

Learning-guided sampling: finite classes and one-sided reliability

Samuel Mausberg

October 14, 2026

Abstract

We study sampling from an unknown member of a known finite class using approximate prefix masses. Under pointwise guidance with factor $9/4$ and error $1/16$, the worst-case query complexity is $\min\{M, n\sqrt{\log M}, n^2\}$ up to logarithms. We show that one-sided safety and bounded stopped reciprocal capacity can change the cost of a specified class from logarithmic to linear, with exact linear caps at fixed error and for exact sampling. For a known background plus an unknown atom of weight α , we determine the worst-case rate over backgrounds as $\Theta(1 + \min\{\alpha M, \alpha^{-1} \log(1 + 2^n \alpha^2)\})$ at error at most $\alpha/4$. A uniform background admits a sharper rate. The upper bounds derive membership tests and atom recovery from the guide; independent fallback sampling controls the complete output law. A composition theorem separates global certification from difficult fixed-error sampling. We also give a self-contained quadratic upper bound under weak guidance. The general fixed-error finite-class gap remains open. The numbered results are formalized in Lean 4, with the cited finite-class upper bound as an explicit hypothesis.

1 Introduction

How much does knowing a finite class help when an approximate guide is the only source of information about its unknown target? Golowich, Moitra, and Rohatgi [1] introduced a finite-class framework for this question and obtained a sampling algorithm whose dependence on class size is through $\sqrt{\log M}$. We prove a matching worst-case lower bound up to logarithmic factors; for factor $R = 9/4$, binary strings and error $1/16$, this answers their Problem 1.4. We then investigate what remains when pointwise accuracy is replaced by one-sided safety and a stopped moment bound.

Sampling can be cheaper than identification. An unknown component of mass α may only need to be found on an α fraction of executions. This saves queries only when those executions can be selected through the available oracle and every other branch has the correct output law. Our atom constructions give a fixed-error separation between the two guidance models on the same class, and a sharp dependence on the known background's prefix geometry.

1.1 Model

For a probability distribution μ on $\{0, 1\}^n$, write $\mu(u) = \sum_{x \succeq u} \mu(x)$. The algorithm knows a class $\mathcal{F}$ of at most M distributions and queries one fixed function $h : \{0, 1\}^{\leq n} \rightarrow \mathbb{R}_{\geq 0}$. The target $\mu \in \mathcal{F}$ is unknown. We require

$$h(\emptyset) = 1, \quad h(x) = \mu(x) \quad (|x| = n). \quad (1)$$

The root is known without a query. Every call to h is charged, including preprocessing and unsuccessful executions. A query bound is a deterministic cap on every execution. Accuracy

means $\text{TV}(\mathcal{L}(X), \mu) \leq \delta$ for the unconditional output law. Computation on known data, including candidate evaluation and sampling, is free. Randomness is also free; no bit-complexity bound is asserted. All logarithms are natural unless a base is displayed, and $\tilde{O}, \tilde{\Omega}, \tilde{\Theta}$ suppress logarithmic factors in the growing parameters.

A pointwise factor- R guide satisfies $\mu/R \leq h \leq R\mu$. The weaker model is

$$h \leq R\mu, \quad \mathcal{C}(\mu, h) := \max_{A \text{ prefix-free}} \sum_{u \in A} \frac{\mu(u)^2}{h(u)} \leq C. \quad (2)$$

A positive numerator over zero has infinite cost; $0/0$ contributes zero. Thus finite capacity and safety force h and μ to have the same support at every prefix. The maximum includes cuts of mixed depths. No additive measure realizing this bound is supplied.

Write $Q_{\text{pw}}(\mathcal{F}, \delta)$ and $Q_{\text{safe}}(\mathcal{F}, \delta)$ for the minimax capped query complexities of a specified class, with factor $R = 9/4$ in the first model and $R = C = 9/4$ in the second. Write $Q_{\text{pw}}(n, M, \delta)$ and $Q_{\text{safe}}(n, M, \delta)$ for their worst cases over classes of size at most M on n bits. All constants called fixed are independent of the target and of n, M, α .

1.2 Pointwise guidance

Theorem 1.1 (Pointwise characterization). *For $n, M \geq 2$, let*

$$K(n, M) = \min\{M, n\sqrt{\log M}, n^2\}, \quad L_n = \log(2n). \quad (3)$$

There are absolute constants $c_0, c_1, c_2 > 0$ such that

$$\frac{c_0}{L_n} K(n, M) \leq Q_{\text{pw}}(n, M, 1/16) \leq c_1 \min\{M, n\sqrt{\log M} \log(2 + \log M), n^2\} \leq c_2 L_n K(n, M). \quad (4)$$

For all sufficiently large n ,

$$c_0 \frac{n^{3/2}}{\sqrt{\log(2n)}} \leq Q_{\text{pw}}(n, \lfloor e^n \rfloor, 1/16) \leq c_1 n^{3/2} \log(2n). \quad (5)$$

A singleton class requires no queries.

The lower bound uses a hidden table shared across output blocks. Its size controls the class cardinality, and finding table entries controls query cost, even for arbitrary long prefix queries. Appendix B gives the full proof. The finite-class upper bound is due to Golowich, Moitra, and Rohatgi [1, Theorem 3.2]; their Theorem 4.12 supplies $\Omega(n)$ when $M = \lfloor e^n \rfloor$, and Lemma B.9 restates its comb with root value one. Exact identification costs at most $M - 1$ queries. Appendix C proves the quadratic upper bound under the weaker assumptions (2) as well.

Theorem 2.1 separates the models on a single class at fixed error. Theorem 3.1 determines the atom-sampling rate under a uniform background, and Theorem 3.6 the worst-case rate over known backgrounds. Sections 4 and 5 quantify a different obstruction: obtaining a certificate valid at every prefix can cost much more than producing a sample.

1.3 Formal verification

Every theorem, lemma, proposition and corollary of this paper is formalized in Lean 4 with Mathlib and checked by the Lean kernel. Algorithms are modeled as adaptive randomized query trees with exact output laws and a cap on every execution, and the lower bounds quantify over all of them.

The finite-class upper bound of [1, Theorem 3.2], used in Theorem 1.1 and Corollary 5.3, enters the formalization as an explicit hypothesis. The formal proof of Lemma B.8 bounds the Hellinger affinity of the output laws of the stopped batch process, which gives the stated bound for every event that an algorithm can decide. The bounds written with $\bar{O}$ in (42) are formalized as the explicit caps behind them. The code and a map from each result to its Lean statement are available at <https://github.com/SamMausberg/learning-guided-sampling>.

2 A fixed-error separation

The pointwise and safe models can have very different costs on a specified class, even when the accuracy and reliability constants are fixed. The construction distributes a heavy background over sixteen prefix subtrees. A safe guide can conceal the atom along a long path inside one subtree. Under pointwise accuracy, one query decides whether a prefix on that path contains the atom.

Theorem 2.1 (Fixed-error separation). *For every integer $\ell \geq 1$, there is a known class $\mathcal{H}_\ell$ of $M = 16\ell$ full-support distributions on $n = \ell + 4$ bits such that*

$$Q_{\text{pw}}(\mathcal{H}_\ell, 1/16) = \Theta(\log(2 + \ell)), \quad Q_{\text{pw}}(\mathcal{H}_\ell, 0) \leq 4 + \lceil \log_2 \ell \rceil, \quad (6)$$

$$Q_{\text{safe}}(\mathcal{H}_\ell, 1/16) = 3\ell, \quad Q_{\text{safe}}(\mathcal{H}_\ell, 0) = 4\ell. \quad (7)$$

The safe lower bounds use fixed positive guides with $h \leq (4/3)\mu$ and stopped reciprocal capacity at most $263/120 < 9/4$.

For these parameters, $K(n, M) = \Theta(\ell)$. Thus the safe lower bound agrees with the worst-case pointwise benchmark, while exceeding the pointwise cost of this particular class exponentially.

Lemma 2.2 (Public prefixes). *Let ν have full support, let $\mu_x = (1 - \alpha)\nu + \alpha\delta_x$, with $0 < \alpha \leq 1/4$, and let $\mathcal{B}$ be any set of proper prefixes. Set $h_x(u) = \nu(u)$ for $u \in \mathcal{B}$ and $h_x(u) = \mu_x(u)$ elsewhere. Then $h_x \leq (4/3)\mu_x$, the leaves are exact, and, writing $a = 1 - \alpha$,*

$$\mathcal{C}(\mu_x, h_x) = \max \left\{ 1, \max_{\substack{u \in \mathcal{B} \\ u \preceq x}} \left(1 + \alpha - 2\alpha^2 - a\alpha\nu(u) + \frac{\alpha^2}{\nu(u)} \right) \right\}. \quad (8)$$

An empty inner maximum is omitted.

Proof. Safety follows from $a\nu \leq \mu_x$. Complete a prefix-free set with uncovered leaves. Off the path to x , the node contribution $\mu_x(u)^2/h_x(u)$ is at most $\mu_x(u)$, so replacing that node by its descendant leaves does not decrease the sum. The remaining cut either consists of exact nodes, with value one, or selects one public ancestor u of x and all leaves outside it. Its value is $1 - \mu_x(u) + \mu_x(u)^2/\nu(u)$. Substituting $\mu_x(u) = a\nu(u) + \alpha$ gives (8). Each displayed value is attained by that cut. $\square$

Proof of Theorem 2.1. For $y \in \{0, 1\}^4$ and $1 \leq j \leq \ell$, put

$$x_{y,j} = y0^{j-1}10^{\ell-j}, \quad \nu = \frac{15}{16}(U_4 \otimes \delta_{0^\ell}) + \frac{1}{16}U_n, \quad \mu_{y,j} = \frac{3}{4}\nu + \frac{1}{4}\delta_{x_{y,j}}.$$

These 16ℓ targets form $\mathcal{H}_\ell$. Use the guide of Lemma 2.2, with public answers at every prefix of length at most four and at each proper prefix $y0^t$. Every such ancestor of an atom has $\nu(u) \geq 15/256$. Since $\alpha = 1/4$,

$$\mathcal{C}(\mu_{y,j}, h_{y,j}) \leq \frac{9}{8} + \frac{(1/4)^2}{15/256} = \frac{263}{120} < \frac{9}{4}.$$

Every other query is either a complete string $y0^\ell$, whose exact answer is public because every atom contains a one, or specifies the four-bit word y and a first-one position in its suffix. Its answer is public unless that pair equals the hidden (y, j) . It is simulated by one equality test, including complete-string queries. Fixing the random seed and following all negative answers shows that q tests predict a uniform pair with probability at most $(q+1)/(16\ell)$: at most q pairs are tested, and the common negative transcript can guess one more. On the other hand, δ -accurate sampling gives

$$\mathbb{P}(X = x_{y,j}) \geq \frac{1}{4} + \frac{3}{64 \cdot 2^n} - \delta.$$

Consequently

$$q \geq 16\ell \left(\frac{1}{4} - \delta + \frac{3}{64 \cdot 2^n} \right) - 1. \quad (9)$$

The positive last term and integrality give $q \geq 3\ell$ at $\delta = 1/16$ and $q \geq 4\ell$ at $\delta = 0$. For matching caps, test a uniformly selected set of q distinct candidate leaves. On a hit output that atom; otherwise draw independently from ν . The output law is $(1 - q/M)\nu + (q/M)\delta_{x_{y,j}}$. Taking $q = 3\ell$ or $q = 4\ell$ proves (7) for every legal guide.

