Rounding Meets Approximate Model Counting

Jiong Yang(✉) and Kuldeep S. Meel

National University of Singapore, Singapore, Singapore
jiong@comp.nus.edu.sg

Abstract. The problem of model counting, also known as #SAT, is to compute the number of models or satisfying assignments of a given Boolean formula F. Model counting is a fundamental problem in computer science with a wide range of applications. In recent years, there has been a growing interest in using hashing-based techniques for approximate model counting that provide (ε, δ)-guarantees: i.e., the count returned is within a $(1+\varepsilon)$-factor of the exact count with confidence at least $1-\delta$. While hashing-based techniques attain reasonable scalability for large enough values of δ, their scalability is severely impacted for smaller values of δ, thereby preventing their adoption in application domains that require estimates with high confidence.

The primary contribution of this paper is to address the Achilles heel of hashing-based techniques: we propose a novel approach based on *rounding* that allows us to achieve a significant reduction in runtime for smaller values of δ. The resulting counter, called ApproxMC6 (The resulting tool ApproxMC6 is available open-source at https://github.com/meelgroup/approxmc), achieves a substantial runtime performance improvement over the current state-of-the-art counter, ApproxMC. In particular, our extensive evaluation over a benchmark suite consisting of 1890 instances shows ApproxMC6 solves 204 more instances than ApproxMC, and achieves a $4\times$ speedup over ApproxMC.

1 Introduction

Given a Boolean formula F, the problem of model counting is to compute the number of models of F. Model counting is a fundamental problem in computer science with a wide range of applications, such as control improvisation [13], network reliability [9,28], neural network verification [2], probabilistic reasoning [5,11,20,21], and the like. In addition to myriad applications, the problem of model counting is a fundamental problem in theoretical computer science. In his seminal paper, Valiant showed that #SAT is #P-complete, where #P is the set of counting problems whose decision versions lie in NP [28]. Subsequently, Toda demonstrated the theoretical hardness of the problem by showing that every problem in the entire polynomial hierarchy can be solved by just one call to a #P oracle; more formally, $\mathsf{PH} \subseteq \mathsf{P}^{\#\mathsf{P}}$ [27].

Given the computational intractability of #SAT, there has been sustained interest in the development of approximate techniques from theoreticians and

© The Author(s) 2023
C. Enea and A. Lal (Eds.): CAV 2023, LNCS 13965, pp. 132–162, 2023.
https://doi.org/10.1007/978-3-031-37703-7_7

practitioners alike. Stockmeyer introduced a randomized hashing-based technique that provides (ε, δ)-guarantees (formally defined in Sect. 2) given access to an NP oracle [25]. Given the lack of practical solvers that could handle problems in NP satisfactorily, there were no practical implementations of Stockmeyere's hashing-based techniques until the 2000s [14]. Building on the unprecedented advancements in the development of SAT solvers, Chakraborty, Meel, and Vardi extended Stockmeyer's framework to a scalable (ε, δ)-counting algorithm, ApproxMC [7]. The subsequent years have witnessed a sustained interest in further optimizations of the hashing-based techniques for approximate counting [5,6,10,11,17–19,23,29,30]. The current state-of-the-art technique for approximate counting is a hashing-based framework called ApproxMC, which is in its fourth version, called ApproxMC4 [22,24].

The core theoretical idea behind the hashing-based framework is to use 2-universal hash functions to partition the solution space, denoted by $\mathsf{sol(F)}$ for a formula F, into *roughly equal small* cells, wherein a cell is considered *small* if it contains solutions less than or equal to a pre-computed threshold, thresh. An NP oracle (in practice, a SAT solver) is employed to check if a cell is small by enumerating solutions one-by-one until either there are no more solutions or we have already enumerated $\mathsf{thresh} + 1$ solutions. Then, we randomly pick a cell, enumerate solutions within the cell (if the cell is small), and scale the obtained count by the number of cells to obtain an estimate for $|\mathsf{sol(F)}|$. To amplify the confidence, we rely on the standard *median technique*: repeat the above process, called ApproxMCCore, multiple times and return the median. Computing the median amplifies the confidence as for the median of t repetitions to be outside the desired range (i.e., $\left[\frac{|\mathsf{sol(F)}|}{1+\varepsilon}, (1+\varepsilon)|\mathsf{sol(F)}|\right]$), it should be the case that at least half of the repetitions of ApproxMCCore returned a wrong estimate.

In practice, every subsequent repetition of ApproxMCCore takes a similar time, and the overall runtime increases linearly with the number of invocations. The number of repetitions depends logarithmically on δ^{-1}. As a particular example, for $\epsilon = 0.8$, the number of repetitions of ApproxMCCore to attain $\delta = 0.1$ is 21, which increases to 117 for $\delta = 0.001$: a significant increase in the number of repetitions (and accordingly, the time taken). Accordingly, it is no surprise that empirical analysis of tools such as ApproxMC has been presented with a high delta (such as $\delta = 0.1$). On the other hand, for several applications, such as network reliability, and quantitative verification, the end users desire estimates with high confidence. Therefore, the design of efficient counting techniques for small δ is a major challenge that one needs to address to enable the adoption of approximate counting techniques in practice.

The primary contribution of our work is to address the above challenge. We introduce a new technique called *rounding* that enables dramatic reductions in the number of repetitions required to attain a desired value of confidence. The core technical idea behind the design of the *rounding* technique is based on the following observation: Let L (resp. U) refer to the event that a given invocation of ApproxMCCore under (resp. over)-estimates $|\mathsf{sol(F)}|$. For a median estimate to be wrong, either the event L happens in half of the invocations of ApproxMCCore or the event U happens in half of the invocations of ApproxMCCore. The number of repetitions depends on $\max(\Pr[L], \Pr[U])$. The current algorithmic design (and ensuing analysis) of ApproxMCCore provides a weak upper bound on $\max\{\Pr[L], \Pr[U]\}$: in particular, the bounds on $\max\{\Pr[L], \Pr[U]\}$ and $\Pr[L \cup U]$ are almost identical. Our key technical contribution is to design a new procedure, ApproxMC6Core, based on the rounding technique that allows us to obtain significantly better bounds on $\max\{\Pr[L], \Pr[U]\}$.

The resulting algorithm, called ApproxMC6, follows a similar structure to that of ApproxMC: it repeatedly invokes the underlying core procedure ApproxMC6Core and returns the median of the estimates. Since a single invocation of ApproxMC6Core takes as much time as ApproxMCCore, the reduction in the number of repetitions is primarily responsible for the ensuing speedup. As an example, for $\varepsilon = 0.8$, the number of repetitions of ApproxMC6Core to attain $\delta = 0.1$ and $\delta = 0.001$ is just 5 and 19, respectively; the corresponding numbers for ApproxMC were 21 and 117. An extensive experimental evaluation on 1890 benchmarks shows that the rounding technique provided $4\times$ speedup than the state-of-the-art approximate model counter, ApproxMC. Furthermore, for a given timeout of 5000 s, ApproxMC6 solves 204 more instances than ApproxMC and achieves a reduction of 1063 s in the PAR-2 score.

The rest of the paper is organized as follows. We introduce notation and preliminaries in Sect. 2. To place our contribution in context, we review related works in Sect. 3. We identify the weakness of the current technique in Sect. 4 and present the rounding technique in Sect. 5 to address this issue. Then, we present our experimental evaluation in Sect. 6. Finally, we conclude in Sect. 7.

2 Notation and Preliminaries

Let F be a Boolean formula in conjunctive normal form (CNF), and let $\mathsf{Vars}(F)$ be the set of variables appearing in F. The set $\mathsf{Vars}(F)$ is also called the *support* of F. An assignment σ of truth values to the variables in $\mathsf{Vars}(F)$ is called a *satisfying assignment* or *witness* of F if it makes F evaluate to true. We denote the set of all witnesses of F by $\mathsf{sol(F)}$. Throughout the paper, we will use n to denote $|\mathsf{Vars}(F)|$.

The *propositional model counting problem* is to compute $|\mathsf{sol(F)}|$ for a given CNF formula F. A *probably approximately correct* (or PAC) counter is a probabilistic algorithm $\mathsf{ApproxCount}(\cdot, \cdot, \cdot)$ that takes as inputs a formula F, a tolerance parameter $\varepsilon > 0$, and a confidence parameter $\delta \in (0, 1]$, and returns an (ε, δ)-estimate c, i.e., $\mathsf{Pr}\left[\frac{|\mathsf{sol(F)}|}{1+\varepsilon} \leq c \leq (1+\varepsilon)|\mathsf{sol(F)}|\right] \geq 1-\delta$. PAC guarantees are also sometimes referred to as (ε, δ)-guarantees.

A closely related notion is projected model counting, where we are interested in computing the cardinality of $\mathsf{sol(F)}$ projected on a subset of variables $\mathcal{P} \subseteq \mathsf{Vars}(F)$. While for clarity of exposition, we describe our algorithm in the context of model counting, the techniques developed in this paper are applicable to projected model counting as well. Our empirical evaluation indeed considers such benchmarks.

2.1 Universal Hash Functions

Let $n, m \in \mathbb{N}$ and $\mathcal{H}(n,m) \triangleq \{h : \{0,1\}^n \rightarrow \{0,1\}^m\}$ be a family of hash functions mapping $\{0,1\}^n$ to $\{0,1\}^m$. We use $h \xleftarrow{R} \mathcal{H}(n,m)$ to denote the probability space obtained by choosing a function h uniformly at random from $\mathcal{H}(n,m)$. To measure the quality of a hash function we are interested in the set of elements of $\mathsf{sol}(\mathsf{F})$ mapped to α by h, denoted $\mathsf{Cell}_{\langle F,h,\alpha\rangle}$ and its cardinality, i.e., $|\mathsf{Cell}_{\langle F,h,\alpha\rangle}|$. We write $\Pr[Z : \Omega]$ to denote the probability of outcome Z when sampling from a probability space Ω. For brevity, we omit Ω when it is clear from the context. The expected value of Z is denoted $\mathsf{E}[Z]$ and its variance is denoted $\sigma^2[Z]$.

Definition 1. *A family of hash functions $\mathcal{H}(n,m)$ is strongly 2-universal if $\forall x, y \in \{0,1\}^n$, $\alpha \in \{0,1\}^m$, $h \xleftarrow{R} \mathcal{H}(n,m)$,*

$$\mathsf{Pr}[h(x) = \alpha] = \frac{1}{2^m} = \mathsf{Pr}[h(x) = h(y)]$$

For $h \xleftarrow{R} \mathcal{H}(n,n)$ and $\forall m \in \{1,...,n\}$, the m^{th} prefix-slice of h, denoted $h^{(m)}$, is a map from $\{0,1\}^n$ to $\{0,1\}^m$, such that $h^{(m)}(y)[i] = h(y)[i]$, for all $y \in \{0,1\}^n$ and for all $i \in \{1,...,m\}$. Similarly, the m^{th} prefix-slice of $\alpha \in \{0,1\}^n$, denoted $\alpha^{(m)}$, is an element of $\{0,1\}^m$ such that $\alpha^{(m)}[i] = \alpha[i]$ for all $i \in \{1,...,m\}$. To avoid cumbersome terminology, we abuse notation and write $\mathsf{Cell}_{\langle F,m\rangle}$(resp. $\mathsf{Cnt}_{\langle F,m\rangle}$) as a short-hand for $\mathsf{Cell}_{\langle F,h^{(m)},\alpha^{(m)}\rangle}$ (resp. $|\mathsf{Cell}_{\langle F,h^{(m)},\alpha^{(m)}\rangle}|$). The following proposition presents two results that are frequently used throughout this paper. The proof is deferred to Appendix A.

Proposition 1. *For every $1 \le m \le n$, the following holds:*

$$\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right] = \frac{|\mathsf{sol}(\mathsf{F})|}{2^m} \tag{1}$$

$$\sigma^2\left[\mathsf{Cnt}_{\langle F,m\rangle}\right] \le \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right] \tag{2}$$

The usage of prefix-slice of h ensures monotonicity of the random variable, $\mathsf{Cnt}_{\langle F,m\rangle}$, since from the definition of prefix-slice, we have that for every $1 \le m < n$, $h^{(m+1)}(y) = \alpha^{(m+1)} \Rightarrow h^{(m)}(y) = \alpha^{(m)}$. Formally,

Proposition 2. *For every $1 \le m < n$, $\mathsf{Cell}_{\langle F,m+1\rangle} \subseteq \mathsf{Cell}_{\langle F,m\rangle}$*

2.2 Helpful Combinatorial Inequality

Lemma 1. *Let $\eta(t,m,p) = \sum_{k=m}^{t} \binom{t}{k} p^k (1-p)^{t-k}$ and $p < 0.5$, then*

$$\eta(t, \lceil t/2 \rceil, p) \in \Theta\left(t^{-\frac{1}{2}}\left(2\sqrt{p(1-p)}\right)^t\right)$$

