Online learning under relative entropy: sharp rates and the price of reference access

Anonymous authors
Paper under double-blind review

Abstract

We study online binary classification when each input law, conditional on the full preceding feedback history, has KL divergence at most one from a fixed reference measure. When the reference is known, we prove the sharp worst-case minimax expected regret $\Theta_d(\sqrt{T\log(eT)})$ at every fixed positive VC dimension bound d, improving the previous $O_d(T^{2/3})$ guarantee. The upper uses one KL budget across all cover resolutions; a matching threshold lower bound shows that the logarithm is necessary. When the reference is unknown, we determine the sharp dependence on reference samples and total-variation approximation accuracy. Even thresholds can require expected regret of order $T/\log(eT)$ despite polynomially many samples and inverse-polynomial approximation error. A bounded forward Rényi divergence of fixed order greater than one recovers the known-reference order. For every fixed k, online inputs alone recover this horizon order uniformly over all unknown full-dimensional log-concave references on $\mathbb{R}^k$. The general upper bounds use proper statistical strategies that choose a hypothesis before the current input; the lower bounds allow arbitrary predictions after the input and adaptive reference-sample requests.

1 Introduction

Let $\mathcal{F}$ be a class of binary classifiers with finite VC dimension, and let μ be a fixed probability measure on the input space. We measure regret relative to the best member of $\mathcal{F}$ on the observed input-label sequence. We study online prediction under the condition $\mathrm{KL}(P_t\|\mu) \le 1$, where P_t is the next-input law conditional on the full preceding feedback history. The input laws and labels may change adaptively while μ remains fixed. This constraint permits online inputs to visit regions much more often than independent reference samples do. For example, if $\mu(A) = e^{-a}$ with $a \ge 1$, the mixture $P = (1-a^{-1})\mu + a^{-1}\mu(\cdot \mid A)$ has $\mathrm{KL}(P\|\mu) \le a^{-1}\log(1/\mu(A)) = 1$ by convexity, while $P(A) \ge 1/a$. We determine the regret when the learner knows μ, and then ask how much unlabeled reference data or approximate knowledge of μ is needed to retain that rate.

The smoothed online model interpolates between independent and adversarial inputs (Rakhlin et al., 2011; Block et al., 2022). Block and Polyanskiy (2023) extend bounded-density smoothing to f-divergence constraints through approximate rejection sampling. At fixed positive VC dimension bound d and conditional KL at most one, Theorem 10 of their revised paper (Block and Polyanskiy, 2024) gives a statistical upper bound of order $O_d(T^{2/3})$ and asks whether the horizon dependence can be improved. Theorem 18 of Blanchard et al. (2026) gives the same order. We answer this statistical question by proving the sharp worst-case rate

$$\Theta_d\Big(\sqrt{T\log(eT)}\Big).$$

The upper uses a proper learner that chooses its hypothesis before the current input and controls disagreements at all cover resolutions with one KL charge. The lower permits arbitrary predictions after the input is revealed and shows that the logarithm is necessary even for thresholds under a known reference. We also determine the sharp dependence on VC dimension and the total conditional KL budget. Appendix S20 gives the precise substitutions into the cited prior bounds.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The next question concerns access to the reference. A learner may know that a suitable μ exists without knowing its density, CDF, or quantiles. Suppose it receives at most m independent samples from μ and a known measure ν with $\mathrm{TV}(\mu, \nu) \le \delta$, where $0 < \delta \le 1$. At every fixed positive dimension bound d, we prove the worst-case agnostic regret law

$$\Theta_d\left(\sqrt{T\log(eT)} + \frac{T}{\log(eT(m+1)/\delta)}\right).$$

The first term is the known-reference regret; the second is the additional cost of limited reference information. For fixed d, this law retains the known-reference order precisely when

$$\log(eT(m+1)/\delta) = \Omega\left(\sqrt{T/\log(eT)}\right).$$

If $1/\delta$ is at most polynomial in T, the sample allowance must be at least exponential in $\sqrt{T/\log(eT)}$. With polynomially many samples, $1/\delta$ must instead be at least exponential in the same quantity. The lower already holds for ordinary thresholds and permits unlimited computation, online label feedback, and adaptive reference-sample requests. For realizable thresholds, the information term remains and the statistical term is $\sqrt{T}$.

A forward moment bound gives a different answer to the access question. If a supplied measure satisfies $D_{1+\eta}(\mu\|\nu) \le 1$, every fixed $\eta > 0$ recovers the known-reference order without reference samples. Online inputs alone recover the known-reference horizon order uniformly over unknown full-dimensional log-concave references on $\mathbb{R}^k$, with a factor $O(\sqrt{k})$ in the general VC bound. Thus the reference shape may be unknown as well as its mean and covariance. For Gaussian references, $O(k^2)$ independent reference samples remove this dimension factor. Without them, a longer prefix of $O(k^2 + \log T)$ online inputs gives a dimension-independent leading factor at an additive $O(k^2)$ regret cost. The statistical term absorbs $\log T$. With known covariance, k samples suffice; the learner can instead use online inputs for calibration, paying for k initial rounds. The proofs control KL cost under a calibrated reference, without estimating the unknown density accurately. These consequences show why the worst-case sample barrier need not hold under a shape constraint on the reference. We also determine the sharp joint sample-and-moment rate, uniformly for $0 < \eta \le 1$, in Appendix S13. Polynomially many samples leave the worst-case horizon order of the moment-only guarantee unchanged, uniformly in η. Improving that order by combining both sources requires superpolynomially many samples.

Building on unknown-reference learnability and the Cover strategy (Blanchard, 2025; Blanchard et al., 2026), we determine the sharp dependence on samples, approximation accuracy, and moment order. The information lower bounds use fixed references with narrow density spikes that reference samples can miss. The general upper bounds provide measurable proper statistical strategies, without a runtime guarantee for arbitrary class and measure encodings. We give a finite-bit threshold learner for the known-reference law. For unknown one-dimensional log-concave references, we also give a raw-input implementation with the expected bit costs stated in Section 6.

2 Model and known reference

All measurable spaces, including round-state spaces, are standard Borel. The known nonempty class $\mathcal{F}$ is finite or countable and consists of measurable maps $X \to \{0, 1\}$, except for explicit threshold statements with their specified parameter sets. The fixed reference is known in the statistical model and unknown in the reference-information model below. The adversary may observe preceding learner randomness. The current input-label law may depend on the preceding history and on the learner's announced distribution, but not on its fresh private hypothesis draw.

A proper learner chooses a random $f_t \in \mathcal{F}$ before receiving x_t and then predicts $f_t(x_t)$. An unrestricted learner may instead predict any $p_t \in [0, 1]$ after receiving x_t and before the label is revealed. Write $\ell(p, y) = y + (1 - 2y)p$ and $\ell_f(x, y) = \mathbf{1}\{f(x) \ne y\}$. Here $\ell(p, y)$ is the expected mistake loss when label one is predicted with probability p. The expected regret is

$$\mathbb{E}\,\mathrm{Reg}_T = \mathbb{E}\left[\sum_{t=1}^{T} \ell(p_t, y_t) - \inf_{f\in\mathcal{F}} \sum_{t=1}^{T} \ell_f(x_t, y_t)\right].$$

For randomized binary predictions we average over the fresh private draw. The hindsight infimum is inside the expectation. The agnostic model permits arbitrary labels. A history is realizable if some classifier in $\mathcal{F}$ fits all its labels.

For integers $1 \le d \le T$, write $A_{d,T} = \sqrt{dT \log(eT/d)}$.

2.1 Known reference

Let P_t be the next-input law conditional on the full preceding history, including preceding learner randomness and adversary state, and put $c_t = \mathrm{KL}(P_t \| \mu)$. The learner need not observe these laws or costs. For $0 \le K \le 1$, require $c_t \le K$ at every round.

Write $V^{\mathrm{all}}_{T,\mathrm{KL}}(d, K)$ and $V^{\mathrm{prop}}_{T,\mathrm{KL}}(d, K)$ for the worst-case minimax expected regrets of unrestricted and proper learners, respectively, under the per-round conditional constraint $\mathrm{KL}(P_t \| \mu) \le K$. The worst case is over classes of VC dimension at most d and fixed references; the learner knows the class and reference.

Theorem 2.1. *There are universal constants $c, C > 0$ such that, for all integers $1 \le d \le T$ and all $0 \le K \le 1$,*

$$\begin{aligned} c\left(\sqrt{dT} + \sqrt{dKT \log(e + T/d)}\right) &\le V^{\mathrm{all}}_{T,\mathrm{KL}}(d, K) \\ &\le V^{\mathrm{prop}}_{T,\mathrm{KL}}(d, K) \\ &\le C\left(\sqrt{dT} + \sqrt{dKT \log(e + T/d)}\right). \end{aligned}$$

For each d, the lower bound holds on one fixed countable threshold-product class of VC dimension d under one fixed reference, independent of T and K.

In particular, at $K = 1$ the minimax rate is $\Theta(A_{d,T})$, with universal constants for $1 \le d \le T$. At fixed positive d, this is $\Theta(\sqrt{T \log(eT)})$. Theorem 2.1 also describes the transition to the independent-input rate $\Theta(\sqrt{dT})$ as K tends to zero. The two terms become comparable at $K \asymp 1/\log(e + T/d)$. The consecutive Appendices S2–S5 give the complete proof and its auxiliary results.

The statistical strategy has exact mathematical access to the known class and reference. Theorem S6.1 in Appendix S6 extends the result to a total budget $\sum_t c_t \le B$, including $B > T$. Its proper upper also holds when only $\mathbb{E}\sum_t c_t \le B$ is assumed. We use this extension when changing the reference measure in Section 5.

Conditioning on the full feedback history is essential. For example, under uniform μ on $\{0, 1\}$, draw x_1 and all labels as independent fair bits, and set $x_{t+1} = y_t$. The marginal input path has law μ^T and zero KL, but each input after the first has conditional KL $\log 2$ given past labels. The total full-history cost is $(T - 1)\log 2$.

3 Why one information budget controls many scales

The prior upper bound compared in Appendix S20 is optimized over one approximation parameter and contains the term $\epsilon T + \sqrt{dT/\epsilon}$ in addition to $O(\sqrt{dT \log(eT)})$. Optimizing over ϵ gives order $T^{2/3}$ at fixed d. We instead keep several cover resolutions and bound their KL costs together. A proper reference cover $H_a \subseteq \mathcal{F}$ has $\log |H_a| = O(da)$ and, for every $f \in \mathcal{F}$, contains an h with $\mu\{f \ne h\} \le e^{-a}$. Successive representatives disagree on reference measure $O(e^{-a})$. The next lemma bounds all their high-density contributions by one KL charge. Write $Pg = \int g\, dP$ and $\mu g = \int g\, d\mu$. In the application, each q_j is a weighted average of disagreement indicators between consecutive cover representatives; the weights may depend on the preceding history. The density in the next lemma is used only in the analysis; the learner does not observe the input laws or their densities.

Lemma 3.1. *Let $a_j = 2^{j-1} a_1$, where $a_1 \ge 4$, and let $0 \le q_j \le 1$ be measurable with $\mu q_j \le 2e^{-a_j}$. For $P \ll \mu$ with density r,*

$$\sum_j a_j P\Big(q_j \mathbf{1}\{r \ge e^{a_j/2}\}\Big) \le 8\,\mathrm{KL}(P\|\mu), \qquad P\Big(q_j \mathbf{1}\{r < e^{a_j/2}\}\Big) \le 2e^{-a_j/2}.$$

Proof. At a fixed point with density r, the sum of dyadic a_j satisfying $a_j \le 2\log r$ is at most $4\log r$. Whenever this sum is nonzero, $r \ge e^2$. On that set, $r\log r \le 2(r\log r - r + 1)$. The latter integrand is nonnegative everywhere and its integral is $\mathrm{KL}(P\|\mu)$. Multiplying the pointwise bound by r, integrating, and using $q_j \le 1$ proves the first inequality. For the second, bound r by $e^{a_j/2}$ inside the integral and use the reference disagreement bound. □

A separate entropy estimate at each resolution would charge the same $\mathrm{KL}(P\|\mu)$ once per level. Lemma 3.1 bounds their sum even though the disagreement functions need not be nested or disjoint.

The proper minimax reduction in Appendix S3 reduces the regret bound to an auxiliary game with fresh independent fair signs. We bound this game using exponential weights on the increments between consecutive cover representatives. At $K = 1$, use covers at errors e^{-u_j}, where u_j doubles up to $u_J \asymp \sqrt{T/d}$. There are $J = O(\log(e + T/d))$ levels. Writing $a_j = u_{j-1}$, the class of increments between consecutive representatives has log cardinality $O(da_j)$, and every increment has mean square at most $2e^{-a_j}$ under μ. A second-order exponential-weights bound at rate proportional to λa_j costs $O(d/\lambda)$ per level for the class size. Bounding each high-density term separately by the same KL budget gives the weaker bound $O(\lambda TJ)$. The lemma bounds their combined variance cost by $O(\lambda T)$. The base cover, capped rates, low-density terms, and terminal disagreement contribute $O(\sqrt{dT})$. Thus the bound is

$$C\left(\sqrt{dT} + \frac{dJ}{\lambda} + \lambda T\right).$$

Taking $\lambda = \sqrt{dJ/T}$ gives $CA_{d,T}$. Appendix S5 combines the minimax reduction with these estimates to obtain the proper learner. Appendix S6 gives the total-budget refinement.

The logarithm occurs for thresholds with a known reference. At binary-search stage j, draw the midpoint with probability $q_j = 1/(8j)$ and draw from μ otherwise. Call the draws from the midpoint component active, and let n_j be the number of rounds in stage j. The fixed reference assigns each depth-j midpoint mass $[2^{j+1}j(j+1)]^{-1}$, whose negative logarithm is at most $3j$. KL convexity therefore bounds the stage cost by $3jq_j \le 1$ (Appendix S5). Every round has a fresh fair label, so each learner has expected loss $T/2$. After a stage, retain the half interval agreeing with the majority of its active labels. One threshold in the final interval realizes these choices; background labels do not affect its selection. For the allocation below and sufficiently large T, its expected savings are at least $c\sum_{j\le L}\sqrt{n_j q_j}$. For $L = \lfloor T^{1/4}\rfloor$, $H_L = \sum_{j\le L} 1/j$, and $n_j = \lfloor T/(2jH_L)\rfloor$, this sum has order $\sqrt{T\log T}$ for large T. Here the reference is known and the comparator is agnostic. The unknown-reference penalty below holds even for realizable labels from a target fixed before play.

4 Reference samples and approximation

4.1 The sample and approximation law

Now require $c_t \le 1$ at every full preceding history, including observed reference samples. The learner may receive iid unlabeled samples from μ, but no density, CDF, quantile, or conditional-sampling oracle for it. A supplied approximation consists of a known mathematical measure ν and an announced $0 < \delta \le 1$ with $\mathrm{TV}(\mu,\nu) \le \delta$. Exact knowledge of ν does not give the learner exact knowledge of μ.

Theorem 4.1. *For every fixed positive integer d there are $c_d, C_d > 0$ such that the following holds for integers $T \ge \max\{512, d\}$ and $m \ge 0$ and every $0 < \delta \le 1$. For every known nonempty countable measurable binary class $\mathcal{F}$ with $\mathrm{VC}(\mathcal{F}) \le d$ and every supplied Borel probability ν, there is a measurable proper pre-input strategy depending only on $T, d, m, \delta, \mathcal{F}, \nu$ and observed history. For every fixed Borel reference μ satisfying $\mathrm{TV}(\mu,\nu) \le \delta$ and every input-label protocol obeying $\mathrm{KL}(P_t\|\mu) \le 1$ at every full preceding history and the current-round privacy convention, this strategy takes a deterministic number of at most m iid μ samples before play and satisfies*

$$\mathbb{E}\,\mathrm{Reg}_T \le C_d\left(A_{d,T} + \frac{T}{\log(eT(m+1)/\delta)}\right).$$

In the worst case over classes and supplied ν, every unrestricted learner has expected regret at least c_d times the same expression. The lower already uses rational thresholds and uniform ν, with a fixed

atomless full-support μ. It permits adaptive post-input reference requests and unlimited computation. The same order holds for real uniformly expected call allowances $m \geq 0$.

The lower-bound reference is fixed before play and may depend on T, m, δ and the learner. The expected allowance is uniform over the announced admissible reference family at its fixed δ. Theorem S10.1 gives the complete proof and sample-access model. The rate is sharp for fixed d. At $\delta = 0$, $\mu = \nu$ exactly and the known-reference statistical theorem applies separately.

The realizable comparison has a smaller statistical floor. For ordinary thresholds $h_\theta(x) = \mathbf{1}\{x \geq \theta\}$, $0 \leq \theta \leq 1$, with supplied $\nu = U$ uniform on $[0, 1]$, Theorem S11.12 gives

$$\Theta\left(\sqrt{T} + \frac{T}{\log(eT(m+1)/\delta)}\right)$$

worst-case mistakes for every $T \geq 512$, $m \geq 0$, and $0 < \delta \leq 1$. Its upper permits arbitrary Borel references; its lower has the same stronger prediction and adaptive-call allowances. For both laws, $\delta = 1$ gives the sample-only problem, including zero samples.

4.2 Why samples can miss predictive information

We explain the information term in Theorem 4.1. The complete lower, including expected allowances, is in Appendix S11.10. The reference is fixed before play.

Write $s = \log(eT(m+1)/\delta)$ and choose

$$q = \frac{1}{32s}, \qquad \alpha = \frac{\delta}{64(m+1)(T+1)^2}.$$

Inside one interval of width $(T+1)^{-2}$, independent fair bits specify a nested threshold path. At each depth, an active band separates the two possible future target intervals. Let Q_j be uniform on the depth-j active band, and let θ_* lie in the final retained interval. Define

$$\mu = (1-\alpha)U + \frac{\alpha}{T}\sum_{j=1}^{T} Q_j.$$

This one measure contains all future bands. It is atomless, has full support, and gives weight $1-\alpha$ to the uniform component. The same construction satisfies not only $\mathrm{TV}(\mu, U) \leq \delta$ but also

$$\mathrm{KL}(\mu\|U) \leq \delta, \qquad \mathrm{KL}(U\|\mu) \leq \delta, \qquad \mu \geq (1-\delta)U.$$

The band widths are allowed to become very small.

An online input is drawn from $(1-q)U + qQ_j$ at the current depth. Its label is $h_{\theta_*}(x)$, and the depth advances only on an active draw. The conditional input KL is at most one at every full preceding history:

$$\mathrm{KL}((1-q)U + qQ_j\|\mu) \leq 2\alpha + q\log(T/\alpha) < 1.$$

Active draws depend on previously revealed bits, not the next bit.

Couple reference samples by first choosing their mixture components. With a probability bounded below by an absolute constant, all potential reference samples come from U, and all online uniform-background draws avoid the small initial path interval. This common event is independent of the hidden path bits and active-draw coins. On it, the full reference-sample transcript gives no path bit, background labels give no future bit, and the next active label is fair even after its input and any permitted adaptive sample requests. Each active draw therefore costs an unrestricted predictor expected loss $1/2$. The expected number of active draws is qT, giving the term cT/s.

This argument uses the entire potential sample list, not a bound on one prespecified query time. The expected-call version truncates an arbitrary learner at a larger fixed cap under the same announced δ, couples the autonomous online process, and bounds the truncation error by Markov's inequality. The separate uniform-reference statistical witness gives $c\sqrt{T\log(eT)}$. The two witnesses need not coincide: their maximum bounds half their sum.

Theorem S12.1 gives a $T/\log(eT)$ eventual almost-sure lower on one fixed reference for each anytime learner with a uniformly polynomial expected sample allowance, at TV accuracy $\delta = 1/4$; Appendix S13 gives the precise statement. These information lower bounds encode the target path in narrow density spikes.

4.3 Upper reductions from the available information

The general upper combines three strategies. They all select proper hypotheses before the current input. Only the empirical-tree strategy uses true-reference samples.

No samples. Cover keeps one hypothesis for each distinct labeling realized by $\mathcal{F}$ on the inputs in completed blocks and restarts exponential weights on these representatives (Blanchard, 2025). A density-clipping reduction and a uniform hindsight bound give $C_d T/\log(eT)$ under expected total KL at most T, without knowing the reference. Appendix S9 states the exploration theorem and its full-history form; Appendix S11.7 gives a direct threshold proof and executable learner.

Reference samples. Two independent samples totaling $O_d(j2^j)$ draws supply a finite proper witness class and estimates of its pairwise disagreement distances. A tree of these representatives gives regret at most $C_d A_{d,T} + 2T/(j\log 2)$. Choosing the largest useful resolution allowed by m gives $C_d[A_{d,T} + T/\log(eT(m+1))]$. The no-sample strategy covers allowances too small for this construction. The statistical learner uses measurable least-index representatives. Appendix S10 proves the bound for every allowance.

A known approximate reference. Choose a finite proper cover under ν and partition the input space by its evaluation patterns, keeping reference-null cells. Adding δ to each cell's ν mass and normalizing gives a known positive reference on this finite space. Total variation bounds the true mass of each cell by its original ν mass plus δ. The finite known-reference theorem then controls regret against the cover, and a uniform disagreement bound controls its error against the full class. The dual-VC bound (Assouad, 1983) bounds the number of cells and gives $C_d[A_{d,T}+T/\log(eT/\delta)]$. Appendix S10 supplies the construction and all accuracy regimes, including atoms and reference-null cells.