Now consider any pointwise factor-9/4 guide. At a prefix $y0^t$, $0 \leq t < \ell$, the background mass lies between $15/256$ and $1/16$. If the prefix excludes the atom, then

$$h(y0^t) \leq \frac{9}{4} \frac{3}{4} \frac{1}{16} = \frac{27}{256} < \frac{1}{8}.$$

If it contains the atom, then

$$h(y0^t) \geq \frac{4}{9} \left(\frac{3}{4} \frac{15}{256} + \frac{1}{4} \right) = \frac{301}{2304} > \frac{1}{8}.$$

Select four of the sixteen four-bit groups with equal inclusion probability $1/4$. Query them until a hit. Within a hit group, a query to $y0^t$ tests exactly whether $j > t$, so binary search recovers j in at most $\lceil \log_2 \ell \rceil$ further calls. Output the atom on a hit, and an independent ν draw otherwise. The atom branch has probability $1/4$, giving the exact target law with the cap in (6).

For its lower bound choose the legal guide $h = \mu_{y,j}$. Each prefix has at most two answers over the M targets. A fixed-seed depth- q decision tree has at most 2^q transcripts and predicts a uniform pair with probability at most $2^q/M$. At error $1/16$, this forces $2^q > 3\ell$. Averaging over the seed proves the same bound for randomized algorithms. $\square$

3 Sampling an unknown atom

For a uniform background, one query to a sufficiently fine prefix group decides whether it contains the atom. The same comparison recovers the atom after a hit.

For $1 \leq k \leq n$, $M = 2^k$, and $0 < \alpha \leq 1/4$, consider

$$\mathcal{F}_{n,k,\alpha} = \{(1 - \alpha)U_n + \alpha\delta_{\theta 0^{n-k}} : \theta \in \{0, 1\}^k\}. \quad (10)$$

The mixture weight, candidate strings, and background are known. Only θ and the fixed legal guide are unknown.

Theorem 3.1 (Unknown atom, uniform background). *Let $0 \leq \delta \leq \alpha/4$ and $R = C = 9/4$. Uniformly over $n \geq k \geq 1$ and $0 < \alpha \leq 1/4$,*

$$Q_{\text{safe}}(\mathcal{F}_{n,k,\alpha}, \delta) = \Theta\left(1 + \min\{\alpha M, \alpha^{-1} + \log_2 M\}\right). \quad (11)$$

The upper bound holds for exact sampling. The lower bound allows arbitrary adaptive prefix queries and arbitrary outputs. The implicit constants are absolute.

For fixed $\alpha > 0$, the rate is $\Theta(1 + \log M)$ as M grows. For $M \asymp \alpha^{-2}$, it is $\Theta(\alpha^{-1})$.

3.1 A membership test

Lemma 3.2 (Membership test). *Suppose a known prefix group u either contains an entire hidden component of mass $\alpha > 0$, or contains none of it. In the latter case its target mass is a known number b . Assume*

$$h(u) \leq R\mu(u), \quad \mu(u)^2 \leq Ch(u), \quad C > 0, \quad Rb < \alpha^2/C.$$

Then the group contains the component exactly when $h(u) \geq \alpha^2/C$.

Proof. In the positive case, $\mu(u) \geq \alpha$, so $h(u) \geq \alpha^2/C$. In the negative case, $h(u) \leq Rb < \alpha^2/C$. The reciprocal inequality follows from the singleton cut $\{u\}$ in (2). $\square$

Write $a = 1 - \alpha$, and let d be the smallest nonnegative integer with

$$G = 2^d \geq \frac{2RCa}{\alpha^2}. \quad (12)$$

If $d \leq k$, partition strings by their first d bits. Every such group has background mass a/G . Every descendant prefix has no larger background mass. Thus the test in Lemma 3.2 is valid at the group and at every deeper key prefix. A hit permits recovery of each remaining key bit with one query to its zero child. The known padding needs no queries.

The test compares one guide value with a known threshold, so recovery accumulates no error. The upper algorithm uses only the singleton cuts of (2); the lower-bound guides satisfy the full stopped-capacity bound.

3.2 Search with thinning

Let $\beta = \alpha$ for an exact sample, or $\beta = \alpha - \delta$ for error at most δ . Set $q = \lceil \beta G \rceil$. Choose a random set of q groups, independently of the eventual background draw, so that each group is included with probability q/G . A uniformly shifted block of q consecutive group indices suffices.

Query these groups until a hit or exhaustion. On a hit, flip a coin with success probability $\beta G/q$. Only on coin success recover the remaining key bits and output the atom. Every other branch outputs a fresh sample from U_n . The probability of the atom-output branch is exactly

$$\frac{q}{G} \frac{\beta G}{q} = \beta. \quad (13)$$

Consequently the complete output law is

$$(1 - \beta)U_n + \beta\delta_{\theta 0^{n-k}}, \quad (14)$$

whose distance to the target is $(\alpha - \beta)(1 - 2^{-n}) \leq \delta$. The deterministic cap is $q + k - d$; the expected number of recovery queries is $\beta(k - d)$. All unsuccessful tests count toward q .

There is also an exact-leaf alternative. Search $\lceil \beta M \rceil$ candidate atoms with the same inclusion and coin rule. A leaf query detects the extra mass exactly. Thus, when $d \leq k$,

$$Q_{\text{safe}}(\mathcal{F}_{n,k,\alpha}, \delta) \leq \min\{\lceil \beta M \rceil, \lceil \beta 2^d \rceil + k - d\}. \quad (15)$$

When $d > k$, use the first term alone. Since $2^d < 4RCa/\alpha^2$, the exact upper bound in Theorem 3.1 follows.

3.3 Lower bounds

First take the legal guide $h = \mu_\theta$. At every prefix it has at most two possible answers as θ varies. A fixed-seed decision tree of depth q therefore has at most 2^q transcripts. Its output predicts a uniform key with probability at most $2^q/M$. Accuracy requires prediction probability at least $\alpha + a/M - \delta$, hence

$$q \geq \log_2((\alpha - \delta)M + a). \quad (16)$$

Zero queries are impossible for $\delta \leq \alpha/4$: two targets have distance $\alpha > 2\delta$.

For the search obstruction, assume $k \geq 2$ and put

$$s = \min\{k - 1, \lfloor \log_2(\alpha^{-2}) \rfloor\}, \quad H = 2^s.$$

Restrict to the H targets whose atoms are $\vartheta 0^{n-s}$, $\vartheta \in \{0, 1\}^s$. Set the guide to $2^{-|u|}$ through depth s and to the exact target below that depth. This is fixed, positive, exact at leaves, and safe since $h \leq \mu/a \leq (4/3)\mu$.

Its stopped capacity is

$$1 + \alpha - 2\alpha^2 - \frac{a\alpha}{H} + \alpha^2 H \leq \frac{17}{8}. \quad (17)$$

To see this, replace every node off the distinguished path by its descendant leaves. Such a replacement does not decrease the cut sum because $\mu \leq h$ there. A cut meeting the distinguished path at depth $t \leq s$ has value

$$a + 2a\alpha - a\alpha 2^{-t} + \alpha^2 2^t,$$

which increases with t . Below depth s the guide is exact. Evaluating at s , using $\alpha^2 H \leq 1$ and $\alpha - 2\alpha^2 \leq 1/8$, proves (17).

Each query through depth s is public. Every longer query specifies one complete coarse key; its answer is public unless this key equals ϑ . It is therefore simulated by one equality test. Following the all-negative transcript shows that q adaptive equality queries predict a uniform key with probability at most $(q + 1)/H$. Accuracy forces

$$q \geq (\alpha - \delta)H - \alpha \geq \frac{3}{8} \min\{\alpha M, \alpha^{-1}\} - \frac{1}{4}. \quad (18)$$

To combine the bounds, set $A = \alpha M$, $B = \alpha^{-1} \geq 4$, and $L = \log_2 B \leq B/2$. Equations (16) and (18), together with $q \geq 1$, imply

$$\min\{A, B\} \leq 4q, \quad k \leq q + L + 1.$$

If $A \leq B$, the desired minimum is at most $4q$. If $A > B$, then $B \leq 4q$ and $k \leq 4q$, so $\min\{A, B+k\} \leq 8q$. Thus the right side of (11) is at most $9q$. The case $k = 1$ follows from $q \geq 1$. This completes the proof.

3.4 Arbitrary atom locations

The upper mechanism does not require a complete cube of possible keys.

Proposition 3.3. *Let $S \subseteq \{0, 1\}^n$ be any known set of $M \geq 2$ distinct strings, and let $\mu_x = (1 - \alpha)U_n + \alpha\delta_x$ for unknown $x \in S$. Under (2), exact sampling uses*

$$O(1 + \min\{\alpha M, RC/\alpha + \log M\}) \quad (19)$$

guide queries. The constant is absolute for $R, C \geq 1$ and $0 < \alpha \leq 1/4$.

Proof. Use the depth d in (12) when $d \leq n$. Search only its $G' \leq 2^d$ groups containing candidates, each with inclusion probability q/G' for $q = \lceil \alpha G' \rceil$. On a hit, proceed with probability $\alpha G'/q$. Lemma 3.2 still tests membership. On the selected atom-output branch, recover x using balanced prefix tests inside its group.

For any remaining set of $N \geq 2$ candidate strings in that group, some descendant prefix contains between $N/3$ and $2N/3$ candidates. Follow a largest child from the group root until its count first falls to at most $2N/3$; that child has more than $N/3$ candidates. Query it and retain the indicated side. Every such prefix has background mass at most $a/2^d$, so the test remains exact. After at most $\lceil \log_{3/2} M \rceil$ queries only one candidate remains. The atom branch has probability α regardless of the hidden location. Compare this cap with the $\lceil \alpha M \rceil$ exact-leaf alternative. If $d > n$, the latter already has cap at most $2RCa/\alpha + 1$. $\square$

Section 3.6 constructs the corresponding light-prefix partition for arbitrary known backgrounds and determines its worst-case cost.

