σ assigns one variable x_a in $C_{\ell k/2}$ to FALSE, then either*

1. *σ assigns a variable x_b in $C_{(\ell+1)k/2}$ to FALSE, or*
2. *All variables in $C_{(\ell+1)k/2}$ are assigned to TRUE in σ, and there exist variables $x_c \in C_{(\ell+2)k/2}$ and $x_d \in C_{\ell k/2} \setminus \{x_a\}$ that are assigned to FALSE in σ.*

Proof of the Claim. Suppose we are not in the first case, so we will show that σ assigns FALSE to $x_c \in C_{(\ell+2)k/2}$ and $x_d \in C_{\ell k/2} \setminus \{x_a\}$. Since σ satisfies Ψ_0, it satisfies at least one of $W_{\ell k/2,a}$ and $\Pi_{(\ell+1)k/2}$. Since variables in $C_{(\ell+1)k/2}$ are all assigned to TRUE by the assumption that we are not in the first case, σ does not satisfy $\Pi_{(\ell+1)k/2}$. If σ satisfies $W_{\ell k/2,a}$

then it assigns a variable x_d to FALSE where $d \in \Xi_{\ell k/2} \setminus \{a\}$. Also, the set $\{j \in \Xi_{\ell k/2} : \sigma(x_j) = \text{FALSE}\}$ is not empty. Let $j = \max\{j \in \Xi_{\ell k/2} : \sigma(x_j) = \text{FALSE}\}$. Note that σ does not satisfy $W_{j,j}$, so for σ to satisfy $W_{j,j} \vee \Pi_{j+(k/2)}$, it must satisfy $\Pi_{j+(k/2)}$, which means σ assigns a variable $x_c \in C_{j+(k/2)}$ to FALSE. Since $x_c \notin C_{(\ell+1)k/2}$ and $C_{j+(k/2)} \subseteq C_{(\ell+1)k/2} \cup C_{(\ell+2)k/2}$, we establish that $x_c \in C_{(\ell+2)k/2}$. This concludes the proof of the claim.

We now show how to use the claim to prove the lemma. Fix an assignment $\sigma \neq \sigma^+$ that satisfies Ψ_0. Fix an index $j(0)$ so that $x_{j(0)}$ is assigned FALSE by σ. By symmetry of Ψ_0, we could assume $j(0) \in \Xi_0$. Let $\ell(0) = 0$ and $G(0) = \{x_{j(0)}\}$. Consider three sequences $\{j(t)\}_{t\geq 0}$, $\{\ell(t)\}_{t\geq 0}$ and $\{G(t)\}_{t\geq 0}$ defined recursively as follows. For every positive integer t, applying the claim to $a = j(t) \in \Xi_{\ell(t)k/2}$, in the first case we let

$$\ell(t+1) = \ell(t) + 1, \quad j(t+1) = b \in \Xi_{(\ell(t)+1)k/2} \quad \text{and} \quad G(t+1) = G(t) \cup \{x_b\};$$

in the latter case we let

$$\ell(t+1) = \ell(t) + 2, \quad j(t+1) = c \in \Xi_{(\ell(t)+2)k/2} \quad \text{and} \quad G(t+1) = G(t) \cup \{x_c, x_d\}.$$

By induction on t (with base case $t = 0$) we conclude that

(i) $j(t) \in \Xi_{\ell(t)k/2}$ is assigned to FALSE,
(ii) σ assigns all variables in $G(t)$ to FALSE, and
(iii) $\ell(t) + 2 \geq \ell(t+1) \geq \ell(t) + 1$ for $t < T$, where

$$T := \min\{t > 0 : \ell(t) = 2n/k - 1 \text{ or } 2n/k\}$$

is a stopping time of $\{j(t), \ell(t), G(t)\}_{t\geq 0}$.

By construction, for all $l \in \{0, 1, \ldots, \frac{2n}{k} - 1\}$, $G(T) \cap (C_{lk/2} \cup C_{(l+1)k/2})$ is not empty. Hence we have proved the first part of the lemma.