Before play, choose the sample or approximation strategy with the larger denominator, $s_m = \log(eT(m+1))$ or $s_\delta = \log(eT/\delta)$. Since $\log(eT(m+1)/\delta) \le 2\max\{s_m, s_\delta\}$, this gives the joint upper in Theorem 4.1. For this TV law, choosing between the two strategies already attains the joint order.

5 Reference regularity and calibration

For $\eta > 0$ and $\mu \ll \nu$, the forward Rényi divergence is

$$D_{1+\eta}(\mu\|\nu) = \frac{1}{\eta}\log\int\left(\frac{d\mu}{d\nu}\right)^{1+\eta} d\nu;$$

it is infinite if μ is not absolutely continuous with respect to ν. We supply ν and the order while keeping μ unknown. The lower in Theorem 4.1 permits narrow density spikes. Suppose the learner knows ν and the fixed unknown reference satisfies $D_{1+\eta}(\mu\|\nu) \le R$. The classical entropy-transfer inequality (Altschuler and Chewi, 2026) gives

$$\mathrm{KL}(P_t\|\nu) \le \frac{1+\eta}{\eta}\,\mathrm{KL}(P_t\|\mu) + D_{1+\eta}(\mu\|\nu).$$

For fixed $\eta > 0$ and R, Theorem S6.1 therefore recovers the known-reference order without any samples from μ. A fixed density ceiling $\mu \le M\nu$ is also sufficient. The sample/TV lower construction in Section 4.2 also permits $D_\infty(U\|\mu) \le \delta$ together with $\mathrm{KL}(\mu\|U) \le \delta$. The direction and order of the reference promise both matter. Appendix S14 proves these deductions.

A regular reference family can therefore avoid the sample barrier of Theorem 4.1. For a Gaussian reference, independent reference samples recover the known-reference rate with a universal constant, without bounds on the mean or covariance. With online inputs alone, Cauchy calibration gives $C\sqrt{k}A_{d,T}$. A longer $O(k^2 + \log T)$ prefix gives $C[k^2 + A_{d,T}]$: quadratic comparisons fit a Gaussian with failure probability at most $1/T$, and failed fits contribute at most one to expected regret (Theorem S15.4). The horizon order is unchanged, but the leading factor is dimension-independent after this initial cost. The corollary also records sharper bounds when the covariance is known.

Corollary 5.1. *Let k, T, d be positive integers with $d \le T$, and let $\mathcal{F}$ be a known nonempty countable class of measurable binary functions on $\mathbb{R}^k$ with $\mathrm{VC}(\mathcal{F}) \le d$. There is a measurable proper pre-input learner, independent of a and Σ, which takes $n = 128(k+1)^2 + 1$ iid samples from $\mu = N(a, \Sigma)$ before play and satisfies $\mathbb{E}\,\mathrm{Reg}_T \le CA_{d,T}$, with universal C, for every $a \in \mathbb{R}^k$, every positive-definite $k \times k$ matrix Σ, and every protocol obeying the current-round privacy convention and $\mathrm{KL}(P_t \| \mu) \le 1$ conditional on the full preceding history, including these samples. Without independent reference samples, there is a measurable proper pre-input learner, independent of a and Σ, satisfying under the same input constraint and privacy convention*

$$\mathbb{E}\,\mathrm{Reg}_T \le \min\{T, C\sqrt{k}\,A_{d,T}, C[k^2 + A_{d,T}]\}.$$

If Σ is supplied, $n = k$ independent reference samples suffice for $CA_{d,T}$ regret, and without such samples the bound improves to $\min\{T, k + CA_{d,T}\}$. These are statistical existence results, with universal C.

Only the reference is Gaussian; the online inputs may have any laws satisfying the conditional KL constraint. The proof is in Appendix S15.

The statistical logarithm remains necessary even for a known $N(0,1)$ reference: rational thresholds have minimax expected regret at least $\sqrt{T \log(eT)}/256$ for $T \ge 512$ (Proposition S15.1).

The calibration argument also permits unknown log-concave shape. A density is log-concave if its logarithm is concave, with $\log 0 = -\infty$. The class includes uniform measures on unknown convex bodies with nonempty interior.

Corollary 5.2. *Let k, T, d be positive integers with $d \le T$, and let $\mathcal{F}$ be a known nonempty countable class of measurable binary functions on $\mathbb{R}^k$ with $\mathrm{VC}(\mathcal{F}) \le d$. There is a measurable proper pre-input learner using no independent reference samples, and independent of μ, such that*

$$\mathbb{E}\,\mathrm{Reg}_T \le \min\{T, C\sqrt{k}\,A_{d,T}\}$$

for every fixed full-dimensional log-concave reference μ and every protocol obeying the current-round privacy convention and $\mathrm{KL}(P_t \| \mu) \le 1$ conditional on the full preceding history. The constant C is universal.

Only the reference is assumed log-concave; the online input laws may be dependent and non-log-concave. Appendix S15.6 proves the bound uniformly over the unknown shape, mean and covariance.

The dependence on approximate shape can also be determined. For fixed d, one proper pre-input learner using no reference samples satisfies, uniformly in dimension k,

$$\mathbb{E}\,\mathrm{Reg}_T \le C_d \left(\sqrt{kT \log(eT)} + \frac{T}{\log(eT/\delta)} \right) \qquad (0 < \delta \le 1)$$

whenever $\mathrm{TV}(\mu, \rho) \le \delta$ for some unknown full-dimensional log-concave measure ρ on $\mathbb{R}^k$, under the same full-history conditional KL constraint. The learner receives none of μ, ρ, δ. The same learner has the bound $C_d\sqrt{kT \log(eT)}$ at $\delta = 0$, and the rate is sharp in the worst case for fixed d, k (Theorem S23.1). In particular, inverse-polynomial TV error still permits regret of order $T/\log(eT)$. The proof calibrates from a prefix with a controlled fraction of corruptions and charges unobserved corrupted rounds in the proper prediction game.

Appendix S13 states the full joint sample-and-moment law in Theorem S13.1, uniformly for $0 < \eta \le 1$, including η tending to zero with T. The theorem allows arbitrary fixed references satisfying the moment promise and includes matching threshold lower bounds with adaptive sample requests. For every fixed $p \ge 0$, allowances $m \le T^p$ leave the no-sample worst-case horizon order unchanged, uniformly in η, with constants depending on d, p. The appendix gives a regime with superpolynomially many samples in which combining both sources improves on the smaller of the sample-only and zero-sample moment guarantees.

6 Finite-bit threshold implementations

The general upper bounds allow exact mathematical access to a class and a measure. For ordinary thresholds, the implementations below charge for input bits and arithmetic under explicit binary encodings.

Under known uniform μ, Theorem S18.1 attains the total-budget rate of Theorem S6.1 for every supplied integer KL allowance $b \geq 1$. The learner is proper, uses a finite dyadic grid, and competes with every real threshold. For $T \geq 2$ and $b < T^2$, it reads $O(\sqrt{\max\{T, b\}})$ binary digits per input and uses polynomially many bit operations. At per-round KL one, taking $b = T$ gives the sharp regret rate with $O(\sqrt{T})$ input digits per round. Appendices S18 and S19 give the algorithm and its access and cost conventions.

For an unknown nondegenerate log-concave reference on $\mathbb{R}$, Theorem S22.1 gives $O(\sqrt{T \log(eT)})$ expected regret against all real thresholds when each full-history conditional input KL is at most one. It chooses finite rational thresholds before the current input and has expected bit work polynomial in T and $L_\mu = 1 + \log_2(1 + |a| + \sigma) + \log_2(1 + \sigma^{-1})$, where a and σ are the reference mean and standard deviation. These parameters are not supplied. The learner uses raw binary input streams and no reference samples, CDF or quantile oracle. Appendix S22 specifies the encoding, charged access and output bits, and almost-sure termination.

Efficient interval updates and proportional sampling for flat exponential weights are established methods (Cohen-Addad and Kanade, 2017; Mirfakhar et al., 2026). We use these ideas to implement learners with the statistical guarantees proved here.

7 Related work and limitations

Constructive chaining and hierarchical entropy are established tools in online learning (Gaillard and Gerchinovitz, 2015; Coester and Lee, 2019; Yang et al., 2022). Our known-reference proof uses a shared KL charge across resolutions. The upper without reference samples adapts Cover and density clipping (Blanchard, 2025; Blanchard et al., 2026). The empirical disagreement tree also builds on earlier trace-acquisition methods (Attias et al., 2025; Ben-David, 2015).

For $D_{1+\eta}(\mu\|\nu) \leq 1$ without reference samples, Theorem 18 of Blanchard et al. (2026) gives

$$O_d\Big(\min\{T/\log(eT), T^{2/3}\eta^{-1/3}\}\Big)$$

uniformly for $0 < \eta \leq 1$ at fixed d. Theorem S13.1 in Appendix S13 determines the sharp transition and joint sample dependence under one fixed unknown reference. Appendix S20 supplies the comparison, including its full-history argument. These comparisons concern the displayed upper-bound guarantees.

Approximate-rejection lower bounds (Block and Polyanskiy, 2024) require the output to be one of the reference samples. Our prediction lower bounds allow arbitrary outputs and online label feedback. Voitovych et al. (2026) study trajectory KL and conditional simulation. Changing only the input kernels in their joint comparison law gives the unrestricted post-input bound $O(\sqrt{dT} + \sqrt{TB})$ under our full-history budget (Appendix S21). This bound is linear at $B = T$, and their matching lower uses dependent references.

The regret upper bounds hold in expectation at a fixed horizon. For general classes the learners are measurable statistical strategies; the computational guarantees apply to thresholds with the stated input and parameter encodings. The realizable information law is proved for thresholds. The finite-horizon lower references may depend on the horizon and sample allowance, but are fixed before play. Theorem S12.1 instead fixes one reference at positive TV accuracy and gives an eventual almost-sure lower. The reference-information lower constructions use fine density spikes. The log-concave consequence shows that this penalty need not persist when the unknown reference satisfies a shape constraint. The general calibration bounds are sufficient and do not optimize the dependence on k; their proofs use exact real-valued observations. Section 6 gives a separate raw-bit threshold implementation. The sample/TV and sample/moment laws are sharp for fixed VC dimension under their stated information constraints.

Reproducibility statement

Complete proofs and the precise access assumptions are given in the appendices. The guide in Appendix S1 identifies the proofs of the two main minimax laws, the Gaussian and log-concave con-

sequences, and the access extensions. The threshold algorithms, input encodings and bit costs are specified in Appendices S18, S19 and S22.

AI-use disclosure

The author designed and directed the research workflow, set research priorities, evaluated proposed directions, and checked the mathematical results. AI tools assisted with formulating mathematical claims, developing and checking proofs, literature searches, manuscript drafting and revision, and code development and verification. The author takes responsibility for the final content of this paper.

References

Jason M. Altschuler and Sinho Chewi. Shifted composition III: Local error framework for KL divergence. *Foundations of Computational Mathematics*, 2026. https://doi.org/10.1007/s10208-026-09753-x.

Hassan Ashtiani, Shai Ben-David, Nicholas J. A. Harvey, Christopher Liaw, Abbas Mehrabian, and Yaniv Plan. Near-optimal sample complexity bounds for robust learning of Gaussian mixtures via compression schemes. *Journal of the ACM*, 67(6), Article 32, 1–42, 2020. https://doi.org/10.1145/3417994.

Patrick Assouad. Densité et dimension. *Annales de l'Institut Fourier*, 33(3):233–282, 1983. https://www.numdam.org/articles/10.5802/aif.938/.

Idan Attias, Steve Hanneke, and Arvind Ramaswami. Tradeoffs between mistakes and ERM oracle calls in online and transductive online learning. arXiv:2506.00135v1, 2025. https://arxiv.org/abs/2506.00135v1.

Shai Ben-David. 2 notes on classes with Vapnik–Chervonenkis dimension 1. arXiv:1507.05307v1, 2015. https://arxiv.org/abs/1507.05307v1.

Moïse Blanchard, Abhishek Shetty, and Alexander Rakhlin. Characterizing online and private learnability under distributional constraints via generalized smoothness. In *Proceedings of Thirty Ninth Conference on Learning Theory*, PMLR 336:729–759, 2026. https://proceedings.mlr.press/v336/blanchard26a.html.

Moïse Blanchard. Agnostic smoothed online learning without knowledge of the base measure. arXiv:2410.05124v3, 2025. https://arxiv.org/abs/2410.05124v3.

Adam Block and Yury Polyanskiy. The sample complexity of approximate rejection sampling with applications to smoothed online learning. In *Proceedings of the Thirty Sixth Conference on Learning Theory*, volume 195 of PMLR, pages 228–273, 2023. https://proceedings.mlr.press/v195/block23a.html.

Adam Block and Yury Polyanskiy. The sample complexity of approximate rejection sampling with applications to smoothed online learning. arXiv:2302.04658v3, 2024. https://arxiv.org/abs/2302.04658v3.

Adam Block and Max Simchowitz. Efficient and near-optimal smoothed online learning for generalized linear functions. *Advances in Neural Information Processing Systems*, 35, 2022. https://proceedings.neurips.cc/paper_files/paper/2022/hash/31455446488c433ef29495ab44d4f53c-Abstract-Conference.html. The cited determinant anti-concentration result is Proposition 41 of the published supplementary material.

Adam Block, Yuval Dagan, Noah Golowich, and Alexander Rakhlin. Smoothed online learning is as easy as statistical learning. In *Proceedings of the Thirty Fifth Conference on Learning Theory*, volume 178 of PMLR, pages 1716–1786, 2022. https://proceedings.mlr.press/v178/block22a.html.

Sergey Bobkov and Mokshay Madiman. The entropy per coordinate of a random vector is highly constrained under convexity conditions. *IEEE Transactions on Information Theory*, 57(8):4940–4954, 2011. https://doi.org/10.1109/TIT.2011.2158475. The entropy–supremum inequality is Proposition I.2 of https://arxiv.org/abs/1006.2883v2.

Zachary Chase, Bogdan Chornomaz, Steve Hanneke, Shay Moran, and Amir Yehudayoff. Dual VC dimension obstructs sample compression by embeddings. arXiv:2405.17120v1, 2024. https://arxiv.org/abs/2405.17120v1.

Christian Coester and James R. Lee. Pure entropic regularization for metrical task systems. In *Proceedings of the Thirty-Second Conference on Learning Theory*, PMLR 99:835–848, 2019. https://proceedings.mlr.press/v99/coester19a.html.

Vincent Cohen-Addad and Varun Kanade. Online optimization of smoothed piecewise constant functions. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, PMLR 54:412–420, 2017. https://proceedings.mlr.press/v54/cohen-addad17a.html.

Pierre Gaillard and Sébastien Gerchinovitz. A chaining algorithm for online nonparametric regression. In *Proceedings of the 28th Conference on Learning Theory*, PMLR 40:764–796, 2015. https://proceedings.mlr.press/v40/Gaillard15.html.

Nika Haghtalab, Tim Roughgarden, and Abhishek Shetty. Smoothed analysis of online and differentially private learning. *Advances in Neural Information Processing Systems*, 33, 2020. https://proceedings.neurips.cc/paper/2020/hash/685bfde03eb646c27ed565881917c71c-Abstract.html.

Nika Haghtalab, Tim Roughgarden, and Abhishek Shetty. Smoothed analysis with adaptive adversaries. *Journal of the ACM*, 71(3), 2024. https://doi.org/10.1145/3656638. Proof statements use the author manuscript, https://people.eecs.berkeley.edu/~nika/pubs/smoothed_online_ACM.pdf.

Boaz Klartag and Joseph Lehec. Affirmative resolution of Bourgain's slicing problem using Guan's bound. *Geometric and Functional Analysis*, 35:1147–1168, 2025. https://doi.org/10.1007/s00039-025-00718-w. We use the published general log-concave consequence in equation (6).

László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358, 2007. https://doi.org/10.1002/rsa.20135. The marginal, density and moment statements use Theorems 3.1, 3.9(e) and 3.17 and Lemma 3.4(a) of the Newton Institute preprint NI02055, https://api.newton.ac.uk/website/v0/events/preprints/NI02055.

Amirmahdi Mirfakhar, Maria-Florina Balcan, and Hedyeh Beyhaghi. Efficient online proportional sampling with applications to smoothed online learning. arXiv:2607.10963v1, 2026. https://arxiv.org/abs/2607.10963v1.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Stochastic and constrained adversaries. arXiv:1104.5070, 2011. https://arxiv.org/abs/1104.5070.

Peter J. Rousseeuw and Katrien Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, 1999. https://doi.org/10.1080/00401706.1999.10485670. The subset-determinant definition appears on page 213.

Norbert Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145–147, 1972.

Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

Sasha Voitovych, Abhishek Shetty, Noah Golowich, and Alexander Rakhlin. Learning with simulators: No regret in a computationally bounded world. arXiv:2606.13576v1, 2026. https://arxiv.org/abs/2606.13576v1.

Lin Yang, Yu-Zhen Janice Chen, Mohammad H. Hajiesmaili, Mark Herbster, and Don Towsley. Hierarchical learning algorithms for multi-scale expert problems. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 6(2), article 34, 1–29, 2022. https://doi.org/10.1145/3530900.

Yannis G. Yatracos. Rates of convergence of minimum distance estimators and Kolmogorov's entropy. *The Annals of Statistics*, 13(2):768–774, 1985. https://doi.org/10.1214/aos/1176349553.

S1 Guide to the proofs

The table gives the proof paths for the two main minimax laws and the reference calibration consequences. The consecutive Appendices S2–S5 contain the complete known-reference proof and its auxiliary results. The implementation and prior-comparison sections are separate from this proof path.

Result	Proof and supporting sections
Known-reference rate, Theorem 2.1	Appendix S5, using the rare-set bound in Appendix S2, the proper-game and cover arguments in Appendix S3, and the fixed-reference construction in Appendix S4.
Sample/TV law, Theorem 4.1	Appendix S10 assembles the upper and lower. The no-sample and empirical-tree uppers are in Appendices S9 and S8; the information lower is in Appendix S11.10.
Reference calibration, Corollaries 5.1 and 5.2; approximate shape, Theorem S23.1	Appendix S15 controls transfer cost and uses the fixed-cost strategies of Lemmas S6.4 and S15.2. Appendices S15.4 and S15.6 calibrate a Cauchy reference from online inputs under Gaussian and log-concave reference assumptions, respectively. Appendix S15.5 fits a Gaussian from an $O(k^2 + \log T)$ prefix with failure probability at most $1/T$ and obtains $C[k^2 + A_{d,T}]$ regret. Proposition S15.1 transports the statistical lower to a known standard Gaussian reference. Appendix S23 proves the sharp approximate-shape law using a proper strategy for hidden corruptions and calibration from a partially corrupted prefix.

Appendix S13 states the joint-law extension and records the persistence qualifications. Appendices S16 and S17 prove Theorem S13.1, using the total-budget bound in Appendix S6, the tree potential in Appendix S7, and the reference-regularity lower bound in Appendix S14. Appendix S12 proves persistence on one fixed reference under uniformly expected polynomial reference calls.

Appendix S18 states the known-reference finite-bit guarantee, and Appendix S19 gives its complete implementation. Appendix S22 gives the raw-input threshold implementation for an unknown one-dimensional log-concave reference, including calibration and expected bit costs. Appendices S20 and S21 give the quantitative comparisons with the prior statistical bounds and simulator potential.

S2 Uniform control at the last resolution

In these preliminary lemmas, $\mathcal{F}$ is a nonempty finite or countable class of measurable binary functions with VC dimension at most an integer $d \geq 1$, and μ is a fixed probability measure.

We give the VC facts and rare-set estimate used in the upper bound. For a binary class $\mathcal{A}$, let $\Pi_{\mathcal{A}}(n)$ be the maximum number of its traces on n points. The Sauer bound (Sauer, 1972) is

$$\Pi_{\mathcal{F}}(n) \leq \sum_{i=0}^{\min\{d,n\}} \binom{n}{i} \leq (en/d)^d \qquad (n \geq d \geq 1).$$

The last inequality follows by multiplying the sum by $(d/n)^d$, bounding each summand by its term in $(1 + d/n)^n$, and using $(1 + d/n)^n \leq e^d$. Repeated observations cannot increase the number of traces.

Lemma S2.1. *For $0 < \delta \leq 1$, there is a finite proper δ-cover $C \subseteq \mathcal{F}$ in the distance $d_\mu(f, g) = \mu\{f \neq g\}$ such that*

$$\log |C| \leq 2d \log(4e/\delta).$$

For $\mathcal{G} \subseteq \mathcal{F} \triangle \mathcal{F}$ and $n \geq d$, we have $\Pi_{\mathcal{G}}(n) \leq (en/d)^{2d}$.