Corollary 3.4 (Pointwise rate). *For the class (10), factor $R = 9/4$, and $0 \leq \delta \leq \alpha/4$,*

$$Q_{\text{pw}}(\mathcal{F}_{n,k,\alpha}, \delta) = \Theta(1 + \log_2(1 + \alpha M)).$$

The upper bound again holds for exact sampling.

Proof. Use the smallest power $G = 2^d \geq 2R^2 a/\alpha$. At depth d and below, an occupied prefix has $h \geq \alpha/R$ and an empty one has $h \leq Ra/G \leq \alpha/(2R)$. Search $\lceil \alpha G \rceil = O(1)$ groups and recover the remaining $k - d$ key bits on the atom branch. If $d > k$, the exact-leaf search costs $O(1)$. Otherwise $k - d \leq \log_2(\alpha M)$. Equation (16) and the zero-query obstruction prove the lower bound. $\square$

3.5 Dependence on the background

The dependence on the background is necessary, even when it is known and has full support.

Proposition 3.5. *For every $M \geq 2$ and $0 < \alpha \leq 1/4$, there is a known full-support background ν on M bits and a known set S of M atom locations such that, for $\mathcal{H} = \{(1 - \alpha)\nu + \alpha\delta_x : x \in S\}$ and $0 \leq \delta \leq \alpha/4$,*

$$Q_{\text{pw}}(\mathcal{H}, \delta) = \Theta(1 + \alpha M), \quad Q_{\text{safe}}(\mathcal{H}, \delta) = \Theta(1 + \alpha M).$$

The implicit constants are absolute. The lower-bound guides are already pointwise factor $4/3$.

Proof. Set $n = M$, $\nu = \frac{3}{4}\delta_{0^n} + \frac{1}{4}U_n$, and $x_j = 0^{j-1}10^{n-j}$ for $1 \leq j \leq M$. At a proper all-zero prefix use the public answer $h_j(u) = \nu(u)$. At every other prefix, including leaves, use the exact target mass. Write $a = 1 - \alpha$. On a proper all-zero prefix, $\nu(u) \geq 3/4$ and

$$a\nu(u) \leq \mu_j(u) \leq a\nu(u) + \alpha, \quad \frac{h_j(u)}{\mu_j(u)} \leq \frac{1}{a} \leq \frac{4}{3}, \quad \frac{\mu_j(u)}{h_j(u)} \leq a + \frac{4\alpha}{3} \leq \frac{13}{12}.$$

Thus the guide is pointwise factor $4/3$, and its stopped capacity is at most $4/3$.

An all-zero query has a public answer. Every other query specifies its first-one position, and is simulated by at most one equality test of j . The output event $\{x_j\}$ has target probability $\alpha + a/(4 \cdot 2^n)$. Equality-query counting therefore gives

$$q \geq M \left(\alpha - \delta + \frac{a}{4 \cdot 2^n} \right) - 1.$$

Together with the zero-query obstruction, this is $\Omega(1 + \alpha M)$. Searching $\lceil \alpha M \rceil$ exact candidate leaves, outputting the atom on a hit and a fresh ν draw otherwise, with the same thinning coin as above, gives an exact sample for every legal guide. $\square$

3.6 Arbitrary known backgrounds

The geometry of the background determines how many light prefix groups are available. Accounting for that geometry gives a sharp worst-case extension of Theorem 3.1.

Let $Q_{\text{atom}}(n, M, \alpha, \delta)$ be the supremum of

$$Q_{\text{safe}}(\{(1 - \alpha)\nu + \alpha\delta_x : x \in S\}, \delta)$$

over known backgrounds ν on $\{0, 1\}^n$ and known sets S of M distinct atoms. Thus ν , S , and α are available without guide queries.

Theorem 3.6 (Unknown atom, known background). *For $n \geq 1$, $2 \leq M \leq 2^n$, $0 < \alpha \leq 1/4$, and $0 \leq \delta \leq \alpha/4$,*

$$Q_{\text{atom}}(n, M, \alpha, \delta) = \Theta\left(1 + \min\left\{\alpha M, \alpha^{-1} \log(1 + 2^n \alpha^2)\right\}\right). \quad (20)$$

The upper bound holds for exact sampling from every such class. The lower bound holds even when the known background has full support. The constants are absolute.

In particular, for fixed $\alpha > 0$ the worst-case rate is $\Theta_\alpha(1 + \min\{M, n\})$. Single-atom classes therefore cannot witness a superlinear fixed-error lower bound in the string length.

Lemma 3.7 (Background partition). *Let $R, C \geq 1$, $a = 1 - \alpha$, and $\tau = \alpha^2/(2RCa)$. There is a known partition of the candidates into G prefix groups, each either of background mass at most τ or a single leaf, with*

$$G \leq \min\left\{M, 1 + \sum_{t=0}^{n-1} \min\{2^t, \tau^{-1}\}\right\}. \quad (21)$$

An exact sampler under (2) has deterministic cap

$$\min\left\{\lceil \alpha M \rceil, \lceil \alpha G \rceil + \lceil \log_{3/2} M \rceil\right\}. \quad (22)$$

The partition and all candidate-set operations use only known data.

Proof. Starting at the root, split a prefix when its background mass exceeds τ and its depth is less than n . The leaves of the resulting finite binary tree partition $\{0, 1\}^n$. At depth t there are at most $\min\{2^t, \tau^{-1}\}$ internal nodes, since their disjoint masses sum to at most one. The number of leaves is one plus the number of internal nodes. Discard groups containing no candidate to obtain (21).

For a light group and every descendant prefix, the background contribution is at most $a\tau$, so Lemma 3.2 tests atom membership with one query. At a leaf, exactness detects the additional mass α . Choose $q = \lceil \alpha G \rceil$ groups with equal inclusion probability q/G . On a hit, proceed with probability $\alpha G/q$. Within a light group, balanced prefix tests reduce the remaining candidate count by a factor at most $2/3$ at each step, as in Proposition 3.3. A leaf group already identifies the atom. Every other branch draws freshly from ν . The atom branch has probability α , and its complement has output law ν , proving exactness and the second cap in (22). Direct leaf search proves the first cap. $\square$

Proof of Theorem 3.6. Write $B = \alpha^{-1} \geq 4$ and $z = 2^n \alpha^2$. For the upper bound fix $R = C = 9/4$. If $z \leq 1$, leaf search costs at most $1 + \alpha M \leq 1 + Bz = O(1 + B \log(1 + z))$. If $z \geq 1$, summing the geometric part and then the constant part in (21) gives

$$G = O(B^2(1 + \log z)).$$

Moreover $\log_2 M \leq n = 2\log_2 B + \log_2 z \leq B + \log_2 z$. Equation (22) therefore bounds the cost by $O(1 + B\log(1 + z))$, as well as by $1 + \alpha M$. This proves the upper bound uniformly, including the transition $z \asymp 1$.

For the lower bound, let

$$d = \min\{n - 1, \lfloor \log_2 B^2 \rfloor\}, \quad H = 2^d, \quad \ell = n - d, \quad K = H\ell.$$

Use the full-support background and K possible atom locations

$$\nu = \frac{15}{16}(U_d \otimes \delta_{0^\ell}) + \frac{1}{16}U_n, \quad x_{y,j} = y0^{j-1}10^{\ell-j} \quad (y \in \{0,1\}^d, 1 \leq j \leq \ell).$$

Here $K \leq 2^n$. Choose any $m = \min\{M, K\}$ of these locations; if $M > K$, add arbitrary distinct strings to make a class of exactly M targets. On the selected m targets use public background answers through depth d and at every proper prefix $y0^t$, with exact answers elsewhere. On any additional targets an exact guide suffices.

Every public ancestor of a selected atom has background mass at least $15/(16H)$. By Lemma 2.2, safety has factor at most $4/3$, and

$$\mathcal{C}(\mu_x, h_x) \leq 1 + \alpha - 2\alpha^2 + \frac{16\alpha^2 H}{15} \leq \frac{9}{8} + \frac{16}{15} = \frac{263}{120} < \frac{9}{4}.$$

Every nonpublic query tests at most one pair (y, j) . The all-negative transcript argument on the uniform prior over the m selected atoms gives

$$q \geq m(\alpha - \delta) - 1 \geq \frac{3}{4}\alpha m - 1.$$

Independently, $q \geq 1$: two distinct targets have distance $\alpha > 2\delta$. Hence the lower bound is $\Omega(1 + \alpha \min\{M, K\})$.

It remains to compare αK with $B\log(1 + z)$. When $d = n - 1$, one has $z \leq 2$ and $\alpha K = Bz/2 \asymp B\log(1 + z)$. Otherwise $B^2/2 < H \leq B^2$, and $\ell = n - \lfloor \log_2 B^2 \rfloor$ differs from $\log_2 z$ by less than one. Since $\ell \geq 1$, this again gives $\alpha K \asymp B\log(1 + z)$. The constants are uniform. Taking the minimum with αM proves (20). $\square$

Proposition 3.8 (Pointwise counterpart). *Define $Q_{\text{atom,pw}}$ by replacing Q_{safe} with Q_{pw} in the definition of Q_{atom} . Under the parameter assumptions of Theorem 3.6,*

$$Q_{\text{atom,pw}}(n, M, \alpha, \delta) = \Theta(1 + \min\{\alpha M, \log(1 + 2^n \alpha)\}). \quad (23)$$

The upper bound holds for exact sampling; full-support backgrounds suffice for the lower bound.

Proof. Let η be uniform on the known candidate set and put $w = (\nu + \eta)/2$. Split the prefix tree using the known mass w and threshold $\tau = \alpha/(4R^2(1 - \alpha))$. A nonleaf terminal group has $\nu(u) \leq 2\tau$. Consequently an empty group has $h(u) \leq \alpha/(2R)$, whereas an occupied group has $h(u) \geq \alpha/R$. All descendant tests have the same gap. Moreover the group contains at most $2\tau M \leq \alpha M$ candidates, so recovery costs $O(1 + \log(1 + \alpha M))$ queries. Terminal leaves are tested exactly.

Writing $z = 2^n \alpha$, the same heavy-node count gives $G = O(\alpha^{-1}(1 + \log z))$ when $z \geq 1$. The same independent search and thinning rule has cap $O(1 + \log z)$, including recovery because $\alpha M \leq z$. When $z < 1$, direct leaf search has constant cap. Combining with the $\lceil \alpha M \rceil$ leaf cap proves the upper bound.