Next we will show that $|G(T)| \geq \frac{2n}{k}$. For this, observe that for $t < T$,

$$|G(t+1)| = |G(t)| + \ell(t+1) - \ell(t).$$

Since $|G(0)| = 1$, $\ell(0) = 0$ and $\ell(T) \geq \frac{2n}{k} - 1$, it holds that $|G(T)| \geq 2n/k$. □

Proof of Lemma 3.1. In the case of Ψ_0, first note that for all $i \in N$, σ^+ satisfies all instances of $W_{i,\ell} \vee \Pi_{i+k/2}$, so σ^+ satisfies Ψ_0. Also, if $\sigma \neq \sigma^+$, then the lemma follows immediately from Lemma 3.2. The case of Ψ_1 holds analogously. □

The names of the indices of the variables of Ψ_0 are not very important, and when we generalise the construction in Lemma 3.4 it will be useful to consider an arbitrary permutation of them. Here is the notation that we will use. Let π be any permutation of $N = \{0, \ldots, n-1\}$. We use the notation $\pi(i)$ to denote the element in N that i is mapped to by π. We will construct a formula Ψ_0^π. Taking id to be the identity permutation on N, the formula Ψ_0 that was already defined is Ψ_0^{id}.

For $i \in \mathbb{N}$, let

$$\Xi_i^\pi = \{\pi(i + j \pmod n) \mid j \in \{0, \ldots, k/2 - 1\}\}$$

and let $C_i^\pi = \{x_\ell \mid \ell \in \Xi_i^\pi\}$. For $\ell \in \Xi_i^\pi$, let $W_{i,\ell}^\pi$ be the length-$(k/2)$ clause with variable set C_i^π in which x_ℓ appears positively and all other variables are negated. Let Π_i^π be the length-$(k/2)$ clause with variable set C_i^π in which all variables are negated. Then

$$\Psi_0^\pi := \wedge_{i\in N, \ell\in\Xi_i}(W_{i,\ell}^\pi \vee \Pi_{i+k/2}^\pi).$$

The proof that $\Psi_0^\pi \in \Phi_{n,d,k}$ for any $d \geq k^2/2$ is exactly the same as the case $\pi = \text{id}$. The formula Ψ_1^π is obtained from Ψ_0^π by negating all of the literals.

The following corollary follows immediately from the proof of Lemma 3.2 (by renaming the indices using π) and the fact that $C_0, C_{k/2}, \ldots, C_{2n/k-1}$ are disjoint sets.

Corollary 3.3. *Let $k \geq 4$ be an even integer and let $n \geq k$ be a multiple of $k/2$. Let π be a permutation of N. If $\sigma \neq \sigma^+$ satisfies Ψ_0^π then there exists a subset $\mathcal{M}^\pi(\sigma) \subseteq \{0, 1, \ldots, 2n/k - 1\}$ of size at least n/k such that for all $\ell \in \mathcal{M}^\pi(\sigma)$, σ assigns at least one variable in $C_{\ell k/2}^\pi$ to* FALSE.

Remark. While Corollary 3.3 does not guarantee the uniqueness of $\mathcal{M}^\pi(\sigma)$, in what follows we define $\mathcal{M}^\pi(\sigma)$ to be the lexicographically smallest set among all the smallest sets satisfying the corollary. Since there are finitely many such sets and they can be lexicographically ordered, $\mathcal{M}^\pi(\sigma)$ is a unique and well-defined set for given π and σ.

Lemma 3.1 provides formula Ψ_0 with a GAP property applying to the number of FALSEs in its satisfying assignments. In the next lemma, we use Ψ_0 to build a larger formula that amplifies the GAP property to make the gap arbitrarily large.

Lemma 3.4. *Let $k \geq 4$ be an even integer, let $b > 1$ be a real number, and let d be an integer satisfying*

$$d \geq k^3\big(1-\frac{1}{b}\big)^{-k/2}.$$

Let $n \geq k$ be a sufficiently large multiple of $k/2$. Then there is a formula $\Psi_2 \in \Phi_{n,k,d}$ such that if an assignment σ satisfies Ψ_2 then either

1. *σ has n TRUEs, or*
2. *σ has at least n/b FALSEs.*

Similarly (by symmetry), there is a formula $\Psi_3 \in \Phi_{n,k,d}$ such that, for any satisfying assignment σ of Ψ_3, either σ has n FALSEs, or σ has at least n/b TRUEs.