Proof. We will derive both an upper and a matching lower bound for $\eta(t, \lceil t/2 \rceil, p)$. We begin by deriving an upper bound: $\eta(t, \lceil t/2 \rceil, p) = \sum_{k=\lceil \frac{t}{2} \rceil}^{t} \binom{t}{k} p^k (1-p)^{t-k} \leq \binom{t}{\lceil t/2 \rceil} \sum_{k=\lceil \frac{t}{2} \rceil}^{t} p^k (1-p)^{t-k} \leq \binom{t}{\lceil t/2 \rceil} \cdot (p(1-p))^{\lceil \frac{t}{2} \rceil} \cdot \frac{1}{1-2p}$ $\leq \frac{1}{\sqrt{2\pi}} \cdot \frac{t}{\sqrt{\left(\frac{t}{2}-0.5\right)\left(\frac{t}{2}+0.5\right)}} \cdot \left(\frac{t}{t-1}\right)^t \cdot e^{\frac{1}{12t} - \frac{1}{6t+6} - \frac{1}{6t-6}} \cdot t^{-\frac{1}{2}} 2^t \cdot (p(1-p))^{\frac{t}{2}} \cdot (p(1-p))^{\frac{1}{2}} \cdot \frac{1}{1-2p}$. The last inequality follows Stirling's approximation. As a result, $\eta(t, \lceil t/2 \rceil, p) \in \mathcal{O}\left(t^{-\frac{1}{2}} \left(2\sqrt{p(1-p)}\right)^t\right)$. Afterwards; we move on to deriving a matching lower bound: $\eta(t, \lceil t/2 \rceil, p) = \sum_{k=\lceil \frac{t}{2} \rceil}^{t} \binom{t}{k} p^k (1-p)^{t-k} \geq \binom{t}{\lceil t/2 \rceil} p^{\lceil \frac{t}{2} \rceil} (1-p)^{t - \lceil \frac{t}{2} \rceil} \geq \frac{1}{\sqrt{2\pi}} \cdot \frac{t}{\sqrt{\left(\frac{t}{2}-0.5\right)\left(\frac{t}{2}+0.5\right)}} \cdot \left(\frac{t}{t+1}\right)^t \cdot e^{\frac{1}{12t} - \frac{1}{6t+6} - \frac{1}{6t-6}} \cdot t^{-\frac{1}{2}} 2^t \cdot (p(1-p))^{\frac{t}{2}} \cdot p^{\frac{1}{2}} (1-p)^{-\frac{1}{2}} \cdot \frac{1}{1-2p}$. The last inequality again follows Stirling's approximation. Hence, $\eta(t, \lceil t/2 \rceil, p) \in \Omega\left(t^{-\frac{1}{2}} \left(2\sqrt{p(1-p)}\right)^t\right)$. Combining these two bounds, we conclude that $\eta(t, \lceil t/2 \rceil, p) \in \Theta\left(t^{-\frac{1}{2}} \left(2\sqrt{p(1-p)}\right)^t\right)$. □

3 Related Work

The seminal work of Valiant established that #SAT is #P-complete [28]. Toda later showed that every problem in the polynomial hierarchy could be solved by just a polynomial number of calls to a #P oracle [27]. Based on Carter and Wegman's seminal work on universal hash functions [4], Stockmeyer proposed a probabilistic polynomial time procedure, with access to an NP oracle, to obtain an (ε, δ)-approximation of F [25].

Built on top of Stockmeyer's work, the core theoretical idea behind the hashing-based approximate solution counting framework, as presented in Algorithm 1 (ApproxMC [7]), is to use 2-universal hash functions to partition the solution space (denoted by $\mathsf{sol}(\mathsf{F})$ for a given formula F) into *small* cells of *roughly equal* size. A cell is considered *small* if the number of solutions it contains is less than or equal to a pre-determined threshold, thresh. An NP oracle is used to determine if a cell is small by iteratively enumerating its solutions until either there are no more solutions or thresh + 1 solutions have been found. In practice, an SAT solver is used to implement the NP oracle. To ensure a polynomial number of calls to the oracle, the threshold, thresh, is set to be polynomial in the input parameter ε at Line 1. The subroutine ApproxMCCore takes the formula F and thresh as inputs and estimates the number of solutions at Line 7. To determine the appropriate number of cells, i.e., the value of m for $\mathcal{H}(n, m)$, ApproxMCCore uses a search procedure at Line 3 of Algorithm 2. The estimate is calculated as the number of solutions in a randomly chosen cell, scaled by the number of cells, i.e., 2^m at Line 5. To improve confidence in the estimate, ApproxMC performs multiple runs of the ApproxMCCore subroutine at Lines 5–9 of Algorithm 1. The final count is computed as the median of the estimates obtained at Line 10.

Algorithm 1. $\mathsf{ApproxMC}(F, \varepsilon, \delta)$

$\mathsf{thresh} \leftarrow 9.84 \left(1 + \frac{\varepsilon}{1+\varepsilon}\right) \left(1 + \frac{1}{\varepsilon}\right)^2$;
$Y \leftarrow \mathsf{BoundedSAT}(F, \mathsf{thresh})$;
if $(|Y| < \mathsf{thresh})$ **then return** $|Y|$;
$t \leftarrow \lceil 17 \log_2(3/\delta) \rceil$; $C \leftarrow \mathsf{emptyList}$; $\mathsf{iter} \leftarrow 0$;
repeat
    $\mathsf{iter} \leftarrow \mathsf{iter} + 1$;
    $\mathsf{nSols} \leftarrow \mathsf{ApproxMCCore}(F, \mathsf{thresh})$;
    $\mathsf{AddToList}(C, \mathsf{nSols})$;
until $(\mathsf{iter} \geq t)$;
$\mathsf{finalEstimate} \leftarrow \mathsf{FindMedian}(C)$;
return $\mathsf{finalEstimate}$;

Algorithm 2. $\mathsf{ApproxMCCore}(F, \mathsf{thresh})$

1: Choose h at random from $\mathcal{H}(n, n)$;
2: Choose α at random from $\{0, 1\}^n$;
3: $m \leftarrow \mathsf{LogSATSearch}(F, h, \alpha, \mathsf{thresh})$;
4: $\mathsf{Cnt}_{\langle F,m \rangle} \leftarrow \mathsf{BoundedSAT}\left(F \wedge \left(h^{(m)}\right)^{-1}\left(\alpha^{(m)}\right), \mathsf{thresh}\right)$;
5: **return** $(2^m \times \mathsf{Cnt}_{\langle F,m \rangle})$;

In the second version of ApproxMC [8], two key algorithmic improvements are proposed to improve the practical performance by reducing the number of calls to the SAT solver. The first improvement is using galloping search to more efficiently find the correct number of cells, i.e., LogSATSearch at Line 3 of Algorithm 2. The second is using linear search over a small interval around the previous value of m before resorting to the galloping search. Additionally, the third and fourth versions [22,23] enhance the algorithm's performance by effectively dealing with CNF formulas conjuncted with XOR constraints, commonly used in the hashing-based counting framework. Moreover, an effective preprocessor named Arjun [24] is proposed to enhance ApproxMC's performance by constructing shorter XOR constraints. As a result, the combination of Arjun and ApproxMC4 solved almost all existing benchmarks [24], making it the current state of the art in this field.

In this work, we aim to address the main limitation of the ApproxMC algorithm by focusing on an aspect that still needs to be improved upon by previous developments. Specifically, we aim to improve the core algorithm of ApproxMC, which has remained unchanged.

4 Weakness of ApproxMC

As noted above, the core algorithm of ApproxMC has not changed since 2016, and in this work, we aim to address the core limitation of ApproxMC. To put our contribution in context, we first review ApproxMC and its core algorithm, called

$\mathsf{ApproxMCCore}$. We present the pseudocode of $\mathsf{ApproxMC}$ and $\mathsf{ApproxMCCore}$ in Algorithms 1 and 2, respectively. $\mathsf{ApproxMCCore}$ may return an estimate that falls outside the PAC range $\left[\frac{|\mathsf{sol}(\mathsf{F})|}{1+\varepsilon}, (1+\varepsilon)|\mathsf{sol}(\mathsf{F})|\right]$ with a certain probability of error. Therefore, $\mathsf{ApproxMC}$ repeatedly invokes $\mathsf{ApproxMCCore}$ (Lines 5– 9) and returns the median of the estimates returned by $\mathsf{ApproxMCCore}$ (Line 10), which reduces the error probability to the user-provided parameter δ.

Let Error_t denote the event that the median of t estimates falls outside $\left[\frac{|\mathsf{sol}(\mathsf{F})|}{1+\varepsilon}, (1+\varepsilon)|\mathsf{sol}(\mathsf{F})|\right]$. Let L denote the event that an invocation $\mathsf{ApproxMCCore}$ returns an estimate less than $\frac{|\mathsf{sol}(\mathsf{F})|}{1+\varepsilon}$. Similarly, let U denote the event that an individual estimate of $|\mathsf{sol}(\mathsf{F})|$ is greater than $(1+\varepsilon)|\mathsf{sol}(\mathsf{F})|$. For simplicity of exposition, we assume t is odd; the current implementation of t indeed ensures that t is odd by choosing the smallest odd t for which $\Pr[\mathsf{Error}_t] \leq \delta$.

In the remainder of the section, we will demonstrate that reducing $\max\{\Pr[L], \Pr[U]\}$ can effectively reduce the number of repetitions t, making the small-δ scenarios practical. To this end, we will first demonstrate the existing analysis technique of $\mathsf{ApproxMC}$ leads to loose bounds on $\Pr[\mathsf{Error}_t]$. We then present a new analysis that leads to tighter bounds on $\Pr[\mathsf{Error}_t]$.

The existing combinatorial analysis in [7] derives the following proposition:

Proposition 3.

$$Pr[\mathsf{Error}_t] \leq \eta(t, \lceil t/2 \rceil, Pr[L \cup U])$$

where $\eta(t, m, p) = \sum_{k=m}^{t} \binom{t}{k} p^k (1-p)^{t-k}$.

Proposition 3 follows from the observation that if the median falls outside the PAC range, at least $\lceil t/2 \rceil$ of the results must also be outside the range. Let $\eta(t, \lceil t/2 \rceil, \Pr[L \cup U]) \leq \delta$, and we can compute a valid t at Line 4 of $\mathsf{ApproxMC}$.

Proposition 3 raises a question: can we derive a tight upper bound for $\Pr[\mathsf{Error}_t]$? The following lemma provides an affirmative answer to this question.

Lemma 2. *Assuming t is odd, we have:*

$$Pr[\mathsf{Error}_t] = \eta(t, \lceil t/2 \rceil, Pr[L]) + \eta(t, \lceil t/2 \rceil, Pr[U])$$

Proof. Let I_i^L be an indicator variable that is 1 when $\mathsf{ApproxMCCore}$ returns a nSols less than $\frac{|\mathsf{sol}(\mathsf{F})|}{1+\varepsilon}$, indicating the occurrence of event L in the i-th repetition. Let I_i^U be an indicator variable that is 1 when $\mathsf{ApproxMCCore}$ returns a nSols greater than $(1+\varepsilon)|\mathsf{sol}(\mathsf{F})|$, indicating the occurrence of event U in the i-th repetition. We aim first to prove that $\mathsf{Error}_t \Leftrightarrow \left(\sum_{i=1}^{t} I_i^L \geq \lceil \frac{t}{2} \rceil\right) \vee \left(\sum_{i=1}^{t} I_i^U \geq \lceil \frac{t}{2} \rceil\right)$. We will begin by proving the right ($\Rightarrow$) implication. If the median of t estimates violates the PAC guarantee, the median is either less than $\frac{|\mathsf{sol}(\mathsf{F})|}{1+\varepsilon}$ or greater than $(1+\varepsilon)|\mathsf{sol}(\mathsf{F})|$. In the first case, since half of the estimates are less than the median, at least $\lceil \frac{t}{2} \rceil$ estimates are less than $\frac{|\mathsf{sol}(\mathsf{F})|}{1+\varepsilon}$. Formally, this implies $\sum_{i=1}^{t} I_i^L \geq \left\lceil \frac{t}{2} \right\rceil$. Similarly, in the case that the median is greater than $(1+\varepsilon)|\mathsf{sol}(\mathsf{F})|$, since half of the estimates are greater than the median, at least $\left\lceil \frac{t}{2} \right\rceil$ estimates are greater than $(1+\varepsilon)|\mathsf{sol}(\mathsf{F})|$, thus formally implying $\sum_{i=1}^{t} I_i^U \geq \left\lceil \frac{t}{2} \right\rceil$. On the other hand, we prove the left ($\Leftarrow$) implication. Given $\sum_{i=1}^{t} I_i^L \geq \left\lceil \frac{t}{2} \right\rceil$, more than half of the estimates are less than $\frac{|\mathsf{sol}(\mathsf{F})|}{1+\varepsilon}$, and therefore the median is less than $\frac{|\mathsf{sol}(\mathsf{F})|}{1+\varepsilon}$, violating the PAC guarantee. Similarly, given $\sum_{i=1}^{t} I_i^U \geq \left\lceil \frac{t}{2} \right\rceil$, more than half of the estimates are greater than $(1+\varepsilon)|\mathsf{sol}(\mathsf{F})|$, and therefore the median is greater than $(1+\varepsilon)|\mathsf{sol}(\mathsf{F})|$, violating the PAC guarantee. This concludes the proof of $\mathsf{Error}_t \Leftrightarrow \left(\sum_{i=1}^{t} I_i^L \geq \left\lceil \frac{t}{2} \right\rceil\right) \vee \left(\sum_{i=1}^{t} I_i^U \geq \left\lceil \frac{t}{2} \right\rceil\right)$. Then we obtain:

$$
\begin{aligned}
\Pr\left[\mathsf{Error}_t\right] &= \Pr\left[\left(\sum_{i=1}^{t} I_i^L \geq \lceil t/2 \rceil\right) \vee \left(\sum_{i=1}^{t} I_i^U \geq \lceil t/2 \rceil\right)\right] \\
&= \Pr\left[\left(\sum_{i=1}^{t} I_i^L \geq \lceil t/2 \rceil\right)\right] + \Pr\left[\left(\sum_{i=1}^{t} I_i^U \geq \lceil t/2 \rceil\right)\right] \\
&\quad - \Pr\left[\left(\sum_{i=1}^{t} I_i^L \geq \lceil t/2 \rceil\right) \wedge \left(\sum_{i=1}^{t} I_i^U \geq \lceil t/2 \rceil\right)\right]
\end{aligned}
$$

Given $I_i^L + I_i^U \leq 1$ for $i = 1, 2, ..., t$, $\sum_{i=1}^{t}(I_i^L + I_i^U) \leq t$ is there, but if $\left(\sum_{i=1}^{t} I_i^L \geq \lceil t/2 \rceil\right) \wedge \left(\sum_{i=1}^{t} I_i^U \geq \lceil t/2 \rceil\right)$ is also given, we obtain $\sum_{i=1}^{t}(I_i^L + I_i^U) \geq t+1$ contradicting $\sum_{i=1}^{t}(I_i^L + I_i^U) \leq t$; Hence, we can conclude that $\Pr\left[\left(\sum_{i=1}^{t} I_i^L \geq \lceil t/2 \rceil\right) \wedge \left(\sum_{i=1}^{t} I_i^U \geq \lceil t/2 \rceil\right)\right] = 0$. From this, we can deduce:

$$
\begin{aligned}
\Pr\left[\mathsf{Error}_t\right] &= \Pr\left[\left(\sum_{i=1}^{t} I_i^L \geq \lceil t/2 \rceil\right)\right] + \Pr\left[\left(\sum_{i=1}^{t} I_i^U \geq \lceil t/2 \rceil\right)\right] \\
&= \eta(t, \lceil t/2 \rceil, \Pr\left[L\right]) + \eta(t, \lceil t/2 \rceil, \Pr\left[U\right])
\end{aligned}
$$

□

Though Lemma 2 shows that reducing $\Pr[L]$ and $\Pr[U]$ can decrease the error probability, it is still uncertain to what extent $\Pr[L]$ and $\Pr[U]$ affect the error probability. To further understand this impact, the following lemma is presented to establish a correlation between the error probability and t depending on $\Pr[L]$ and $\Pr[U]$.