For the lower bound use the background and locations from Theorem 3.6, now with

$$d = \min\{n - 1, \lfloor \log_2(1/(4\alpha)) \rfloor\}, \quad H = 2^d, \quad \ell = n - d, \quad K = H\ell.$$

At every public ancestor, $\nu(u) \geq 15/(16H)$ and $\alpha H \leq 1/4$. Thus $h/\mu \leq 4/3$ and $\mu/h \leq 1 + 4/15 < 4/3$ there; elsewhere the guide is exact. Equality-query counting gives $\Omega(1 + \alpha \min\{M, K\})$ as before. If $d = n - 1$, then $z \leq 1/2$ and $\alpha K = z/2 \asymp \log(1 + z)$. Otherwise $1/(8\alpha) < H \leq 1/(4\alpha)$, $z > 1/2$, and ℓ differs from $\log_2 z + 2$ by less than one. Hence $\alpha K \asymp \log(1 + z)$ in this case as well. $\square$

4 Precision and certification

The atom rate contains an inverse-accuracy regime. The following construction gives exact caps there with a positive, level-normalized guide and bounded local growth. It also quantifies the cost of constructing a global certificate.

For $R > 1$, define the stopped additive defect by

$$d_R(u) = \max\{0, h(u) - R\mu(u), \mu(u) - Rh(u)\}, \quad \mathcal{D}_R(\mu, h) = \max_{A \text{ prefix-free}} \sum_{u \in A} d_R(u). \quad (24)$$

A coherent correction is a known nonnegative additive measure λ with $\lambda(u) \geq d_R(u)$ at every prefix. Its representation must permit evaluation without hidden calls to h . Such a correction of mass B yields a pointwise factor- R guide

$$\bar{h} = \frac{h + \lambda/(R-1)}{1 + B/(R-1)} \quad \text{for} \quad \bar{\mu} = \frac{\mu + \lambda/(R-1)}{1 + B/(R-1)}.$$

Indeed, adding $\lambda/(R-1)$ to each side of $h \leq R\mu + \lambda$ and $\mu \leq Rh + \lambda$ gives both multiplicative inequalities. The leaves remain exact and $\text{TV}(\bar{\mu}, \mu) \leq B/(R-1+B)$.

Theorem 4.1 (Precision and certification). *For every integer $m \geq 3$, put*

$$M = 4^m, \quad n = 3m + 2, \quad \alpha = 2^{-m}, \quad \delta_m = \alpha/4.$$

There is an explicitly represented class $\mathcal{F}_m$ of M full-support targets for which

$$m \leq Q_{\text{pw}}(\mathcal{F}_m, \delta_m) \leq Q_{\text{pw}}(\mathcal{F}_m, 0) \leq m + 5, \quad (25)$$

$$Q_{\text{safe}}(\mathcal{F}_m, \delta_m) = 3 \cdot 2^{m-2}, \quad Q_{\text{safe}}(\mathcal{F}_m, 0) = 2^m. \quad (26)$$

The safe lower bound uses fixed guides that are positive, have mass one at every depth, and satisfy $h(ub) \leq 2h(u)$ and $h \leq 2\mu$.

On this hard guide family the minimum coherent correction mass, at factor $9/4$, is

$$d_m = \alpha - \frac{5/4 + \alpha}{M} > 0. \quad (27)$$

For $B \geq d_m$, the largest uniformly guaranteed probability of returning a fully known valid correction of mass at most B , using at most q guide queries, is exactly

$$\min \left\{ 1, \frac{q + \lfloor B/d_m \rfloor}{M} \right\}. \quad (28)$$

In particular, producing an optimal-mass correction with certainty requires exactly $M - 1$ queries. Success is evaluated against the hidden target; a randomized output need not announce whether it is valid. All statements allow arbitrary adaptive prefix queries and arbitrary computation.

At any fixed positive sampling tolerance, this class eventually requires zero queries: its uniform background lies within α of every target. The theorem concerns shrinking error. Section 5 transfers the certificate obstruction to a fixed-error sampling task.

4.1 The hard guide

Set $k = 2m$ and $a = 1 - \alpha$. Index targets by $\theta \in \{0, 1\}^k$, and let $x_\theta = \theta 0^{m+2}$. Define

$$\mu_\theta = aU_n + \alpha\delta_{x_\theta}. \quad (29)$$

For a prefix of length $t \leq k$, let $h_\theta(u) = 2^{-t}$. For $t = k + j$, $1 \leq j \leq m + 2$, put

$$\beta_j = \frac{2^{j-1}}{2M}, \quad h_\theta(u) = (1 - \beta_j)2^{-t} + \beta_j \mathbf{1}\{u \preceq x_\theta\}. \quad (30)$$

The guide is fixed before the interaction. Its complete formula and the class may be known to the algorithm; only θ is hidden.

Lemma 4.2 (Properties of the hard guide). *The guide (30) is positive, level-normalized, exact at leaves, and satisfies $h \leq 2\mu$ and $h(ub) \leq 2h(u)$. Moreover,*

$$\mathcal{C}(\mu_\theta, h_\theta) = C_m := 2 + \alpha - 2\alpha^2 - \alpha^3 + \alpha^4 \leq \frac{17}{8} < \frac{9}{4}, \quad \mathcal{D}_{9/4}(\mu_\theta, h_\theta) = d_m. \quad (31)$$

The measure $d_m\delta_{x_\theta}$ is an optimal coherent correction.

Proof. The level sums equal one by construction. Since $0 \leq \beta_j \leq \alpha \leq 1/2$, positivity holds and $h \leq \mu/a \leq 2\mu$. At the final depth $\beta_{m+2} = \alpha$, proving exactness.

Before depth k , child values are half their parents. At the first padding step the distinguished child has value $1/M - 1/(4M^2)$, at most its parent's $1/M$. At subsequent steps β doubles. If the parent is distinguished and $v = 2^{-t}$, its distinguished child's value is

$$(1 - 2\beta)v/2 + 2\beta \leq 2((1 - \beta)v + \beta).$$

Every other child comparison is bounded by one half or has a larger parent denominator. This proves local growth.

For the reciprocal capacity, complete any prefix-free set with uncovered leaves. Off the distinguished path, $\mu \leq h$, so its node contribution μ^2/h is at most its total leaf contribution μ . Replacing every such node by leaves can only increase the sum. Thus the maximum is exactly

$$\max_{0 \leq t \leq n} F_t, \quad F_t = 1 - \mu(u_t) + \frac{\mu(u_t)^2}{h(u_t)}, \quad u_t = x_{\theta, 1:t}. \quad (32)$$

For $t \leq k$, writing $p = 2^{-t}$ gives

$$F_t = a + 2a\alpha - a\alpha p + \frac{\alpha^2}{p}.$$

Both variable terms increase as p decreases, so the maximum through depth k is $F_k = C_m$. At the first padding step, direct subtraction gives

$$F_k - F_{k+1} = \frac{\alpha(\alpha + 1)(2\alpha^3 - \alpha^2 - 6\alpha + 4)}{2(\alpha + 2)} \geq 0.$$

For $0 \leq \alpha \leq 1/2$, the last polynomial is at least $4 - 1/4 - 3 = 3/4$. At later padding steps, $h(u_t) \geq 1/M$ and $\alpha \leq \mu(u_t) \leq \mu(u_k)$. Since $Mv^2 - v$ increases for $v \geq 1/M$, these steps also have $F_t \leq F_k$. Finally, $-\alpha^3 + \alpha^4 \leq 0$ and $\alpha - 2\alpha^2 \leq 1/8$, proving the bound on C_m .

The overestimation defect is zero. All other nonzero demands lie on the distinguished path, so the maximum cut demand is its largest node demand. Up to depth k , the underestimation expression is $\alpha - (9/4 - a)2^{-t}$, maximized at k with value d_m . For $j \geq 2$, the guide is no smaller and the target mass is no larger than at depth k . At $j = 1$, the difference between the depth- k and depth- $(k + 1)$ expressions is

$$\frac{a}{2M} - \frac{9/4}{4M^2} > 0.$$

This proves the defect identity. Every demand is dominated by $d_m \delta_{x_\theta}$, and the depth- k demand forces any dominating additive measure to have mass at least d_m . $\square$

4.2 Safe sampling caps

Every query to h_θ is simulated by at most one equality test of a proposed complete key. Queries of length at most k are public. A longer prefix specifies a key; if its padding is not all zero its answer is public, and otherwise the answer is the public baseline in (30) plus β_j exactly when that key is θ . Arbitrary long queries and exact leaf queries are included.

Lemma 4.3 (Equality-query counting). *An adaptive randomized algorithm making at most q equality queries predicts a uniformly random key among M possibilities with probability at most $(q + 1)/M$.*

Proof. Fix the internal randomness and follow the all-negative transcript. It tests at most q distinct keys. Every other key produces that same transcript, whose final guess can be correct for at most one of them. Positive transcripts contribute at most the tested keys. Average over the internal randomness. $\square$

A draw from (29) has its first k bits equal to θ with probability $\alpha + a/M$. A δ -accurate sampler therefore predicts the key with probability at least $\alpha + a/M - \delta$. Lemma 4.3 implies

$$q \geq \alpha(M - 1) - \delta M. \quad (33)$$

For $\delta = \alpha/4$ its right side is $3 \cdot 2^{m-2} - \alpha$; for $\delta = 0$ it is $2^m - \alpha$. Taking integer ceilings proves the two lower bounds in (26).

For matching upper bounds, choose q distinct keys uniformly and query their complete strings x_θ . Exact leaves reveal whether a key was found. On a hit output that string; otherwise output a fresh draw from U_n . For any guide satisfying (1), its output law is

$$(1 - q/M)U_n + (q/M)\delta_{x_\theta}. \quad (34)$$

At $q = 3 \cdot 2^{m-2}$, its TV distance is $(\alpha/4)(1 - 2^{-n}) \leq \delta_m$. At $q = 2^m$ the law is exact. This proves the safe sampling portion of Theorem 4.1 uniformly over all legal guides.

4.3 Pointwise sampling caps

The following sampler applies to every pointwise factor-9/4 guide for (29). Put $d = m + 3$ and partition $\{0, 1\}^n$ by its first d bits into $G = 2^d = 8/\alpha$ equal groups. A group prefix u not containing x_θ has

$$h(u) \leq \frac{9}{4} \frac{a}{G} < \frac{\alpha}{3},$$

whereas a group containing x_θ has $h(u) \geq 4\alpha/9 > \alpha/3$. The same test works at every deeper key prefix.