Proof. Fix k, b, d and n as in the statement of the lemma. Let $N = \{0, \ldots, n-1\}$. The variables of Ψ_2 are $\{x_i \mid i \in N\}$. We will use the following notation. For any set Γ of permutations of N, let Ψ_2^Γ be the formula with variables $\{x_i \mid i \in N\}$ and clauses $\bigcup_{\pi\in\Gamma}\{W^\pi_{i,\ell} \vee \Pi^\pi_{i+k/2} \mid i \in N, \ell \in \Xi^\pi_i\}$. Let

$$J_* := \max\left\{2, \left\lfloor \frac{2k}{b}\big(1-\frac{1}{b}\big)^{-k/2}\right\rfloor\right\}$$

and let Γ_{J_*} be a set of J_* permutations of N, each chosen independently and uniformly at random. Let the permutations in Γ_{J_*} be denoted $\pi_1, \ldots, \pi_{J_*}$. For each positive integer $J \leq J_*$, let $\Gamma_J = \{\pi_1, \ldots, \pi_J\}$. The formula that we will construct is $\Psi_2 := \Psi_2^{\Gamma_{J_*}}$. Since the degree of each formula Ψ_0^π is at most $k^2/2$, the degree of Ψ_2 is at most $d_* := J_* k^2/2$. Note that

$$d_* = J_* \cdot \frac{k^2}{2} \leq \max\left\{k^2, \frac{k^3}{b}\big(1-\frac{1}{b}\big)^{-k/2}\right\} \leq k^3(1-\frac{1}{b})^{-k/2}.$$

We will show that, with positive probability over the choice of Γ_{J_*}, the formula $\Psi_2 = \Psi_2^{J_*}$ satisfies the requirements in the lemma statement. Since σ^+ satisfies every formula Ψ_0^π, it suffices to show that, with positive probability, every satisfying assignment $\sigma \neq \sigma^+$ of $\Psi_2^{J_*}$ has at least n/b variables assigned to FALSE.

For any set Γ of permutations of N, let $\Omega_n^-(\Gamma) \subseteq \Omega_n$ be the set of assignments $\sigma \neq \sigma^+$ that satisfy Ψ_2^Γ. Recall that $m(\sigma)$ is the number of variables that are assigned to TRUE by σ. Let $\mathrm{MinFalse}(\Gamma) = \min\{n - m(\sigma) \mid \sigma \in \Omega_n^-(\Gamma)\}$, so that $\mathrm{MinFalse}(\Gamma)$ is the minimum number of FALSE variables in any $\sigma \in \Omega_n^-(\Gamma)$. Since $\Omega_n^-(\Gamma_1) \supseteq \Omega_n^-(\Gamma_2) \supseteq \cdots \supseteq \Omega_n^-(\Gamma_{J_*})$, we have $\mathrm{MinFalse}(\Gamma_1) \leq \cdots \leq \mathrm{MinFalse}(\Gamma_{J_*})$. It suffices to show that, with positive probability over the choice of Γ_{J_*}, $\mathrm{MinFalse}(\Gamma_{J_*}) \geq n/b$.

Using the fact that n is sufficiently large, we will show that for all $t \in [J_* - 1]$ and all Γ_t such that $\mathrm{MinFalse}(\Gamma_t) < n/b$,

$$\mathbb{E}_{\pi_{t+1}}\big[\mathrm{MinFalse}(\Gamma_t \cup \{\pi_{t+1}\})\big] \geq \min\left\{\mathrm{MinFalse}(\Gamma_t) + \frac{3n}{4k}\big(1-\frac{1}{b}\big)^{k/2}, \frac{n}{b}\right\}. \tag{19}$$