Proof. Consider M functions with pairwise distance greater than δ. If $\log M \le d$, the desired cardinality bound holds. Otherwise choose $n = \lceil 3 \log M/\delta \rceil$ independent μ-samples. Two fixed functions agree on all samples with probability at most $e^{-n\delta} \le M^{-3}$. A union bound over the pairs is less than one, so some sample distinguishes all M functions. Here $n > d$ and $n \le 4 \log M/\delta$. The Sauer bound gives $M \le (4e \log M/(d\delta))^d$. Set $z = \log M/d > 1$. Then $z \le \log(4e/\delta) + \log z$. If $z \ge 2$, use $\log z \le z/2$; if $1 < z < 2$, use $\log(4e/\delta) > 1$. Both cases give $z \le 2 \log(4e/\delta)$. A greedy construction of a separated subset must therefore stop after finitely many choices, and the resulting subset is a proper cover. A disagreement trace is determined by two traces of $\mathcal{F}$, proving the second assertion. □

Lemma S2.2. *Let $\mathcal{G}$ be a nonempty countable class of measurable indicators with $\sup_{g\in\mathcal{G}} \mu g \le \delta$, where $0 < \delta \le 1$. For an integer $N \ge 1$ and independent $W_1, \dots, W_N$ with law μ, put $Z_N = \sup_{g\in\mathcal{G}} \sum_{i=1}^N g(W_i)$ and $m = N + \lceil N/\delta \rceil$. Then, for every $\lambda \ge 0$,*

$$\mathbb{E}e^{\lambda Z_N} \le 2\Pi_{\mathcal{G}}(m) \exp\bigl(5N\delta(e^\lambda - 1)\bigr).$$

Proof. Draw an independent ghost sample of size $n' = \lceil N/\delta \rceil$. Since Z_N takes values in $\{0, \dots, N\}$, its supremum is attained. Choose the first maximizer in an enumeration of $\mathcal{G}$. Conditional on the original sample, the maximizer's expected ghost count is at most $n'\delta \le N+1 \le 2N$. With conditional probability at least $1/2$, that count is at most $4N$. It follows that

$$\mathbb{E}e^{\lambda Z_N} \le 2\mathbb{E} \max_{g:\ \text{ghost count of } g \le 4N} \exp\left(\lambda \sum_{i=1}^N g(W_i)\right),$$

where an empty maximum is zero.

Generate the two samples equivalently by drawing an independent pool of m observations and then selecting uniformly N positions for the original sample. Conditional on the pool, each trace of $\mathcal{G}$ is a fixed subset A of the positions. A trace in the displayed maximum has $|A| \le 5N$. Bound the maximum by the sum over all such traces, of which there are at most $\Pi_{\mathcal{G}}(m)$. For a fixed trace of size s, let H count its positions in the selected subset and set $b = e^\lambda - 1$. Expanding $(1+b)^H$ and using the factorial moments of sampling without replacement gives

$$\mathbb{E}e^{\lambda H} = \sum_{j=0}^N \binom{N}{j} b^j \frac{(s)_j}{(m)_j} \le \sum_{j=0}^N \binom{N}{j} b^j (s/m)^j \le \exp(Nbs/m).$$

Here $(s)_j/(m)_j = 0$ for $j > s$; otherwise each successive ratio is at most s/m. Since $s \le 5N$ and $m \ge N/\delta$, the last expression is at most $\exp(5N\delta b)$. Sum over the traces and remove the conditioning. □

For $\mathcal{G} \subseteq \mathcal{F} \triangle \mathcal{F}$, $N \ge d$, $\delta = e^{-u}$, and $u \ge 1$, the preceding lemma with $\lambda = u/2$ gives

$$\log \mathbb{E}e^{uZ_N/2} \le \log 2 + 2du + 2d \log(3eN/d) + 5Ne^{-u/2}. \tag{1}$$

Indeed $m \le 3Ne^u$, and Lemma S2.1 bounds the growth function.

S3 Proper games and signed covers

For a family $\mathcal{U}$ of input laws, define the constrained signed complexity

$$\mathfrak{R}_T(\mathcal{F}, \mathcal{U}) = \sup \mathbb{E} \sup_{f\in\mathcal{F}} \sum_{t=1}^T \varepsilon_t f(x_t).$$

In the supremum, an arbitrary $P_t \in \mathcal{U}$ is selected from the preceding history, then $x_t \sim P_t$ is drawn, and then an independent fair sign ε_t is drawn. The history may include earlier auxiliary randomness and observations. The current sign is independent of the current input and all preceding variables.

Lemma S3.1. *Suppose that X and $\mathcal{F}$ are finite and $\mathcal{U}$ is a nonempty compact convex set of input laws. The proper minimax expected classification regret is at most $2\mathfrak{R}_T(\mathcal{F}, \mathcal{U})$.*

Proof. Let $\mathcal{D}$ be the compact convex set of joint input-label laws whose input marginal belongs to $\mathcal{U}$. The state $L \in \mathbb{R}^{\mathcal{F}}$ records cumulative comparator losses. Write $V_t(L)$ for the value of the remaining game, including the final negative comparator loss. Then $V_{T+1}(L) = -\min_f L(f)$ and backward induction gives

$$V_t(L) = \min_{q \in \Delta(\mathcal{F})} \max_{D \in \mathcal{D}} \left\{ \sum_f q(f) D\ell_f + \mathbb{E}_{S \sim D} V_{t+1}(L + \ell_S) \right\}.$$

The fresh hypothesis is hidden from D, so the first term is its expected current loss. The family of allowed laws on future rounds is independent of the state. The objective is continuous and affine in each variable on compact convex sets. The finite-dimensional minimax theorem (Sion, 1958) interchanges the two extrema, giving

$$V_t(L) = \max_{D \in \mathcal{D}} \left\{ \min_f D\ell_f + \mathbb{E}_{S \sim D} V_{t+1}(L + \ell_S) \right\}.$$

For a fixed horizon there are finitely many reachable states and observations, so backward induction supplies measurable proper strategies. Iteration yields

$$\begin{aligned} V_1(0) &= \sup_{D_t} \mathbb{E} \left[\sum_t \min_f D_t \ell_f - \min_f \sum_t \ell_f(S_t) \right] \\ &\le \sup_{D_t} \mathbb{E} \sup_f \sum_t \big(D_t \ell_f - \ell_f(S_t) \big). \end{aligned}$$

The last inequality holds pathwise since each $\min_g D_t \ell_g \le D_t \ell_f$.

For a fixed adaptive law, condition on its entire realized path and draw independent tangent samples S'_t from the corresponding D_t. Conditional Jensen gives

$$\mathbb{E} \sup_f \sum_t (D_t \ell_f - \ell_f(S_t)) \le \mathbb{E} \sup_f \sum_t (\ell_f(S'_t) - \ell_f(S_t)).$$

Generate the same joint law as follows. At time t, draw two conditionally independent samples S_t^+, S_t^- from D_t, which depends on the previously selected real history. Draw a fresh fair sign ε_t. Select S_t^+ as real if $\varepsilon_t = 1$ and S_t^- otherwise; the other sample is tangent. Conditional on the preceding real history, selected and unselected samples are independent with law D_t. Future real laws depend only on the selected history. Induction therefore gives the same joint real and tangent law, including the stated conditional law after the full real path is fixed. The signed difference is

$$\sum_t \varepsilon_t \big(\ell_f(S_t^-) - \ell_f(S_t^+) \big).$$

Its supremum is at most the sum of the two branch suprema. For either branch, write $S_t^\pm = (X_t^\pm, Y_t^\pm)$ and $\ell_f(S_t^\pm) = Y_t^\pm + (1 - 2Y_t^\pm) f(X_t^\pm)$. The expected sign sum of the first term is zero. Conditional on both current samples, $\varepsilon_t(1 - 2Y_t^\pm)$ is a fair sign independent of those samples. Replacing the old sign by this sign preserves nonanticipation: the old sign is recovered from the new sign and the current label before the next round. Conditional on the enlarged preceding history, each branch input still has a law in $\mathcal{U}$. Each branch complexity is thus at most $\mathfrak{R}_T(\mathcal{F}, \mathcal{U})$. Reversing one branch's sign leaves this bound unchanged. Substitution proves the lemma. □

The next finite weighting inequality will be applied at each resolution.

Lemma S3.2. *Let H be a finite nonempty set, let $z_t(h) \in [-1, 1]$, and let $0 < \eta \le 1/2$. Start with uniform weights and multiply the weight of h after round t by $e^{\eta z_t(h)}$. If w_t is the normalized weight before round t, then every $h \in H$ satisfies*

$$\sum_t z_t(h) \le \sum_t \sum_{g \in H} w_t(g) z_t(g) + \frac{\log |H|}{\eta} + \eta \sum_t \sum_{g \in H} w_t(g) z_t(g)^2.$$

Proof. For $|v| \le 1/2$, the Taylor series gives $e^v \le 1 + v + v^2$. For example, the absolute sum of its terms of order at least two is at most $v^2 \sum_{k\ge 2}(1/2)^{k-2}/k! < v^2$. Let W_t be the total unnormalized weight, with initial total one. Then

$$\log(W_{t+1}/W_t) \le \eta \sum_g w_t(g) z_t(g) + \eta^2 \sum_g w_t(g) z_t(g)^2,$$

using $\log(1+a) \le a$ for $a > -1$ after the exponential bound. The quantity inside that logarithm is positive since it bounds a sum of positive exponentials from above. Summing and using $W_{T+1} \ge |H|^{-1} \exp(\eta \sum_t z_t(h))$ proves the assertion. □

Lemma S3.3. *Let $1 \le d \le T$ be integers, let $\mathcal{F}$ be a nonempty countable measurable binary class of VC dimension at most d, and let μ be a fixed probability measure. For any integer $J \ge 1$, put*

$$u_j = 4 \cdot 2^j \quad (0 \le j \le J), \qquad a_j = u_{j-1} \quad (1 \le j \le J).$$

There are finite proper e^{-u_j}-covers C_j and fixed maps $\pi_j : \mathcal{F} \to C_j$ such that the nonempty increment classes

$$H_j = \{g - h : g \in C_j,\ h \in C_{j-1}, \mu\{g \ne h\} \le 2e^{-a_j}\}$$

satisfy

$$\log|C_0| \le Cd, \qquad \log|H_j| \le Cda_j, \qquad |h| \le 1, \quad \mu h^2 \le 2e^{-a_j} \quad (h \in H_j), \tag{2}$$

where C is universal. Fix any rates $0 < \eta_j \le 1/2$. Consider any adaptive input process followed each round by a fresh fair sign ε_t, independent of that input and the complete preceding history. For each H_j, let $w_{t,j}$ be the normalized exponential weights before time t, starting uniformly and updated with rewards $\varepsilon_t h(X_t)$ at rate η_j. Define

$$q_{t,j}(x) = \sum_{h \in H_j} w_{t,j}(h) h(x)^2.$$

These functions are predictable, lie in $[0,1]$, and satisfy $\mu q_{t,j} \le 2e^{-a_j}$ at every history. Moreover,

$$\mathbb{E} \sup_{f\in\mathcal{F}} \sum_{t=1}^T \varepsilon_t f(X_t) \le C\sqrt{dT} + \sum_{j=1}^J \frac{\log|H_j|}{\eta_j} + \mathbb{E}\sum_{t=1}^T \sum_{j=1}^J \eta_j q_{t,j}(X_t) + \mathbb{E} \sup_{f\in\mathcal{F}} \sum_{t=1}^T \mathbf{1}\{f(X_t) \ne \pi_J f(X_t)\}. \tag{3}$$

No restriction on the input laws is imposed in this lemma.

Proof. Take the covers from Lemma S2.1, fix an enumeration of each finite cover, and let $\pi_j f$ be its first member within e^{-u_j} of f. The triangle inequality gives

$$\mu\{\pi_j f \ne \pi_{j-1} f\} \le e^{-u_j} + e^{-u_{j-1}} \le 2e^{-a_j}.$$

Thus every adjacent approximant increment belongs to H_j, and H_j is nonempty. The cover lemma bounds $\log|C_j| \le 2d(\log(4e) + u_j)$. Since $u_j = 2a_j$ and $a_j \ge 4$, $\log|H_j| \le \log|C_j| + \log|C_{j-1}| \le Cda_j$. The base bound follows from $u_0 = 4$. A difference of binary functions has absolute value at most one and square equal to their disagreement indicator. This proves (2) and the bounds on $q_{t,j}$.

Apply Lemma S3.2 to every H_j with $z_t(h) = \varepsilon_t h(X_t)$. The resulting inequalities hold simultaneously for all members of this finite class. Their weighted linear terms have expectation zero: the weights depend only on the preceding history, and the current sign is fair even after conditioning on that history and X_t. Their weighted square terms are exactly $q_{t,j}(X_t)$.

Apply the same lemma to C_0, with rewards $\varepsilon_t g(X_t)$ and rate $\eta_0 = \min\{1/2, \sqrt{d/T}\}$. The weighted square contribution is at most $T\eta_0$ and the linear contribution again has zero expectation. Its entropy and square costs satisfy

$$\frac{\log|C_0|}{\eta_0} + T\eta_0 \le C\sqrt{dT}.$$

Indeed, if $\eta_0 = \sqrt{d/T}$ this follows directly; otherwise $d \ge T/4$, and $Cd + T/2 \le C'\sqrt{dT}$ using $d \le T$.

For every $f \in \mathcal{F}$, pathwise,

$$f = \pi_0 f + \sum_{j=1}^{J} (\pi_j f - \pi_{j-1} f) + (f - \pi_J f).$$

Insert the simultaneous finite weighting inequalities for these increments and for $\pi_0 f$. The resulting linear, entropy and square terms do not depend on f. The remaining signed sum is at most $\sum_t \mathbf{1}\{f(X_t) \neq \pi_J f(X_t)\}$. Take the supremum over all f in this pathwise inequality, then the expectation, and use the zero-mean linear terms and the base bound. This proves (3). Countability makes the terminal and signed suprema measurable. All weights and their square functions are measurable finite functions of preceding observations. □

S4 A fixed reference for lower bounds

We use one reference and class for all horizons and all KL budgets. Let

$$X_1 = \{0\} \cup \left\{ x_{l,v} = \frac{2v+1}{2^l} : l \geq 1,\ 0 \leq v < 2^{l-1} \right\}.$$

Assign reference mass

$$\mu_1(0) = \frac{3}{4}, \qquad \mu_1(\{x_{l,v}\}) = \frac{1}{2^{l+1} l(l+1)}.$$

The total mass on level l is $1/[4l(l+1)]$, whose sum is $1/4$. On $X = \{1, \dots, d\} \times X_1$, let μ choose the component uniformly and then sample from μ_1. Let $\mathcal{F}$ consist of

$$f_\theta(i, x) = \mathbf{1}\{x \geq \theta_i\},$$

where each $\theta_i \in (0, 1)$ is rational and is not a dyadic rational. This class is countable and has VC dimension d. It shatters one midpoint in each component by independently choosing thresholds on either side. Any set of $d+1$ points has two in the same component, whose labels cannot be chosen independently by a threshold.

Lemma S4.1. *Let $n \geq 0$ be an integer and $0 \leq p \leq 1$. Let S be the sum of n independent variables taking values $1, -1, 0$ with respective probabilities $p/2, p/2, 1-p$, and put $v = np$. Then*

$$\mathbb{E}|S| \geq \frac{1}{2} \min\{v, \sqrt{v}\}.$$

Proof. Independence, zero means and symmetry give

$$\mathbb{E}S^2 = v, \qquad \mathbb{E}S^4 = np + 3n(n-1)p^2 \leq v + 3v^2.$$

For the fourth moment, only terms in which every index occurs an even number of times survive: there are n single-index terms with fourth moment p, and $6\binom{n}{2}$ two-index terms with expectation p^2. Hölder's inequality gives

$$\mathbb{E}|S|^2 = \mathbb{E}\big(|S|^{2/3}|S|^{4/3}\big) \leq (\mathbb{E}|S|)^{2/3} (\mathbb{E}|S|^4)^{1/3}.$$

For $v > 0$, rearranging yields

$$\mathbb{E}|S| \geq \frac{v^{3/2}}{\sqrt{v + 3v^2}} = \frac{v}{\sqrt{1 + 3v}} \geq \frac{1}{2} \min\{v, \sqrt{v}\}.$$

Use $1 + 3v \leq 4$ when $v \leq 1$ and $1 + 3v \leq 4v$ when $v \geq 1$. For $v = 0$, including $n = 0$ and $p = 0$, the stated lower bound is zero and holds directly. □

S5 Relative-entropy budgets tending to zero

We prove Theorem 2.1, including universal constants as K tends to zero.

Lemma S5.1. *Let $a_j = 2^{j-1}a_1$, where $a_1 \geq 4$, and suppose $0 \leq q_j \leq 1$ and $\mu q_j \leq 2e^{-a_j}$. For $r = dP/d\mu$, write*

$$B_j = \int_{r<e^{a_j/2}} r q_j \, d\mu, \qquad H_j = \int_{r\geq e^{a_j/2}} r q_j \, d\mu.$$

Then

$$B_j \leq b_j := 2e^{-a_j/2}, \qquad \sum_j a_j H_j \leq 8\,\mathrm{KL}(P\|\mu).$$

Proof. The first bound follows by bounding r on the integration set. For the second, geometric summation gives $\sum_{j:a_j\leq 2\log r} a_j q_j \leq 4\log r$. There is no contributing index unless $r \geq e^2$. The nonnegative function $F(r) = r\log r - r + 1$, with $F(0) = 1$, satisfies $F(r) \geq r\log r/2$ for $r \geq e^2$. Therefore

$$\sum_j a_j H_j \leq 4\int_{r\geq e^2} r\log r \, d\mu \leq 8\int F(r)\, d\mu = 8\,\mathrm{KL}(P\|\mu).$$

If $\mathrm{KL}(P\|\mu) = 0$, all nonnegative high-density terms vanish. □

Lemma S5.2. *Suppose $1 \leq d \leq T$, $u \geq 4\log(eT/d)$, and $\mathcal{G} \subseteq \mathcal{F} \triangle \mathcal{F}$ is countable with $\mu g \leq e^{-u}$ for every $g \in \mathcal{G}$. If the conditional input laws satisfy $\mathrm{KL}(P_t\|\mu) \leq K$, then*

$$\mathbb{E}\sup_{g\in\mathcal{G}} \sum_{t=1}^{T} g(x_t) \leq Cd + \frac{2KT}{u}$$

for a universal C.

Proof. Conditioning only on the input history replaces the detailed conditional input law by a mixture. Convexity of $r\log r$ preserves the KL bound under that mixture. Factoring the joint density into conditional densities and integrating its logarithm gives relative entropy at most KT between the input-path law Q and μ^T. The negative part of each $r\log r$ integral is finite, and each conditional KL is bounded, so the chain-rule integration is valid.

For any bounded Z, tilting a reference law P by the density $e^Z/\mathbb{E}_P e^Z$ and using nonnegativity of relative entropy gives $\mathbb{E}_Q Z \leq \mathrm{KL}(Q\|P) + \log \mathbb{E}_P e^Z$. Apply this inequality to $Z = (u/2)\sup_g \sum_t g(x_t)$ and use (1) with $N = T$. The resulting bound is

$$\frac{2KT}{u} + 4d + \frac{2\log 2 + 4d\log(3eT/d) + 10Te^{-u/2}}{u}.$$

For $R = T/d \geq 1$, the assumption on u gives $Te^{-u/2} \leq dR(eR)^{-2} \leq d$ and $\log(3eR)/u \leq C$. This proves the lemma. □

Proof of the upper bound in Theorem 2.1. Use the signed complexity and weighting inequality of Section S3. Set $u_j = 2^{j+2}$ and choose J least such that $U = u_J \geq 256\sqrt{T/d}$. Then $U < 512\sqrt{T/d}$. Take proper e^{-u_j}-covers, use $a_j = u_{j-1}$ for the increment classes, and define the predictable weighted squared increments $q_{t,j}$ as in Lemma S3.3. We have $\log|H_j| \leq Cda_j$ and $\mu q_{t,j} \leq 2e^{-a_j}$ at every history. The base cover has logarithmic cardinality at most Cd and contributes at most $C\sqrt{dT}$, because $d \leq T$.

For $K > 0$, choose

$$\lambda = \sqrt{\frac{dJ}{KT}}, \qquad \eta_j = \min\left\{\frac{1}{2}, \lambda a_j, \sqrt{\frac{da_j}{Tb_j}}\right\}, \qquad b_j = 2e^{-a_j/2}.$$

For $K = 0$, omit the middle term in the minimum and all subsequent terms involving λ. The entropy penalties satisfy

$$\begin{aligned}\sum_j \frac{Cda_j}{\eta_j} &\leq C\sum_j \left(da_j + \frac{d}{\lambda} + \sqrt{dTa_jb_j}\right) \\ &\leq C\left(dU + \frac{dJ}{\lambda} + \sqrt{dT}\right).\end{aligned}$$

The sum $\sum_j a_j$ is at most U. The sum $\sum_j \sqrt{a_j b_j}$ has a universal bound: the a_j are separated by at least four, so compare it with $\sum_{n\geq 4} \sqrt{2(n+1)}e^{-n/4}$, which converges by the ratio test.