Choose eight distinct groups uniformly, independently of any proposed output, and query their prefixes. If none passes the threshold, output a fresh U_n draw. On a hit, recover the remaining $k - d = m - 3$ key bits: at each step query the zero child, choosing zero if its value exceeds $\alpha/3$ and one otherwise. Output the recovered string x_θ with its known padding. The query cap is $8 + (m - 3) = m + 5$.

The hit probability is $8/G = \alpha$. The unconditional output law is therefore exactly

$$(1 - \alpha)U_n + \alpha\delta_{x_\theta}.$$

For the lower bound, choose the legal pointwise guide $h = \mu_\theta$. Each prefix has at most two possible answers over the keys. A deterministic q -query decision tree has at most 2^q transcripts, and its final key guess can be correct for at most 2^q keys. Randomization preserves the bound. Consequently

$$2^q \geq M(\alpha - \delta) + a. \quad (35)$$

At $\delta = \alpha/4$, this is strictly larger than 2^{m-1} , so $q \geq m$. The exact case even gives $q \geq m + 1$. This completes (25).

4.4 The certificate tradeoff

For the hard family, a valid coherent correction must put mass at least d_m on the disjoint depth- k prefix naming the true key. A single additive measure of mass B can therefore certify at most $\lfloor B/d_m \rfloor$ keys on a transcript that does not identify one.

Proof of the certification assertion in Theorem 4.1. Fix the procedure's randomness and follow its all-negative equality transcript. At most q keys are tested. For every remaining key it returns the same fully specified measure, valid for at most $\lfloor B/d_m \rfloor$ of them. Including positive transcripts, average success is at most $(q + \lfloor B/d_m \rfloor)/M$. A uniform guarantee cannot exceed this average. Averaging over randomness proves the upper bound on success.

For attainment, if $q \geq M - 1$, identify the key and return its optimal correction. Otherwise query a uniformly random set of q distinct keys. On a hit return $d_m\delta_{x_\theta}$. Otherwise choose uniformly up to $\lfloor B/d_m \rfloor$ untested keys and put mass d_m on each of their distinguished strings. For every selected key, this measure dominates its path demands by Lemma 4.2. Its success probability is the right side of (28), uniformly in the hidden key. If fewer than q queries already determine the key, the procedure may finish early. $\square$

For $B < d_m$, no valid certificate exists. For $B \geq d_m$ and $0 \leq \zeta \leq 1$, success at least $1 - \zeta$ is possible exactly when

$$q + \lfloor B/d_m \rfloor \geq \lceil (1 - \zeta)M \rceil.$$

Since $d_m \geq \alpha/2$ for $m \geq 3$, a fixed budget $B \geq 0$ gives the necessary bound

$$q \geq (1 - \zeta)M - 2B\sqrt{M}. \quad (36)$$

5 Composition

The small spike can be placed after any pointwise-guided sampling problem. This preserves the reliability constants and transfers that problem's fixed-error sampling cost. The certificate lower bound of Theorem 4.1 carries over. For outer and inner distributions ν, σ with guides g, h , define

$$\mu = \nu \otimes \sigma, \quad H(u) = g(u) \ (|u| \leq d), \quad H(xv) = \nu(x)h(v) \ (|x| = d). \quad (37)$$

Exact outer leaves and $h(\emptyset) = 1$ make the definitions agree at $|u| = d$.

Lemma 5.1 (Capacity under concatenation). *Let g, h have root value one, exact leaves, and the same prefix supports as distributions ν on $d \geq 1$ bits and σ on $n \geq 1$ bits. Write $C_{\text{out}} = \mathcal{C}(\nu, g) < \infty$ and $C_{\text{in}} = \mathcal{C}(\sigma, h) < \infty$. For the product guide defined in (37),*

$$\mathcal{C}(\nu \otimes \sigma, H) = C_{\text{in}} + \max_{A \subseteq \{0,1\}^{<d} \text{ prefix-free}} \sum_{u \in A} \left(\frac{\nu(u)^2}{g(u)} - C_{\text{in}} \nu(u) \right) \leq C_{\text{out}} + C_{\text{in}} - 1. \quad (38)$$

The empty set is allowed. If $g \geq \nu/R$, the stronger bound $\mathcal{C}(\nu \otimes \sigma, H) \leq \max\{R, C_{\text{in}}\}$ holds. Repeated concatenation therefore adds at most the capacities' excesses above one.

Proof. Fix the proper outer prefixes A selected by a cut. In every outer leaf subtree not covered by A , an optimal inner cut contributes $C_{\text{in}} \nu(x)$. Filling all those subtrees gives the displayed equality, including zero-mass regions. Exact outer leaves imply

$$\sum_{u \in A} \frac{\nu(u)^2}{g(u)} + 1 - \nu(A) \leq C_{\text{out}}.$$

Since $C_{\text{in}} \geq 1$ (the root cut has value one), substitution proves the additive upper bound. If $\nu(u)^2/g(u) \leq R\nu(u)$, the same formula gives the bound by the maximum. The assertion for repeated concatenation follows by induction. $\square$

Theorem 5.2 (Sequential composition). *Let $\mathcal{P}$ be a known finite class on $\{0,1\}^d$. Let $\nu \in \mathcal{P}$ have an arbitrary pointwise factor- R guide g with exact leaves. Let σ on $\{0,1\}^n$ have a safe guide $h \leq R\sigma$, exact leaves, and stopped capacity at most C . For the product guide (37),*

$$H \leq R\mu, \quad \mathcal{C}(\mu, H) \leq \max\{R, C\}, \quad \mathcal{D}_R(\mu, H) = \mathcal{D}_R(\sigma, h). \quad (39)$$

Now set $R = 9/4$ and take $(\sigma, h) = (\mu_\theta, h_\theta)$ from Section 4, with $M = 4^m$ and $\alpha = 2^{-m}$. Let $Q_\otimes(\mathcal{P}, m, \delta)$ denote the minimax query complexity of the class $\{\nu \otimes \mu_\theta : \nu \in \mathcal{P}, \theta \in \{0,1\}^{2m}\}$ over precisely the product guides (37). For $\alpha < \delta < 1$,

$$Q_{\text{pw}}(\mathcal{P}, \delta) \leq Q_\otimes(\mathcal{P}, m, \delta) \leq Q_{\text{pw}}(\mathcal{P}, \delta - \alpha). \quad (40)$$

Every product instance has minimum coherent correction mass d_m . A procedure with cap q returning a fully known correction of mass at most $B \geq d_m$ has uniformly guaranteed success at most

$$\min \left\{ 1, \frac{q + \lfloor B/d_m \rfloor}{M} \right\}. \quad (41)$$

Optimal-mass certification with certainty has query complexity between $M - 1$ and $M + |\mathcal{P}| - 2$.

Proof. Safety and exact leaves follow directly. Lemma 5.1 gives the capacity bound, including cuts of mixed depths and zero-mass subtrees.

The outer additive defects vanish. Below an outer leaf x , every defect is $\nu(x)$ times the corresponding inner defect. Each cut therefore has demand at most $\mathcal{D}_R(\sigma, h) \sum_x \nu(x)$. Taking a maximizing inner cut in every positive-mass outer subtree attains equality. An optimal inner correction λ gives the product correction $\nu \otimes \lambda$.

For the upper bound in (40), sample ν to error $\delta - \alpha$ using outer queries, and append an independent free draw from U_n . The product error is at most δ , since $\text{TV}(\sigma, U_n) \leq \alpha$. For the lower bound, fix θ publicly. Any query to H can be simulated using at most one query to g : a long query uses $g(x)h_\theta(v)$. Projection of the output onto its first d bits gives a δ -accurate sample from ν .

For certification, fix ν and g publicly instead. Every query to H is simulated by at most one equality query about θ . At each prefix $x\theta$, a valid correction needs mass at least $\nu(x)d_m$. Summing over outer leaves, it needs mass at least d_m on the set of strings whose inner key is θ . Those sets are disjoint for different keys. The all-negative transcript thus certifies at most $\lfloor B/d_m \rfloor$ untested keys, proving (41). For the upper bound, identify ν with at most $|\mathcal{P}| - 1$ distinguishing queries at depth d . Choose a known positive-mass outer leaf, identify the inner key with at most $M - 1$ queries, and return $\nu \otimes d_m \delta_{x_\theta}$. $\square$

Corollary 5.3 (Exponential certificate cost). *For all sufficiently large m , there is a full-support class on $4m + 2$ bits, of logarithmic cardinality $\Theta(m)$, and a specified family of safe guides with $R = C = 9/4$, for which*

$$Q_{\otimes}(\mathcal{P}, m, 1/16) = \tilde{\Theta}(m^{3/2}), \quad \mathcal{D}_{9/4}(\mu, H) = d_m \rightarrow 0.$$

Producing an optimal-mass correction with certainty requires $\Theta(4^m)$ queries. More generally, a fixed mass budget B and success probability $1 - \zeta$ require

$$q \geq (1 - \zeta)4^m - 2B2^m.$$

The sampling rates concern the specified product guides.

Proof. Take the class of Theorem B.1 on $d = m$ bits, with the parameters of Appendix B.4; it has at most $\lfloor e^m \rfloor$ members. It may be made full-support by replacing each target and its guide by their $7/8$ mixture with the uniform law and its exact prefix masses. The guide remains pointwise factor $9/4$, and each query to the smoothed guide is simulated by one original query. In the hard-prior proof, the target root predicts the hidden bit with probability at least $3/5$. Smoothing makes this at least $47/80$; error $1/16$ still implies advantage $1/40$ and hence transcript distance at least $1/20$. The proof's bound $65/2048 < 1/20$ for its small-query regime remains contradictory. Thus the same order of lower bound survives smoothing.

Apply Theorem 5.2 and use $\alpha \leq 1/32$. Its upper bound is at most $Q_{\text{pw}}(\mathcal{P}, 1/32)$, which is $O(m^{3/2} \log(2m))$ by [1, Theorem 3.2] at error $1/32$. The lower bound is $\Omega(m^{3/2}/\sqrt{\log(2m)})$. The total length is $m + (3m + 2)$ and the class has between 4^m and $4^m e^m$ members. The certificate bounds follow from $M - 1 \leq q \leq M + e^m - 2$ for certainty and $d_m \geq \alpha/2$ for fixed B . $\square$

The fixed-error sampling lower bound comes from the outer task; the inner component, which carries the certificate cost, is negligible at fixed sampling accuracy.