By Lemma 3.2 (using symmetry to establish the statement for π_1 rather than for the identity permutation), $\mathrm{MinFalse}(\Gamma_1) \geq 2n/k \geq (3n/(4k))(1-1/b)^{k/2}$. Thus by (19), it follows that

$$\begin{aligned}
\mathbb{E}_{\Gamma_{J_*}}[\mathrm{MinFalse}(\Gamma_{J_*})] &\geq \min\left\{J_* \cdot \frac{3n}{4k}\big(1-\frac{1}{b}\big)^{k/2}, \frac{n}{b}\right\} \\
&= \min\left\{\max\left\{2, \left\lfloor \frac{2k}{b}\big(1-\frac{1}{b}\big)^{-k/2}\right\rfloor\right\} \cdot \frac{3n}{4k}\big(1-\frac{1}{b}\big)^{k/2}, \frac{n}{b}\right\} \\
&\geq \min\left\{2 \cdot \frac{3n}{4k} \cdot \frac{k}{2b}, \frac{4k}{3b}\big(1-\frac{1}{b}\big)^{-k/2} \cdot \frac{3n}{4k}\big(1-\frac{1}{b}\big)^{k/2}, \frac{n}{b}\right\} \geq \frac{n}{b}.
\end{aligned}$$

The second to last inequality needs some explanation. Let $x = (2k/b)(1-1/b)^{-k/2}$. If $x \geq 2$ then $\lfloor x \rfloor \geq x - x/3 = 2x/3$, and this is applied in the middle term of the final min. On the other hand, if $x < 2$ so that the maximum is taken at 2, then $(1-1/b)^{k/2} > k/b$ and the first term of the minimum is

$$2 \cdot \frac{3n}{4k}(1-1/b)^{k/2} > 2 \cdot \frac{3n}{4k}\frac{k}{b} = \frac{3n}{2b} \geq \frac{n}{b}.$$

Since $\mathrm{MinFalse}(\Gamma_{J_*})$ is bounded from above by n, the conclusion $\mathbb{E}_{\Gamma_{J_*}}[\mathrm{MinFalse}(\Gamma_{J_*})] \geq n/b$ implies that, with positive probability, $\mathrm{MinFalse}(\Gamma_{J_*}) \geq n/b$, completing the proof of the lemma.

It remains to prove the lower bound in (19). We start with some notation. For every assignment $\sigma \in \Omega_n \setminus \{\sigma^+\}$ let $F(\sigma)$ be the set of indices of variables that are assigned FALSE by σ. For any set Γ of permutations of N, and any $\sigma \in \Omega_n^-(\Gamma)$, let $S(\sigma,\Gamma)$ be the smallest (and lexicographically least, amongst the smallest) non-empty subset of $F(\sigma)$ such that the assignment σ' with $F(\sigma') = S(\sigma,\Gamma)$ satisfies Ψ_2^Γ. Clearly, $|S(\sigma,\Gamma)| \geq \text{MinFalse}(\Gamma)$. For any non-empty set $S \subseteq N$, any permutation π of N, and any set $M \subseteq \{0,1,\ldots,\frac{2n}{k}-1\}$, let

$$\Omega_n^-(\pi,\Gamma,S,M) = \{\sigma \in \Omega_n^-(\Gamma \cup \{\pi\}) \mid S(\sigma,\Gamma) = S, \mathcal{M}^\pi(\sigma) = M\}.$$

If $\Omega_n^-(\pi,\Gamma,S,M) \neq \emptyset$ let $\text{ExtMinFalse}(\pi,\Gamma,S,M) = \min\{n - m(\sigma), \sigma \in \Omega_n^-(\pi,\Gamma,S,M)\}$. (Otherwise, we do not define $\Omega_n^-(\pi,\Gamma,S,M)$.)