Lemma 3. *Let $p_{max} = \max\{Pr[L], Pr[U]\}$ and $p_{max} < 0.5$, we have*

$$
Pr\left[\mathsf{Error}_t\right] \in \Theta\left(t^{-\frac{1}{2}}\left(2\sqrt{p_{max}(1-p_{max})}\right)^t\right)
$$

Proof. Applying Lemmas 1 and 2, we have

$$\Pr\left[\mathsf{Error}_t\right] \in \Theta\left(t^{-\frac{1}{2}}\left(\left(2\sqrt{\Pr\left[L\right]\left(1-\Pr\left[L\right]\right)}\right)^t + \left(2\sqrt{\Pr\left[U\right]\left(1-\Pr\left[U\right]\right)}\right)^t\right)\right)$$
$$= \Theta\left(t^{-\frac{1}{2}}\left(2\sqrt{p_{max}(1-p_{max})}\right)^t\right)$$

□

In summary, Lemma 3 provides a way to tighten the bound on $\Pr[\mathsf{Error}_t]$ by designing an algorithm such that we can obtain a tighter bound on p_{max} in contrast to previous approaches that relied on obtaining a tighter bound on $\Pr[L \cup U]$.

5 Rounding Model Counting

In this section, we present a *rounding*-based technique that allows us to obtain a tighter bound on p_{max}. On a high-level, instead of returning the estimate from one iteration of the underlying core algorithm as the number of solutions in a randomly chosen cell multiplied by the number of cells, we *round* each estimate of the model count to a value that is more likely to be within $(1+\varepsilon)$-bound. While counter-intuitive at first glance, we show that rounding the estimate reduces $\max\left\{\Pr\left[L\right], \Pr\left[U\right]\right\}$, thereby resulting in a smaller number of repetitions of the underlying algorithm.

We present ApproxMC6, a *rounding*-based approximate model counting algorithm, in Sect. 5.1. Section 5.2 will demonstrate how ApproxMC6 decreases $\max\left\{\Pr\left[L\right], \Pr\left[U\right]\right\}$ and the number of estimates. Lastly, in Sect. 5.3, we will provide proof of the theoretical correctness of the algorithm.

5.1 Algorithm

Algorithm 3 presents the procedure of ApproxMC6. ApproxMC6 takes as input a formula F, a tolerance parameter ε, and a confidence parameter δ. ApproxMC6 returns an (ε, δ)-estimate c of $|\mathsf{sol(F)}|$ such that $\mathsf{Pr}\left[\frac{|\mathsf{sol(F)}|}{1+\varepsilon} \le c \le (1+\varepsilon)|\mathsf{sol(F)}|\right] \ge 1-\delta$. ApproxMC6 is identical to ApproxMC in its initialization of data structures and handling of base cases (Lines 1–4).

In Line 5, we pre-compute the rounding type and rounding value to be used in ApproxMC6Core. configRound is implemented in Algorithm 5; the precise choices arise due to technical analysis, as presented in Sect. 5.2. Note that, in configRound, $\mathsf{Cnt}_{\langle F,m\rangle}$ is *rounded up* to roundValue for $\varepsilon < 3$ ($\mathsf{roundUp} = 1$) but *rounded* to roundValue for $\varepsilon \ge 3$ ($\mathsf{roundUp} = 0$). Rounding up means we assign roundValue to $\mathsf{Cnt}_{\langle F,m\rangle}$ if $\mathsf{Cnt}_{\langle F,m\rangle}$ is less than roundValue and, otherwise, keep $\mathsf{Cnt}_{\langle F,m\rangle}$ unchanged. Rounding means that we assign roundValue to $\mathsf{Cnt}_{\langle F,m\rangle}$ in all cases. ApproxMC6 computes the number of repetitions necessary to lower error probability down to δ at Line 6. The implementation of computeIter is presented

Algorithm 3. $\mathsf{ApproxMC6}(F, \varepsilon, \delta)$

1: $\mathsf{thresh} \leftarrow 9.84\left(1+\frac{\varepsilon}{1+\varepsilon}\right)\left(1+\frac{1}{\varepsilon}\right)^2$;
2: $Y \leftarrow \mathsf{BoundedSAT}(F, \mathsf{thresh})$;
3: **if** $(|Y| < \mathsf{thresh})$ **then return** $|Y|$;
4: $C \leftarrow \mathsf{emptyList}; \mathsf{iter} \leftarrow 0$;
5: $(\mathsf{roundUp}, \mathsf{roundValue}) \leftarrow \mathsf{configRound}(\varepsilon)$
6: $t \leftarrow \mathsf{computeIter}(\varepsilon, \delta)$
7: **repeat**
8:     $\mathsf{iter} \leftarrow \mathsf{iter} + 1$;
9:     $\mathsf{nSols} \leftarrow \mathsf{ApproxMC6Core}(F, \mathsf{thresh}, \mathsf{roundUp}, \mathsf{roundValue})$;
10:     $\mathsf{AddToList}(C, \mathsf{nSols})$;
11: **until** $(\mathsf{iter} \geq t)$;
12: finalEstimate $\leftarrow \mathsf{FindMedian}(C)$;
13: **return** finalEstimate ;

in Algorithm 6 following Lemma 2. The iterator keeps increasing until the tight error bound is no more than δ. As we will show in Sect. 5.2, $\Pr[L]$ and $\Pr[U]$ depend on ε. In the loop of Lines 7–11, ApproxMC6Core repeatedly estimates $|\mathsf{sol}(\mathsf{F})|$. Each estimate nSols is stored in List C, and the median of C serves as the final estimate satisfying the (ε, δ)-guarantee.

Algorithm 4 shows the pseudo-code of ApproxMC6Core. A random hash function is chosen at Line 1 to partition $\mathsf{sol}(\mathsf{F})$ into *roughly equal* cells. A random hash value is chosen at Line 2 to randomly pick a cell for estimation. In Line 3, we search for a value m such that the cell picked from 2^m available cells is *small* enough to enumerate solutions one by one while providing a good estimate of $|\mathsf{sol}(\mathsf{F})|$. In Line 4, a bounded model counting is invoked to compute the size of the picked cell, i.e., $\mathsf{Cnt}_{\langle F,m\rangle}$. Finally, if roundUp equals 1, $\mathsf{Cnt}_{\langle F,m\rangle}$ is rounded up to roundValue at Line 6. Otherwise, roundUp equals 0, and $\mathsf{Cnt}_{\langle F,m\rangle}$ is rounded to roundValue at Line 8. Note that *rounding up* returns roundValue only if $\mathsf{Cnt}_{\langle F,m\rangle}$ is less than roundValue. However, in the case of *rounding*, roundValue is always returned no matter what value $\mathsf{Cnt}_{\langle F,m\rangle}$ is.

For large ε ($\varepsilon \geq 3$), ApproxMC6Core returns a value that is independent of the value returned by BoundedSAT in line 4 of Algorithm 4. However, observe the value depends on m returned by LogSATSearch [8], which in turn uses BoundedSAT to find the value of m; therefore, the algorithm's run is not independent of all the calls to BoundedSAT. The technical reason for correctness stems from the observation that for large values of ε, we can always find a value of m such that $2^m \times c$ (where c is a constant) is a $(1+\varepsilon)$-approximation of $|\mathsf{sol}(\mathsf{F})|$. An example, consider $n = 7$ and let $c = 1$, then a $(1+3)$-approximation of a number between 1 and 128 belongs to $[1, 2, 4, 8, 16, 32, 64, 128]$; therefore, returning an answer of the form $c \times 2^m$ suffices as long as we are able to search for the right value of m, which is accomplished by LogSATSearch. We could skip the final call to BoundedSAT in line 4 of ApproxMC6Core for large values of ε, but the actual computation of BoundedSAT comes with LogSATSearch.

Algorithm 4. $\mathsf{ApproxMC6Core}(F, \mathsf{thresh}, \mathsf{roundUp}, \mathsf{roundValue})$

1: Choose h at random from $\mathcal{H}(n, n)$;
2: Choose α at random from $\{0, 1\}^n$;
3: $m \leftarrow \mathsf{LogSATSearch}(F, h, \alpha, \mathsf{thresh})$;
4: $\mathsf{Cnt}_{\langle F,m\rangle} \leftarrow \mathsf{BoundedSAT}\left(F \wedge \left(h^{(m)}\right)^{-1}\left(\alpha^{(m)}\right), \mathsf{thresh}\right)$;
5: **if** $\mathsf{roundUp} = 1$ **then**
6: **return** $(2^m \times \max\{\mathsf{Cnt}_{\langle F,m\rangle}, \mathsf{roundValue}\})$;
7: **else**
8: **return** $(2^m \times \mathsf{roundValue})$;

Algorithm 5. $\mathsf{configRound}(\varepsilon)$

1: **if** $(\varepsilon < \sqrt{2} - 1)$ **then return** $(1, \frac{\sqrt{1+2\varepsilon}}{2}\mathsf{pivot})$;
2: **else if** $(\varepsilon < 1)$ **then return** $(1, \frac{\mathsf{pivot}}{\sqrt{2}})$;
3: **else if** $(\varepsilon < 3)$ **then return** $(1, \mathsf{pivot})$;
4: **else if** $(\varepsilon < 4\sqrt{2} - 1)$ **then return** $(0, \mathsf{pivot})$;
5: **else**
6: **return** $(0, \sqrt{2}\mathsf{pivot})$;

5.2 Repetition Reduction

We will now show that ApproxMC6Core allows us to obtain a smaller $\max\{\Pr[L], \Pr[U]\}$. Furthermore, we show the large gap between the error probability of ApproxMC6 and that of ApproxMC both analytically and visually.

The following lemma presents the upper bounds of $\Pr[L]$ and $\Pr[U]$ for ApproxMC6Core. Let $\mathsf{pivot} = 9.84\left(1 + \frac{1}{\varepsilon}\right)^2$ for simplicity.

Lemma 4. *The following bounds hold for* ApproxMC6*:*

$$Pr[L] \leq \begin{cases} 0.262 & \text{if } \varepsilon < \sqrt{2} - 1 \\ 0.157 & \text{if } \sqrt{2} - 1 \leq \varepsilon < 1 \\ 0.085 & \text{if } 1 \leq \varepsilon < 3 \\ 0.055 & \text{if } 3 \leq \varepsilon < 4\sqrt{2} - 1 \\ 0.023 & \text{if } \varepsilon \geq 4\sqrt{2} - 1 \end{cases}$$

$$Pr[U] \leq \begin{cases} 0.169 & \text{if } \varepsilon < 3 \\ 0.044 & \text{if } \varepsilon \geq 3 \end{cases}$$

The proof of Lemma 4 is deferred to Sect. 5.3. Observe that Lemma 4 influences the choices in the design of configRound (Algorithm 5). Recall that $\max\{\Pr[L], \Pr[U]\} \leq 0.36$ for ApproxMC (Appendix C), but Lemma 4 ensures $\max\{\Pr[L], \Pr[U]\} \leq 0.262$ for ApproxMC6. For $\varepsilon \geq 4\sqrt{2} - 1$, Lemma 4 even delivers $\max\{\Pr[L], \Pr[U]\} \leq 0.044$.

Algorithm 6. computeIter(ε, δ)

1: iter $\leftarrow 1$;
2: **while** (η(iter, $\lceil$iter$/2\rceil$, $\Pr_\varepsilon[L]$) $+ \eta$(iter, $\lceil$iter$/2\rceil$, $\Pr_\varepsilon[U]$) $> \delta$) **do**
3: iter $\leftarrow$ iter $+ 2$;
4: **return** iter;

The following theorem analytically presents the gap between the error probability of ApproxMC6 and that of ApproxMC[1].