6 The fixed-error gap

For $R = C = 9/4$ and fixed $0 < \delta \leq 1/16$, the general bounds are

$$\tilde{\Omega}(K(n, M)) \leq Q_{\text{safe}}(n, M, \delta) \leq \tilde{O}(\min\{M, n \log M, n^2\}). \quad (42)$$

The lower bound follows from the pointwise submodel. Exact identification, the frontier learner in Appendix A, and Theorem C.1 give the upper bounds with deterministic caps. The unresolved question is whether $\tilde{O}(K(n, M))$ holds in the safe model, or whether some classes require asymptotically more. Theorem 2.1 rules out transferring the *class-specific* pointwise cost unchanged. Since its parameters have $M = \Theta(n)$, it leaves the worst-case benchmark intact.

The atom theorem narrows the search for harder examples. For fixed α , every known-background single-atom class has an $O_\alpha(n)$ sampler. Its proof works because a known mass partition and the

singleton consequence of capacity supply exact local tests. An arbitrary finite class has no such decomposition. The composition theorem preserves a hard sampling task while adding a costly global certificate, but its certificate-hard component is negligible at fixed error.

A possible route to the upper bound is to replace the pointwise learner’s two-sided candidate tests by tests that preserve the target using only safety. For a first-occurrence frontier on which $\mu(u)/h(u) > L$, capacity bounds its target mass by C/L . That observation alone does not justify trimming the guide: the learner may query a low-mass prefix and eliminate the target, changing all later output branches. The frontier learner of Appendix A handles these adaptive histories at a cost proportional to $n \log M$, up to logarithms; the square-root dependence remains the missing step.

6.1 Related work

Block and Polyanskiy [2, Theorems 3 and 5] characterize approximate-rejection-sampling rates up to logarithmic factors under divergence bounds when the output is selected from independent reference samples. Uniform Race [3, Corollary 1 and Appendix D] achieves inverse-budget error under a second-moment bound without an input acceptance threshold. These results already establish inverse-accuracy phenomena in that access model. Here arbitrary adaptive prefix queries and outputs are allowed. Our atom bounds charge for obtaining the local information that a mixture sampler needs; the mixture identity itself is elementary.

Poisson branching and independent-offspring particle methods predate this work [4, 5]. Appendix C specializes them to absolute prefix masses and proves the unconditional error and hard cap directly.

AI disclosure

GPT-6 and Claude were used to assist with mathematical proofs, writing, and formalization.

A A safe frontier learner

Assume $0 < \delta < 1$, $M \geq 2$, and (1) and (2). Define

$$J_0 = 1 + \lceil \log_2 M \rceil, \quad p_0 = \frac{\delta^2}{32RC}, \quad T = \left\lceil p_0^{-1} \log \frac{4J_0}{\delta} \right\rceil, \quad J = \left\lceil 4R \log \frac{4}{\delta} \right\rceil.$$

The query cap is $2n(J_0T + J)$, with TV error at most δ .

Maintain a version space G containing μ , and its uniform mixture η . A scan draws $X \sim \eta$ for free and follows its path, querying both children at each internal prefix. Stop at the first parent having a child v with $h(v) > 2R\eta(v)$. Retain only candidates with $\phi(v) \geq h(v)/R$. The target survives by safety, and fewer than half the candidates remain, because every retained candidate has mass strictly greater than twice the mixture mass. A singleton is sampled directly.

For a fixed stage let A be its possible first-violation parents. This is a prefix-free set, with $h(u) \leq 2R\eta(u)$ at every $u \in A$. Writing $p = \eta(A)$ and $b = \mu(A)$, weighted Cauchy–Schwarz gives

$$b^2 \leq \left(\sum_{u \in A} \frac{\mu(u)^2}{h(u)} \right) \left(\sum_{u \in A} h(u) \right) \leq 2RCp.$$

Zero terms are omitted. All leaves outside these subtrees satisfy $\mu(x) = h(x) \leq 2R\eta(x)$.

Run at most T independent scans per stage. A violation updates the class. If all pass, freeze it. Conditional on the entire history before a stage, its mixture and frontier are fixed. If $p > p_0$, a false

pass has probability at most $e^{-Tp_0} \leq \delta/(4J_0)$. Union bound over the capped number of stages gives probability at most $\delta/4$ of freezing at such a stage. Otherwise $b \leq \delta/4$.

Use J fresh scans for production. Reject stopped scans, and accept a good leaf with probability $\mu(x)/(2R\eta(x))$, available from its leaf query. Its accepted law is μ conditioned on good leaves; a trial accepts with probability $(1-b)/(2R) \geq 1/(4R)$. Exhaustion has probability at most $e^{-J/(4R)} \leq \delta/4$. A fixed fallback and the three error contributions give TV at most $3\delta/4$. Each scan costs at most $2n$ calls.

B Proof of the pointwise characterization

This appendix proves Theorem 1.1, including the simulation of arbitrary prefix queries for the lower bound. Here $Q(n, M) = Q_{\text{pw}}(n, M, 1/16)$ and

$$(4/9)\mu(u) \leq h(u) \leq (9/4)\mu(u), \quad h(x) = \mu(x) \quad (|x| = n). \quad (43)$$

A table with r^2 rows hides each label behind one of ℓ locations. Reusing the table in r output blocks gives length about $r\ell$, logarithmic class size $O(r^2 \log \ell)$, and query cost $\Omega(r^2 \ell)$.

Theorem B.1 (Parameterized lower bound). *Let $r \geq 8$ be a power of two and let $\ell \geq 2$ be an integer. Put*

$$N = 1 + r(2 \log_2 r + \ell), \quad M_0 = 2(2\ell)^{2r^2}.$$

There is a class of at most M_0 distributions on $\{0, 1\}^N$, with one fixed legal oracle for each distribution, on which every sampler with error at most $1/16$ on every instance has query cap

$$q > 2^{-16}(\ell - 1)r^2. \quad (44)$$

The same lower bound holds on $\{0, 1\}^n$ for every $n \geq N$.

B.1 The hard class

Fix the parameters of Theorem B.1, and set

$$W = r^2, \quad k = \log_2 W = 2 \log_2 r, \quad d = k + \ell, \quad m_{\pm} = W/2 \pm r.$$

An instance θ consists of a bit B , two location tables $I_{a,j} \in [\ell]$, and two label tables $Y_{a,j} \in \{0, 1\}$, where $a \in \{0, 1\}$ and $j \in [W]$. The label table at root B has exactly m_+ ones; the other has exactly m_- ones.

For the hard prior, B is uniform. Conditional on B , the two label tables are independently uniform among vectors of their specified weights. All $2W$ locations are independent and uniform on $[\ell]$, independently of the labels and B .

Use k bits to encode a row $j \in [W]$. Among the ℓ -bit strings, define

$$c = 1^\ell, \quad s_i = 1^{i-1}0^{\ell-i+1} \quad (i \in [\ell]).$$

A block is a pair $z = (j, s)$ of a row address and an ℓ -bit string. Define its weight at root a by

$$w_a(j, s) = \begin{cases} 1, & s = c, \\ 1 + Y_{a,j}/2, & s = s_{I_{a,j}}, \\ 0, & \text{otherwise.} \end{cases} \quad (45)$$

Write

$$A_{\pm} = 2W + \frac{m_{\pm}}{2} = \frac{9W}{4} \pm \frac{r}{2}, \quad A_{\circ} = \frac{9W}{4}, \quad Z = A_+^r + A_-^r. \quad (46)$$

For a given instance, let $A_a = A_+$ when $a = B$, and $A_a = A_-$ otherwise. Indeed, $\sum_z w_a(z) = A_a$.

For a complete string $x = (a, z_1, \dots, z_r)$ of length $N = 1 + rd$, set

$$\mu_{\theta}(x) = \frac{1}{Z} \prod_{t=1}^r w_a(z_t). \quad (47)$$

The total weight is Z , so this is a distribution. Conditional on its first bit, its blocks are independent. They use identical weights from one reused table.

Lemma B.2. *The number of targets is at most $2(2\ell)^{2W}$, and every target satisfies*

$$\mu_{\theta}(X_1 = B) \geq \frac{3}{5}. \quad (48)$$

The normalizer Z is independent of the instance.

Proof. The parameter count is

$$2 \binom{W}{m_+} \binom{W}{m_-} \ell^{2W} \leq 2(2\ell)^{2W}.$$

A target determines its tables: holding all but one block at common strings exposes each weight $w_a(j, s)$ through (47). Thus distinct parameter instances give distinct targets. Both root normalizers occur once in Z , regardless of B and the table contents. The odds of the first bit being B are

$$\left(\frac{A_+}{A_-} \right)^r = \left(\frac{1 + 2/(9r)}{1 - 2/(9r)} \right)^r.$$

The inequality $\log((1+x)/(1-x)) \geq 2x$ for $0 \leq x < 1$ gives a logarithm of at least $4/9$. Finally, $e^{4/9} \geq 1 + 4/9 + (4/9)^2/2 = 125/81 > 3/2$. The probability is therefore at least $3/5$. $\square$

B.1.1 The oracle

Set $h_{\theta}(\emptyset) = 1$. For a nonempty prefix u , parse its first bit as a , followed by t complete blocks and a possibly empty partial block v , with $|v| < d$. Let

$$P = \prod_{b=1}^t w_a(z_b).$$

If $P = 0$, set $h_{\theta}(u) = 0$. Otherwise use the following definitions.

At a block boundary, set

$$h_{\theta}(u) = P A_{\circ}^{r-t} / Z \quad (v = \emptyset). \quad (49)$$

For a nonempty partial address v with $|v| < k$, let $J = 2^{k-|v|}$ be its number of compatible rows, and set

$$h_{\theta}(u) = P \left(\frac{9}{4} J \right) A_{\circ}^{r-t-1} / Z. \quad (50)$$

When $|v| \geq k$, write $v = (j, s)$ with $|s| < \ell$, allowing $s = \emptyset$. Define

$$g_a(j, s) = \begin{cases} 3/2, & s = 1^{|s|}, \\ 1 + Y_{a,j}/2, & s = 1^{i-1}0^{|s|-i+1} \text{ for some } i \leq |s|, I_{a,j} = i, \\ 0, & \text{otherwise,} \end{cases} \quad (51)$$

and set

$$h_\theta(u) = P g_a(j, s) A_\circ^{r-t-1} / Z. \quad (52)$$

The dependence of g_a on θ is implicit, as in w_a . These formulas define one map before any interaction takes place.

Lemma B.3 (Oracle validity). *For every instance and every prefix,*

$$\frac{9}{20} \mu_\theta(u) \leq h_\theta(u) \leq \frac{20}{9} \mu_\theta(u).$$

At complete strings $h_\theta = \mu_\theta$. Hence the oracle satisfies (43).