We are now ready to prove the lower bound in (19). Given a fixed $t \in [J_* - 1]$ and a fixed Γ_t such that $\text{MinFalse}(\Gamma_t) < n/b$, consider the distribution of $\text{MinFalse}(\Gamma_t \cup \{\pi_{t+1}\})$ (under the random choice of π_{t+1}). We will find it convenient to also fix S and M. We will show the following condition (20) for every $M \subseteq \{0,1,\ldots,\frac{2n}{k}-1\}$ with $|M| \geq n/k$ and every non-empty set $S \subseteq N$ with $|S| \geq \text{MinFalse}(\Gamma_t)$ such that $\Omega_n^-(\pi,\Gamma,S,M) \neq \emptyset$,

$$\mathbb{E}_{\pi_{t+1}}\big[\text{ExtMinFalse}(\pi_{t+1},\Gamma_t,S,M)\big] \geq \min\left\{\text{MinFalse}(\Gamma_t) + \frac{3n}{4k}\big(1-\frac{1}{b}\big)^{k/2}, \frac{n}{b}\right\}, \tag{20}$$

proving (19).

If $|S| \geq n/b$ then trivially $\text{ExtMinFalse}(\pi_{t+1},\Gamma_t,S,M) \geq n/b$ so suppose $|S| < n/b$. For every non-negative integer $\ell < 2n/k$, let $Y_\ell(S)$ be the indicator for the event that the random permutation π_{t+1} makes the intersection $\Xi_{\ell k/2}^{\pi_{t+1}} \cap S$ empty. Note that the sets in $\{\Xi_{\ell k/2}^{\pi_{t+1}} \mid 0 \leq \ell < 2n/k\}$ are disjoint, so by Corollary 3.3, any formula $\sigma \in \Omega_n^-(\pi_{t+1},\Gamma_t,S,M)$ has at least $|S| + \sum_{\ell \in M} Y_\ell(S)$ variables assigned to FALSE. So we need only show that the expectation of $|S| + \sum_{\ell \in M} Y_\ell(S)$ (under the choice of the random permutation π_{t+1}) is at least the right-hand-side of (20). First, for every non-negative integer $\ell < 2n/k$, note that

$$\begin{aligned}\mathbb{E}_{\pi_{t+1}}[Y_\ell(S)] = \Pr_{\pi_{t+1}}[Y_\ell(S)=1] &= \prod_{r=0}^{k/2-1}\frac{n-|S|-r}{n-r} \\ &\geq \big(1-\frac{|S|}{n-k}\big)^{k/2} > \big(1-\frac{1}{b(1-k/n)}\big)^{k/2} \geq \frac{3}{4}\big(1-\frac{1}{b}\big)^{k/2},\end{aligned}$$

where the last equality holds when n is large compared to k. We conclude that

$$\mathbb{E}_{\pi_{t+1}}\sum_{\ell \in M} Y_\ell(S) \geq \frac{3n}{4k}\cdot\big(1-\frac{1}{b}\big)^{k/2}.$$

Therefore, we conclude (20) from $|S| \geq \text{MinFalse}(\Gamma_t)$. □

Before proving Theorem 1.2, we provide the following more general result from which Theorem 1.2 is a corollary.

Theorem 3.5. *For all $\beta \neq 0$, let $\alpha = \alpha(|\beta|)$ be any real in $(0,1)$ satisfying*

$$|\beta| + \frac{\alpha}{1-\alpha}\ln\alpha + \ln(1-\alpha) > 0. \tag{21}$$

Let $k \geq 4$ be an even integer and let $\beta^* \in \mathbb{R}$ be such that $|\beta^*| \geq |\beta|$. Let n be a multiple of $k/2$ that is large enough. If

$$d \geq k^3\alpha^{-\frac{k}{2}}, \tag{22}$$

then there exists a formula $\Phi \in \Phi_{n,k,d}$ such that it is impossible to estimate β^* from a sample $\sigma \sim \Pr_{\Phi,\beta^*}$ with high probability.

Proof of Theorem 3.5. We first assume $\beta^* > 0$ and let Ψ_2 be the formula given by Lemma 3.4. Let $\alpha = \alpha(|\beta|) \in (0,1)$ be a constant satisfying (21). To see the existence of α, note that

$$f(\alpha) := -\frac{\alpha}{1-\alpha}\ln\alpha - \ln(1-\alpha)$$

is an increasing bijective function from $(0,1)$ to $(0,\infty)$. By (21), we have $\beta \geq f(\alpha)$. We will show that for any $\beta^* \geq \beta$, samples from both $\Pr_{\Psi_2,\beta^*}$ will be σ^+ with probability $1-e^{-C_1 n}$ for some $C_1 > 0$. Hence, not only does one-shot learning fail with high probability, but even $e^{C_1 n/2}$ many independent samples provide no additional information with high probability.