Theorem 1. *For* $\sqrt{2}-1 \leq \varepsilon < 1$,

$$Pr[\mathsf{Error}_t] \in \begin{cases} \mathcal{O}\left(t^{-\frac{1}{2}} 0.75^t\right) & \textit{for } \mathsf{ApproxMC6} \\ \mathcal{O}\left(t^{-\frac{1}{2}} 0.96^t\right) & \textit{for } \mathsf{ApproxMC} \end{cases}$$

Proof. From Lemma 4, we obtain $p_{max} \leq 0.169$ for ApproxMC6. Applying Lemma 3, we have

$$\Pr[\mathsf{Error}_t] \in \mathcal{O}\left(t^{-\frac{1}{2}}\left(2\sqrt{0.169(1-0.169)}\right)^t\right) \subseteq \mathcal{O}\left(t^{-\frac{1}{2}} 0.75^t\right)$$

For ApproxMC, combining $p_{max} \leq 0.36$ (Appendix C) and Lemma 3, we obtain

$$\Pr[\mathsf{Error}_t] \in \mathcal{O}\left(t^{-\frac{1}{2}}\left(2\sqrt{0.36(1-0.36)}\right)^t\right) = \mathcal{O}\left(t^{-\frac{1}{2}} 0.96^t\right)$$

□

Figure 1 visualizes the large gap between the error probability of ApproxMC6 and that of ApproxMC. The x-axis represents the number of repetitions (t) in ApproxMC6 or ApproxMC. The y-axis represents the upper bound of error probability in the log scale. For example, as $t = 117$, ApproxMC guarantees that with a probability of 10^{-3}, the median over 117 estimates violates the PAC guarantee. However, ApproxMC6 allows a much smaller error probability that is at most 10^{-15} for $\sqrt{2}-1 \leq \varepsilon < 1$. The smaller error probability enables ApproxMC6 to repeat fewer repetitions while providing the same level of theoretical guarantee. For example, given $\delta = 0.001$ to ApproxMC, i.e., $y = 0.001$ in Fig. 1, ApproxMC requests 117 repetitions to obtain the given error probability. However, ApproxMC6 claims that 37 repetitions for $\varepsilon < \sqrt{2}-1$, 19 repetitions for $\sqrt{2}-1 \leq \varepsilon < 1$, 17 repetitions for $1 \leq \varepsilon < 3$, 7 repetitions for $3 \leq \varepsilon < 4\sqrt{2}-1$, and 5 repetitions for $\varepsilon \geq 4\sqrt{2}-1$ are sufficient to obtain the same level of error probability. Consequently, ApproxMC6 can obtain 3×, 6×, 7×, 17×, and 23× speedups, respectively, than ApproxMC.

[1] We state the result for the case $\sqrt{2}-1 \leq \varepsilon < 1$. A similar analysis can be applied to other cases, which leads to an even bigger gap between ApproxMC6 and ApproxMC.

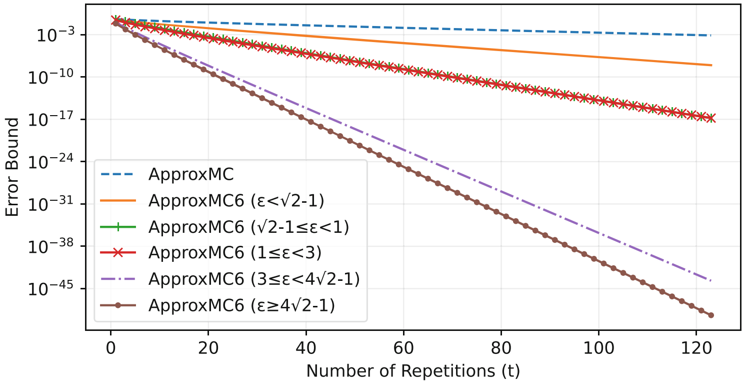

Fig. 1. Comparison of error bounds for ApproxMC6 and ApproxMC.

5.3 Proof of Lemma 4 for Case $\sqrt{2}-1 \leq \varepsilon < 1$

We provide full proof of Lemma 4 for case $\sqrt{2}-1 \leq \varepsilon < 1$. We defer the proof of other cases to Appendix D.

Let T_m denote the event $\left(\mathsf{Cnt}_{\langle F,m\rangle} < \mathsf{thresh}\right)$, and let L_m and U_m denote the events $\left(\mathsf{Cnt}_{\langle F,m\rangle} < \frac{\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right]}{1+\varepsilon}\right)$ and $\left(\mathsf{Cnt}_{\langle F,m\rangle} > \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right](1+\varepsilon)\right)$, respectively. To ease the proof, let U'_m denote $\left(\mathsf{Cnt}_{\langle F,m\rangle} > \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right](1+\frac{\varepsilon}{1+\varepsilon})\right)$, and thereby $U_m \subseteq U'_m$. Let $m^* = \lfloor \log_2 |\mathsf{sol(F)}| - \log_2(\mathsf{pivot}) + 1 \rfloor$ such that m^* is the smallest m satisfying $\frac{|\mathsf{sol(F)}|}{2^m}(1+\frac{\varepsilon}{1+\varepsilon}) \leq \mathsf{thresh} - 1$.

Let us first prove the lemmas used in the proof of Lemma 4.

Lemma 5. *For every* $0 < \beta < 1$, $\gamma > 1$, *and* $1 \leq m \leq n$, *the following holds:*

1. $Pr\left[\mathsf{Cnt}_{\langle F,m\rangle} \leq \beta \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right]\right] \leq \frac{1}{1+(1-\beta)^2\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right]}$
2. $Pr\left[\mathsf{Cnt}_{\langle F,m\rangle} \geq \gamma \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right]\right] \leq \frac{1}{1+(\gamma-1)^2\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right]}$

Proof. Statement 1 can be proved following the proof of Lemma 1 in [8]. For statement 2, we rewrite the left-hand side and apply Cantelli's inequality: $\Pr\left[\mathsf{Cnt}_{\langle F,m\rangle} - \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right] \geq (\gamma-1)\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right]\right] \leq \frac{\sigma^2\left[\mathsf{Cnt}_{\langle F,m\rangle}\right]}{\sigma^2\left[\mathsf{Cnt}_{\langle F,m\rangle}\right]+((\gamma-1)\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right])^2}$. Finally, applying Eq. 2 completes the proof. □

Lemma 6. *Given* $\sqrt{2}-1 \leq \varepsilon < 1$, *the following bounds hold:*

1. $Pr[T_{m^*-3}] \leq \frac{1}{62.5}$
2. $Pr[L_{m^*-2}] \leq \frac{1}{20.68}$
3. $Pr[L_{m^*-1}] \leq \frac{1}{10.84}$
4. $Pr[U'_{m^*}] \leq \frac{1}{5.92}$

Proof. Following the proof of Lemma 2 in [8], we can prove statements 1, 2, and 3. To prove statement 4, replacing γ with $(1+\frac{\varepsilon}{1+\varepsilon})$ in Lemma 5 and employing $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right] \geq \mathsf{pivot}/2$, we obtain $\Pr\left[U'_{m^*}\right] \leq \frac{1}{1+\left(\frac{\varepsilon}{1+\varepsilon}\right)^2 \mathsf{pivot}/2} \leq \frac{1}{5.92}$. □

Now we prove the upper bounds of $\Pr[L]$ and $\Pr[U]$ in Lemma 4 for $\sqrt{2}-1 \leq \varepsilon < 1$. The proof for other ε is deferred to Appendix D due to the page limit.

Lemma 4. *The following bounds hold for* ApproxMC6:

$$\Pr[L] \leq \begin{cases} 0.262 & \text{if } \varepsilon < \sqrt{2}-1 \\ 0.157 & \text{if } \sqrt{2}-1 \leq \varepsilon < 1 \\ 0.085 & \text{if } 1 \leq \varepsilon < 3 \\ 0.055 & \text{if } 3 \leq \varepsilon < 4\sqrt{2}-1 \\ 0.023 & \text{if } \varepsilon \geq 4\sqrt{2}-1 \end{cases}$$

$$\Pr[U] \leq \begin{cases} 0.169 & \text{if } \varepsilon < 3 \\ 0.044 & \text{if } \varepsilon \geq 3 \end{cases}$$

Proof. We prove the case of $\sqrt{2}-1 \leq \varepsilon < 1$. The proof for other ε is deferred to Appendix D. Let us first bound $\Pr[L]$. Following LogSATSearch in [8], we have

$$\Pr[L] = \left[\bigcup_{i \in \{1,\ldots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap L_i\right)\right] \tag{3}$$

Equation 3 can be simplified by three observations labeled $O1, O2$ and $O3$ below.

$O1$: $\forall i \leq m^*-3, T_i \subseteq T_{i+1}$. Therefore,

$$\bigcup_{i \in \{1,\ldots,m^*-3\}} (\overline{T_{i-1}} \cap T_i \cap L_i) \subseteq \bigcup_{i \in \{1,\ldots,m^*-3\}} T_i \subseteq T_{m^*-3}$$

$O2$:]For $i \in \{m^*-2, m^*-1\}$, we have

$$\bigcup_{i \in \{m^*-2,m^*-1\}} (\overline{T_{i-1}} \cap T_i \cap L_i) \subseteq L_{m^*-2} \cup L_{m^*-1}$$

$O3$: $\forall i \geq m^*$, since rounding $\mathsf{Cnt}_{\langle F,i\rangle}$ up to $\frac{\mathsf{pivot}}{\sqrt{2}}$ and $m^* \geq \log_2|\mathsf{sol(F)}| - \log_2(\mathsf{pivot})$, we have $2^i \times \mathsf{Cnt}_{\langle F,i\rangle} \geq 2^{m^*} \times \frac{\mathsf{pivot}}{\sqrt{2}} \geq \frac{|\mathsf{sol(F)}|}{\sqrt{2}} \geq \frac{|\mathsf{sol(F)}|}{1+\varepsilon}$. The last inequality follows from $\varepsilon \geq \sqrt{2}-1$. Then we have $\mathsf{Cnt}_{\langle F,i\rangle} \geq \frac{\mathsf{E}\left[\mathsf{Cnt}_{\langle F,i\rangle}\right]}{1+\varepsilon}$. Therefore, $L_i = \emptyset$ for $i \geq m^*$ and we have

$$\bigcup_{i \in \{m^*,\ldots,n\}} (\overline{T_{i-1}} \cap T_i \cap L_i) = \emptyset$$

Following the observations $O1$, $O2$, and $O3$, we simplify Eq. 3 and obtain

$$\Pr[L] \leq \Pr[T_{m^*-3}] + \Pr[L_{m^*-2}] + \Pr[L_{m^*-1}]$$

Employing Lemma 6 gives $\Pr[L] \leq 0.157$.

Now let us bound $\Pr[U]$. Similarly, following LogSATSearch in [8], we have

$$\Pr[U] = \left[\bigcup_{i \in \{1,\ldots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap U_i \right) \right] \tag{4}$$

We derive the following observations $O4$ and $O5$.

$O4$: $\forall i \leq m^* - 1$, since $m^* \leq \log_2 |\mathsf{sol(F)}| - \log_2(\mathsf{pivot}) + 1$, we have $2^i \times \mathsf{Cnt}_{\langle F,i \rangle} \leq 2^{m^*-1} \times \mathsf{thresh} \leq |\mathsf{sol(F)}| \left(1 + \frac{\varepsilon}{1+\varepsilon}\right)$. Then we obtain $\mathsf{Cnt}_{\langle F,i \rangle} \leq \mathsf{E}\left[\mathsf{Cnt}_{\langle F,i \rangle}\right] \left(1 + \frac{\varepsilon}{1+\varepsilon}\right)$. Therefore, $T_i \cap U_i' = \emptyset$ for $i \leq m^* - 1$ and we have

$$\bigcup_{i \in \{1,\ldots,m^*-1\}} \left(\overline{T_{i-1}} \cap T_i \cap U_i \right) \subseteq \bigcup_{i \in \{1,\ldots,m^*-1\}} \left(\overline{T_{i-1}} \cap T_i \cap U_i' \right) = \emptyset$$

$O5$: $\forall i \geq m^*$, $\overline{T_i}$ implies $\mathsf{Cnt}_{\langle F,i \rangle} > \mathsf{thresh}$, and then we have $2^i \times \mathsf{Cnt}_{\langle F,i \rangle} > 2^{m^*} \times \mathsf{thresh} \geq |\mathsf{sol(F)}| \left(1 + \frac{\varepsilon}{1+\varepsilon}\right)$. The second inequality follows from $m^* \geq \log_2 |\mathsf{sol(F)}| - \log_2(\mathsf{pivot})$. Then we obtain $\mathsf{Cnt}_{\langle F,i \rangle} > \mathsf{E}\left[\mathsf{Cnt}_{\langle F,i \rangle}\right] \left(1 + \frac{\varepsilon}{1+\varepsilon}\right)$. Therefore, $\overline{T_i} \subseteq U_i'$ for $i \geq m^*$. Since $\forall i, \overline{T_i} \subseteq \overline{T_{i-1}}$, we have

$$\begin{aligned}
\bigcup_{i \in \{m^*,\ldots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap U_i \right) &\subseteq \bigcup_{i \in \{m^*+1,\ldots,n\}} \overline{T_{i-1}} \cup \left(\overline{T_{m^*-1}} \cap T_{m^*} \cap U_{m^*} \right) \\
&\subseteq \overline{T_{m^*}} \cup \left(\overline{T_{m^*-1}} \cap T_{m^*} \cap U_{m^*} \right) \\
&\subseteq \overline{T_{m^*}} \cup U_{m^*} \\
&\subseteq U_{m^*}'
\end{aligned} \tag{5}$$

Remark that for $\sqrt{2} - 1 \leq \varepsilon < 1$, we round $\mathsf{Cnt}_{\langle F,m^* \rangle}$ up to $\frac{\mathsf{pivot}}{\sqrt{2}}$, and we have $2^{m^*} \times \frac{\mathsf{pivot}}{\sqrt{2}} \leq |\mathsf{sol(F)}|(1+\varepsilon)$, which means *rounding* doesn't affect the event U_{m^*}; therefore, Inequality 5 still holds.