Apply Lemma S5.1 at each history. The variance contribution is at most

$$\begin{aligned}\sum_{t,j} \eta_j \mathbb{E} q_{t,j}(x_t) &\leq T \sum_j \eta_j b_j + \lambda \sum_t \mathbb{E} \sum_j a_j H_{t,j} \\ &\leq \sqrt{dT} \sum_j \sqrt{a_j b_j} + 8\lambda KT \\ &\leq C\sqrt{dT} + 8\lambda KT.\end{aligned}$$

For $K = 0$, the high-density contribution is zero and the calculation holds with only its background term. The fresh sign makes each predictable weighted reward have expectation zero. Telescoping the comparator and applying Lemma S5.2 to the terminal disagreement give

$$\mathfrak{R}_T \leq C\left(\sqrt{dT} + dU + \sqrt{dKTJ} + d + KT/U\right).$$

The choice of U satisfies the terminal lemma's hypothesis since $\log(eR) \leq 2\sqrt{R}$. Also $dU \leq 512\sqrt{dT}$, $KT/U \leq \sqrt{dT}/256$ for $K \leq 1$, and $J \leq C\log(e + T/d)$. Thus

$$\mathfrak{R}_T \leq C\left(\sqrt{dT} + \sqrt{dKT\log(e + T/d)}\right). \tag{4}$$

For a finite space and class, the KL sublevel set is compact, convex, and contains μ, including when $K = 0$. Lemma S3.1 bounds proper regret by twice (4). For a countable class, take a proper e^{-v}-cover with

$$v \geq \max\{4\log(eT/d), KT/d, 1\}$$

and partition the space by its evaluation patterns. Jensen's inequality for $r\log r$ shows that KL decreases under this partition and under the conditional mixtures obtained from the smaller history. Apply the finite proper strategy. Lemma S5.2 bounds the difference of the two hindsight comparator losses by Cd, since $KT/v \leq d$. This is absorbed into $C\sqrt{dT}$. The strategy is measurable and chooses hypotheses from the original class before receiving the input. □

Proof of the lower bound in Theorem 2.1. Use the fixed countable threshold class and reference in Section S4. For $K > 0$, run binary-search epochs with lengths n_l. Let $z_{l,i}$ be component i's current midpoint at epoch l, and let Q_l choose uniformly from these d midpoints. Set

$$q_l = \frac{K}{8l}, \qquad P_t = (1 - q_l)\mu + q_l Q_l.$$

Generate the mixture with an independent Bernoulli indicator B_t of mean q_l. On $B_t = 1$ use Q_l, and otherwise use μ. Give every round a fresh fair label. At the end of each epoch, retain the half interval agreeing with the majority on the informative rounds $B_t = 1$ in that component. Background labels do not enter any interval update. Padding rounds use μ and fresh fair labels. The process ignores learner actions.

KL convexity gives, conditionally on the entire past including latent mixture indicators,

$$\begin{aligned}\mathrm{KL}(P_t\|\mu) \leq q_l\,\mathrm{KL}(Q_l\|\mu) &= \frac{q_l}{d}\sum_i \log\frac{1}{\mu_1(\{z_{l,i}\})} \\ &\leq 3lq_l \leq K.\end{aligned}$$

Here $2^{l+1}l(l+1) \leq 2^{3l}$, so the logarithm of the reciprocal atom mass is at most $3l$. Padding has KL zero.

Every learner has expected loss $T/2$, since each current label is fair even after the current input and prediction are fixed. The final retained intervals contain a comparator agreeing with all informative majorities. Choose its threshold vector as a fixed function of the informative indicators, components, and labels. This comparator is independent of every background label, even after the background inputs and all informative variables are fixed: the process and its updates never use background

labels. Its expected background loss is therefore half their expected count. On informative rounds its saving is $\frac{1}{2}\sum_{l,i}|S_{l,i}|$, where $S_{l,i}$ is the informative label imbalance. The second- and fourth-moment calculation in Lemma S4.1, now with probabilities $q_l/(2d), q_l/(2d), 1-q_l/d$, gives

$$\mathbb{E}\,\mathrm{Reg}_T \geq \frac{1}{4}\sum_l \min\{n_l q_l, \sqrt{d n_l q_l}\}. \tag{5}$$

This comparator is permitted in agnostic regret; it need not be determined by the learner's observations alone.

Write $R = T/d$ and $z = KR$. For z larger than a sufficiently large universal constant, choose

$$L = \lfloor z^{1/4} \rfloor, \qquad H_L = \sum_{l=1}^{L} \frac{1}{l}, \qquad S_L = \frac{K}{8} H_L, \qquad n_l = \left\lfloor \frac{T q_l}{2 S_L} \right\rfloor.$$

The lengths sum to at most $T/2$. Each quantity inside a floor is at least $T/(2LH_L) \geq z/[2z^{1/4}(1+\log z)]$, which exceeds two for all sufficiently large z. Consequently

$$\frac{n_l q_l}{d} \geq \frac{TK}{32 d l^2 H_L} \geq \frac{\sqrt{z}}{32 H_L} \geq 1$$

after increasing that universal threshold. The square-root branch in (5) applies, and

$$\mathbb{E}\,\mathrm{Reg}_T \geq \frac{1}{8}\sqrt{dTS_L} \geq c\sqrt{dKT\log z}. \tag{6}$$

The last inequality uses $L \geq z^{1/4}/2$ and $H_L \geq \log(L+1) \geq c\log z$ for large z.

The same class and reference give a $c\sqrt{dT}$ lower bound at KL zero. Use independent reference inputs and independent fair labels. In each component compare thresholds $1/3$ and $2/3$, whose disagreement probability under μ_1 is a fixed $\delta_0 \geq 1/8$. Minimizing over their 2^d coordinatewise combinations saves half the sum of the d imbalance magnitudes. Each magnitude has variance $T\delta_0/d$ and the fourth-moment bound already proved. Since $T/d \geq 1$, this yields

$$\mathbb{E}\,\mathrm{Reg}_T \geq cd\min\{T\delta_0/d, \sqrt{T\delta_0/d}\} \geq c'\sqrt{dT}. \tag{7}$$

Independence between components is not needed when summing these expectations.

If $z \leq \sqrt{R}$, then $K \leq R^{-1/2}$, so $K\log(e+R)$ is universally bounded and (7) controls both terms in the theorem. If $z > \sqrt{R}$ and z exceeds the threshold for (6), then $\log z > \frac{1}{2}\log R$. For $R \geq 2$, the shift lower bound controls the required second term. The range $1 \leq R < 2$ is covered by the baseline. If $z > \sqrt{R}$ but z is below the fixed threshold, then $R < z^2$ is universally bounded and the baseline again suffices. The maximum of the two valid lower bounds is at least half their sum. This proves the theorem with universal constants. □

S6 A TOTAL CONDITIONAL KL BUDGET

For integers $1 \leq d \leq T$ and $B \geq 0$, define

$$\Phi_{d,T}(B) = \begin{cases} \sqrt{dT} + \sqrt{dB\log(e+T/d)}, & 0 \leq B \leq T, \\ \sqrt{dT} + \sqrt{dB\log(eT/\sqrt{dB})}, & T < B \leq T^2/d, \\ T, & B > T^2/d. \end{cases} \tag{8}$$

The middle interval is empty when $d = T$. Adjacent expressions agree within universal constant factors. Let $\overline{V}_T^{\mathrm{all}}(d,B)$ and $\overline{V}_T^{\mathrm{prop}}(d,B)$ be the worst-case statistical minimax expected regrets under the pathwise constraint $\sum_t c_t \leq B$ almost surely, with VC dimension at most d. The strategy may depend on d, T, B, μ, and the class.

Theorem S6.1. *There are universal constants $c, C > 0$ such that, for all integers $1 \leq d \leq T$ and all $B \geq 0$,*

$$c\Phi_{d,T}(B) \leq \overline{V}_T^{\mathrm{all}}(d,B) \leq \overline{V}_T^{\mathrm{prop}}(d,B) \leq C\Phi_{d,T}(B).$$

The proper upper remains valid under the weaker constraint $\mathbb{E}\sum_t c_t \leq B$. For each d, the pathwise lower holds on one fixed countable threshold-product class of VC dimension d under one fixed reference, independent of T and B.

The learner need not observe the conditional input laws or their costs. When $d/T \to 0$, $\Phi_{d,T}(B) = o(T)$ exactly for $B = o(T^2/d)$.

We prove Theorem S6.1. The upper bound uses the expected sum of full-history conditional KL costs. This permits finite partitions without assuming that a pathwise remaining budget survives conditional averaging.

S6.1 A PENALIZED PROPER GAME

First suppose that X and $\mathcal{F}$ are finite and $\mu(x) > 0$ for every $x \in X$. For $\tau > 0$, define

$$\mathcal{S}_\tau = \sup \mathbb{E}\left[\sup_{f\in\mathcal{F}} \sum_{t=1}^{T} \varepsilon_t f(x_t) - \tau \sum_{t=1}^{T} \mathrm{KL}(P_t\|\mu)\right].$$

The supremum is over adaptive input processes with arbitrary predictable auxiliary state. At each time, P_t is the input law given that preceding state, and the fresh fair sign is independent of the input and state.

Lemma S6.2. *For every $\tau > 0$, there is a proper learner such that every adversary satisfies*

$$\mathbb{E}\,\mathrm{Reg}_T \le 2\mathcal{S}_\tau + 2\tau\mathbb{E}\sum_{t=1}^{T} \mathrm{KL}(P_t\|\mu).$$

The learner need not observe the conditional input laws or their costs.

Proof. Consider the auxiliary game with payoff $\mathrm{Reg}_T - 2\tau\sum_t \mathrm{KL}(P_t\|\mu)$. Let L record the cumulative comparator losses and put $W_{T+1}(L) = -\min_f L(f)$. Backward induction gives

$$W_t(L) = \min_{q\in\Delta(\mathcal{F})} \max_{D} \Big\{ \sum_f q(f) D\ell_f - 2\tau\,\mathrm{KL}(D_X\|\mu) + \mathbb{E}_{S\sim D} W_{t+1}(L + \ell_S) \Big\},$$

where D ranges over the simplex of joint input-label laws. Full support makes its KL term finite and continuous. The objective is affine in q and concave in D, because negative KL is concave and the continuation term is affine in D. The finite-dimensional minimax theorem (Sion, 1958) interchanges the extrema on these compact convex domains. The minimizing choices of q give a proper strategy depending only on L. There are finitely many reachable L for a fixed horizon. Past penalty payments are additive constants and restrict no future action in this auxiliary game, so they need not be observed.

Iterating the interchanged recursion and using $\min_g D_t\ell_g \le D_t\ell_f$ for every f gives

$$W_1(0) \le \sup_{D_t} \mathbb{E}\left[\sup_f \sum_t (D_t\ell_f - \ell_f(S_t)) - 2\tau\sum_t c_t\right], \qquad c_t = \mathrm{KL}((D_t)_X\|\mu).$$

For a fixed adaptive law, draw tangent samples from the corresponding D_t, independently conditional on the complete realized real path. Conditional Jensen bounds the centered supremum by the expected supremum of tangent-minus-real losses. The cost term depends on the real path and is unchanged.

Generate the same joint law by drawing two conditionally independent samples S_t^+, S_t^- from D_t and then a fresh fair sign ε_t. Select S_t^+ as real when $\varepsilon_t = 1$, and select S_t^- otherwise. The unselected sample is tangent. Future laws depend on the selected real history. Conditional on the preceding real history, the selected and unselected samples are independent with common law D_t. Induction therefore gives the asserted real/tangent joint law.

The tangent-minus-real expression is

$$\sup_f \sum_t \varepsilon_t\big(\ell_f(S_t^-) - \ell_f(S_t^+)\big).$$

Split this supremum into the two branch suprema and assign the penalty $\tau \sum_t c_t$ to each branch. For either branch, $\ell_f(X, Y) = Y + (1 - 2Y)f(X)$. The signed label-only term has expectation zero. Conditional on both current samples, multiplying the fresh sign by $1 - 2Y$ leaves a fair sign. The old sign is recovered from the transformed sign and current label before the next round.

More explicitly, take the preceding filtration to contain all past paired samples and signs and the selected real history. Given that filtration, each current branch input has law $(D_t)_X$, because D_t uses the selected real history and the new samples are drawn independently. After the sign transformation, store both current samples and recover the old sign; the next preceding filtration contains the same information. Thus each branch is an admissible process in the definition of $\mathcal{S}_\tau$, with conditional cost exactly c_t. Reversing one branch sign does not change this conclusion. Each penalized branch expectation is at most $\mathcal{S}_\tau$, so $W_1(0) \le 2\mathcal{S}_\tau$. The strategy attaining this game bound proves the lemma. □

S6.2 The charge and the final approximation

Write $C_* = \mathbb{E} \sum_t c_t$ for an arbitrary adaptive process with finite expected cost. At dyadic scales $a_j \ge 4$, let $q_{t,j}$ be predictable, with $0 \le q_{t,j} \le 1$ and $\mu q_{t,j} \le 2e^{-a_j}$. Split $P_t q_{t,j}$ according to the density cutoff $r_t \ge e^{a_j/2}$:

$$P_t q_{t,j} = B_{t,j} + H_{t,j}, \qquad B_{t,j} \le b_j := 2e^{-a_j/2}.$$

Lemma S5.1 gives

$$\sum_j a_j H_{t,j} \le 8c_t, \qquad \mathbb{E} \sum_{t,j} a_j H_{t,j} \le 8C_*. \tag{9}$$

The same assertion holds for any subset of scales by setting the other $q_{t,j}$ equal to zero.

Lemma S6.3. *Suppose $1 \le d \le T$, $u \ge 4 \log(eT/d)$, and $\mathcal{G} \subseteq \mathcal{F} \triangle \mathcal{F}$ is countable with $\mu g \le e^{-u}$ for every $g \in \mathcal{G}$. Then every process of expected total conditional KL cost $C_* < \infty$ satisfies*

$$\mathbb{E} \sup_{g \in \mathcal{G}} \sum_{t=1}^{T} g(x_t) \le Cd + \frac{2C_*}{u}.$$

Proof. Conditioning only on the input history replaces the full-history input laws by mixtures. Convexity of KL gives

$$\mathbb{E}\, \mathrm{KL}(\mathrm{Law}(x_t \mid x_{<t}) \| \mu) \le \mathbb{E} c_t.$$

The chain rule therefore bounds the divergence of the input-path law from μ^T by C_*. Finite expected cost and integrability of the negative parts justify the density-logarithm integrations.

For bounded Z, tilting a reference law Q by $e^Z / \mathbb{E}_Q e^Z$ and using nonnegativity of KL gives $\mathbb{E}_P Z \le \mathrm{KL}(P\|Q) + \log \mathbb{E}_Q e^Z$. Apply this with $Z = (u/2) \sup_g \sum_t g(x_t)$ and use (1). The result is

$$\frac{2C_*}{u} + 4d + \frac{2 \log 2 + 4d \log(3eT/d) + 10Te^{-u/2}}{u}.$$

For $R = T/d \ge 1$, the hypothesis gives $Te^{-u/2} \le dR(eR)^{-2} \le d$ and $\log(3eR)/u \le C$. This proves the lemma. □

S6.3 Coarse and fine resolutions

Proof of the upper bound in Theorem S6.1. At $B = 0$, every nonnegative conditional cost is zero almost surely, even under the expected-cost convention. The proper upper bound in Theorem 2.1 at $K = 0$ is $C\sqrt{dT}$. For $B \ge T^2/d$, regret is at most T. Assume $0 < B \le T^2/d$.

Use resolutions $u_j = 4 \cdot 2^j$, $0 \le j \le J$, with J least such that

$$U = u_J \ge 256 \sqrt{\max\{T, B\}/d}.$$

Then $U < 512 \sqrt{\max\{T, B\}/d}$. Put

$$a_j = u_{j-1} \quad (1 \le j \le J), \qquad A_0 = 4 \max\{1, B/T\}.$$

Call a scale coarse if $a_j \le A_0$ and fine otherwise. Let M be the number of fine scales. We have $U/A_0 \ge 64$, so $M \ge 1$.

Use the covers and adjacent increment classes of Lemma S3.3. They satisfy

$$\log |H_j| \le C d a_j, \qquad |h| \le 1, \qquad \mu h^2 \le 2e^{-a_j} \quad (h \in H_j).$$

Apply Lemma S3.2 to signed rewards $\varepsilon_t h(x_t)$. Its predictable weighted linear terms have mean zero. Its weighted square functions are the $q_{t,j}$ above. The base cover contributes at most $C\sqrt{dT}$.

For a coarse scale, use $\eta_j = \min\{1/2, \sqrt{da_j/T}\}$. Since every conditional square moment is at most one, its contribution is at most $C(\sqrt{dTa_j} + da_j)$. The dyadic sums are at most

$$C\left(\sqrt{dTA_0} + dA_0\right). \tag{10}$$

For a fine scale, set

$$\lambda = \sqrt{dM/B}, \qquad \eta_j = \min\left\{\frac{1}{2}, \lambda a_j, \sqrt{\frac{da_j}{Tb_j}}\right\}, \qquad b_j = 2e^{-a_j/2}.$$

The entropy penalties sum to at most

$$C\sum_{j \text{ fine}}\left(da_j + \frac{d}{\lambda} + \sqrt{dTa_jb_j}\right) \le C\left(dU + \frac{dM}{\lambda} + \sqrt{dT}\right).$$

Here $\sum_j a_j \le U$. The sum $\sum_j \sqrt{a_jb_j}$ is at most the universal constant

$$\sum_{n\ge 4}\sqrt{2(n+1)}e^{-n/4}.$$

By (9), the fine-scale variance contribution is at most

$$T\sum_{j \text{ fine}}\eta_j b_j + \lambda\mathbb{E}\sum_{t,j \text{ fine}} a_j H_{t,j} \le C\sqrt{dT} + 8\lambda C_*.$$

Telescope every comparator through the covers and bound its final signed difference by the supremum of the unsigned disagreement counts. Lemma S6.3 applies at U, since $\log(eR) \le 2\sqrt{R}$ for $R = T/d \ge 1$. The resulting bound, valid for every signed process, is

$$\mathbb{E}\sup_f \sum_t \varepsilon_t f(x_t) \le A_* + (8\lambda + 2/U)C_*, \tag{11}$$

where

$$A_* = C\left(\sqrt{dT} + \sqrt{dTA_0} + dA_0 + dU + \frac{dM}{\lambda} + d\right).$$

Thus $\mathcal{S}_\tau \le A_*$ for $\tau = 8\lambda + 2/U$. Lemma S6.2 gives, for every process with expected cost at most B,

$$\begin{aligned}\mathbb{E}\,\mathrm{Reg}_T &\le 2A_* + 2(8\lambda + 2/U)B \\ &\le C\left(\sqrt{d\max\{T,B\}} + \sqrt{dBM}\right).\end{aligned} \tag{12}$$

Indeed, dU, B/U, and dA_0 are at most $C\sqrt{d\max\{T,B\}}$; the last of these uses $B \le T^2/d$. Also $dM/\lambda = \lambda B = \sqrt{dBM}$.

If $B \le T$, then $M \le C\log(e + T/d)$. If $T < B \le T^2/d$, put $r = T/\sqrt{dB} \ge 1$. Since $U/A_0 \le 128r$, we have $M \le C\log(er)$. Substitution in (12) proves the finite-class bound.

For a countable class, choose a finite proper e^{-v}-cover, where

$$v \ge \max\{4\log(eT/d), B/d, 1\},$$

and partition the input space by its evaluation patterns. Discard reference-null cells, which admissible inputs never visit. Apply the finite proper strategy to the induced class and reference. Data processing and conditional averaging imply that the expected sum of the induced conditional KL costs is at most the original expected sum. A pathwise budget need not survive this averaging, but (12) already applies under the expected-cost convention.

The two hindsight comparator losses differ by at most the supremum of the terminal disagreement counts. Lemma S6.3 bounds its expectation by $Cd + 2B/v \le C'd$, absorbed into $C\sqrt{dT}$. The finite strategy depends on finitely many comparator-loss states and the measurable evaluation partition, and selects hypotheses in the original class. It is therefore a measurable proper learner. This proves the upper bound without access to the conditional KL expenditure. □

The penalized game also gives a single strategy whose guarantee is affine in the actual expected cost. We use it when the reference is estimated before play and its estimation error is not known to the learner.

Lemma S6.4. *Let $1 \le d \le T$ be integers, let ν be a known reference, and let $\mathcal{F}$ be a known nonempty countable measurable binary class with $\mathrm{VC}(\mathcal{F}) \le d$ in the model of Section 2. There is a measurable proper pre-input strategy, depending only on $T, d, \mathcal{F}, \nu$, such that every protocol satisfying the current-round privacy convention and*

$$C_* := \mathbb{E}\sum_{t=1}^{T} \mathrm{KL}(P_t \| \nu) < \infty$$

has

$$\mathbb{E}\,\mathrm{Reg}_T \le C A_{d,T} + C\frac{A_{d,T}}{T} C_*$$

for a universal C. The same strategy works for every C_* and does not observe the input laws or their costs.