Proof. Let $b_a(v)$ be the sum of weights of blocks extending the partial block v . The exact prefix mass is

$$\mu_\theta(u) = \begin{cases} P A_a^{r-t} / Z, & v = \emptyset, \\ P b_a(v) A_a^{r-t-1} / Z, & v \neq \emptyset. \end{cases} \quad (53)$$

For $x = 2/(9r) \leq 1/9$ and $0 \leq s \leq r$,

$$r \log(1+x) \leq 2/9, \quad -r \log(1-x) \leq \frac{rx}{1-x} \leq 1/4.$$

Since $e^{1/4} < 4/3$, it follows that

$$\frac{3}{4} \leq (A_\circ/A_a)^s \leq \frac{4}{3}. \quad (54)$$

For an incomplete address, $2J \leq b_a(v) \leq (5/2)J$. Replacing it by $(9/4)J$ changes it by a multiplicative factor at most $9/8$ in either direction.

For a fully specified address and an all-ones partial string s , the common block is always compatible. The special block is either also compatible or already excluded. Consequently, the exact local mass is one of $1, 2, 5/2$. The value $3/2$ approximates it within a factor $5/3$. For a partial string containing a zero, the local mass is exactly the second or third case of (51).

Zero masses agree in every case. Combining the local factor at most $5/3$ with (54) gives the factor $20/9 < 9/4$. At a complete string $t = r$ and $v = \emptyset$, formula (49) is precisely (47). $\square$

B.2 Simulating prefix queries

A long prefix can specify special blocks in several rows. We bound such queries by giving the algorithm a stronger oracle.

Definition B.4 (First-failure batch oracle). A batch specifies an ordered list of at most r guesses of the form $I_{a,j} = i$. The oracle tests them in order. A success reports the corresponding label $Y_{a,j}$. The first failure is reported and ends the batch. If all guesses succeed, all corresponding labels are reported. Previously learned information remains available to the algorithm.

Thus one batch can expose several labels. Its cost is one, regardless of its length.

Lemma B.5 (Simulation). *Every query to h_θ can be answered using at most one batch, with no additional access to the hidden instance. This includes queries to complete strings.*

Proof. Parse the query as in Appendix B.1. A complete common block has known weight one. A complete block with a zero either has the form s_i and demands the equality $I_{a,j} = i$, or has weight zero for every instance. A partial block with a zero behaves in the same way. A partial address or an all-ones partial block has the fixed coefficient in (50) or (51).

Collect the demanded equalities. There are at most r , since a nonempty partial block leaves at most $r - 1$ complete blocks. Repeated consistent guesses for a row can be combined; two different guesses for one row force the answer to be zero. Invalid string patterns also force zero. Guesses already decided by the transcript require no test.

For the remaining list, a failure means that a factor of the queried mass is zero. If every guess succeeds, the batch reveals every label needed to calculate the product P and the possible final special factor. The coefficients $A_\circ$ and Z are known and instance-independent. This computes the exact oracle answer. The additional individual labels and the identity of the first failed guess only give the batch algorithm more information. $\square$

B.2.1 Discovering rows

A row is called discovered when its label has been reported following a successful location test. For an undiscovered row, let $f_{a,j}$ be the number of distinct excluded locations.

Lemma B.6 (Conditional law). *Fix $B = b$ and condition on an enhanced transcript. The locations of undiscovered rows are independent and uniform on their remaining locations. They are independent of the remaining labels. If table a has exposed t_a labels, of which s_a equal one, then its remaining labels are uniform subject to having $m_{a,b} - s_a$ ones, where $m_{a,b} = m_+$ for $a = b$ and m_- otherwise.*

Proof. First fix the internal randomness. For a fixed transcript, every query and the order of tests are fixed. Its consistency conditions specify, in each row, either an exact discovered location or a set of excluded locations. Separately they specify the labels at discovered rows. The prior on locations is a product measure, independent of the two fixed-weight label measures. Thus the event defining the transcript factors into location constraints and observed-label constraints. Conditioning gives the asserted laws. Restoring the independent random seed in the transcript leaves the argument unchanged. $\square$

Lemma B.7 (Search potential). *For any adaptive algorithm making at most q batches, let D be its number of discovered rows. Under either value of B ,*

$$\mathbb{E}_b D \leq \frac{2q}{\ell - 1}. \quad (55)$$

The same bound holds if the process is stopped partway through a batch.

Proof. Consider the nonnegative potential

$$\Phi = \frac{\ell - 1}{2} D + \frac{1}{2} \sum_{\text{undiscovered } (a,j)} f_{a,j}. \quad (56)$$

Initially it is zero. Suppose a fresh guess is made in an undiscovered row with f excluded locations, leaving $d' = \ell - f$ possibilities. By Lemma B.6, success has conditional probability $1/d'$. On success the potential increases by

$$\frac{\ell - 1}{2} - \frac{f}{2} = \frac{d' - 1}{2}.$$

On failure it increases by $1/2$. Its conditional expected increase is therefore

$$\frac{1}{d'} \frac{d' - 1}{2} + \frac{d' - 1}{d'} \frac{1}{2} = \frac{d' - 1}{d'},$$

which is the conditional probability of failure. Tests whose outcomes were already known give no new information and have potential increment zero, still bounded in expectation by their failure probability.

There is at most one failure per batch. Summing the conditional inequalities over the at most qr elementary tests, including only active tests in a stopped process, gives $\mathbb{E}_b \Phi \leq q$. The first term in (56) now proves (55). $\square$

B.3 An information bound

Let T_b be the full enhanced transcript under $B = b$, including the random seed. We next show that few batches leave the two transcript laws close. Stopping is needed because a nearly complete fixed-weight label table can reveal B with certainty.

Lemma B.8. *For every algorithm with batch cap q , put*

$$\eta = \frac{q}{(\ell - 1)W}.$$

Then

$$\text{TV}(T_0, T_1) \leq 8\sqrt{\eta} + 32\eta. \quad (57)$$

Proof. Stop the process immediately upon its $L = W/8$ -th discovery, or when it otherwise ends. Let T_b^* denote this stopped transcript. Before a new label is revealed, let t be the number of previously discovered labels in its table, and s the number of their ones. Here $t < L$. By Lemma B.6, the label has success probabilities, in some order under the two hypotheses,

$$p = \frac{m_+ - s}{W - t}, \quad p' = \frac{m_- - s}{W - t}.$$

For $r \geq 8$, $m_{\pm} \in [3W/8, 5W/8]$. Since $t \leq W/8$ and $0 \leq s \leq t$, both probabilities lie in $[1/4, 3/4]$, and

$$|p - p'| = \frac{2r}{W - t} \leq \frac{16}{7r} < \frac{3}{r}. \quad (58)$$

The elementary inequality $\log x \leq x - 1$ implies

$$D_{\text{KL}}(\text{Ber}(p) \parallel \text{Ber}(p')) \leq \frac{(p - p')^2}{p'(1 - p')} \leq \frac{48}{W} \leq \frac{64}{W}.$$

All location-test outcomes have identical conditional probabilities under the two hypotheses at a common history. Only newly revealed labels contribute divergence. The chain rule, followed by Lemma B.7, gives

$$D_{\text{KL}}(T_0^* \parallel T_1^*) \leq \frac{64}{W} \mathbb{E}_0 D^* \leq \frac{128q}{(\ell - 1)W} = 128\eta. \quad (59)$$

All stopped histories have common support: any assignment to at most $W/8$ exposed labels in either table is compatible with both fixed weights. The random seed contributes zero initial divergence.

Pinsker's inequality gives $\text{TV}(T_0^*, T_1^*) \leq 8\sqrt{\eta}$. By Markov's inequality and Lemma B.7, under each hypothesis the probability that the stopping threshold is reached is at most

$$\frac{\mathbb{E}_b D}{L} \leq 16\eta.$$

Full and stopped transcripts agree outside this event. Formally, embed stopped transcripts in the full transcript space by padding the unrevealed continuation with a fixed cemetery symbol. Coupling each full process to its stopped version and using the triangle inequality adds at most 16η for each hypothesis. This proves (57). $\square$

For completeness, the information inequalities used above have elementary finite-space proofs. The chain rule follows by writing a joint likelihood ratio as the product of its conditional likelihood ratios and taking expectations. To obtain Pinsker's inequality, take the event E on which one mass function exceeds the other and apply the log-sum inequality to E and its complement. The resulting divergence between $\text{Ber}(p)$ and $\text{Ber}(p')$ is at least $2(p - p')^2$: its second derivative with respect to p is $1/[p(1 - p)] \geq 4$, and it and its first derivative vanish at $p = p'$. Thus $D_{\text{KL}}(P\|Q) \geq 2 \text{TV}(P, Q)^2$. Conditioning on the independent random seed gives the same statements for randomized algorithms.

Proof of Theorem B.1. Use the known class from Appendix B.1 and its fixed oracles. If a sampler were valid on every instance, its first output bit would satisfy

$$\mathbb{P}_\theta(X_1 = B) \geq \frac{3}{5} - \frac{1}{16} = \frac{1}{2} + \frac{3}{80}$$

for every θ , by Lemma B.2. This remains true after averaging over the hard prior. Hence the first output bit would distinguish the uniform bit B with success at least $1/2 + 3/80$.

For any two transcript laws, the success of a test of a uniform bit is at most $(1 + \text{TV}(T_0, T_1))/2$. This follows directly by applying the definition of total variation to the event on which the test outputs zero. Thus the sampler would require transcript distance at least $3/40$.

Lemma B.5 simulates its queries with the same cap in the stronger model. If $q \leq 2^{-16}(\ell - 1)W$, Lemma B.8 instead gives

$$\text{TV}(T_0, T_1) \leq \frac{8}{256} + \frac{32}{65536} = \frac{65}{2048} < \frac{3}{40},$$

a contradiction. The class size and oracle validity were already proved.

To extend from N to $n \geq N$, append deterministic zeros to every target string. For a prefix of length at most N , retain the old oracle. For a longer prefix, return the old exact leaf probability when its padding bits are all zero, and return zero otherwise. This is legal, exact at the new leaves, and every new query can be simulated with at most one old query. Taking the first N output bits cannot increase total variation distance. $\square$

B.4 Choice of parameters

B.4.1 The exponential-size regime

For sufficiently large n , let r be the largest power of two at most

$$\sqrt{\frac{n}{4 \log(2n)}}, \quad \ell = \left\lfloor \frac{n-1}{r} \right\rfloor - 2 \log_2 r.$$

Then $r \geq 8$, $N \leq n$, and

$$r \geq \frac{1}{4} \sqrt{\frac{n}{\log(2n)}}, \quad \ell - 1 \geq \frac{n}{2r}. \quad (60)$$

For the second inequality, observe that $r \leq n/[16 \log(2n)]$ once $n \geq 64 \log(2n)$, while $2 \log_2 r \leq 3 \log(2n)$. The floor loses less than $1 + 1/r$, and these bounds give the claim.