Following the observations $O4$ and $O5$, we simplify Eq. 4 and obtain

$$\Pr[U] \leq \Pr[U_{m^*}']$$

Employing Lemma 6 gives $\Pr[U] \leq 0.169$. □

The breakpoints in ε of Lemma 4 arise from how we use rounding to lower the error probability for events L and U. Rounding up counts can lower $\Pr[L]$ but may increase $\Pr[U]$. Therefore, we want to round up counts to a value that doesn't affect the event U. Take $\sqrt{2} - 1 \leq \varepsilon < 1$ as an example; we round up the count to a value such that L_{m^*} becomes an empty event with zero probability while U_{m^*} remains unchanged. To make L_{m^*} empty, we have

$$2^{m^*} \times \mathsf{roundValue} \geq 2^{m^*} \times \frac{1}{1+\varepsilon}\mathsf{pivot} \geq \frac{1}{1+\varepsilon}|\mathsf{sol(F)}| \tag{6}$$

where the last inequality follows from $m^* \geq \log_2 |\mathsf{sol(F)}| - \log_2(\mathsf{pivot})$. To maintain U_{m^*} unchanged, we obtain

$$2^{m^*} \times \mathsf{roundValue} \leq 2^{m^*} \times \frac{1+\varepsilon}{2}\mathsf{pivot} \leq (1+\varepsilon)|\mathsf{sol(F)}| \tag{7}$$

where the last inequality follows from $m^* \leq \log_2 |\mathsf{sol(F)}| - \log_2(\mathsf{pivot}) + 1$. Combining Eqs. 6 and 7 together, we obtain

$$2^{m^*} \times \frac{1}{1+\varepsilon}\mathsf{pivot} \leq 2^{m^*} \times \frac{1+\varepsilon}{2}\mathsf{pivot}$$

which gives us $\varepsilon \geq \sqrt{2} - 1$. Similarly, we can derive other breakpoints.

6 Experimental Evaluation

It is perhaps worth highlighting that both ApproxMCCore and ApproxMC6Core invoke the underlying SAT solver on identical queries; the only difference between ApproxMC6 and ApproxMC lies in what estimate to return and how often ApproxMCCore and ApproxMC6Core are invoked. From this viewpoint, one would expect that theoretical improvements would also lead to improved runtime performance. To provide further evidence, we perform extensive empirical evaluation and compare ApproxMC6's performance against the current state-of-the-art model counter, ApproxMC [22]. We use Arjun as a pre-processing tool. We used the latest version of ApproxMC, called ApproxMC4; an entry based on ApproxMC4 won the Model Counting Competition 2022.

Previous comparisons of ApproxMC have been performed on a set of 1896 instances, but the latest version of ApproxMC is able to solve almost all the instances when these instances are pre-processed by Arjun. Therefore, we sought to construct a new comprehensive set of 1890 instances derived from various sources, including Model Counting Competitions 2020–2022 [12,15,16], program synthesis [1], quantitative control improvisation [13], quantification of software properties [26], and adaptive chosen ciphertext attacks [3]. As noted earlier, our technique extends to projected model counting, and our benchmark suite indeed comprises 772 projected model counting instances.

Experiments were conducted on a high-performance computer cluster, with each node consisting of 2xE5-2690v3 CPUs featuring 2×12 real cores and 96GB of RAM. For each instance, a counter was run on a single core, with a time limit of 5000 s and a memory limit of 4 GB. To compare runtime performance, we use the PAR-2 score, a standard metric in the SAT community. Each instance is assigned a score that is the number of seconds it takes the corresponding tool to complete execution successfully. In the event of a timeout or memory out, the score is the doubled time limit in seconds. The PAR-2 score is then calculated as the average of all the instance scores. We also report the speedup of ApproxMC6 over ApproxMC4, calculated as the ratio of the runtime of ApproxMC4 to that of ApproxMC6 on instances solved by both counters. We set δ to 0.001 and ε to 0.8.

Specifically, we aim to address the following research questions:

RQ 1. How does the runtime performance of ApproxMC6 compare to that of ApproxMC4?

RQ 2. How does the accuracy of the counts computed by ApproxMC6 compare to that of the exact count?

Summary. In summary, ApproxMC6 consistently outperforms ApproxMC4. Specifically, it solved 204 additional instances and reduced the PAR-2 score by 1063 s in comparison to ApproxMC4. The average speedup of ApproxMC6 over ApproxMC4 was 4.68. In addition, ApproxMC6 provided a high-quality approximation with an average observed error of 0.1, much smaller than the theoretical error tolerance of 0.8.

6.1 RQ1. Overall Performance

Figure 2 compares the counting time of ApproxMC6 and ApproxMC4. The x-axis represents the index of the instances, sorted in ascending order of runtime, and the y-axis represents the runtime for each instance. A point (x, y) indicates that a counter can solve x instances within y seconds. Thus, for a given time limit y, a counter whose curve is on the right has solved more instances than a counter on the left. It can be seen in the figure that ApproxMC6 consistently outperforms ApproxMC4. In total, ApproxMC6 solved 204 more instances than ApproxMC4.

Table 1 provides a detailed comparison between ApproxMC6 and ApproxMC4. The first column lists three measures of interest: the number of solved instances, the PAR-2 score, and the speedup of ApproxMC6 over ApproxMC4. The second and third columns show the results for ApproxMC4 and ApproxMC6, respectively. The second column indicates that ApproxMC4 solved 998 of the 1890 instances and achieved a PAR-2 score of 4934. The third column shows that ApproxMC6 solved 1202 instances and achieved a PAR-2 score of 3871. In comparison, ApproxMC6 solved 204 more instances and reduced the PAR-2 score by 1063 s in comparison to ApproxMC4. The geometric mean of the speedup for ApproxMC6 over ApproxMC4 is 4.68. This speedup was calculated only for instances solved by both counters.

6.2 RQ2. Approximation Quality

We used the state-of-the-art probabilistic exact model counter Ganak to compute the exact model count and compare it to the results of ApproxMC6. We collected statistics on instances solved by both Ganak and ApproxMC6. Figure 3 presents results for a subset of instances. The x-axis represents the index of instances

Table 1. The number of solved instances and PAR-2 score for ApproxMC6 versus ApproxMC4 on 1890 instances. The geometric mean of the speedup of ApproxMC6 over ApproxMC4 is also reported.

	ApproxMC4	ApproxMC6
# Solved	998	1202
PAR-2 score	4934	3871
Speedup	—	4.68

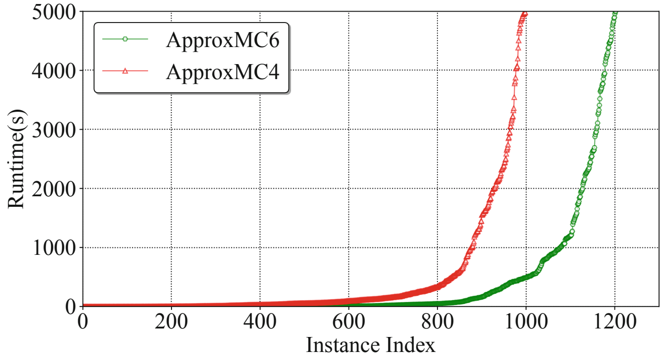

Fig. 2. Comparison of counting times for ApproxMC6 and ApproxMC4.

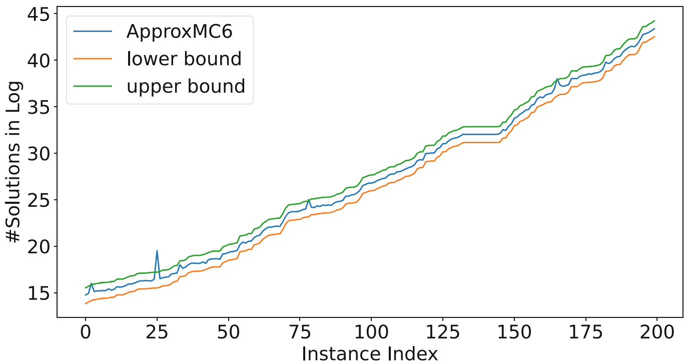

Fig. 3. Comparison of approximate counts from ApproxMC6 to exact counts from Ganak.

sorted in ascending order by the number of solutions, and the y-axis represents the number of solutions in a log scale. Theoretically, the approximate count from ApproxMC6 should be within the range of $|\mathsf{sol}(\mathsf{F})| \cdot 1.8$ and $|\mathsf{sol}(\mathsf{F})|/1.8$ with probability 0.999, where $|\mathsf{sol}(\mathsf{F})|$ denotes the exact count returned by Ganak. The range is indicated by the upper and lower bounds, represented by the curves $y = |\mathsf{sol}(\mathsf{F})| \cdot 1.8$ and $y = |\mathsf{sol}(\mathsf{F})|/1.8$, respectively. Figure 3 shows that the approximate counts from ApproxMC6 fall within the expected range $[|\mathsf{sol(F)}|/1.8, |\mathsf{sol(F)}| \cdot 1.8]$ for all instances except for four points slightly above the upper bound. These four outliers are due to a bug in the preprocessor Arjun that probably depends on the version of the C++ compiler and will be fixed in the future. We also calculated the observed error, which is the mean relative difference between the approximate and exact counts in our experiments, i.e., $\max\{\mathsf{finalEstimate}/|\mathsf{sol(F)}| - 1, |\mathsf{sol(F)}|/\mathsf{finalEstimate} - 1\}$. The overall observed error was 0.1, which is significantly smaller than the theoretical error tolerance of 0.8.

7 Conclusion

In this paper, we addressed the scalability challenges faced by ApproxMC in the smaller δ range. To this end, we proposed a *rounding*-based algorithm, ApproxMC6, which reduces the number of estimations required by 84% while providing the same (ε, δ)-guarantees. Our empirical evaluation on 1890 instances shows that ApproxMC6 solved 204 more instances and achieved a reduction in PAR-2 score of 1063 s. Furthermore, ApproxMC6 achieved a 4× speedup over ApproxMC on the instances both ApproxMC6 and ApproxMC could solve.

Acknowledgements. This work was supported in part by National Research Foundation Singapore under its NRF Fellowship Programme [NRF-NRFFAI1-2019-0004], Ministry of Education Singapore Tier 2 Grant [MOE-T2EP20121-0011], and Ministry of Education Singapore Tier 1 Grant [R-252-000-B59-114]. The computational work for this article was performed on resources of the National Supercomputing Centre, Singapore https://www.nscc.sg. We are thankful to Yash Pote for the insightful early discussions that helped shape the idea. We are grateful to Tim van Bremen for his detailed feedback on the early drafts of the paper. We sincerely appreciate the anonymous reviewers for their constructive comments to enhance this paper.

A Proof of Proposition 1

Proof. For $\forall y \in \{0,1\}^n, \alpha^{(m)} \in \{0,1\}^m$, let $\gamma_{y,\alpha^{(m)}}$ be an indicator variable that is 1 when $h^{(m)}(y) = \alpha^{(m)}$. According to the definition of strongly 2-universal function, we obtain $\forall x, y \in \{0,1\}^n, \mathsf{E}\left[\gamma_{y,\alpha^{(m)}}\right] = \frac{1}{2^m}$ and $\mathsf{E}\left[\gamma_{x,\alpha^{(m)}} \cdot \gamma_{y,\alpha^{(m)}}\right] = \frac{1}{2^{2m}}$. To prove Eq. 1, we obtain

$$\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right] = \mathsf{E}\left[\sum_{y \in \mathsf{sol(F)}} \gamma_{y,\alpha^{(m)}}\right] = \sum_{y \in \mathsf{sol(F)}} \mathsf{E}\left[\gamma_{y,\alpha^{(m)}}\right] = \frac{|\mathsf{sol(F)}|}{2^m}$$

To prove Eq. 2, we derive

$$\begin{aligned}
\mathsf{E}\left[\mathsf{Cnt}^2_{\langle F,m\rangle}\right] &= \mathsf{E}\left[\sum_{y\in\mathsf{sol(F)}} \gamma^2_{y,\alpha^{(m)}} + \sum_{x\neq y\in\mathsf{sol(F)}} \gamma_{x,\alpha^{(m)}}\cdot\gamma_{y,\alpha^{(m)}}\right] \\
&= \mathsf{E}\left[\sum_{y\in\mathsf{sol(F)}} \gamma_{y,\alpha^{(m)}}\right] + \sum_{x\neq y\in\mathsf{sol(F)}} \mathsf{E}\left[\gamma_{x,\alpha^{(m)}}\cdot\gamma_{y,\alpha^{(m)}}\right] \\
&= \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right] + \frac{|\mathsf{sol(F)}|(|\mathsf{sol(F)}|-1)}{2^{2m}}
\end{aligned}$$

Then, we obtain

$$\begin{aligned}
\sigma^2\left[\mathsf{Cnt}_{\langle F,m\rangle}\right] &= \mathsf{E}\left[\mathsf{Cnt}^2_{\langle F,m\rangle}\right] - \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right]^2 \\
&= \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right] + \frac{|\mathsf{sol(F)}|(|\mathsf{sol(F)}|-1)}{2^{2m}} - \left(\frac{|\mathsf{sol(F)}|}{2^m}\right)^2 \\
&= \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right] - \frac{|\mathsf{sol(F)}|}{2^{2m}} \\
&\leq \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m\rangle}\right]
\end{aligned}$$

□

B Weakness of Proposition 3

The following proposition states that Proposition 3 provides a loose upper bound for $\Pr\left[\mathsf{Error}_t\right]$.