Proof. If $d = T$, any fixed hypothesis has regret at most $T = A_{d,T}$. Assume $d < T$ and first take a finite space and class with a full-support reference. Use the preceding upper construction with the fixed tuning value $B_0 = T$. Its scales satisfy

$$256\sqrt{T/d} \le U < 512\sqrt{T/d}, \qquad A_0 = 4, \qquad 1 \le M \le C\log(eT/d).$$

The last bound follows by counting the dyadic scales up to U. Set $\lambda = \sqrt{dM/T}$ and $\tau = 8\lambda + 2/U$ as in that construction. Then

$$\tau \le C A_{d,T}/T,$$
$$A_* \le C\left(\sqrt{dT} + dU + \frac{dM}{\lambda} + d\right) \le C A_{d,T}.$$

Indeed, $dU \le 512\sqrt{dT}$, $dM/\lambda = \sqrt{dTM}$, and $d \le \sqrt{dT}$. Equation (11) holds for every actual finite expected cost, before any bound by the tuning value is imposed. Thus it gives $\mathcal{S}_\tau \le A_*$ for this fixed τ. All input costs on the finite full-support space are finite and bounded. Lemma S6.2 supplies one strategy with regret at most $2A_* + 2\tau C_*$ for every adversary, proving the finite case.

For a countable class, fix a finite proper e^{-v}-cover $H \subseteq \mathcal{F}$ under ν, where

$$v = \max\{4\log(eT/d), T/d, 1\}.$$

For each $f \in \mathcal{F}$, fix $\pi f \in H$ with $\nu\{f \ne \pi f\} \le e^{-v}$, and partition the input space by the evaluation patterns of H. Discard reference-null cells and assign an arbitrary fixed continuation there to define the strategy everywhere. Finite expected cost implies $P_t \ll \nu$ almost surely, so these cells are never visited except on a null event. The induced reference has full support. Its class has VC dimension at most d: a shattered set of cells would give a shattered set of original points by choosing one point in each cell.

Apply the fixed finite strategy to this partition and select the corresponding hypotheses in H. Data processing and conditional averaging bound the induced expected KL cost by C_*, as in the preceding upper proof. Pathwise, regret against $\mathcal{F}$ is at most regret against H plus

$$\sup_{f\in\mathcal{F}} \sum_t \mathbf{1}\{f(x_t) \ne \pi f(x_t)\}.$$

Lemma S6.3 bounds the expectation of this countable disagreement supremum by $Cd + 2C_*/v$. Since $d \le A_{d,T}$ and $1/v \le d/T \le A_{d,T}/T$, these terms are absorbed into the claimed bound. The evaluation partition is measurable and the finite strategy uses finitely many comparator-loss states, so the lifted strategy is measurable and proper. Its cover, rates and strategy were all chosen independently of C_*. □

S6.4 Fixed-reference midpoint epochs

We use the class and reference in Section S4, independently of T and B. The zero-cost baseline is (7). For $B \le T$, the lower bound in Theorem 2.1, with $K = B/T$, satisfies summed cost at most B on every path. This gives the first range of Theorem S6.1. For larger budgets we use the following allocation of midpoint epochs.

Lemma S6.5. *Choose integers $1 \le a \le L$ and positive integer lengths n_l, $a \le l \le L$, with $\sum_l n_l \le T$. There is an adversary on the fixed class and reference such that, on every path,*

$$\sum_t c_t \le 3 \sum_{l=a}^{L} l n_l,$$

and every unrestricted learner satisfies

$$\mathbb{E}\,\mathrm{Reg}_T \ge \frac{1}{4} \sum_{l=a}^{L} \min\{n_l, \sqrt{d n_l}\}.$$

Proof. In each component choose a fixed dyadic interval of depth $a-1$ before play. At level l, use n_l rounds, each choosing uniformly among the components' current midpoints and receiving a fresh fair label. At the end of the epoch retain, in each component, the half interval on which a threshold agrees with its majority label. Break ties by a fixed rule. The process ignores learner actions. Choosing an initial interval requires no preceding rounds; it restricts the thresholds used by the construction.

At level l, the density at a current midpoint is the reciprocal of its one-component reference mass. The component probabilities cancel. By the reference definition,

$$\frac{1}{\mu_1(\{x_{l,v}\})} = 2^{l+1} l(l+1) \le 2^{3l}.$$

The inequality holds at $l = 1$. Its left-side ratio from l to $l+1$ is $2(l+2)/l \le 8$, proving induction. Taking natural logarithms shows that each conditional epoch cost is at most $3l$. Use independent reference inputs and fair labels for all remaining padding rounds, which have zero cost. This proves the pathwise budget bound.

Choose a final rational nondyadic threshold in each retained open interval as a fixed function of the epoch variables. It agrees with every epoch majority and is independent of padding labels, including after all epoch variables and padding inputs are fixed. Every learner has expected loss $T/2$. The comparator saves half the sum of the absolute epoch label imbalances, and has no expected advantage on padding.

A component imbalance is a sum of independent variables with probabilities $1/(2d), 1/(2d), 1-1/d$ for values $1, -1, 0$. It has variance $v = n_l/d$ and fourth moment at most $v + 3v^2$. Hölder gives

$$\mathbb{E}|S| \ge \frac{(\mathbb{E}S^2)^{3/2}}{(\mathbb{E}S^4)^{1/2}} \ge \frac{1}{2} \min\{v, \sqrt{v}\}.$$

Summing over components and epochs proves the assertion. □

Proof of the remaining lower bounds. Assume $T \le B \le T^2/d$ and put $r = T/\sqrt{dB} \ge 1$. First suppose $r \ge R_0 = e^{20}$. Set

$$a = \lceil B/T \rceil, \qquad L = \lfloor a\sqrt{r} \rfloor, \qquad H = \sum_{l=a}^{L} \frac{1}{l}, \qquad n_l = \left\lfloor \frac{B}{100 H l^2} \right\rfloor.$$

We have $B/T \le a \le 2B/T$ and $L \ge a\sqrt{r}/2$. Comparison with a logarithmic integral gives

$$\tfrac{1}{2}\log r - \log 2 \le H \le 1 + \tfrac{1}{2}\log r.$$

Thus $H \asymp \log(er)$ and $H \ge 1$. The time used is at most

$$\frac{B}{100H}\sum_{l=a}^{\infty}\frac{1}{l^2} \le \frac{B}{50Ha} \le \frac{T}{50H} \le T,$$

since the series is at most $1/a^2 + 1/a \le 2/a$. The pathwise cost in Lemma S6.5 is at most $3B/100$.

At every used level,

$$\frac{B}{100Hl^2} \ge \frac{B}{100Ha^2 r} \ge \frac{dr}{400H} \ge 2d.$$

For the last inequality, use $H \le 1 + \frac{1}{2}\log r$. The ratio $r/(1 + \frac{1}{2}\log r)$ is increasing for $r \ge 1$ and already exceeds 800 at e^{20}. The floors therefore retain at least half each unrounded length and give $n_l \ge d$. The square-root branch of the epoch bound applies, giving

$$\mathbb{E}\,\mathrm{Reg}_T \ge \frac{1}{4}\sqrt{\frac{dB}{200H}}\sum_{l=a}^{L}\frac{1}{l} = \frac{\sqrt{dBH}}{4\sqrt{200}} \ge c\sqrt{dB\log(er)}.$$

Together with the zero-cost baseline, this proves the claimed lower rate when $r \ge R_0$.

For bounded r, a simpler allocation gives $c\sqrt{dB}$. If $B/d \ge 48$, set

$$L = \left\lfloor \sqrt{B/(12d)} \right\rfloor, \qquad a = 1, \qquad n_l = d \quad (1 \le l \le L).$$

Then $L \ge \sqrt{B/(12d)}/2$. The time is $dL \le \sqrt{dB/12} \le T$, and the cost is at most

$$\frac{3d}{2}L(L+1) \le 3dL^2 \le B/4.$$

The regret is at least $dL/4 \ge c\sqrt{dB}$. If $B/d < 48$, then $B/T \le B/d < 48$, so the zero-cost baseline already has this order. For $1 \le r < R_0$, the factor $\sqrt{\log(er)}$ is at most $\sqrt{21}$. This proves the remaining middle-range lower bound.

At $B_0 = T^2/d$ this construction has $r = 1$ and regret at least cT. The same adversary is admissible for every larger budget. Thus monotonicity gives the final range. All constructions use the same class and reference for each dimension, completing the proof. □

S7 A PROPER COVER-TREE IMPLEMENTATION

We replace the finite game's state enumeration by a recursive potential on a reference-metric cover tree. Constructive chaining (Gaillard and Gerchinovitz, 2015), multiscale conditional entropy (Coester and Lee, 2019), and hierarchical expert selection (Yang et al., 2022) are established methods. We use a hierarchical entropy potential whose local loss changes are controlled by predictable reference disagreements. The resulting algorithm admits the same KL charge as the statistical proof. Its internal computation is polynomial in the acquired finite table. The costs of implementing supplied ERM and reference-sampling oracles are separate from this internal bound.

Theorem S7.1. *Let $\mathcal{F}$ be a nonempty finite binary class, supplied as a table of m functions on s inputs, with $\mathrm{VC}(\mathcal{F}) \le d$ and $1 \le d \le T$. Let q be a supplied strictly positive rational reference law on these inputs, and let b be a positive integer with $b < T^2/d$. There is a proper learner with internal arithmetic and bit complexity polynomial in m, s, T and the supplied rational bit length, and constants $A \ge 0$ and $c_b > 0$, depending only on d, T, b, such that*

$$\mathbb{E}\,\mathrm{Reg}_T(\mathcal{F}) \le A + c_b C_* + 1, \qquad A + c_b b \le C\Phi_{d,T}(b). \tag{13}$$

Here $C_* < \infty$ is the actual expected total full-history conditional input KL cost relative to q, and C is universal. The learner selects its current function before the current input and keeps its fresh selection private until the labeled example is generated. Conditional labels are unrestricted under this privacy convention. The learner need not observe the input laws or their KL costs.

The first inequality holds without assuming $C_* \le b$. This affine form will permit conditioning on random preprocessing.

S7.1 Constructing the tree

Put $u_j = 4 \cdot 2^j$ for $0 \le j \le J$, where J is least such that

$$U = u_J \ge 256\sqrt{\max\{T, b\}/d}.$$

Then $U < 512\sqrt{\max\{T, b\}/d} \le 512T$. Set

$$a_j = u_{j-1} \quad (1 \le j \le J), \qquad \delta_j = 2^{-6 \cdot 2^j}.$$

The radii are rational, $\delta_{j+1} = \delta_j^2$, and $\delta_j \le e^{-u_j}$ because $6 \log 2 > 4$.

For every j, traverse the supplied function list and retain a function exactly when its q disagreement distance from all retained functions exceeds δ_j. The resulting C_j is a proper δ_j-cover and a separated set. The packing argument in Lemma S2.1 gives

$$\log |C_0| \le Cd, \qquad \log |C_j| \le Cdu_j. \tag{14}$$

For each center in C_j, $j \ge 1$, choose its first center within δ_{j-1} in C_{j-1} as parent. Connect C_0 to a dummy root. Occurrences of the same function at different levels are different nodes. The leaves are C_J; prune every node with no descendant in C_J. Pruning preserves retained edge distances and cannot increase a level's cardinality. Every root-to-leaf path contains one node from each level $0, \dots, J$. Write h_v for a nonroot node's center.

If f is a leaf and v is its level $j-1$ ancestor, then

$$q\{f \ne h_v\} \le \sum_{l=j-1}^{J-1} \delta_l \le 2\delta_{j-1} \le 2e^{-a_j}. \tag{15}$$

For the middle inequality, if $0 < \delta \le 1/64$, then $\sum_{k \ge 0} \delta^{2^k} \le \sum_{k \ge 0} \delta^{k+1} = \delta/(1-\delta) \le 2\delta$. Every construction uses only distances computed from the supplied table.

S7.2 Nondecreasing split rates

Set $A_0 = 4 \max\{1, b/T\}$. Scale $j \ge 1$ is coarse when $a_j \le A_0$ and fine otherwise. Let M be the number of fine scales. If $b \le T$, then $U/A_0 \ge 64\sqrt{T/d} \ge 64$. If $b > T$, then $U/A_0 \ge 64T/\sqrt{db} > 64$. Thus $M \ge 1$. Define

$$\begin{aligned}
\eta_0 &= \min\{1/2, \sqrt{d/T}\}, \\
\eta_j &= \min\{1/2, \sqrt{da_j/T}\} && \text{at coarse scales,} \\
\lambda &= \sqrt{dM/b}, \qquad \gamma_j = 2^{1-a_j/2}, \\
\eta_j &= \min\left\{\frac{1}{2}, \lambda a_j, \sqrt{\frac{da_j}{T\gamma_j}}\right\} && \text{at fine scales.}
\end{aligned} \tag{16}$$

Here $\gamma_j \ge 2e^{-a_j/2}$ is rational, and

$$\sum_j \sqrt{a_j \gamma_j} \le \sum_{n \ge 4} \sqrt{2(n+1)}\, 2^{-n/4} < \infty. \tag{17}$$

These rates satisfy

$$0 < \eta_0 \le \eta_1 \le \cdots \le \eta_J \le \frac{1}{2}. \tag{18}$$

The coarse rates increase with a_j. Each uncapped fine term also increases, since γ_j decreases. At the transition write $a_f = 2a_c$, where a_c is the last coarse scale. Since $\gamma_f \le 2$,

$$\sqrt{\frac{da_f}{T\gamma_f}} \ge \sqrt{\frac{da_c}{T}}, \qquad \frac{(\lambda a_f)^2}{da_c/T} = \frac{2MTa_f}{b} > 8M \ge 1.$$

The strict inequality uses $a_f > A_0 \ge 4b/T$. Taking minima with $1/2$ preserves the ordering. Finally $a_1 = 4 \le A_0$ and $\eta_1 \ge \eta_0$, proving (18).

S7.3 The recursive potential

Let L_f be the cumulative loss of leaf f. Set $V_f(L) = L_f$ at a leaf. For an internal node v whose children are at level j, give its split the rate $\eta(v) = \eta_j$; the dummy root has rate η_0. Define

$$V_v(L) = -\frac{1}{\eta(v)} \log \sum_{c \text{ child of } v} e^{-\eta(v) V_c(L)},$$
$$p(c \mid v; L) = \frac{e^{-\eta(v) V_c(L)}}{\sum_{c' \text{ child of } v} e^{-\eta(v) V_{c'}(L)}}. \tag{19}$$

Write $V = V_{\text{root}}$, and let $p_f(L)$ be the product of the conditional probabilities along the path to f. Differentiation gives $\nabla V(L) = p(L)$. Before each input, compute the node values from the leaves upward and privately sample a leaf by its conditional probabilities. After the label, update all leaf losses from the table. We first analyze exact probabilities, then give a finite implementation.

An unnormalized soft minimum is at most its smallest argument. Consequently $V(L) \le \min_f L_f$. It is also at least its smallest argument minus the logarithm of its number of terms divided by its rate. Applying this inequality successively from the leaves upward at $L = 0$ gives

$$-V(0) \le \frac{\log |C_0|}{\eta_0} + \sum_{j=1}^{J} \frac{\log |C_j|}{\eta_j}. \tag{20}$$

Each number of children is bounded by its level cardinality. The soft minima in (19) are unnormalized; no additional prior factor is inserted in their sums.

S7.4 Stability through predictable disagreements

Fix any leaf loss vector $\ell \in [0,1]^{|C_J|}$ and consider $L + z\ell$ for $0 \le z \le 1$. Put

$$g_v(z) = \frac{d}{dz} V_v(L + z\ell), \qquad H_v(z) = -\frac{d^2}{dz^2} V_v(L + z\ell).$$

At leaves, $g_f = \ell_f$ and $H_f = 0$. Differentiating (19) gives

$$g_v = \sum_c p(c \mid v) g_c,$$
$$H_v = \sum_c p(c \mid v) H_c + \eta(v) \operatorname{Var}_{c \sim p(\cdot \mid v)}(g_c). \tag{21}$$

In particular, every g_v lies in $[0,1]$. If p_v is the probability of reaching v, expansion yields

$$H_{\text{root}} = \sum_{v \text{ internal}} p_v \eta(v) \operatorname{Var}_{c \sim p(\cdot \mid v)}(g_c). \tag{22}$$

For a root-to-leaf path,

$$\frac{d}{dz} \log p_f(L + z\ell) = \sum_{v \text{ on path}} \eta(v)(g_v - g_{\text{child}})$$
$$= \eta_0 g_{\text{root}} + \sum_{\substack{v \text{ nonroot internal} \\ v \text{ on path}}} \big(\eta(v) - \eta(\text{parent}(v))\big) g_v - \eta_J \ell_f. \tag{23}$$

All coefficients in the positive part are nonnegative by (18), and their sum is $\eta_J \le 1/2$. The derivative is therefore in $[-\eta_J, \eta_J]$, so

$$p_f(L + z\ell) \le e^{z/2} p_f(L). \tag{24}$$

This bound has no factor depending exponentially on the tree depth.

Now let $\ell_f = \mathbf{1}\{f(x) \ne y\}$ at an arbitrary labeled input. For a nonroot internal node v, variance is no larger than squared deviation about the loss ℓ_{h_v} of its center. Conditional Jensen gives

$$\operatorname{Var}_{c \sim p(\cdot \mid v)}(g_c) \le \sum_c p(c \mid v)(g_c - \ell_{h_v})^2$$
$$\le \mathbb{E}_{f \mid v}(\ell_f - \ell_{h_v})^2 = \mathbb{E}_{f \mid v} \mathbf{1}\{f(x) \ne h_v(x)\}. \tag{25}$$

The last equality uses binary predictions and labels, not a condition on the label law.

At time t, define for $j \geq 1$ the predictable function

$$q_{t,j}(x) = \sum_{f \in C_J} p_f(L_{t-1}) \mathbf{1}\{f(x) \neq h_{\mathrm{anc}_{j-1}(f)}(x)\}.$$

It lies in $[0,1]$ and has reference mean at most $2e^{-a_j}$ by (15). Use the root variance bound one. Equations (22)–(25) give

$$H_{\mathrm{root}}(L_{t-1} + z\ell_t) \leq e^{1/2} \left(\eta_0 + \sum_j \eta_j q_{t,j}(x_t) \right).$$

Integral Taylor expansion now yields

$$\begin{aligned} p(L_{t-1}) \cdot \ell_t - \big(V(L_t) - V(L_{t-1})\big) &= \int_0^1 (1-z) H_{\mathrm{root}}(L_{t-1} + z\ell_t)\, dz \\ &\leq \eta_0 + \sum_j \eta_j q_{t,j}(x_t), \end{aligned} \tag{26}$$

since $e^{1/2}/2 < 1$. This holds for every cumulative-loss history and labeled example. In particular, the charging functions do not depend on a future optimal comparator.

S7.5 Affine full-history KL accounting

Condition on the complete preceding history, including completed learner randomness and adversary state. The current private selection has law $p(L_{t-1})$ and is independent of the current example under that conditioning. Its expected current loss conditional further on the example is therefore $p(L_{t-1}) \cdot \ell_t$. Summing (26), telescoping V, and using $V(L_T) \leq \min_f L_T(f)$ gives

$$\mathbb{E}\,\mathrm{Reg}_T(C_J) \leq -V(0) + T\eta_0 + \mathbb{E} \sum_{t,j} \eta_j q_{t,j}(X_t). \tag{27}$$

The argument allows the adversary to use previous selections and public preprocessing. It uses neither a minimax interchange nor a restriction to fair labels.

At a coarse scale, $P_t q_{t,j} \leq 1$. By (14) and (16), its entropy and variance contributions are at most $C(\sqrt{dTa_j} + da_j)$. The base scale costs at most $C\sqrt{dT}$, since $d \leq T$. Summing coarse dyadic scales gives

$$C\big(\sqrt{dTA_0} + dA_0\big). \tag{28}$$

At a fine scale split $P_t q_{t,j}$ according to $dP_t/dq \geq e^{a_j/2}$. The lower-density part is at most $2e^{-a_j/2} \leq \gamma_j$. Write $H_{t,j}$ for the remaining part. Equation (9) applies to these predictable functions:

$$\sum_{j \text{ fine}} a_j H_{t,j} \leq 8\,\mathrm{KL}(P_t \| q).$$

The fine variance terms are consequently bounded by

$$T \sum_{j \text{ fine}} \eta_j \gamma_j + \lambda \mathbb{E} \sum_{t,j \text{ fine}} a_j H_{t,j} \leq C\sqrt{dT} + 8\lambda C_*. \tag{29}$$

Here we used $\eta_j \leq \sqrt{da_j/(T\gamma_j)}$ and (17). The fine entropy terms in (20) are at most

$$C \sum_{j \text{ fine}} \left(da_j + \frac{d}{\lambda} + \sqrt{dTa_j\gamma_j} \right) \leq C \left(dU + \frac{dM}{\lambda} + \sqrt{dT} \right). \tag{30}$$

We used $u_j = 2a_j$ and $\sum_j a_j \leq U$.

Choose $\pi(f) \in C_J$ within $\delta_J \leq e^{-U}$ of every $f \in \mathcal{F}$. The two hindsight losses differ by at most

$$\sup_{f \in \mathcal{F}} \sum_t \mathbf{1}\{f(X_t) \neq \pi(f)(X_t)\}.$$

These indicators belong to $\mathcal{F} \triangle \mathcal{F}$. Since $U \geq 256\sqrt{T/d} \geq 4\log(eT/d)$, Lemma S6.3 bounds the displayed expectation by $Cd + 2C_*/U$. Combining this with (27)–(30) proves the exact-probability affine inequality with

$$c_b = 8\lambda + 2/U, \qquad A = C\left(\sqrt{dT} + \sqrt{dTA_0} + dA_0 + dU + \frac{dM}{\lambda} + d\right). \tag{31}$$

The rates use the supplied b, whereas the inequality pays the actual expected cost C_*.