Moreover,

$$\log M_0 \leq \log 2 + 2r^2 \log(2n) \leq \log 2 + n/2.$$

This is smaller than $\log \lfloor e^n \rfloor$ for all sufficiently large n . Theorem B.1 and (60) give the lower bound in (5). The finite-class upper bound gives its other side.

B.4.2 A small-class lower bound

We include the comb argument from the small-class part of [1, Theorem 4.12], with root value one and its cap made explicit.

Lemma B.9. *For every $n, M \geq 2$,*

$$Q(n, M) \geq \frac{3}{16} \min\{M, n\}.$$

Proof. Put $S = \min\{M, n\}$. For $0 \leq i < S$, set $v_i = 1^i 0^{n-i}$ and let $v_n = 1^n$. Define

$$\mu_i = \frac{1}{2} \delta_{v_i} + \frac{1}{2} \delta_{v_n}.$$

The oracle is $1/2$ on nonempty prefixes of either support string and zero elsewhere; set its value at the empty prefix to one. It is exact at leaves and is a factor-two approximation.

Fix the random seed and run the algorithm on a reference map that is $1/2$ on nonempty all-ones prefixes and zero elsewhere, with value one at the empty prefix. Every query off this common path can test at most one index i . If an index is never tested, the true oracle gives exactly the reference transcript. Consequently, among the S targets, at most q can have a different transcript, and the single reference output can equal v_i for at most one further index. Averaging over uniform i and then over the seed gives

$$\mathbb{P}(X = v_i) \leq \frac{q+1}{S}.$$

Validity requires this average to be at least $1/2 - 1/16 = 7/16$, so $q \geq 7S/16 - 1$. For $S \geq 4$ this is at least $3S/16$. For $S = 2, 3$, zero queries are impossible: two of the targets have total variation distance $1/2$, and no single output law is within $1/16$ of both. Thus $q \geq 1 \geq 3S/16$ also in these cases. $\square$

B.4.3 Proof of the lower bound

Lower bound in Theorem 1.1. Write $H = \log M$ and $L = \log(2n)$. Suppose first that $n \geq 1024$, and define

$$R_* = \min \left\{ \sqrt{\frac{H}{4L}}, \frac{n}{16L} \right\}.$$

When $R_* \geq 8$, take r to be the largest power of two at most R_* and let

$$\ell = \left\lfloor \frac{n-1}{r} \right\rfloor - 2 \log_2 r.$$

As above, $N \leq n$ and $\ell - 1 \geq n/(2r)$. Also $H \geq 256L$, so

$$\log M_0 \leq \log 2 + 2r^2 L \leq \log 2 + H/2 \leq H.$$

Since $r \geq R_*/2$, Theorem B.1 yields

$$Q(n, M) \geq c_0 \min \left\{ n\sqrt{\frac{H}{L}}, \frac{n^2}{L} \right\} \geq \frac{c_0}{L} K(n, M) \quad (61)$$

for an absolute constant $c_0 > 0$.

When $R_* < 8$, the inequality $n/(16L) \geq 8$ for $n \geq 1024$ implies $H < 256L$. If $M \leq n$, Lemma B.9 immediately suffices. If $M > n$, then

$$K(n, M)/L \leq n\sqrt{H}/L < 16n/\sqrt{L} \leq 16n,$$

and the same lemma again suffices. Finally, for $2 \leq n < 1024$, one has $Q(n, M) \geq 1$ and $K(n, M) \leq n^2$, so reducing the universal constant covers this finite range of n , uniformly in M . $\square$

B.4.4 Upper bounds

Proposition B.10. *For every $n, M \geq 2$,*

$$Q(n, M) \leq \min \left\{ M - 1, c_3 n \sqrt{\log M} \log(2 + \log M), c_3 n^2 \right\} \quad (62)$$

for an absolute constant c_3 .

Proof. For exact identification, maintain the candidates consistent with all queried leaf values. Whenever two distinct distributions remain, choose a leaf where their probabilities differ and query it. At least one candidate is eliminated. After at most $M - 1$ queries, the remaining candidates are one distribution, from which sampling is free.

For the second bound, apply [1, Theorem 3.2] with ratio 9/4, binary alphabet, and error 1/16. Its stated guarantee already bounds the number of queries on every execution; its proof implements the cap using a query tail bound.

The last bound follows from Theorem C.1, since a pointwise factor-9/4 guide satisfies safety and stopped capacity with $R = C = 9/4$. $\square$

Upper bound in Theorem 1.1. If $K(n, M) = M$, use the first bound in (62). If $K(n, M) = n^2$, use its last bound. In the remaining case $K(n, M) = n\sqrt{\log M}$, so $\log M \leq n^2$ and $\log(2 + \log M) = O(\log(2n))$. The middle bound in (62) then proves (4). A singleton class needs no oracle calls. $\square$

C A quadratic upper bound

Poisson tree particle filters have unbiased unnormalized estimators [4, Section 3 and Proposition 4.1]. Golowich, Chen, Rohatgi, et al. [5, Appendix E] analyze an independent-offspring formulation in their inference model. We give the complete argument for the present prefix oracle. It uses absolute guide masses and includes extinction, normalization, and a deterministic query cap.

Theorem C.1 (Poisson branching). *Let $n \geq 1$. Suppose $h(\emptyset) = 1$, h has the same support as μ at every prefix, and its leaves are exact. Let public finite bounds A, B satisfy*

$$\sum_{t=0}^{n-1} \sum_{|u|=t} h(u) \leq A, \quad \sum_{t=1}^n \sum_{|u|=t} \frac{\mu(u)^2}{h(u)} \leq B. \quad (63)$$

For $0 < \delta < 1$, a sampler with unconditional error at most δ has deterministic cap

$$P = \lceil B/\delta^2 \rceil, \quad q = \lceil 4PA/\delta \rceil. \quad (64)$$

In particular, one-sided safety $h \leq R\mu$ and stopped capacity at most C give $A = Rn$, $B = Cn$, and $O(RCn^2/\delta^3 + Rn/\delta)$ queries.

Proof. Start one particle at the root. A particle at a nonleaf u has, independently for each $b \in \{0, 1\}$, a Poisson number of children at ub with mean $h(ub)/h(u)$. Query both children when processing the particle; retain the parent's value from its creation. The root value is known. Zero-valued children receive no particles. Finite depth and finite means make every generated tree finite almost surely. We count repeated calls separately.

Let $N(u)$ be the number of particles at u . Induction gives $\mathbb{E}N(u) = h(u)$. If Y is the counting measure of terminal particles and $L = Y(\{0, 1\}^n)$, exact leaves imply

$$\mathbb{E}Y(x) = \mu(x), \quad \mathbb{E}L = 1. \quad (65)$$

The expected query count of a tree is $2 \sum_{t < n} \sum_{|u|=t} h(u) \leq 2A$.

Define $z(u) = \mu(u)/h(u)$ on the positive support, and zero otherwise. With respect to the generated populations,

$$Z_t = \sum_{|u|=t} N(u)z(u)$$

is a martingale: a parent particle at u has expected child contribution $\sum_b \mu(ub)/h(u) = z(u)$. Independent Poisson offspring and $\mathbb{E}N(u) = h(u)$ give

$$\mathbb{E}[(Z_t - Z_{t-1})^2] = \sum_{|v|=t} \frac{\mu(v)^2}{h(v)}.$$

Since $Z_0 = 1$ and $Z_n = L$, orthogonality of the martingale increments proves $\text{Var } L \leq B$.

Generate P independent trees and pool their leaves. Let T be the total leaf count. Conditional on the pool, let ρ be its uniform empirical probability measure if $T > 0$, and a fixed point mass if $T = 0$. Draw the output from ρ . By (65), including extinction,

$$\mathbb{E} \left[\frac{T}{P} \rho \right] = \mu.$$

Consequently the untruncated output law satisfies

$$\text{TV}(\mathbb{E}\rho, \mu) \leq \frac{1}{2} \mathbb{E} \left| 1 - \frac{T}{P} \right| \leq \frac{1}{2} \sqrt{\text{Var}(T/P)} \leq \frac{1}{2} \sqrt{B/P} \leq \frac{\delta}{2}.$$

The expected total number of queries is at most $2PA$. Abort before making query $q + 1$ and return the fixed fallback. Markov's inequality bounds this event by $2PA/q \leq \delta/2$. Coupling capped and uncapped executions adds at most $\delta/2$ to their output distance. Thus the cap in (64) includes every call in successful, extinct, and aborted trees.

Finally, safety bounds each level sum of h by R , and each level is an antichain, so capacity bounds each reciprocal sum by C . This proves the stated choices of A and B . $\square$

References

- [1] Noah Golowich, Ankur Moitra, and Dhruv Rohatgi. *The Power of Test-Time Training for Approximate Sampling*. arXiv:2606.11437v1, 2026. <https://arxiv.org/abs/2606.11437v1>.
- [2] Adam Block and Yury Polyanskiy. The Sample Complexity of Approximate Rejection Sampling With Applications to Smoothed Online Learning. In *Proceedings of the 36th Conference on Learning Theory, Proceedings of Machine Learning Research* **195**:228–273, 2023. <https://proceedings.mlr.press/v195/block23a.html>.
- [3] Seiyun Shin, Juhyeong Pang, and Kwang-Sung Jun. *Uniform Race: Parameter-Free Approximate Rejection Sampling*. arXiv:2609.34639v1, 2026. <https://arxiv.org/abs/2609.34639v1>.
- [4] Tomasz Cąkała, Błażej Miasojedow, and Wojciech Niemiro. *Particle MCMC with Poisson Resampling: Parallelization and Continuous Time Models*. *Journal of Computational and Graphical Statistics* **30**(3):671–684, 2021. Cited by its arXiv numbering, arXiv:1707.01660v2. <https://arxiv.org/abs/1707.01660v2>.
- [5] Noah Golowich, Fan Chen, Dhruv Rohatgi, Raghav Singhal, Carles Domingo-Enrich, Dylan J. Foster, and Akshay Krishnamurthy. *Reject, Resample, Repeat: Understanding Parallel Reasoning in Language Model Inference*. arXiv:2603.07887v1, 2026. <https://arxiv.org/abs/2603.07887v1>.