Proposition 4. *Assuming t is odd, we have:*

$$Pr\left[\mathsf{Error}_t\right] < \eta(t, \lceil t/2\rceil, Pr\left[L\cup U\right])$$

Proof. We will now construct a case counted by $\eta(t, \lceil t/2\rceil, \Pr\left[L\cup U\right])$ but not contained within the event Error_t. Let I_i^L be an indicator variable that is 1 when $\mathsf{ApproxMCCore}$ returns a nSols less than $\frac{|\mathsf{sol(F)}|}{1+\varepsilon}$, indicating the occurrence of event L in the i-th repetition. Let I_i^U be an indicator variable that is 1 when $\mathsf{ApproxMCCore}$ returns a nSols greater than $(1+\varepsilon)|\mathsf{sol(F)}|$, indicating the occurrence of event U in the i-th repetition. Consider a scenario where $I_i^L = 1$ for $i = 1, 2, ..., \left\lceil \frac{t}{4}\right\rceil$, $I_j^U = 1$ for $j = \left\lceil\frac{t}{4}\right\rceil + 1, ..., \left\lceil\frac{t}{2}\right\rceil$, and $I_k^L = I_k^U = 0$ for $k > \left\lceil\frac{t}{2}\right\rceil$. $\eta(t, \lceil t/2\rceil, \Pr\left[L\cup U\right])$ represents $\sum_{i=1}^t (I_i^L \vee I_i^U) \geq \left\lceil\frac{t}{2}\right\rceil$. We can see that this case is included in $\sum_{i=1}^t (I_i^L \vee I_i^U) \geq \left\lceil\frac{t}{2}\right\rceil$ and therefore counted by $\eta(t, \lceil t/2\rceil, \Pr\left[L\cup U\right])$ since there are $\left\lceil\frac{t}{2}\right\rceil$ estimates outside the PAC range. However, this case means that $\left\lceil\frac{t}{4}\right\rceil$ estimates fall within the range less than $\frac{|\mathsf{sol(F)}|}{1+\varepsilon}$ and $\left\lceil\frac{t}{2}\right\rceil - \left\lceil\frac{t}{4}\right\rceil$ estimates fall within the range greater than $(1+\varepsilon)|\mathsf{sol(F)}|$, while the remaining $\left\lfloor\frac{t}{2}\right\rfloor$ estimates correctly fall within the range $\left[\frac{|\mathsf{sol(F)}|}{1+\varepsilon}, (1+\varepsilon)|\mathsf{sol(F)}|\right]$.

Therefore, after sorting all the estimates, ApproxMC6 returns a correct estimate since the median falls within the PAC range $\left[\frac{|\mathsf{sol}(\mathsf{F})|}{1+\varepsilon}, (1+\varepsilon)|\mathsf{sol}(\mathsf{F})|\right]$. In other words, this case is out of the event Error_t. In conclusion, there is a scenario that is out of the event Error_t, undesirably included in expression $\sum_{i=1}^{t}(I_i^L \vee I_i^U) \geq \lceil \frac{t}{2} \rceil$ and counted by $\eta(t, \lceil t/2 \rceil, \Pr[L \cup U])$, which means $\Pr[\mathsf{Error}_t]$ is strictly less than $\eta(t, \lceil t/2 \rceil, \Pr[L \cup U])$. □

C Proof of $p_{max} \leq 0.36$ for ApproxMC

Proof. We prove the case of $\sqrt{2}-1 \leq \varepsilon < 1$. Similarly to the proof in Sect. 5.3, we aim to bound $\Pr[L]$ by the following equation:

$$\Pr[L] = \left[\bigcup_{i \in \{1,\ldots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap L_i\right)\right] \qquad (3 \text{ revisited})$$

which can be simplified by three observations labeled $O1, O2$ and $O3$ below.

$O1: \forall i \leq m^* - 3, T_i \subseteq T_{i+1}$. Therefore,

$$\bigcup_{i \in \{1,\ldots,m^*-3\}} (\overline{T_{i-1}} \cap T_i \cap L_i) \subseteq \bigcup_{i \in \{1,\ldots,m^*-3\}} T_i \subseteq T_{m^*-3}$$

$O2$: For $i \in \{m^*-2, m^*-1\}$, we have

$$\bigcup_{i \in \{m^*-2,m^*-1\}} (\overline{T_{i-1}} \cap T_i \cap L_i) \subseteq L_{m^*-2} \cup L_{m^*-1}$$

$O3$: $\forall i \geq m^*$, $\overline{T_i}$ implies $\mathsf{Cnt}_{\langle F,i \rangle} > \mathsf{thresh}$ and then we have $2^i \times \mathsf{Cnt}_{\langle F,i \rangle} > 2^{m^*} \times \mathsf{thresh} \geq |\mathsf{sol}(\mathsf{F})| \left(1 + \frac{\varepsilon}{1+\varepsilon}\right)$. The second inequality follows from $m^* \geq \log_2 |\mathsf{sol}(\mathsf{F})| - \log_2(\mathsf{pivot})$. Then we obtain $\left(\mathsf{Cnt}_{\langle F,i \rangle} > \mathsf{E}\left[\mathsf{Cnt}_{\langle F,i \rangle}\right]\left(1 + \frac{\varepsilon}{1+\varepsilon}\right)\right)$. Therefore, $\overline{T_i} \subseteq U_i'$ for $i \geq m^*$. Since $\forall i, \overline{T_i} \subseteq \overline{T_{i-1}}$, we have

$$\begin{aligned}
\bigcup_{i \in \{m^*,\ldots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap L_i\right) &\subseteq \bigcup_{i \in \{m^*+1,\ldots,n\}} \overline{T_{i-1}} \cup (\overline{T_{m^*-1}} \cap T_{m^*} \cap L_{m^*}) \\
&\subseteq \overline{T_{m^*}} \cup (\overline{T_{m^*-1}} \cap T_{m^*} \cap L_{m^*}) \\
&\subseteq \overline{T_{m^*}} \cup L_{m^*} \\
&\subseteq U'_{m^*} \cup L_{m^*}
\end{aligned}$$

Following the observations $O1, O2$ and $O3$, we simplify Eq. 3 and obtain

$$\Pr[L] \leq \Pr[T_{m^*-3}] + \Pr[L_{m^*-2}] + \Pr[L_{m^*-1}] + \Pr[U'_{m^*} \cup L_{m^*}]$$

Employing Lemma 2 in [8] gives $\Pr[L] \leq 0.36$. Note that U in [8] represents U' of our definition.

Then, following the $O4$ and $O5$ in Sect. 5.3, we obtain

$$\Pr[U] \leq \Pr[U'_{m^*}]$$

Employing Lemma 6 gives $\Pr[U] \leq 0.169$. As a result, $p_{max} \leq 0.36$. □

D Proof of Lemma 4

We restate the lemma below and prove the statements section by section. The proof for $\sqrt{2}-1 \le \varepsilon < 1$ has been shown in Sect. 5.3.

Lemma 4. *The following bounds hold for* ApproxMC6:

$$\Pr[L] \le \begin{cases} 0.262 & \text{if } \varepsilon < \sqrt{2}-1 \\ 0.157 & \text{if } \sqrt{2}-1 \le \varepsilon < 1 \\ 0.085 & \text{if } 1 \le \varepsilon < 3 \\ 0.055 & \text{if } 3 \le \varepsilon < 4\sqrt{2}-1 \\ 0.023 & \text{if } \varepsilon \ge 4\sqrt{2}-1 \end{cases}$$

$$\Pr[U] \le \begin{cases} 0.169 & \text{if } \varepsilon < 3 \\ 0.044 & \text{if } \varepsilon \ge 3 \end{cases}$$

D.1 Proof of $\Pr[L] \le 0.262$ for $\varepsilon < \sqrt{2}-1$

We first consider two cases: $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right] < \frac{1+\varepsilon}{2}\mathsf{thresh}$ and $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right] \ge \frac{1+\varepsilon}{2}\mathsf{thresh}$, and then merge the results to complete the proof.

Case 1: $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right] < \frac{1+\varepsilon}{2}\,\mathsf{thresh}$

Lemma 7. *Given $\varepsilon < \sqrt{2}-1$, the following bounds hold:*

1. $Pr[T_{m^*-2}] \le \frac{1}{29.67}$
2. $Pr[L_{m^*-1}] \le \frac{1}{10.84}$

Proof. Let's first prove the statement 1. For $\varepsilon < \sqrt{2}-1$, we have $\mathsf{thresh} < (2-\frac{\sqrt{2}}{2})\mathsf{pivot}$ and $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-2\rangle}\right] \ge 2\mathsf{pivot}$. Therefore, $\Pr[T_{m^*-2}] \le \Pr\left[\mathsf{Cnt}_{\langle F,m^*-2\rangle} \le (1-\frac{\sqrt{2}}{4})\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-2\rangle}\right]\right]$. Finally, employing Lemma 5 with $\beta = 1-\frac{\sqrt{2}}{4}$, we obtain $\Pr[T_{m^*-2}] \le \frac{1}{1+(\frac{\sqrt{2}}{4})^2\cdot 2\mathsf{pivot}} \le \frac{1}{1+(\frac{\sqrt{2}}{4})^2\cdot 2\cdot 9.84\cdot(1+\frac{1}{\sqrt{2}-1})^2} \le \frac{1}{29.67}$. To prove the statement 2, we employ Lemma 5 with $\beta = \frac{1}{1+\varepsilon}$ and $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-1\rangle}\right] \ge \mathsf{pivot}$ to obtain $\Pr[L_{m^*-1}] \le \frac{1}{1+(1-\frac{1}{1+\varepsilon})^2\cdot\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-1\rangle}\right]} \le \frac{1}{1+(1-\frac{1}{1+\varepsilon})^2\cdot 9.84\cdot(1+\frac{1}{\varepsilon})^2} = \frac{1}{10.84}$. □

Then, we prove that $\Pr[L] \le 0.126$ for $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right] < \frac{1+\varepsilon}{2}\mathsf{thresh}$.

Proof. We aim to bound $\Pr[L]$ by the following equation:

$$\Pr[L] = \left[\bigcup_{i\in\{1,\ldots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap L_i\right)\right] \qquad (3 \text{ revisited})$$

which can be simplified by the three observations labeled $O1, O2$ and $O3$ below.

$O1: \forall i \leq m^*-2, T_i \subseteq T_{i+1}$. Therefore,

$$\bigcup_{i\in\{1,\ldots,m^*-2\}} (\overline{T_{i-1}} \cap T_i \cap L_i) \subseteq \bigcup_{i\in\{1,\ldots,m^*-2\}} T_i \subseteq T_{m^*-2}$$

$O2$: For $i = m^*-1$, we have

$$\overline{T_{m^*-2}} \cap T_{m^*-1} \cap L_{m^*-1} \subseteq L_{m^*-1}$$

$O3$: $\forall i \geq m^*$, since rounding $\mathsf{Cnt}_{\langle F,i\rangle}$ up to $\frac{\sqrt{1+2\varepsilon}}{2}\mathsf{pivot}$, we have $\mathsf{Cnt}_{\langle F,i\rangle} \geq \frac{\sqrt{1+2\varepsilon}}{2}\mathsf{pivot} \geq \frac{\mathsf{thresh}}{2} > \frac{\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right]}{1+\varepsilon} \geq \frac{\mathsf{E}\left[\mathsf{Cnt}_{\langle F,i\rangle}\right]}{1+\varepsilon}$. The second last inequality follows from $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right] < \frac{1+\varepsilon}{2}\mathsf{thresh}$. Therefore, $L_i = \emptyset$ for $i \geq m^*$ and we have

$$\bigcup_{i\in\{m^*,\ldots,n\}} (\overline{T_{i-1}} \cap T_i \cap L_i) = \emptyset$$

Following the observations $O1, O2$ and $O3$, we simplify Eq. 3 and obtain

$$\Pr[L] \leq \Pr[T_{m^*-2}] + \Pr[L_{m^*-1}]$$

Employing Lemma 7 gives $\Pr[L] \leq 0.126$. □

Case 2: $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right] \geq \frac{1+\varepsilon}{2}$ thresh

Lemma 8. *Given* $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right] \geq \frac{1+\varepsilon}{2}\mathsf{thresh}$, *the following bounds hold:*

1. $Pr[T_{m^*-1}] \leq \frac{1}{10.84}$
2. $Pr[L_{m^*}] \leq \frac{1}{5.92}$

Proof. Let's first prove the statement 1. From $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right] \geq \frac{1+\varepsilon}{2}\mathsf{thresh}$, we can derive $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-1\rangle}\right] \geq (1+\varepsilon)\mathsf{thresh}$. Therefore, $\Pr[T_{m^*-1}] \leq \Pr\left[\mathsf{Cnt}_{\langle F,m^*-1\rangle} \leq \frac{1}{1+\varepsilon}\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-1\rangle}\right]\right]$. Finally, employing Lemma 5 with $\beta = \frac{1}{1+\varepsilon}$, we obtain $\Pr[T_{m^*-1}] \leq \frac{1}{1+(1-\frac{1}{1+\varepsilon})^2\cdot\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-1\rangle}\right]} \leq \frac{1}{1+(1-\frac{1}{1+\varepsilon})^2\cdot(1+\varepsilon)\mathsf{thresh}} = \frac{1}{1+9.84(1+2\varepsilon)} \leq \frac{1}{10.84}$. To prove the statement 2, we employ Lemma 5 with $\beta = \frac{1}{1+\varepsilon}$ and $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right] \geq \frac{1+\varepsilon}{2}\mathsf{thresh}$ to obtain $\Pr[L_{m^*}] \leq \frac{1}{1+(1-\frac{1}{1+\varepsilon})^2\cdot\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right]} \leq \frac{1}{1+(1-\frac{1}{1+\varepsilon})^2\cdot\frac{1+\varepsilon}{2}\mathsf{thresh}} = \frac{1}{1+4.92(1+2\varepsilon)} \leq \frac{1}{5.92}$. □

Then, we prove that $\Pr[L] \leq 0.262$ for $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*\rangle}\right] \geq \frac{1+\varepsilon}{2}\mathsf{thresh}$.