The assumptions $b < T^2/d$ and $d \leq T$ imply

$$dU + b/U + dA_0 + \sqrt{dTA_0} \leq C\sqrt{d\max\{T, b\}}.$$

Also $dM/\lambda = \lambda b = \sqrt{dbM}$. Thus

$$A + c_b b \leq C\left(\sqrt{d\max\{T, b\}} + \sqrt{dbM}\right).$$

If $b \leq T$, then $M \leq C\log(e + T/d)$, giving $A + c_b b \leq C\Phi_{d,T}(b)$. If $T < b < T^2/d$, put $r = T/\sqrt{db} > 1$. Since $U/A_0 \leq 128r$, we have $M \leq C\log(er)$ and obtain the same conclusion from (8). For a separately supplied $b \geq T^2/d$, a fixed proper prediction instead has regret at most T.

S7.6 Finite internal computation

All pairwise disagreement distances can be computed with $O(m^2 s)$ rational operations. Greedy covers and parent links require $O(m^2)$ comparisons per level. There are $J + 1 = O(\log(T + 2))$ levels and at most $m(J + 1) + 1$ nodes. Every cover threshold has denominator bit length at most $6 \cdot 2^J = 3U/2 = O(T)$. Every rate is a minimum of rational numbers or square roots of positive rationals, so the minimum can be selected by squared rational comparisons. By (18),

$$\eta_j \geq \eta_0 \geq \frac{1}{2\sqrt{T}}.$$

Each online round requires $O(mJ)$ node computations and at most m table lookups and integer loss updates. It does not enumerate the $(T + 1)^m$ possible cumulative-loss vectors.

We bound the precision needed for those computations. For real v_i, nonnegativity of KL to the softmax distribution proves

$$-\frac{1}{\eta}\log\sum_i e^{-\eta v_i} = \min_{p\in\Delta}\left(\sum_i p_i v_i - \frac{H(p)}{\eta}\right),$$

where $H(p) = -\sum_i p_i \log p_i$. Recursive substitution expresses each node value as expected descendant leaf loss minus reach-weighted conditional entropies divided by their rates. The entropy chain rule and $1/\eta_j \leq 2\sqrt{T}$ bound the subtraction by $2\sqrt{T}\log m$. Since leaf losses lie in $[0, T]$, every exact node value lies in

$$[-2\sqrt{T}\log m, T]. \tag{32}$$

Put $D = J + 1$ and

$$\zeta = \frac{1}{32(T + 1)^2 D(D + 1)}.$$

Compute a node's soft minimum from its already approximated children with absolute error at most ζ. A soft minimum is 1-Lipschitz in the supremum norm: perturbing each argument by at most h changes its exponential sum by a factor between $e^{-\eta h}$ and $e^{\eta h}$. Induction therefore bounds each node-value error by $D\zeta$.

For a fixed $\eta \leq 1/2$, softmax is 2η-Lipschitz from supremum norm of its arguments to ℓ_1 distance between distributions. Along a direction w, the derivative of probability p_i is $-\eta p_i(w_i - \sum_k p_k w_k)$, whose absolute sum is at most $2\eta\|w\|_\infty$. Using the approximate children thus changes each conditional probability vector by at most $D\zeta$ in ℓ_1.

Approximate this conditional vector by a dyadic vector with an additional ℓ_1 error at most ζ. For at most m entries, first approximate each probability to absolute error $\zeta/(16m)$, clip negative approximations to zero, and normalize. Clipping cannot increase distance from the nonnegative exact

vector. The total preliminary error is at most $\zeta/16$, so the sum is positive and normalization adds at most $\zeta/16$. Choose a power of two $Q \geq 16m/\zeta$. Round down the first $k-1$ normalized probabilities to multiples of $1/Q$ and give the last of the k entries the remainder. The extra ℓ_1 error is at most $2(k-1)/Q \leq \zeta/8$. The total is less than ζ. Sampling such a vector uses exactly $\log_2 Q$ fair bits and finite integer cumulative sums. At most D node choices are needed, with no unbounded rejection loop.

Couple exact and approximate conditional path samplers until their first different branch. The union bound on their conditional mismatch probabilities implies that their leaf distributions have ℓ_1 distance at most the sum of the worst conditional ℓ_1 errors along the D levels. This is at most

$$D(D+1)\zeta \leq \frac{1}{32(T+1)^2}. \tag{33}$$

All current random bits remain private.

The elementary approximations above have polynomial arithmetic and bit complexity. Here are explicit controls on parameter sensitivity. Set $h_m = \lceil \log_2(m+1) \rceil$ and $R_* = T+4Th_m+1$; by (32), all node values and their approximations of error at most one have magnitude at most R_*. For fixed child values, the derivative of their soft minimum with respect to η is $H(p)/\eta^2$. For $\eta \geq 1/(4\sqrt{T})$ its magnitude is at most $16T \log m$. The corresponding softmax vector has rate derivative of ℓ_1 norm at most $2R_*$. Thus approximating a rate to accuracy

$$\frac{\zeta}{128m(T+1)(R_*+1)h_m}$$

keeps its contribution to a soft-minimum error below $\zeta/8$ and to each probability error below $\zeta/(64m)$, and preserves the lower rate bound used in these derivative estimates. Rational bisection on squares achieves this precision in polynomially many bit operations. The exact branch of a rate minimum was already selected by squared rational comparisons, so no transcendental equality must be decided.

For soft minima, subtract the smallest supplied child value before exponentiation. The exponents then have polynomially bounded magnitude, and their sum lies between one and m. Exponential Taylor expansion with range reduction gives certified enclosures in polynomially many operations at any inverse-polynomial requested precision. For logarithms, scale the argument to $[1,2]$ and use

$$\log z = 2\sum_{k\geq 0} \frac{w^{2k+1}}{2k+1}, \qquad w = \frac{z-1}{z+1}.$$

On $[1,2]$, $|w| \leq 1/3$, so truncation needs only a logarithmic number of terms in the reciprocal precision. The same series computes $\log 2$ for the range-reduction correction. The exponential remainder and this geometric tail certify the requested value and probability errors; the displayed magnitude and derivative bounds show that polynomial precision suffices throughout. All rational distances and thresholds have polynomial bit length in the supplied input size and T.

Finally, compare the implemented distribution with the exact gradient at the same actual cumulative-loss vector on each round. Equation (33) bounds the additional expected loss on every current example. The potential inequality holds along these actual vectors regardless of how previous selections were sampled. Define its predictable disagreement functions using the exact gradient evaluated at the actual history; their reference bounds are unchanged. Both distributions are compared at the same actual history. The total additional expected regret is less than one, completing the proof of Theorem S7.1.

S8 A STATISTICAL EMPIRICAL REFERENCE TREE

Let $\mathcal{F}$ be a known nonempty finite or countable class of measurable binary functions with $\mathrm{VC}(\mathcal{F}) \leq d$, where $1 \leq d \leq T$ are integers. The fixed Borel reference μ may be unknown and may have atoms. Assume an announced $B \geq 0$ bounds the expected total full-history input cost, $\mathbb{E}\sum_{t=1}^{T} \mathrm{KL}(P_t \| \mu) \leq B$, with the history including all observed reference samples. The current input and label cannot depend on the learner's fresh private pre-input selection. We construct a measurable proper learner from two independent reference samples. In the nonsaturated branch defined

below, for every supplied resolution between the cutoffs j_0 and j_{stat}, its regret satisfies (55). The sample sizes are N and R in (34) and (35); all draws occur before play. This is a statistical construction, with no runtime or oracle-complexity guarantee for arbitrary class encodings.

The potential in Appendix S7 uses only reference disagreement bounds and finite level cardinalities. We obtain these from the two samples and append every acquired witness as a final leaf. This last level preserves every acquired sample trace.

S8.1 Parameters and measurable acquisition

Fix $1 \le d \le T$ and $B \ge 0$. Supply the learner the positive integer $b = \lceil B \rceil + 1$, so $B+1 \le b \le B+2$. This integer is a supplied parameter of the statistical strategy. If $b \ge T^2/d$, select a fixed member of $\mathcal{F}$ on every round. We justify this branch below. Otherwise put

$$j_0 = \max\{16, \lceil \log_2(T+2) \rceil\}.$$

For a supplied integer resolution $j \ge j_0$, define

$$\rho = 2^{-j}, \qquad u = j \log 2, \qquad N = 256(d+1)^2 j 2^j, \qquad M = (N+1)^d. \tag{34}$$

Thus $\rho \le 1/(T+2)$ and $u \ge \log(T+2)$. Draw independent reference inputs $S_1, \dots, S_N$ and acquire one proper witness for each realizable trace on these sample occurrences. Write $C = \{f_1, \dots, f_m\}$ for the acquired list, with $1 \le m \le M$ by the Sauer bound.

Choose the witnesses by Lemma S10.4. Fix an enumeration of the known countable class, choose its least-index witness for each realized first-sample trace, and order the selected list by those traces. The existence of a trace is a countable union of measurable events, and each value of its least-index selector is measurable. There are only finitely many possible traces, so this gives a measurable finite list, including with repeated inputs. This selection uses no further reference draw. Computing the list for an arbitrary class encoding would also require deciding which traces are realizable.

Independently draw a second reference sample of size

$$R = 8L_R 2^j, \qquad L_R = \min\{l \in \mathbb{N} : 2^l \ge 4M^2/\rho\}. \tag{35}$$

Evaluate every witness on every second-sample input. From these bits compute all empirical disagreement distances $\widehat{D}(f,g)$. This requires R reference draws and mR evaluations. The arithmetic on these acquired bits requires no further class access.

S8.2 Uniform reference geometry

Write $D(f,g) = \mu\{f \ne g\}$. We first show that, except with probability ρ, every original pair agreeing on the first sample has $D(f,g) \le \rho$. The disagreement class has at most $(2N+1)^{2d}$ traces on $2N$ inputs. If a set of mass greater than ρ misses the first sample, choose the first such set in a fixed countable enumeration. Its count in an independent ghost sample of size N is at least $N\rho/2$ with probability at least one half: Chebyshev bounds the failure probability by $4/(N\rho) \le 1/2$. Conditional on the combined $2N$ observations, a random equal split places all the positive positions of a trace having at least $N\rho/2$ positives in the ghost half with probability at most $2^{-N\rho/2} \le e^{-N\rho/4}$. If there are more than N positives this probability is zero. The union bound therefore gives failure probability at most

$$2(2N+1)^{2d} e^{-N\rho/4} \le \rho. \tag{36}$$

For the last inequality, $\log(2N+1) \le 7 + 2d + 2j$. The logarithm of the left side divided by ρ is at most

$$\log 2 + 2d(7 + 2d + 2j) + j \log 2 - 64(d+1)^2 j < 0.$$

For $j \ge 16$, its positive terms are at most $(d^2 + 7d + 2)j$, smaller than the displayed negative coefficient. On this event, let $\pi(f)$ be the first acquired witness sharing f's first-sample trace. Then $\mu\{f \ne \pi(f)\} \le \rho$ simultaneously for every $f \in \mathcal{F}$. This map is used only for analysis.

For a Bernoulli empirical mean $\widehat{p}$ from R independent observations with mean p, exponential Markov at $\log 2$ gives

$$\Pr\{\widehat{p} > 2p + t \text{ or } \widehat{p} < p/2 - t\} \le 2e^{-Rt/2} \qquad (t > 0). \tag{37}$$

Indeed, for the count Z, $\mathbb{E}e^{(\log 2)Z} \le e^{Rp}$ and $\mathbb{E}e^{-(\log 2)Z} \le e^{-Rp/2}$. Substituting the respective thresholds and using $2\log 2 \ge 1$ and $\log 2 \le 1$ proves the two tails. A negative lower threshold gives an empty lower-tail event.

Conditional on the first sample and all its selected witnesses, apply (37) with $t = \rho/2$ to at most m^2 pairwise disagreements. The second sample is independent of those witnesses. Since $R\rho/4 = 2L_R$, its union bound is at most $2M^2e^{-2L_R} \le \rho$. Outside that event,

$$D(f,g) \le 2\widehat{D}(f,g) + \rho, \qquad \widehat{D}(f,g) \le 2D(f,g) + \rho \quad (f,g \in C). \tag{38}$$

Let G be the conjunction of this event and the first-sample event. Then $\Pr(G^c) \le 2\rho$. The second event controls pairwise disagreements within the acquired finite list.

S8.3 The empirical tree and its final level

Put $u_l = 4 \cdot 2^l$, and let H be the largest integer $l \ge 0$ with $2u_l \le j-1$. It exists because $j \ge 16$. For $0 \le l \le H$, set

$$\delta_l = 2^{-2u_l}. \tag{39}$$

These radii satisfy $\delta_l \ge 2\rho$, $\delta_{l+1} = \delta_l^2$, and

$$j/5 \le u_H \le j/2. \tag{40}$$

Maximality gives $u_H > (j-1)/4$, implying the lower bound for $j \ge 16$.

For each l, traverse C and retain a center exactly when its empirical distance from every retained center exceeds δ_l. The resulting C_l is an empirical δ_l-cover of C. For each center at level $l \ge 1$, choose its first center within empirical distance δ_{l-1} in C_{l-1} as parent. Add a dummy root above C_0. Append all witnesses in C as a final leaf level, assigning each its first parent within empirical distance δ_H in C_H. Occurrences at different levels are distinct nodes. Prune internal nodes without final descendants. Every final witness still occurs as one leaf and every leaf path has one node at each retained level.

On G, empirical separation greater than δ_l implies true separation greater than $\delta_l/4$, by (38) and $\rho \le \delta_l/2$. The packing proof of Lemma S2.1 gives

$$\log|C_0| \le Cd, \qquad \log|C_l| \le Cdu_l. \tag{41}$$

A parent edge from level l has true disagreement at most $2\delta_l + \rho \le (5/2)\delta_l$. Thus a leaf f and its level-l ancestor v satisfy

$$\mu\{f \ne h_v\} \le \frac{5}{2}\sum_{k=l}^{H}\delta_k \le 5\delta_l \le 2e^{-u_l}. \tag{42}$$

The sum is bounded by $2\delta_l$ because successive radii square and $\delta_l \le 1/256$. The last inequality follows from $5 \cdot 2^{-2a} \le 2e^{-a}$ for $a \ge 4$.

For nonroot split levels $r = 1, \dots, H+1$, write $a_r = u_{r-1}$. The last split is the split into the full list C. It has the required separate cardinality bound:

$$\log m \le d\log(N+1) \le Cdj \le Cda_{H+1}. \tag{43}$$

For the middle inequality, use $d \le T$, $j \ge \log_2(T+2)$, $\log j \le j$, and (34). For earlier nonroot splits the same bound follows from (41) and $u_r = 2a_r$. Consequently every child level has logarithmic cardinality at most Cda_r. The final full-witness level is needed: stopping at C_H could leave an approximation radius of order $\sqrt{\rho}$.

S8.4 Prediction and actual conditional cost

Set $A_0 = 4\max\{1, b/T\}$. A scale is coarse when $a_r \le A_0$ and fine otherwise. Let M_f be the number of fine scales, which may be zero. Use

$$\begin{aligned} \eta_0 &= \min\{1/2, \sqrt{d/T}\}, \\ \eta_r &= \min\{1/2, \sqrt{da_r/T}\} \qquad \text{at coarse scales.} \end{aligned} \tag{44}$$

If $M_f > 0$, put

$$\lambda = \sqrt{dM_f/b}, \qquad \gamma_r = 2^{1-a_r/2}, \qquad \eta_r = \min\left\{\frac{1}{2}, \lambda a_r, \sqrt{\frac{da_r}{T\gamma_r}}\right\} \tag{45}$$

at fine scales. If $M_f = 0$, set $\lambda = 0$ and evaluate no fine rate. Only in that case interpret dM_f/λ as zero.

Every rate is positive and the split rates are nondecreasing. The coarse and fine sequences separately increase. At their transition $a_f = 2a_c, \gamma_f \le 2$, and

$$\frac{(\lambda a_f)^2}{da_c/T} = \frac{2M_f T a_f}{b} > 8M_f, \qquad \sqrt{\frac{da_f}{T\gamma_f}} \ge \sqrt{\frac{da_c}{T}}.$$

Here $a_f > A_0 \ge 4b/T$. Taking the common half cap preserves this ordering. Since $a_1 = 4 \le A_0$, there is at least one coarse scale and $\eta_1 \ge \eta_0$. This also proves monotonicity when all scales are coarse.

Use the recursive unnormalized soft minimum (19) on this tree, with cumulative losses L_f for its leaves $f \in C$. Its gradient is the product leaf distribution $p(L)$. Before the current input, select an original leaf index using the finite approximation specified below and keep the choice private until the labeled example is generated. After receiving the input, evaluate the selected witness for prediction. Evaluate the remaining witnesses on that input and, after the label, update all leaf losses. Additional evaluations do not revise the current selected index.

We apply the local calculation in (21)–(26), not the finite-reference conclusion of Theorem S7.1. That calculation needs only a finite tree with binary leaf losses, an original function as each nonroot internal center, and nondecreasing split rates at most one half. It does not require a supplied reference law, a finite original input space, or a positive number of fine scales. All its hypotheses hold for the present tree, with last rate η_{H+1}. Its log-path derivative has nonnegative positive coefficients summing to $\eta_{H+1} \le 1/2$, and its variance comparison uses the center on that same leaf path. Therefore, putting

$$q_{t,r}(x) = \sum_{f \in C} p_f(L_{t-1}) \mathbf{1}\{f(x) \ne h_{\mathrm{anc}_{r-1}(f)}(x)\},$$

we have at every actual loss history and labeled example

$$p(L_{t-1}) \cdot \ell_t - [V(L_t) - V(L_{t-1})] \le \eta_0 + \sum_{r=1}^{H+1} \eta_r q_{t,r}(X_t). \tag{46}$$

On G, these predictable functions satisfy $0 \le q_{t,r} \le 1$ and $\mu q_{t,r} \le 2e^{-a_r}$ by (42). The unnormalized comparator and initial-entropy bounds (20) hold with the final child level C.

Fix a good preprocessing outcome s and let

$$C_s = \mathbb{E}\left[\sum_{t=1}^{T} \mathrm{KL}(P_t \| \mu) \,\middle|\, s\right]$$

be the actual conditional expected full-history cost. The conditional cost need not satisfy $C_s \le B$. Current-selection privacy permits the conditional expected loss to be replaced by $p(L_{t-1}) \cdot \ell_t$ and (46) to telescope. The base entropy and variance cost at most $C\sqrt{dT}$. At coarse scales the cost is at most $C\sum_r(\sqrt{dTa_r} + da_r)$, hence at most $C(\sqrt{dTA_0} + dA_0)$ by dyadic summation.

At a fine scale split $P_t q_{t,r}$ at density $e^{a_r/2}$. The low-density part is at most $2e^{-a_r/2} \le \gamma_r$. If $H_{t,r}$ is the high-density part, Lemma S5.1 gives $\sum_r a_r H_{t,r} \le 8\,\mathrm{KL}(P_t \| \mu)$ at each full history. The fine variance cost is at most

$$T\sum_{r \text{ fine}} \eta_r \gamma_r + \lambda \mathbb{E}\left[\sum_{t,r \text{ fine}} a_r H_{t,r} \,\middle|\, s\right] \le C\sqrt{dT} + 8\lambda C_s.$$

We used (17) and $\eta_r \le \sqrt{da_r/(T\gamma_r)}$. By (41) and (43), fine entropy costs at most

$$C \sum_{r \text{ fine}} \left(da_r + \frac{d}{\lambda} + \sqrt{dTa_r\gamma_r} \right) \le C \left(du_H + \frac{dM_f}{\lambda} + \sqrt{dT} \right).$$

All fine terms are absent when $M_f = 0$. Thus the exact-gradient learner has

$$\mathbb{E}[\mathrm{Reg}_T(C) \mid s] \le A + 8\lambda C_s, \tag{47}$$

where the nonnegative deterministic quantity

$$A = C \left(\sqrt{dT} + \sqrt{dTA_0} + dA_0 + du_H + \frac{dM_f}{\lambda} \right) \tag{48}$$

can be chosen uniformly over good preprocessing outcomes. In particular its value and λ do not depend on s.

For the original class, pathwise

$$\mathrm{Reg}_T(\mathcal{F}) \le \mathrm{Reg}_T(C) + \sup_{f \in \mathcal{F}} \sum_{t=1}^{T} \mathbf{1}\{f(X_t) \ne \pi(f)(X_t)\}. \tag{49}$$

Conditional on good preprocessing, the indicators form a countable subset of $\mathcal{F} \triangle \mathcal{F}$, each with reference mass at most ρ. The input-path chain rule and KL convexity under conditional mixtures bound its input-path divergence from μ^T by C_s, as in the proof of Lemma S6.3. Use the unsimplified bound (1) at $u = \log(1/\rho)$. The variational KL inequality gives

$$\begin{aligned} \mathbb{E}\left[\sup_f \sum_t \mathbf{1}\{f(X_t) \ne \pi(f)(X_t)\} \,\middle|\, s \right] &\le \frac{2C_s}{u} + 4d + \frac{2\log 2 + 4d\log(3eT/d) + 10T\sqrt{\rho}}{u} \\ &\le \frac{2C_s}{u} + C\sqrt{dT}. \end{aligned} \tag{50}$$

The last inequality follows from $d \le T$, $u \ge \log(T+2) \ge 1$, and $\rho \le 1/(T+2)$. This use of the MGF does not require $u \ge 4\log(eT/d)$.