Setting $b = (1-\alpha)^{-1}$, condition (22) becomes $d \geq k^3\big(1-\frac{1}{b}\big)^{-k/2}$. Thus, Lemma 3.4 yields

$$\Pr_{\Psi_2,\beta^*}[\sigma = \sigma^+] = \frac{Z_1}{Z_1 + \sum_{\alpha=0}^{1-1/b} Z_\alpha}, \text{ where } Z_\alpha := \sum_{\sigma \in \Omega_n : m(\sigma) = \alpha n} e^{\beta^* \alpha n}\mathbb{1}[\sigma \models \Psi_2]. \tag{23}$$

Using Stirling's approximation $\binom{n}{\alpha n} \leq \frac{(\alpha^\alpha \cdot (1-\alpha)^{1-\alpha})^{-n}}{\sqrt{2\pi n\alpha(1-\alpha)}}$, we obtain that

$$Z_\alpha \leq e^{\beta^*\alpha n}\binom{n}{\alpha n} \leq \exp\left[n \cdot \left(\beta^*\alpha - \ln(\alpha^\alpha \cdot (1-\alpha)^{1-\alpha})\right)\right]. \tag{24}$$

From the definition of α in (21), it follows that

$$\exp\left[n \cdot \left(\beta^*\alpha - \ln(\alpha^\alpha \cdot (1-\alpha)^{1-\alpha})\right)\right] < e^{\beta^* n} \cdot e^{-\Omega(n)} = Z_1 \cdot e^{-\Omega(n)}. \tag{25}$$

Combining (23) with the estimates (24) and (25), we have

$$\Pr_{\Psi_2,\beta^*}[\sigma = \sigma^+] \geq \frac{Z_1}{Z_1 + n \cdot Z_1 \cdot e^{-\Omega(n)}} = \frac{1}{1 + n \cdot e^{-\Omega(n)}} = 1 - e^{-\Omega(n)}.$$

Using the formula Ψ_3, the proof of the case $\beta^* < 0$ is completely analogous. □

Finally, we prove Theorem 1.2 as a special case of Theorem 3.5.

Proof. We give an explicit $\alpha : (1,\infty) \to (0,1)$ such that $\alpha(|\beta^*|)$ satisfies (21) for any $|\beta^*| > 1$:

$$\alpha(\beta) = 1 - \frac{1}{e^{\beta-1}} = \frac{e^\beta - e}{e^\beta}. \tag{26}$$

Thus, we obtain the explicit lower bound (3) of d by plugging (26) to (22).

Next we verify that the choice of α in (26) satisfies (21). As in the proof of Theorem 3.5, we define $f(\alpha) := -\frac{\alpha}{1-\alpha}\ln\alpha - \ln(1-\alpha)$. To verify (21), it suffices to show that $f(\alpha(\beta)) < \beta$ for all $\beta > 1$. After some algebraic simplifications, we arrive at

$$\begin{aligned}
f(\alpha(\beta)) &= -\left[\frac{e^\beta - e}{e^\beta} \cdot \frac{1}{e^{1-\beta}} \cdot \ln\left(\frac{e^\beta - e}{e^\beta}\right) + \ln\frac{1}{e^{\beta-1}}\right] \\
&= -\left[(e^{\beta-1} - 1) \cdot \left(\ln(e^\beta - e) - \beta\right) + 1 - \beta\right] \\
&= \ln(e^\beta - e) - 1 + \beta e^{\beta-1} - e^{\beta-1}\ln(e^\beta - e) \\
&= \beta + \ln(1 - e^{1-\beta}) - 1 + \beta e^{\beta-1} - e^{\beta-1}\left[\beta + \ln(1 - e^{1-\beta})\right] \\
&= \beta + \ln(1 - e^{1-\beta}) - 1 - e^{\beta-1}\ln(1 - e^{1-\beta}) \\
&= \beta - 1 + (1 - e^{\beta-1})\ln(1 - e^{1-\beta})
\end{aligned}$$