Proof. We aim to bound $\Pr[L]$ by the following equation:

$$\Pr[L] = \left[\bigcup_{i\in\{1,\ldots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap L_i\right)\right] \qquad (3 \text{ revisited})$$

which can be simplified by the three observations labeled $O1, O2$ and $O3$ below.

$O1 : \forall i \leq m^* - 1, T_i \subseteq T_{i+1}$. Therefore,

$$\bigcup_{i \in \{1,\ldots,m^*-1\}} (\overline{T_{i-1}} \cap T_i \cap L_i) \subseteq \bigcup_{i \in \{1,\ldots,m^*-1\}} T_i \subseteq T_{m^*-1}$$

$O2$: For $i = m^*$, we have

$$\overline{T_{m^*-1}} \cap T_{m^*} \cap L_{m^*} \subseteq L_{m^*}$$

$O3 : \forall i \geq m^* + 1$, since rounding $\mathsf{Cnt}_{\langle F,i \rangle}$ up to $\frac{\sqrt{1+2\varepsilon}}{2}\mathsf{pivot}$ and $m^* \geq \log_2 |\mathsf{sol(F)}| - \log_2(\mathsf{pivot})$, we have $2^i \times \mathsf{Cnt}_{\langle F,i \rangle} \geq 2^{m^*+1} \times \frac{\sqrt{1+2\varepsilon}}{2}\mathsf{pivot} \geq \sqrt{1+2\varepsilon}|\mathsf{sol(F)}| \geq \frac{|\mathsf{sol(F)}|}{1+\varepsilon}$. Then we have $\left(\mathsf{Cnt}_{\langle F,i \rangle} \geq \frac{\mathsf{E}\left[\mathsf{Cnt}_{\langle F,i \rangle}\right]}{1+\varepsilon}\right)$. Therefore, $L_i = \emptyset$ for $i \geq m^* + 1$ and we have

$$\bigcup_{i \in \{m^*+1,\ldots,n\}} (\overline{T_{i-1}} \cap T_i \cap L_i) = \emptyset$$

Following the observations $O1, O2$ and $O3$, we simplify Eq. 3 and obtain

$$\Pr[L] \leq \Pr[T_{m^*-1}] + \Pr[L_{m^*}]$$

Employing Lemma 8 gives $\Pr[L] \leq 0.262$. □

Combining the Case 1 and 2, we obtain $\Pr[L] \leq \max\{0.126, 0.262\} = 0.262$. Therefore, we prove the statement for $\mathsf{ApproxMC6}$: $\Pr[L] \leq 0.262$ for $\varepsilon < \sqrt{2}-1$.

D.2 Proof of $\Pr[L] \leq 0.085$ for $1 \leq \varepsilon < 3$

Lemma 9. *Given $1 \leq \varepsilon < 3$, the following bounds hold:*

1. $Pr[T_{m^*-4}] \leq \frac{1}{86.41}$
2. $Pr[L_{m^*-3}] \leq \frac{1}{40.36}$
3. $Pr[L_{m^*-2}] \leq \frac{1}{20.68}$

Proof. Let's first prove the statement 1. For $\varepsilon < 3$, we have $\mathsf{thresh} < \frac{7}{4}\mathsf{pivot}$ and $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-4 \rangle}\right] \geq 8\mathsf{pivot}$. Therefore, $\Pr[T_{m^*-4}] \leq \Pr\left[\mathsf{Cnt}_{\langle F,m^*-4 \rangle} \leq \frac{7}{32}\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-4 \rangle}\right]\right]$. Finally, employing Lemma 5 with $\beta = \frac{7}{32}$, we obtain $\Pr[T_{m^*-4}] \leq \frac{1}{1+(1-\frac{7}{32})^2 \cdot 8\mathsf{pivot}} \leq \frac{1}{1+(1-\frac{7}{32})^2 \cdot 8 \cdot 9.84 \cdot (1+\frac{1}{3})^2} \leq \frac{1}{86.41}$. To prove the statement 2, we employ Lemma 5 with $\beta = \frac{1}{1+\varepsilon}$ and $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-3 \rangle}\right] \geq 4\mathsf{pivot}$ to obtain $\Pr[L_{m^*-3}] \leq \frac{1}{1+(1-\frac{1}{1+\varepsilon})^2 \cdot \mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-3 \rangle}\right]} \leq \frac{1}{1+(1-\frac{1}{1+\varepsilon})^2 \cdot 4 \cdot 9.84 \cdot (1+\frac{1}{\varepsilon})^2} = \frac{1}{40.36}$. Following the proof of Lemma 2 in [8] we can prove the statement 3. □

Now let us prove the statement for $\mathsf{ApproxMC6}$: $\Pr[L] \leq 0.085$ for $1 \leq \varepsilon < 3$.

Proof. We aim to bound $\Pr[L]$ by the following equation:

$$\Pr[L] = \left[\bigcup_{i\in\{1,\ldots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap L_i\right)\right] \qquad (3 \text{ revisited})$$

which can be simplified by the three observations labeled $O1, O2$ and $O3$ below.

$O1 : \forall i \leq m^* - 4, T_i \subseteq T_{i+1}$. Therefore,

$$\bigcup_{i\in\{1,\ldots,m^*-4\}} (\overline{T_{i-1}} \cap T_i \cap L_i) \subseteq \bigcup_{i\in\{1,\ldots,m^*-4\}} T_i \subseteq T_{m^*-4}$$

$O2$: For $i \in \{m^* - 3, m^* - 2\}$, we have

$$\bigcup_{i\in\{m^*-3,m^*-2\}} (\overline{T_{i-1}} \cap T_i \cap L_i) \subseteq L_{m^*-3} \cup L_{m^*-2}$$

$O3 : \forall i \geq m^* - 1$, since rounding $\mathsf{Cnt}_{\langle F,i\rangle}$ up to pivot and $m^* \geq \log_2 |\mathsf{sol(F)}| - \log_2(\mathsf{pivot})$, we have $2^i \times \mathsf{Cnt}_{\langle F,i\rangle} \geq 2^{m^*-1} \times \mathsf{pivot} \geq \frac{|\mathsf{sol(F)}|}{2} \geq \frac{|\mathsf{sol(F)}|}{1+\varepsilon}$. The last inequality follows from $\varepsilon \geq 1$. Then we have $\left(\mathsf{Cnt}_{\langle F,i\rangle} \geq \frac{\mathsf{E}\left[\mathsf{Cnt}_{\langle F,i\rangle}\right]}{1+\varepsilon}\right)$. Therefore, $L_i = \emptyset$ for $i \geq m^* - 1$ and we have

$$\bigcup_{i\in\{m^*-1,\ldots,n\}} (\overline{T_{i-1}} \cap T_i \cap L_i) = \emptyset$$

Following the observations $O1, O2$ and $O3$, we simplify Eq. 3 and obtain

$$\Pr[L] \leq \Pr[T_{m^*-4}] + \Pr[L_{m^*-3}] + \Pr[L_{m^*-2}]$$

Employing Lemma 9 gives $\Pr[L] \leq 0.085$. □

D.3 Proof of $\Pr[L] \leq 0.055$ for $3 \leq \varepsilon < 4\sqrt{2} - 1$

Lemma 10. *Given $3 \leq \varepsilon < 4\sqrt{2} - 1$, the following bound hold:*

$$Pr[T_{m^*-3}] \leq \frac{1}{18.19}$$

Proof. For $\varepsilon < 4\sqrt{2} - 1$, we have $\mathsf{thresh} < (2 - \frac{\sqrt{2}}{8})\mathsf{pivot}$ and $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-3\rangle}\right] \geq 4\mathsf{pivot}$. Therefore, $\Pr[T_{m^*-3}] \leq \Pr\left[\mathsf{Cnt}_{\langle F,m^*-3\rangle} \leq (\frac{1}{2} - \frac{\sqrt{2}}{32})\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-3\rangle}\right]\right]$. Finally, employing Lemma 5 with $\beta = \frac{1}{2} - \frac{\sqrt{2}}{32}$, we obtain $\Pr[T_{m^*-3}] \leq \frac{1}{1+(1-(\frac{1}{2}-\frac{\sqrt{2}}{32}))^2\cdot 4\mathsf{pivot}} \leq \frac{1}{1+(1-(\frac{1}{2}-\frac{\sqrt{2}}{32}))^2\cdot 4\cdot 9.84\cdot(1+\frac{1}{4\sqrt{2}-1})^2} \leq \frac{1}{18.19}$. □

Now let us prove the statement for $\mathsf{ApproxMC6}$: $\Pr[L] \leq 0.055$ for $3 \leq \varepsilon < 4\sqrt{2} - 1$.

Proof. We aim to bound $\Pr[L]$ by the following equation:

$$\Pr[L] = \left[\bigcup_{i \in \{1,\ldots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap L_i\right)\right] \tag{3 revisited}$$

which can be simplified by the two observations labeled $O1$ and $O2$ below.

$O1 : \forall i \leq m^* - 3, T_i \subseteq T_{i+1}$. Therefore,

$$\bigcup_{i \in \{1,\ldots,m^*-3\}} (\overline{T_{i-1}} \cap T_i \cap L_i) \subseteq \bigcup_{i \in \{1,\ldots,m^*-3\}} T_i \subseteq T_{m^*-3}$$

$O2 : \forall i \geq m^* - 2$, since rounding $\mathsf{Cnt}_{\langle F,i\rangle}$ to pivot and $m^* \geq \log_2 |\mathsf{sol(F)}| - \log_2(\mathsf{pivot})$, we have $2^i \times \mathsf{Cnt}_{\langle F,i\rangle} \geq 2^{m^*-2} \times \mathsf{pivot} \geq \frac{|\mathsf{sol(F)}|}{4} \geq \frac{|\mathsf{sol(F)}|}{1+\varepsilon}$. The last inequality follows from $\varepsilon \geq 3$. Then we have $\left(\mathsf{Cnt}_{\langle F,i\rangle} \geq \frac{\mathsf{E}\left[\mathsf{Cnt}_{\langle F,i\rangle}\right]}{1+\varepsilon}\right)$. Therefore, $L_i = \emptyset$ for $i \geq m^* - 2$ and we have

$$\bigcup_{i \in \{m^*-2,\ldots,n\}} (\overline{T_{i-1}} \cap T_i \cap L_i) = \emptyset$$

Following the observations $O1$ and $O2$, we simplify Eq. 3 and obtain

$$\Pr[L] \leq \Pr[T_{m^*-3}]$$

Employing Lemma 10 gives $\Pr[L] \leq 0.055$. □

D.4 Proof of $\Pr[L] \leq 0.023$ for $\varepsilon \geq 4\sqrt{2} - 1$

Lemma 11. *Given $\varepsilon \geq 4\sqrt{2} - 1$, the following bound hold:*

$$Pr[T_{m^*-4}] \leq \frac{1}{45.28}$$

Proof. We have $\mathsf{thresh} < 2\mathsf{pivot}$ and $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-4\rangle}\right] \geq 8\mathsf{pivot}$. Therefore, $\Pr[T_{m^*-4}] \leq \Pr\left[\mathsf{Cnt}_{\langle F,m^*-4\rangle} \leq \frac{1}{4}\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*-4\rangle}\right]\right]$. Finally, employing Lemma 5 with $\beta = \frac{1}{4}$, we obtain $\Pr[T_{m^*-4}] \leq \frac{1}{1+(1-\frac{1}{4})^2 \cdot 8\mathsf{pivot}} \leq \frac{1}{1+(1-\frac{1}{4})^2 \cdot 8 \cdot 9.84} \leq \frac{1}{45.28}$. □

Now let us prove the statement for $\mathsf{ApproxMC6}$: $\Pr[L] \leq 0.023$ for $\varepsilon \geq 4\sqrt{2} - 1$.

Proof. We aim to bound $\Pr[L]$ by the following equation:

$$\Pr[L] = \left[\bigcup_{i \in \{1,\ldots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap L_i\right)\right] \tag{3 revisited}$$

which can be simplified by the two observations labeled $O1$ and $O2$ below.

$O1 : \forall i \leq m^* - 4, T_i \subseteq T_{i+1}$. Therefore,

$$\bigcup_{i \in \{1,\ldots,m^*-4\}} (\overline{T_{i-1}} \cap T_i \cap L_i) \subseteq \bigcup_{i \in \{1,\ldots,m^*-4\}} T_i \subseteq T_{m^*-4}$$

$O2 : \forall i \geq m^* - 3$, since rounding $\mathsf{Cnt}_{\langle F,i \rangle}$ to $\sqrt{2}\mathsf{pivot}$ and $m^* \geq \log_2 |\mathsf{sol(F)}| - \log_2 (\mathsf{pivot})$, we have $2^i \times \mathsf{Cnt}_{\langle F,i \rangle} \geq 2^{m^*-3} \times \sqrt{2}\mathsf{pivot} \geq \frac{\sqrt{2}|\mathsf{sol(F)}|}{8} \geq \frac{|\mathsf{sol(F)}|}{1+\varepsilon}$. The last inequality follows from $\varepsilon \geq 4\sqrt{2} - 1$. Then we have $\left(\mathsf{Cnt}_{\langle F,i \rangle} \geq \frac{\mathsf{E}\left[\mathsf{Cnt}_{\langle F,i \rangle}\right]}{1+\varepsilon}\right)$. Therefore, $L_i = \emptyset$ for $i \geq m^* - 3$ and we have

$$\bigcup_{i \in \{m^*-3,\ldots,n\}} (\overline{T_{i-1}} \cap T_i \cap L_i) = \emptyset$$

Following the observations $O1$ and $O2$, we simplify Eq. 3 and obtain

$$\Pr[L] \leq \Pr[T_{m^*-4}]$$

Employing Lemma 11 gives $\Pr[L] \leq 0.023$. □

D.5 Proof of $\Pr[U] \leq 0.169$ for $\varepsilon < 3$

Lemma 12

$$Pr[U'_{m^*}] \leq \frac{1}{5.92}$$

Proof. Employing Lemma 5 with $\gamma = (1 + \frac{\varepsilon}{1+\varepsilon})$ and $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^* \rangle}\right] \geq \mathsf{pivot}/2$, we obtain $\Pr[U'_{m^*}] \leq \frac{1}{1+\left(\frac{\varepsilon}{1+\varepsilon}\right)^2 \mathsf{pivot}/2} \leq \frac{1}{1+9.84/2} \leq \frac{1}{5.92}$. □

Now let us prove the statement for $\mathsf{ApproxMC6}$: $\Pr[U] \leq 0.169$ for $\varepsilon < 3$.