The finite implementation below adds less than one expected loss. Average (47) and (50) on G, using $\mathbb{E}[\mathbf{1}_G C_s] \le B$. On G^c, regret is at most T. We obtain

$$\mathbb{E}\,\mathrm{Reg}_T \le A + 8\lambda B + \frac{2B}{u} + C\sqrt{dT} + 2T\rho + 1. \tag{51}$$

The learner need not identify either good event. There is no change of reference measure and no missing-pattern cost.

S8.5 The statistical resolution cap

Let

$$j_{\text{stat}} = \max\left\{ j_0, \min\{k \in \mathbb{N}_{>0} : dk^2 \ge 4096 \max\{T, b\}\} \right\}. \tag{52}$$

We use only resolutions $j \le j_{\text{stat}}$. There is a finite constant H_d, depending only on d, such that

$$j_{\text{stat}} \le H_d \sqrt{\max\{T, b\}/d}. \tag{53}$$

To see this, put $r = T/d \ge 1$ and use $\log_2(T+2) \le \log_2(d+2) + \log_2 r$ and $\log_2 r \le 2\sqrt{r}$. The constant $H_d = 100 + 4\log_2(d+2)$ suffices, including the ceiling and the lower cutoff 16. It follows that $du_H \le C_d \sqrt{d \max\{T, b\}}$.

Since $b < T^2/d$,

$$dA_0 + \sqrt{dTA_0} \le C\sqrt{d\max\{T, b\}}.$$

If $M_f > 0$, then $dM_f/\lambda = \lambda b = \sqrt{dbM_f}$. For $b \le T$, dyadic scale counting and (53) give $M_f \le C_d \log(e + T/d)$. For $T < b < T^2/d$, put $r_b = T/\sqrt{db} > 1$. Then $u_H/A_0 \le C_d r_b$, giving $M_f \le C_d \log(e r_b)$. Both bounds include $M_f = 0$. Substituting in (48) proves

$$A + 8\lambda b \le C_d \Phi_{d,T}(b). \tag{54}$$

We also have $\Phi_{d,T}(B+2) \le C\Phi_{d,T}(B)$, including at $B = 0$. To prove this directly, put $B_* = T^2/d$ and, for $0 \le x \le B_*$, define

$$L(x) = \log \frac{eT}{\sqrt{d \max\{x, T\}}}, \qquad F(x) = \min\{T, \sqrt{dT} + \sqrt{dxL(x)}\}.$$

Extend $F(x) = T$ for $x > B_*$. When $x \le T$, $L(x) = 1 + \frac{1}{2}\log(T/d)$ is within universal constant factors of $\log(e + T/d)$. When $T < x \le B_*$, it equals the logarithm in (8). Since the uncapped expression there is at most $2T$, F is comparable to $\Phi_{d,T}$ in every regime. The function L is nonincreasing and at least one on $[0, B_*]$. The product $xL(x)$ is nondecreasing: below T it is linear with positive slope, and above T its derivative is $L(x) - 1/2 \ge 1/2$. Consequently F is nondecreasing and, for every $c \ge 1$,

$$F(cx) \le \sqrt{c}\, F(x).$$

If $cx \le B_*$ this follows by decreasing L and scaling both nonnegative terms inside the cap. If $x \le B_* < cx$, then $\sqrt{dx} \ge T/\sqrt{c}$ and $L(x) \ge 1$, so $F(x) \ge T/\sqrt{c}$. If $x > B_*$ both values are T; if $x = 0$ the inequality is immediate. For $x, y \ge 0$ with $M = \max\{x, y\} > 0$, it follows that

$$F(x + y) \le F(2M) \le \sqrt{2}F(M) \le \sqrt{2}\big(F(x) + F(y)\big).$$

The same conclusion holds when $x = y = 0$. Thus $\Phi(B + 2) \le C[\Phi(B) + \Phi(2)]$. For $T \ge 2$, $\log(e + T/d) \le T$ gives $\Phi(2) \le C\sqrt{dT}$. For $T = 1$ we have $d = 1$ and $\Phi(2) = 1$. This baseline is already at most $\Phi(B)$, proving the comparison. Since $b \le B + 2$, monotonicity of the comparable F also gives $\Phi(b) \le C\Phi(B)$. Equations (51) and (54), with $2T\rho < 2$, yield

$$\mathbb{E}\,\mathrm{Reg}_T \le C_d\Phi_{d,T}(B) + \frac{2B}{j \log 2}. \tag{55}$$

This also validates the earlier saturation branch: if $b \ge T^2/d$, then $T \le C\Phi(b) \le C'\Phi(B)$ and a fixed proper witness suffices.

S8.6 Finite probability approximation and measurability

The finite implementation in Appendix S7 applies with depth $D = H + 2$. Its hypotheses are monotone split rates in $[1/(2\sqrt{T}), 1/2]$, a finite tree, and integer leaf losses in $[0, T]$. Neither a supplied numerical reference law nor fine scales are required. Every rate here is a minimum of rational numbers or square roots of positive rationals, with its branch determined by squared rational comparisons. Since $j \le j_{\text{stat}} = O(T + \log(T + 2))$ in the unsaturated branch, all radii and rates have polynomial-length descriptions.

More explicitly, take node tolerance $\zeta = [32(T + 1)^2 D(D + 1)]^{-1}$. Soft-minimum value errors are at most $D\zeta$, and softmax Lipschitzness bounds the resulting conditional probability error by $D\zeta$ in ℓ_1. The normalized dyadic construction preceding (33), with power-of-two denominator at least $16m/\zeta$, adds at most ζ per conditional vector. Sampling uses a fixed finite number of fair bits at each of at most D levels. The final distribution error is at most $D(D + 1)\zeta \le [32(T + 1)^2]^{-1}$. Comparison with the exact gradient is made at the implemented learner's actual cumulative losses, so the pointwise potential and predictable charging functions remain valid there. The total additional expected loss is less than one. Both distributions are compared at the same actual history.

The entropy-chain argument in (32) bounds every subtree value between $-2\sqrt{T}\log m$ and T. The rate sensitivity and exponential/logarithm approximation bounds proved there therefore provide all the required value and probability tolerances with polynomial arithmetic and bit complexity. No exact transcendental probability sampling or equality test is used.

The witness indices are measurable functions of the first sample, and their evaluations on the second sample and online inputs are measurable. All empirical distances, centers, ties and tree operations involve finite lists and fixed index order. The finite probability approximation therefore gives a measurable proper random prediction on every history. Preprocessing may be public before play. The current random bits, index, and any current evaluation transcript remain private until the labeled example is generated, including on bad preprocessing histories and in the saturation branch. The algorithm requires neither conditional laws nor their costs.

S9 Cover with an unknown reference

Let X be standard Borel and let $\mathcal{F} = (f_i)_{i\geq 1}$ be a nonempty finite or countable class of measurable binary functions of VC dimension at most $d \geq 1$. A finite enumeration may be repeated. The fixed Borel reference μ is unavailable: the learner receives no reference samples, densities, set probabilities or conditional reference oracle. The full round-end history $\mathcal{H}_t$ is generated by an initial standard-Borel state and finitely many standard-Borel round variables. It includes preceding inputs, labels, revealed learner choices and any additional preceding adversary information. Put

$$P_t = \mathrm{Law}(X_t \mid \mathcal{H}_{t-1}), \qquad B = \mathbb{E}\sum_{t=1}^{T} \mathrm{KL}(P_t\|\mu) < \infty.$$

The current label may depend on the full preceding history, the input and the announced distribution of the learner, but not on its fresh private hypothesis. Conditional on the preceding history, this hypothesis sample is independent of the current input-label generation. Later rounds may depend on earlier realized choices. Regret is to the best member of $\mathcal{F}$ on the complete realized path.

We use Cover, Algorithm 3 in Section 4.1 of Blanchard (2025). The exploration theorem imported below is Proposition 17 in its Appendix D, with the proof in Section 5.5. Replacing large-density observations by a common dummy is also the device in Lemma 12 and Appendix C of Blanchard et al. (2026). We supply the truncation, conditioning and hindsight reductions needed here. The learner is a measurable statistical strategy. A known enumeration does not give a computable test for the nonexistence of an empirical trace, and no efficient generic cover oracle is assumed or obtained.

Theorem S9.1. *Under the preceding model, there is a proper pre-input rule using neither d, μ nor B such that, for $T \geq 2$,*

$$\mathbb{E}\,\mathrm{Reg}_T \leq \min\left\{T,\ C_d T^{3/4}\log^3(eT) + \frac{8(B+T/e)}{\log T}\right\}. \tag{56}$$

Here C_d depends only on d. Consequently, if $B \leq T$, then

$$\mathbb{E}\,\mathrm{Reg}_T \leq C'_d \frac{T}{\log(eT)}. \tag{57}$$

The latter conclusion also holds at $T = 1$ after enlarging C'_d. In particular it holds under the per-round full-history condition $\mathrm{KL}(P_t\|\mu) \leq 1$ almost surely.

If the VC dimension is zero, all members of the nonempty class are pointwise equal, so selecting any member has regret zero.

S9.1 The imported exploration estimate

Lemma S9.2. *Let Z_t be $\mathcal{H}_t$-measurable, and suppose its conditional law given $\mathcal{H}_{t-1}$ has density at most $1/\sigma$ relative to a fixed probability ρ, where $0 < \sigma \leq 1$. Let deterministic epochs $E_k = (s_{k-1}, s_k]$, $1 \leq k \leq K$, partition $\{1,\dots,T\}$. For $t \in E_k$, define*

$$\gamma_t = \sup_{\substack{f,g\in\mathcal{F}:\\ f(Z_u)=g(Z_u)\ (u\leq s_{k-1})}} \Pr\{f(Z_t) \neq g(Z_t) \mid \mathcal{H}_{t-1}\}.$$

Countability makes this supremum measurable. For every $q \in (0,1]$ and $\delta \in (0,1/2]$, there is a universal constant A such that, with probability at least $1-\delta$,

$$\#\left\{k : \sum_{t\in E_k} \gamma_t \mathbf{1}\{\gamma_t \geq q\} \geq W\right\} \leq \frac{A\log^2(eT)}{q\sigma},$$

$$W = 10(d+1)\log\frac{eT}{\sigma\delta}. \tag{58}$$

Auxiliary information in the full preceding history is permitted.

Proof. Proposition 17 of Blanchard (2025) gives this estimate with threshold

$$d \log\left(\frac{T}{\sigma}\log\frac{1}{\delta}\right) + \log\frac{T}{\delta} + 2$$

and the smaller numerator $\log^2 T$. The displayed W is larger, by $\log\log(1/\delta) \le \log(1/\delta)$ and $\log(eT/(\sigma\delta)) \ge 1$. The universal constant A and the tangent-process and Wills-functional estimates enter through that proposition.

We justify the full-filtration scope by reducing it to the source's revealed-loss model. Condition on the initial state if present. Write $\mathcal{H}_t = \sigma(U_1, \dots, U_t)$, where U_t contains the complete round increment and lies in a standard-Borel space. Let c_t be a Borel embedding of that space into $[0, 1]$, with measurable inverse on its image. In the prediction game of the proposition, use input Z_t and a forecaster that always predicts zero. After generating Z_t, draw U_t from its regular conditional law given the preceding increments and Z_t, and reveal the constant loss function

$$\ell_t(a) = c_t(U_t) \qquad (0 \le a \le 1).$$

These $[0, 1]$-valued, 0-Lipschitz losses are allowed by that model. The revealed loss recovers U_t, so the observable history is exactly $\mathcal{H}_t$. Its input kernels and hence its γ_t are exactly the kernels and quantities in the statement. The imported conclusion depends only on these inputs and kernels, not on the constant losses. Its probability bound is uniform over the initial state, so averaging removes the initial conditioning. This is an analysis reduction and supplies no extra observation to the actual learner. □

Lemma S9.3. *In Lemma S9.2, put* $L = \max_k |E_k|$ *and* $\Gamma = \sum_t \gamma_t$. *Then*

$$\mathbb{E}\Gamma \le 2KW + \frac{2AL}{\sigma}\log^2(eT)\log(eL) + \delta KL. \tag{59}$$

Proof. Write $\Gamma_k = \sum_{t \in E_k} \gamma_t \le L$. For $2W \le u \le L$, take $q = u/(2L)$. If $\Gamma_k > u$, then

$$\sum_{t \in E_k} \gamma_t \mathbf{1}\{\gamma_t \ge q\} \ge \Gamma_k - q|E_k| > u/2 \ge W.$$

Equation (58) therefore implies

$$\mathbb{E}\#\{k : \Gamma_k > u\} \le \frac{2AL\log^2(eT)}{\sigma u} + \delta K.$$

Integrate in u, using the trivial bound K below $2W$. If $2W < L$, the resulting logarithm is at most $\log(eL)$, which proves the assertion. If $2W \ge L$, the bound $\Gamma \le KL \le 2KW$ suffices. For each deterministic u we used a separate expectation bound; Tonelli's theorem, not a simultaneous uncountable union, justifies its integration. □

S9.2 The proper learner and its loss bound

For $T \ge 2$, set

$$K = \lceil\sqrt{T}\rceil, \quad s_k = \lfloor kT/K \rfloor, \quad j = \lceil T^{1/4}\rceil, \quad \beta = 1 - \frac{1}{j}, \quad \eta = -\log\beta.$$

Thus $j \ge 2$ and $1/j \le \eta \le 2/j$. At each epoch boundary select the least-indexed representative of every $\mathcal{F}$-trace on all inputs in completed epochs. Call the resulting finite class S_k. Each trace-existence event is a countable union of measurable events; the event that an index is the first representative is measurable. Selecting the finitely many present traces thus defines a measurable finite subset of the index space. Recognizing that every representative has been found in an arbitrary enumeration is a separate computational question.

Give S_k equal initial weights, multiplying an expert's weight by β after each mistake. Before observing the current input, privately sample an expert from the current weights. For $n \le T$, Sauer's bound (Sauer, 1972) gives

$$|S_k| \le \sum_{i=0}^{d}\binom{n}{i} \le (T+1)^d.$$

Here binomial coefficients with $i > n$ are zero; the last inequality also follows by encoding each subset of size at most d by its increasing entries padded by zeros to length d. Let $\bar{\ell}_t$ be the weighted mean binary loss.

The finite-expert calculation gives, for every $g \in S_k$,

$$\sum_{t \in E_k} \bar{\ell}_t - L_{E_k}(g) \le \frac{\log |S_k|}{\eta} + \frac{\eta |E_k|}{8}.$$

Indeed, the log moment generating function of a $[0,1]$-valued variable has second derivative equal to a tilted variance, at most $1/4$. Integrating twice proves $\log \mathbb{E} e^{-\eta Z} \le -\eta \mathbb{E} Z + \eta^2/8$. Apply this inequality to the normalized current weights, telescope the total weight, and lower bound its final value by the weight of g. Summing epochs defines the common bound

$$H_T = \frac{Kd \log(T+1)}{\eta} + \frac{\eta T}{8}. \tag{60}$$

For each deterministic $f \in \mathcal{F}$, let $g_{k,f}$ be its selected epoch representative. It is measurable before the epoch and agrees with f on all original preceding inputs. Since a binary loss difference is at most a prediction disagreement, pathwise and simultaneously for every f,

$$\sum_t \bar{\ell}_t - L_T(f) \le H_T + \sum_k \sum_{t \in E_k} \mathbf{1}\{g_{k,f}(X_t) \ne f(X_t)\}. \tag{61}$$

S9.3 Truncation with the original history retained

For analysis only, set $R = T^{1/4}$. Choose jointly measurable conditional density versions $r_t = dP_t/d\mu$, using the standard-Borel kernel hypotheses, and put

$$C_t = \mathbf{1}\{r_t(X_t) > R\}, \quad D = \sum_t C_t, \quad Z_t = \begin{cases} X_t, & C_t = 0, \\ *, & C_t = 1. \end{cases}$$

Extend every f by $f(*) = 0$ and put $\rho = (\mu + \delta_*)/2$. The conditional density of Z_t relative to ρ is at most $2R$ on X and at most 2 at $*$. Consequently Lemmas S9.2 and S9.3 apply with $\sigma = 1/(2R)$ and the full original history. Adding a common-valued point does not increase VC dimension.

Since $u \log u \ge -1/e$,

$$\log R \, \Pr(C_t = 1 \mid \mathcal{H}_{t-1}) \le \int_{\{r_t > R\}} r_t \log r_t \, d\mu \le \mathrm{KL}(P_t \| \mu) + 1/e.$$

Thus

$$\mathbb{E} D \le \frac{B + T/e}{\log R}. \tag{62}$$

Agreement on the original preceding inputs implies agreement on the transformed preceding inputs. For fixed f put

$$A_{t,f} = \mathbf{1}\{g_{k,f}(X_t) \ne f(X_t)\}(1 - C_t) \qquad (t \in E_k).$$

Then $\mathbb{E}[A_{t,f} \mid \mathcal{H}_{t-1}] \le \gamma_t$. In particular

$$\mathbb{E}\left[\sum_t \bar{\ell}_t - L_T(f)\right] \le H_T + \mathbb{E}\Gamma + \mathbb{E} D.$$

For the hindsight comparison we need an exponential version. As $A_{t,f}$ is binary,

$$\mathbb{E}[e^{A_{t,f}} \mid \mathcal{H}_{t-1}] = 1 + (e-1)\mathbb{E}[A_{t,f} \mid \mathcal{H}_{t-1}] \le e^{(e-1)\gamma_t}.$$

Iteration yields

$$\mathbb{E} \exp\left(\sum_t A_{t,f} - (e-1)\Gamma\right) \le 1. \tag{63}$$

No step conditions on a truncation-success event. The learner's representatives and weights still use the original data.

S9.4 A FIXED ANALYSIS COVER AND THE HINDSIGHT MAXIMUM

Lemma S9.4. *For any probability ρ and $0 < a \le 1$, a countable binary VC-at-most-d class has a finite proper a-cover H satisfying*

$$\log |H| \le 2d \log \frac{10(d+1)}{a}.$$

Proof. Take $V \ge 2$ functions whose pairwise disagreement probabilities exceed a, put $v = \log V$, and draw $n = \lceil 3v/a \rceil$ iid ρ observations. The chance that any pair has the same trace is at most $V(V-1)e^{-an}/2 < 1$. Some sample therefore distinguishes every function, so

$$V \le (n+1)^d, \qquad v \le d \log \frac{3v+2}{a}.$$

Put $M_0 = 10(d+1)/a$ and $v_0 = 2d \log M_0$. The inequalities $\log M_0 \le M_0$ and $6dM_0 + 2 \le 10(d+1)M_0$ give $(3v_0+2)/a \le M_0^2$. Hence $v_0 \ge d \log((3v_0+2)/a)$. The function $v - d \log((3v+2)/a)$ is strictly increasing for $v \ge v_0 > d$, proving $v \le v_0$. The assertion is also valid for $V = 1$.

Traverse the class enumeration, retaining a function exactly when its distance from every retained function exceeds a. The bound on every finite retained family implies that only finitely many functions are retained. Every skipped function is within distance a of an earlier retained member. The final set is the required cover. This is a mathematical construction, not a computable stopping rule for an arbitrary enumeration. □

Apply the lemma to the extended class with $a = T^{-2}$. Fix this cover H and a first cover representative $\pi(f)$ for every f, independently of the observations. The learner does not know or construct these analysis objects. Summing (63) over $h \in H$ and using Jensen's inequality gives

$$\mathbb{E} \max_{h \in H} \left\{ \sum_t A_{t,h} - (e-1)\Gamma \right\} \le \log |H|.$$

Equation (61) now implies

$$\mathbb{E} \left[\sum_t \bar{\ell}_t - \min_{h \in H} L_T(h) \right] \le H_T + (e-1)\mathbb{E}\Gamma + \log |H| + \mathbb{E}D. \tag{64}$$

The finite-cover maximum controls the comparator uniformly over the whole labeled path.

Let $\mathcal{A} = \{\mathbf{1}\{f \ne \pi(f)\} : f \in \mathcal{F}\}$. Every member has ρ-mean at most a, and the class has at most $(M+1)^{2d}$ traces on M points. A disagreement trace is determined by two original traces; no VC-dimension bound of $2d$ is asserted. For iid $V_1, \dots, V_M$ with law ρ, symmetrization gives

$$\mathbb{E} \sup_{g \in \mathcal{A}} \sum_{i=1}^{M} g(V_i) \le Ma + 4\sqrt{dM \log(M+1)}. \tag{65}$$

To verify the bound, use an independent copy to center the sum. Jensen's inequality and coordinate exchanges in the iid pairs bound the centered supremum by twice the expected supremum of $\sum_i \varepsilon_i g(V_i)$. Conditional on V, the finitely many trace vectors have squared norm at most M, so their signed sums have moment generating functions at most $e^{\lambda^2 M/2}$. Bounding the exponential of the maximum by the sum and optimizing λ gives $2\sqrt{dM \log(M+1)}$ before that factor two. Countable suprema follow by increasing finite subclasses.