Notice for all $0 < x < 1$, by Taylor's expansion,

$$(1 - \frac{1}{x})\ln(1-x) = -(1 - \frac{1}{x}) \cdot \sum_{n=1}^{\infty}\frac{x^n}{n} = 1 + \sum_{n=1}^{\infty}\left(\frac{1}{n+1} - \frac{1}{n}\right)x^n = 1 - \sum_{n=1}^{\infty}\frac{x^n}{n(n+1)} < 1.$$

Hence, by setting $x = e^{1-\beta}$, we have shown $f(\alpha(\beta)) < \beta$. □

CRediT authorship contribution statement

Andreas Galanis: Writing – review & editing, Supervision, Investigation, Conceptualization. **Leslie Ann Goldberg:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation. **Xusheng Zhang:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] J. Besag, Spatial interaction and the statistical analysis of lattice systems, J. R. Stat. Soc. Ser. B, Methodol. 36 (1974) 192–236.
[2] I. Bezáková, A. Blanca, Z. Chen, D. Štefankovič, E. Vigoda, Lower bounds for testing graphical models: colorings and antiferromagnetic Ising models, J. Mach. Learn. Res. 21 (2020).
[3] B.B. Bhattacharya, S. Mukherjee, Inference in Ising models, Bernoulli 24 (2018) 493–525.
[4] B.B. Bhattacharya, K. Ramanan, Parameter estimation for undirected graphical models with hard constraints, IEEE Trans. Inf. Theory 67 (2021) 6790–6809.
[5] A. Blanca, Z. Chen, D. Stefankovič, E. Vigoda, Hardness of identity testing for restricted Boltzmann machines and Potts models, J. Mach. Learn. Res. 22 (2021) 1–56.
[6] G. Bresler, Efficiently learning Ising models on arbitrary graphs, in: Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing, 2015, pp. 771–782.
[7] G. Bresler, D. Gamarnik, D. Shah, Hardness of parameter estimation in graphical models, in: Proceedings of the 28th International Conference on Neural Information Processing Systems, vol. 1, 2014, pp. 1062–1070.
[8] G. Bresler, D. Gamarnik, D. Shah, Structure learning of antiferromagnetic Ising models, in: Advances in Neural Information Processing Systems, Curran Associates, Inc., 2014, pp. 1062–1070.
[9] G. Bresler, D. Gamarnik, D. Shah, Learning graphical models from the Glauber dynamics, IEEE Trans. Inf. Theory 64 (2018) 4072–4080, https://doi.org/10.1109/TIT.2017.2713828.
[10] G. Bresler, D. Nagaraj, Optimal single sample tests for structured versus unstructured network data, in: Proceedings of the 31st Conference on Learning Theory, 2018, pp. 1657–1690.
[11] S. Chatterjee, Estimation in spin glasses: a first step, Ann. Stat. 35 (2007) 1931–1946.
[12] C. Chow, C. Liu, Approximating discrete probability distributions with dependence trees, IEEE Trans. Inf. Theory 14 (1968) 462–467, https://doi.org/10.1109/TIT.1968.1054142.
[13] F. Comets, On consistency of a class of estimators for exponential families of Markov random fields on the lattice, Ann. Stat. 20 (1992) 455–468.
[14] F. Comets, B. Gidas, Asymptotics of maximum likelihood estimators for the Curie-Weiss model, Ann. Stat. 19 (1991) 557–578, https://doi.org/10.1214/aos/1176348111.
[15] Y. Dagan, C. Daskalakis, N. Dikkala, A.V. Kandiros, Learning Ising models from one or multiple samples, in: Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, 2021, pp. 161–168.
[16] C. Daskalakis, N. Dikkala, G. Kamath, Testing Ising models, in: Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, 2018, pp. 1989–2007.
[17] C. Daskalakis, N. Dikkala, I. Panageas, Regression from dependent observations, in: Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, 2019, pp. 881–889.
[18] C. Daskalakis, N. Dikkala, I. Panageas, Logistic regression with peer-group effects via inference in higher-order Ising models, in: AISTATS, 2020, pp. 3653–3663, http://proceedings.mlr.press/v108/daskalakis20a.html.