Proof. We aim to bound $\Pr[U]$ by the following equation:

$$\Pr[U] = \left[\bigcup_{i \in \{1,\ldots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap U_i\right)\right] \quad \text{(4 revisited)}$$

We derive the following observations $O1$ and $O2$.

$O1 : \forall i \leq m^* - 1$, since $m^* \leq \log_2 |\mathsf{sol(F)}| - \log_2 (\mathsf{pivot}) + 1$, we have $2^i \times \mathsf{Cnt}_{\langle F,i \rangle} \leq 2^{m^*-1} \times \mathsf{thresh} \leq |\mathsf{sol(F)}| \left(1 + \frac{\varepsilon}{1+\varepsilon}\right)$. Then we obtain $\left(\mathsf{Cnt}_{\langle F,i \rangle} \leq \mathsf{E}\left[\mathsf{Cnt}_{\langle F,i \rangle}\right] \left(1 + \frac{\varepsilon}{1+\varepsilon}\right)\right)$. Therefore, $T_i \cap U'_i = \emptyset$ for $i \leq m^* - 1$ and we have

$$\bigcup_{i \in \{1,\ldots,m^*-1\}} \left(\overline{T_{i-1}} \cap T_i \cap U_i\right) \subseteq \bigcup_{i \in \{1,\ldots,m^*-1\}} \left(\overline{T_{i-1}} \cap T_i \cap U'_i\right) = \emptyset$$

$O2$: $\forall i \geq m^*$, $\overline{T_i}$ implies $\mathsf{Cnt}_{\langle F,i\rangle} > \mathsf{thresh}$ and then we have $2^i \times \mathsf{Cnt}_{\langle F,i\rangle} > 2^{m^*} \times \mathsf{thresh} \geq |\mathsf{sol(F)}|\left(1+\frac{\varepsilon}{1+\varepsilon}\right)$. The second inequality follows from $m^* \geq \log_2|\mathsf{sol(F)}| - \log_2(\mathsf{pivot})$. Then we obtain $\left(\mathsf{Cnt}_{\langle F,i\rangle} > \mathsf{E}\left[\mathsf{Cnt}_{\langle F,i\rangle}\right]\left(1+\frac{\varepsilon}{1+\varepsilon}\right)\right)$. Therefore, $\overline{T_i} \subseteq U_i'$ for $i \geq m^*$. Since $\forall i, \overline{T_i} \subseteq \overline{T_{i-1}}$, we have

$$\begin{aligned}\bigcup_{i\in\{m^*,\ldots,n\}}\left(\overline{T_{i-1}} \cap T_i \cap U_i\right) &\subseteq \bigcup_{i\in\{m^*+1,\ldots,n\}} \overline{T_{i-1}} \cup (\overline{T_{m^*-1}} \cap T_{m^*} \cap U_{m^*})\\ &\subseteq \overline{T_{m^*}} \cup (\overline{T_{m^*-1}} \cap T_{m^*} \cap U_{m^*})\\ &\subseteq \overline{T_{m^*}} \cup U_{m^*}\\ &\subseteq U'_{m^*} \end{aligned} \tag{8}$$

Remark that for $\varepsilon < \sqrt{2}-1$, we round $\mathsf{Cnt}_{\langle F,m^*\rangle}$ up to $\frac{\sqrt{1+2\varepsilon}}{2}\mathsf{pivot}$ and we have $2^{m^*} \times \frac{\sqrt{1+2\varepsilon}}{2}\mathsf{pivot} \leq |\mathsf{sol(F)}|(1+\varepsilon)$. For $\sqrt{2}-1 \leq \varepsilon < 1$, we round $\mathsf{Cnt}_{\langle F,m^*\rangle}$ up to $\frac{\mathsf{pivot}}{\sqrt{2}}$ and we have $2^{m^*} \times \frac{\mathsf{pivot}}{\sqrt{2}} \leq |\mathsf{sol(F)}|(1+\varepsilon)$. For $1 \leq \varepsilon < 3$, we round $\mathsf{Cnt}_{\langle F,m^*\rangle}$ up to pivot and we have $2^{m^*} \times \mathsf{pivot} \leq |\mathsf{sol(F)}|(1+\varepsilon)$. The analysis means *rounding* doesn't affect the event U_{m^*} and therefore Inequality 8 still holds.

Following the observations $O1$ and $O2$, we simplify Eq. 4 and obtain

$$\Pr[U] \leq \Pr[U'_{m^*}]$$

Employing Lemma 12 gives $\Pr[U] \leq 0.169$. □

D.6 Proof of $\Pr[U] \leq 0.044$ for $\varepsilon \geq 3$

Lemma 13

$$Pr\left[\overline{T_{m^*+1}}\right] \leq \frac{1}{23.14}$$

Proof. Since $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*+1\rangle}\right] \leq \frac{\mathsf{pivot}}{2}$, we have $\Pr\left[\overline{T_{m^*+1}}\right] \leq \Pr\left[\mathsf{Cnt}_{\langle F,m^*+1\rangle} > 2(1+\frac{\varepsilon}{1+\varepsilon})\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*+1\rangle}\right]\right]$. Employing Lemma 5 with $\gamma = 2(1+\frac{\varepsilon}{1+\varepsilon})$ and $\mathsf{E}\left[\mathsf{Cnt}_{\langle F,m^*+1\rangle}\right] \geq \frac{\mathsf{pivot}}{4}$, we obtain $\Pr\left[\overline{T_{m^*+1}}\right] \leq \frac{1}{1+\left(1+\frac{2\varepsilon}{1+\varepsilon}\right)^2\mathsf{pivot}/4} = \frac{1}{1+2.46\cdot\left(3+\frac{1}{\varepsilon}\right)^2} \leq \frac{1}{1+2.46\cdot 3^2} \leq \frac{1}{23.14}$. □

Now let us prove the statement for $\mathsf{ApproxMC6}$: $\Pr[U] \leq 0.044$ for $\varepsilon \geq 3$.

Proof. We aim to bound $\Pr[U]$ by the following equation:

$$\Pr[U] = \left[\bigcup_{i\in\{1,\ldots,n\}}\left(\overline{T_{i-1}} \cap T_i \cap U_i\right)\right] \tag{4 revisited}$$

We derive the following observations $O1$ and $O2$.

$O1$: $\forall i \le m^*+1$, for $3 \le \varepsilon < 4\sqrt{2}-1$, because we round $\mathsf{Cnt}_{\langle F,i\rangle}$ to pivot and have $m^* \le \log_2|\mathsf{sol(F)}| - \log_2(\mathsf{pivot})+1$, we obtain $2^i \times \mathsf{Cnt}_{\langle F,i\rangle} \le 2^{m^*+1} \times \mathsf{pivot} \le 4\cdot|\mathsf{sol(F)}| \le (1+\varepsilon)|\mathsf{sol(F)}|$. For $\varepsilon \ge 4\sqrt{2}-1$, we round $\mathsf{Cnt}_{\langle F,i\rangle}$ to $\sqrt{2}\mathsf{pivot}$ and obtain $2^i \times \mathsf{Cnt}_{\langle F,i\rangle} \le 2^{m^*+1} \times \sqrt{2}\mathsf{pivot} \le 4\sqrt{2}\cdot|\mathsf{sol(F)}| \le (1+\varepsilon)|\mathsf{sol(F)}|$. Then, we obtain $\mathsf{Cnt}_{\langle F,i\rangle} \le \mathsf{E}\left[\mathsf{Cnt}_{\langle F,i\rangle}\right](1+\varepsilon)$. Therefore, $U_i = \emptyset$ for $i \le m^*+1$ and we have

$$\bigcup_{i\in\{1,\dots,m^*+1\}} \left(\overline{T_{i-1}} \cap T_i \cap U_i\right) = \emptyset$$

$O2$: $\forall i \ge m^*+2$, since $\forall i, \overline{T_i} \subseteq \overline{T_{i-1}}$, we have

$$\bigcup_{i\in\{m^*+2,\dots,n\}} \left(\overline{T_{i-1}} \cap T_i \cap U_i\right) \subseteq \bigcup_{i\in\{m^*+2,\dots,n\}} \overline{T_{i-1}} \subseteq \overline{T_{m^*+1}}$$

Following the observations $O1$ and $O2$, we simplify Eq. 4 and obtain

$$\Pr[U] \le \Pr\left[\overline{T_{m^*+1}}\right]$$

Employing Lemma 13 gives $\Pr[U] \le 0.044$. □

References

1. Alur, R., et al.: Syntax-guided synthesis. In: Proceedings of FMCAD (2013)
2. Baluta, T., Shen, S., Shine, S., Meel, K.S., Saxena, P.: Quantitative verification of neural networks and its security applications. In: Proceedings of CCS (2019)
3. Beck, G., Zinkus, M., Green, M.: Automating the development of chosen ciphertext attacks. In: Proceedings of USENIX Security (2020)
4. Carter, J.L., Wegman, M.N.: Universal classes of hash functions. J. Comput. Syst. Sci. (1977)
5. Chakraborty, S., Fremont, D.J., Meel, K.S., Seshia, S.A., Vardi, M.Y.: Distribution-aware sampling and weighted model counting for SAT. In: Proceedings of AAAI (2014)
6. Chakraborty, S., Meel, K.S., Mistry, R., Vardi, M.Y.: Approximate probabilistic inference via word-level counting. In: Proceedings of AAAI (2016)
7. Chakraborty, S., Meel, K.S., Vardi, M.Y.: A scalable approximate model counter. In: Proceedings of CP (2013)
8. Chakraborty, S., Meel, K.S., Vardi, M.Y.: Algorithmic improvements in approximate counting for probabilistic inference: from linear to logarithmic SAT calls. In: Proceedings of IJCAI (2016)
9. Duenas-Osorio, L., Meel, K.S., Paredes, R., Vardi, M.Y.: Counting-based reliability estimation for power-transmission grids. In: Proceedings of AAAI (2017)
10. Ermon, S., Gomes, C.P., Sabharwal, A., Selman, B.: Embed and project: discrete sampling with universal hashing. In: Proceedings of NeurIPS (2013)
11. Ermon, S., Gomes, C.P., Sabharwal, A., Selman, B.: Taming the curse of dimensionality: discrete integration by hashing and optimization. In: Proceedings of ICML (2013)

12. Fichte, J.K., Hecher, M., Hamiti, F.: The model counting competition 2020. ACM J. Exp. Algorithmics (2021)
13. Gittis, A., Vin, E., Fremont, D.J.: Randomized synthesis for diversity and cost constraints with control improvisation. In: Proceedings of CAV (2022)
14. Gomes, C.P., Sabharwal, A., Selman, B.: Model counting: a new strategy for obtaining good bounds. In: Proceedings of AAAI (2006)
15. Hecher, M., Fichte, J.K.: Model counting competition 2021 (2021). https://www.mccompetition.org/2021/mc_description
16. Hecher, M., Fichte, J.K.: Model counting competition 2022 (2022). https://mccompetition.org/2022/mc_description
17. Ivrii, A., Malik, S., Meel, K.S., Vardi, M.Y.: On computing minimal independent support and its applications to sampling and counting. Constraints (2016)
18. Meel, K.S., Akshay, S.: Sparse hashing for scalable approximate model counting: theory and practice. In: Proceedings of LICS (2020)
19. Meel, K.S., et al.: Constrained sampling and counting: universal hashing meets sat solving. In: Proceedings of Workshop on Beyond NP(BNP) (2016)
20. Roth, D.: On the hardness of approximate reasoning. Artif. Intell. (1996)
21. Sang, T., Bearne, P., Kautz, H.: Performing Bayesian inference by weighted model counting. In: Proceedings of AAAI (2005)
22. Soos, M., Gocht, S., Meel, K.S.: Tinted, detached, and lazy CNF-XOR solving and its applications to counting and sampling. In: Proceedings of CAV (2020)
23. Soos, M., Meel, K.S.: Bird: engineering an efficient CNF-XOR sat solver and its applications to approximate model counting. In: Proceedings of AAAI (2019)
24. Soos, M., Meel, K.S.: Arjun: an efficient independent support computation technique and its applications to counting and sampling. In: Proceedings of ICCAD (2022)
25. Stockmeyer, L.: The complexity of approximate counting. In: Proceedings of STOC (1983)
26. Teuber, S., Weigl, A.: Quantifying software reliability via model-counting. In: Proceedings of QEST (2021)
27. Toda, S.: On the computational power of PP and (+)P. In: Proceedings of FOCS (1989)
28. Valiant, L.G.: The complexity of enumeration and reliability problems. SIAM J. Comput. (1979)
29. Yang, J., Chakraborty, S., Meel, K.S.: Projected model counting: beyond independent support. In: Proceedings of ATVA (2022)
30. Yang, J., Meel, K.S.: Engineering an efficient PB-XOR solver. In: Proceedings of CP (2021)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.