[19] W. Feng, H. Guo, Y. Yin, C. Zhang, Fast sampling and counting k-sat solutions in the local lemma regime, J. ACM 68 (2021), https://doi.org/10.1145/3469832.
[20] J. Gaitonde, A. Moitra, E. Mossel, Bypassing the noisy parity barrier: learning higher-order Markov random fields from dynamics, https://arxiv.org/abs/2409.05284, arXiv:2409.05284, 2024.
[21] J. Gaitonde, E. Mossel, A unified approach to learning Ising models: beyond independence and bounded width, in: Proceedings of the 56th Annual ACM Symposium on Theory of Computing, Association for Computing Machinery, 2024, pp. 503–514.
[22] A. Galanis, A. Kalavasis, A.V. Kandiros, Learning hard-constrained models with one sample, in: Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2024, pp. 3184–3196.
[23] C.J. Geyer, E.A. Thompson, Constrained Monte Carlo maximum likelihood for dependent data, J. R. Stat. Soc. Ser. B, Methodol. 54 (1992) 657–699.
[24] P. Ghosal, S. Mukherjee, Joint estimation of parameters in Ising model, Ann. Stat. 48 (2020) 785–810.
[25] B. Gidas, Consistency of maximum likelihood and pseudo-likelihood estimators for Gibbs distributions, in: W. Fleming, P.L. Lions (Eds.), Stochastic Differential Systems, Stochastic Control Theory and Applications, 1988, pp. 129–145.
[26] H. Guo, M. Jerrum, J. Liu, Uniform sampling through the Lovász local lemma, J. ACM 66 (2019), https://doi.org/10.1145/3310131.
[27] L. Hamilton, F. Koehler, A. Moitra, Information theoretic properties of Markov random fields, and their algorithmic applications, in: Advances in Neural Information Processing Systems, 2017, pp. 2460–2469.
[28] K. He, C. Wang, Y. Yin, Deterministic counting Lovász local lemma beyond linear programming, in: Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2023, pp. 3388–3425.
[29] V. Jain, H.T. Pham, T.D. Vuong, Towards the sampling Lovász local lemma, in: 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS), 2022, pp. 173–183.
[30] A. Klivans, R. Meka, Learning graphical models using multiplicative weights, in: 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), 2017, pp. 343–354.
[31] W. Matthews, R. Paturi, Uniquely satisfiable k-sat instances with almost minimal occurrences of each variable, in: O. Strichman, S. Szeider (Eds.), Theory and Applications of Satisfiability Testing – SAT 2010, 2010, pp. 369–374.
[32] A. Moitra, Approximate counting, the Lovász local lemma, and inference in graphical models, J. ACM 66 (2019), https://doi.org/10.1145/3268930.
[33] R. Mukherjee, S. Mukherjee, M. Yuan, Global testing against sparse alternatives under Ising models, Ann. Stat. 46 (2018) 2062–2093.
[34] S. Mukherjee, J. Son, B.B. Bhattacharya, Estimation in tensor Ising models, Inf. Inference 11 (2022) 1457–1500.
[35] N.P. Santhanam, M.J. Wainwright, Information-theoretic limits of selecting binary graphical models in high dimensions, IEEE Trans. Inf. Theory 58 (2012) 4117–4134, https://doi.org/10.1109/TIT.2012.2191659.
[36] M. Vuffray, S. Misra, A.Y. Lokhov, M. Chertkov, Interaction screening: efficient and sample-optimal learning of Ising models, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016, pp. 2603–2611.
[37] C. Wang, Y. Yin, A sampling Lovász local lemma for large domain sizes, in: 2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS), 2024, pp. 129–150.
[38] H. Zhang, G. Kamath, J. Kulkarni, Z.S. Wu, Privately learning Markov random fields, in: Proceedings of the 37th International Conference on Machine Learning, 2020, pp. 11129–11140.