For analysis only, set

$$k = \left\lceil \frac{2 \log T}{\sigma} \right\rceil, \qquad M = Tk.$$

At round t, draw k fresh iid ρ candidates $V_{t,i}$ and k fresh independent uniforms, independently of the preceding history. Accept each candidate with probability σ times the conditional density of Z_t relative to ρ. Use the first accepted candidate; if all are rejected, draw Z_t directly from its required conditional law with an extra independent seed. Conditional on success the accepted value has that law, and the fallback has the same law. Thus the entire conditional law of Z_t is preserved.

Afterwards reconstruct the remaining original round variables using their regular conditional kernel given the original preceding history and Z_t. This preserves the original full joint process, including its learner and label history, and permits fresh candidate pools at the next round. All M candidates are jointly iid ρ, although their acceptance densities are adaptive. Each trial accepts with conditional probability σ. The probability of any fallback is at most $T(1-\sigma)^k \le Te^{-\sigma k} \le 1/T$. Off that event,

$$\sup_{g\in\mathcal{A}} \sum_t g(Z_t) \le \sup_{g\in\mathcal{A}} \sum_{t,i} g(V_{t,i}).$$

On the event, the left side is at most T. Hence

$$\mathbb{E} \sup_{g\in\mathcal{A}} \sum_t g(Z_t) \le Ma + 4\sqrt{dM\log(M+1)} + 1. \tag{66}$$

For every original f, the discarded observations cost at most D, and all other disagreements equal the transformed disagreements. It follows pathwise that

$$\min_{h\in H} L_T(h) - \inf_{f\in\mathcal{F}} L_T(f) \le D + \sup_{g\in\mathcal{A}} \sum_t g(Z_t).$$

Combining this with (64) and (66) gives

$$\mathbb{E}\left[\sum_t \bar{\ell}_t - \inf_{f\in\mathcal{F}} L_T(f)\right] \le H_T + (e-1)\mathbb{E}\Gamma + \log|H| + 2\mathbb{E}D + Ma + 4\sqrt{dM\log(M+1)} + 1. \tag{67}$$

The discarded count occurs exactly twice: in comparison to a fixed analysis-cover expert, and in comparison of that cover to the full hindsight class. It has no epoch or depth multiplier. Privacy makes the learner's expected sampled loss equal to its expected mixed loss at each actual history. This equality is used only for the learner's loss; the same realized comparator infimum is retained. Thus (67) bounds expected regret for the proper learner.

S9.5 Parameters and scope

Proof of Theorem S9.1. In Lemma S9.3, take $\delta = T^{-2}$. Our choices satisfy

$$K \le 2\sqrt{T}, \quad L \le 2\sqrt{T}, \quad \sigma^{-1} = 2T^{1/4},$$
$$W = 10(d+1)\log(2eT^{13/4}) \le 50(d+1)\log(eT).$$

The last inequality follows from $\log(2e) + (13/4)\log T \le 5(1+\log T)$. Consequently

$$\mathbb{E}\Gamma \le C_d\left(\sqrt{T}\log(eT) + T^{3/4}\log^3(eT) + 1\right) \le C_d T^{3/4}\log^3(eT).$$

Since $j \le 2T^{1/4}$, (60) gives

$$H_T \le 4dT^{3/4}\log(T+1) + T^{3/4}/4.$$

The analysis cover has $\log|H| \le 2d\log(10(d+1)T^2)$. Also

$$M \le T + 4T^{5/4}\log T \le 5T^{5/4}\log(eT), \qquad \log(M+1) \le 5\log(eT).$$

Thus $Ma + 4\sqrt{dM\log(M+1)}$ is at most $C_d[1 + T^{5/8}\log(eT)]$, absorbed by the previous bound. Equation (62) gives $2\mathbb{E}D \le 8(B + T/e)/\log T$. Substitution in (67) proves (56); the pathwise bound T supplies its minimum.

If $B \le T$, use $\log(eT)/\log T \le 1 + 1/\log 2$ for $T \ge 2$ and

$$\sup_{T\ge 1} T^{-1/4}\log^4(eT) = 16^4 e^{-15/4} < \infty.$$

For the latter identity, put $u = \log T$; the logarithmic derivative of $e^{-u/4}(1+u)^4$ changes sign at $u = 15$. Every term is therefore bounded by $C_d' T/\log(eT)$. The trivial bound handles $T = 1$. □

This statistical upper follows from the cited Cover exploration theorem and the full-history truncation argument. Implementation cost, matching lowers for individual finite-VC classes, and sharp interpolation for arbitrary B require separate analyses. The explicit threshold construction in Theorem S11.2 retains its separate constructive scope and constants.

S10 Reference information for countable VC classes

Let X and all round-state spaces be standard Borel, with the full-history convention of Section S9. Let $\mathcal{F} = (f_i)_{i\geq 1}$ be a known nonempty finite or countable class of measurable binary functions, with VC dimension at most d, where $1 \leq d \leq T$ and $T \geq 512$. A fixed Borel probability μ is unknown. Every input law conditional on the full preceding history satisfies $\mathrm{KL}(P_t\|\mu) \leq 1$ almost surely. The history includes preceding learner randomness, observed reference draws and preceding adversary information, but not fresh sampling coins. Current labels may depend on this history, the input and the announced learner distribution, but not on its fresh private hypothesis. Expected regret retains the best hindsight member of $\mathcal{F}$ inside the expectation.

A true-reference request returns an unlabeled observation with law μ, independently of all preceding randomness. An integer allowance m bounds the total requests on every history. The upper below requests a deterministic number before play. The lower allows adaptive requests interspersed with play, including after a current input and before prediction, and unrestricted post-input predictions. Every procedure must produce its predictions. An infinite unfinished request sequence is not a strategy.

The learner may also receive a known mathematical Borel probability ν and $0 < \delta \leq 1$ with $\mathrm{TV}(\mu,\nu) \leq \delta$. Knowledge of ν has the same statistical meaning as knowledge of the reference in Theorem S6.1: finite ν-covers and probabilities of specified measurable sets are available as mathematical objects. This access is statistical. Computing these objects from arbitrary encodings of ν or $\mathcal{F}$, including recognizing all nonempty evaluation cells or all realizable sample traces, requires additional computational assumptions. Only draws from the true μ count toward m; no density, CDF, quantile or conditional-sampling oracle for μ is supplied.

Write

$$\begin{aligned} A_{d,T} &= \sqrt{dT\log(eT/d)}, \qquad & s_m &= \log(eT(m+1)), \\ s_\delta &= \log(eT/\delta), & s_{m,\delta} &= \log(eT(m+1)/\delta). \end{aligned} \tag{68}$$

Theorem S10.1. *For each fixed positive integer d, there is a constant C_d such that, for every integer $T \geq \max\{512, d\}$, every integer $m \geq 0$, every $0 < \delta \leq 1$, and every admissible $\mathcal{F}, \nu$, there is a measurable proper pre-input strategy depending only on $T, d, m, \delta, \mathcal{F}, \nu$ and the observed history that satisfies the following bound for every admissible μ and input-label protocol:*

$$\mathbb{E}\,\mathrm{Reg}_T \leq C_d\left(A_{d,T} + \frac{T}{s_{m,\delta}}\right). \tag{69}$$

The strategy uses at most m true-reference draws, all before play. The assertion includes every Borel μ, ν, including atoms and a component of μ singular relative to ν.

There is $c_d > 0$ such that the corresponding worst case over classes of VC dimension at most d and supplied references ν is at least $c_d(A_{d,T} + T/s_{m,\delta})$. The lower already holds for one fixed countable rational-threshold class and ν uniform on $[0,1]$, with an atomless full-support true reference. It permits the stronger prediction and request access specified above.

The constants may depend strongly on d. The lower takes the worst case over classes and supplied references. Sharp dependence on growing dimension and efficient implementation for arbitrary classes remain separate questions. At $\delta = 1$ the strategy can ignore ν. At $\delta = 0$, use the known-reference theorem separately; no logarithm of $1/\delta$ is evaluated. VC dimension zero again has zero regret.

The proof combines the empirical disagreement tree in Appendix S8, the finite statistical upper of Theorem S6.1, and Cover (Blanchard, 2025; Blanchard et al., 2026). We count reference draws separately from the computational work of the construction. The exponential KL-to-accuracy sampling cost in approximate rejection sampling already appears in the corrected version of Block and Polyanskiy (2024). Here the counted resource is true-reference observations used by an online predictor, whose output need not be one of those observations.

S10.1 A finite known reference constructed from an approximation

We first prove the accuracy upper without true-reference samples. The cell count uses the classical dual-VC argument of Assouad (1983, Proposition 2.12); see also Theorem 3 of Chase et al. (2024).

Lemma S10.2. *Put $D = 2^{d+1}$. If $H \subseteq \mathcal{F}$ is finite with $|H| = N$, its nonempty evaluation cells have number*

$$K \le \sum_{i=0}^{D-1} \binom{N}{i} \le (N+1)^D. \tag{70}$$

Proof. The dual evaluation class on H has VC dimension less than D. Otherwise it would shatter D functions indexed by $f_a, a \in \{0,1\}^{d+1}$. For each $j \in \{1, \dots, d+1\}$ choose an input x_j at which $f_a(x_j) = a_j$ for every a; dual shattering supplies the required evaluation pattern. These points are distinct because their coordinate patterns differ. The f_a then realize all labels on the $d+1$ points, contradicting the primal VC bound. Sauer's inequality on this dual class gives the first bound in (70). Encoding subsets of size at most D by their increasing elements padded with zeros to length D gives the second. □

Lemma S10.3. *Under the model of Theorem S10.1, without any true-reference draws, a measurable proper pre-input strategy has*

$$\mathbb{E}\,\mathrm{Reg}_T \le C_d \left(A_{d,T} + \frac{T}{s_\delta} \right). \tag{71}$$

Proof. Choose a finite proper ϵ-cover $H \subseteq \mathcal{F}$ in ν-disagreement distance by Lemma S2.1. Thus $\log |H| \le 2d \log(4e/\epsilon)$. Its nonempty evaluation cells $C_1, \dots, C_K$ are measurable; retain cells of zero ν-mass. Put

$$v_i = \nu(C_i), \qquad p_i = \mu(C_i), \qquad \rho_i = \frac{v_i + \delta}{1 + K\delta}. \tag{72}$$

The vector ρ is a known strictly positive probability. Total variation gives $p_i \le v_i + \delta$ separately on each cell. For an actual conditional input law put $q_i = P_t(C_i)$. Data processing and finite conditional KL imply $q_i = 0$ whenever $p_i = 0$, and on the remaining cells

$$\begin{aligned} \mathrm{KL}(q\|\rho) &= \mathrm{KL}(q\|p) + \sum_{i:q_i>0} q_i \log \frac{p_i}{\rho_i} \\ &\le 1 + \log(1 + K\delta). \end{aligned} \tag{73}$$

The learner never evaluates the unknown p_i, q_i or true densities. In particular an atom in a ν-null cell is not discarded, and no assertion that $\mu \ll \nu$ is made.

Each $h \in H$ descends to a binary function on the finite cell space. This class has VC dimension at most d: one original point from every shattered cell would witness the same shattering in $\mathcal{F}$. Run the finite known-reference statistical learner with reference ρ. Its sampled cell function is the corresponding original $h \in H$, selected before the input. Losses agree at every original point. If $K\delta \le 1$, (73) gives total full-history KL at most $(1 + \log 2)T$. Removing surplus preceding information only mixes conditional laws and cannot increase their KL by convexity. Current-label privacy is unchanged. Hence Theorem S6.1 implies

$$\mathbb{E}\left[\text{learner loss} - \min_{h \in H} L_T(h) \right] \le C A_{d,T}. \tag{74}$$

To check this statistical-rate substitution, at a budget $B' = (1+\log 2)T$ the first or middle expression in (8) is at most $C A_{d,T}$. If $B' > T^2/d$, then $d > T/(1+\log 2)$, so $T \le C A_{d,T}$ also bounds its saturated expression.

For each f, select its first cover representative $\pi(f)$. Then

$$\mu\{f \ne \pi(f)\} \le \nu\{f \ne \pi(f)\} + \delta \le \epsilon + \delta.$$

This is a TV bound on each single disagreement set, not a union bound over all functions. The countable family $\mathcal{A} = \{\mathbf{1}\{f \ne \pi(f)\} : f \in \mathcal{F}\}$ lies in $\mathcal{F} \triangle \mathcal{F}$. Pathwise,

$$\min_{h \in H} L_T(h) - \inf_{f \in \mathcal{F}} L_T(f) \le \sup_{g \in \mathcal{A}} \sum_t g(X_t). \tag{75}$$

Indeed $L_T(\pi(f)) - L_T(f)$ is bounded by its disagreement count for each f, while $\min_H L_T \le L_T(\pi(f))$. The inequality holds simultaneously for every hindsight comparator.

If $\epsilon + \delta \le e^{-v}$ and $v \ge 4\log(eT/d)$, Lemma S6.3, at actual expected budget at most T relative to μ, gives

$$\mathbb{E}\,\mathrm{Reg}_T \le CA_{d,T} + Cd + \frac{2T}{v}. \tag{76}$$

Its path-law KL transfer is relative to μ^T. The artificial ρ is used only for the finite learner, not for the terminal error on the original space.

Set $c_d = 8dD$ and $L_\delta = \log(1/\delta)$. First suppose

$$L_\delta \ge 16c_d \log(eT/d). \tag{77}$$

Take $u = L_\delta/(2c_d)$ and $\epsilon = e^{-u}$. The cover cardinality and Lemma S10.2 imply

$$\log K \le D[\log 2 + 2d(\log(4e) + u)] \le c_d(u+1) = L_\delta/2 + c_d \le L_\delta.$$

For the middle inequality use $\log 2 < 1$, $\log(4e) < 3$ and $d \ge 1$; for the last use $L_\delta \ge 16c_d$. Thus $K\delta \le 1$. Also $L_\delta \ge u$, so $\epsilon + \delta \le 2e^{-u}$. With $v = u - \log 2$, condition (77) gives $u \ge 8\log(eT/d) \ge 8$, and therefore $v \ge u/2 \ge 4\log(eT/d)$. Equation (76) and $d \le A_{d,T}$ yield

$$\mathbb{E}\,\mathrm{Reg}_T \le CA_{d,T} + \frac{8c_dT}{L_\delta}.$$

Since $\log(eT/d) \ge 1$,

$$\log(eT) = \log(eT/d) + \log d \le (1 + \log d)\log(eT/d) \le (1 + \log d)L_\delta.$$

Hence $s_\delta \le (2 + \log d)L_\delta$, proving (71) in this range.

If (77) fails, ignore ν and use Theorem S9.1. Here $s_\delta \le (1 + 16c_d)\log(eT)$, so its bound $C_dT/\log(eT)$ proves (71). This branch includes $\delta = 1$ without division by $\log(1/\delta)$. The choice between branches uses only the supplied T, d, δ and known mathematical objects. □

S10.2 Statistical trace selection and the number of samples

Lemma S10.4. *For every finite sample $z = (z_1, \dots, z_N)$, one can select measurably a proper witness for every realizable $\mathcal{F}$-trace on that sample. The selected class has size at most $(N+1)^d$, including with repeated observations. This selection uses no further reference draw.*

Proof. For every $a \in \{0,1\}^N$, the event $\{z : \exists i,\ (f_i(z_1), \dots, f_i(z_N)) = a\}$ is a countable union of measurable events. On it select the least such index; each value of that selection is again a measurable event. On its complement choose a fixed default index but do not include that trace in the list. There are finitely many candidate traces, and Sauer's bound gives at most $(N+1)^d$ present traces. Ordering these witnesses by their traces gives a measurable finite list. Repeated inputs do not create additional traces. This defines a measurable selection from the whole enumeration. Deciding that a trace is absent from an arbitrary infinite enumeration requires additional computational assumptions. □

The finite witness property in Lemma S10.4 is exactly the acquisition property used in Appendix S8. Evaluating its selected functions on a second independent reference sample and on online inputs is measurable. All subsequent finite distance, tree and tie choices can be resolved by fixed index order. A proper random prediction is generated from fresh uniform variables, or by the finite approximation already given in that appendix. The empirical construction here uses this measurable trace selection; it makes no ERM-access or arbitrary-class runtime assumption.

Lemma S10.5. *Without a supplied approximation and with at most m iid true-reference draws, where $m \ge 0$ is an integer, a measurable proper pre-input strategy satisfies*

$$\mathbb{E}\,\mathrm{Reg}_T \le C_d\left(A_{d,T} + \frac{T}{s_m}\right). \tag{78}$$

The number of draws is deterministic and all are taken before play.

Proof. Use the empirical construction with $B = T$ and $b = T + 1$. If $b \geq T^2/d$, its saturation argument gives $T \leq C\Phi_{d,T}(T) \leq C' A_{d,T}$. A fixed proper member, without samples, therefore suffices. Otherwise set

$$j_0 = \max\{16, \lceil \log_2(T+2) \rceil\}, \qquad j_{\text{stat}} = \max\left\{ j_0, \left\lceil 64\sqrt{(T+1)/d} \right\rceil \right\}. \tag{79}$$

For an integer $j \geq j_0$, the first and second sample sizes in (34) and (35) are exactly

$$\begin{aligned} \rho &= 2^{-j}, & N &= 256(d+1)^2 j 2^j, \quad M = (N+1)^d, \\ L_R &= \left\lceil \log_2 \frac{4M^2}{\rho} \right\rceil, & R &= 8 L_R 2^j. \end{aligned} \tag{80}$$

Use Lemma S10.4 on the first sample and evaluate every selected witness on the independent second sample. There are exactly $N + R$ true-reference draws. Trace selection, pairwise evaluations, tree construction and online predictions are not additional draws.

The following envelope holds on every sample outcome:

$$N + R \leq G_d(j) := 512(d+1)^2 j 2^j. \tag{81}$$

Indeed, $d \leq T$ and $j \geq \log_2(T+2)$ imply $\log_2(d+1) \leq j$. For $j \geq 16$ we have $8 \leq j/2$ and $\log_2 j \leq j$. Thus

$$\log_2 N = 8 + 2\log_2(d+1) + \log_2 j + j \leq (9/2)j, \qquad \log_2(N+1) \leq 5j.$$

Consequently $L_R \leq 3 + 2d \log_2(N+1) + j \leq 12(d+1)j$, so $R \leq 96(d+1)j2^j$ and $N + R \leq 352(d+1)^2 j 2^j \leq G_d(j)$. The envelope uses the deterministic witness bound M, not the realized number of traces.

For $j_0 \leq j \leq j_{\text{stat}}$, the empirical-resolution bound (55) gives

$$\mathbb{E}\,\text{Reg}_T \leq C_d A_{d,T} + \frac{2T}{j \log 2}. \tag{82}$$

Here $\Phi_{d,T}(T) \leq C A_{d,T}$ by $\log(e + T/d) \leq C \log(eT/d)$. The imported proof uses only the proper witness property: the first-sample event controls all agreeing original pairs, and the independent second sample controls all pairs of selected witnesses. Its empirical tree appends every acquired witness as a final leaf; it does not stop at a coarser center level. Its one-step potential calculation needs that finite tree and its nondecreasing rates, not a supplied numerical law of the reference.

For precision about conditioning, on a good preprocessing outcome s let C_s be the actual conditional expected total KL cost. Equations (47) and (50) are affine in C_s. Equation (51) averages them using $\mathbb{E}[\mathbf{1}_G C_s] \leq T$, and charges the bad-sample event by T. It does not assume that an unconditional expected budget remains unchanged after conditioning on preprocessing. The current per-round hypothesis also holds given the full preprocessing history. Atoms and repeated samples are included by the countable disagreement and finite-trace arguments. The finite probability approximation in that appendix adds less than one expected loss, already included in (82). The learner selects its original proper witness privately before the input, without access to the true conditional input law.

Put $K_d = 512(d+1)^2$ and $s = \log(eT)$. For $T \geq 512$,

$$j_0 \leq 4s, \qquad s \leq j \quad (j \geq j_0). \tag{83}$$

For the first inequality, $16 \leq 4s$ and $\lceil \log_2(T+2) \rceil \leq \log_2(T+2) + 1 \leq 4s$. For the second, $j \log 2 \geq \log(T+2)$ and $1 + j \log 2 \leq j$ at $j \geq 16$, giving $1 + \log T \leq j$.

If $m < G_d(j_0)$, ignore the samples and use Theorem S9.1. Integrality gives $m + 1 \leq G_d(j_0)$, so

$$s_m \leq s + \log K_d + \log j_0 + j_0 \log 2 \leq [1 + \log K_d + 4(1 + \log 2)]s.$$

Thus the Cover upper is at most $C'_d T/s_m$. This includes $m = 0$.

If $m \geq G_d(j_0)$, let j_m be the largest integer $j \geq j_0$ with $G_d(j) \leq m$, and use $j = \min\{j_m, j_{\text{stat}}\}$. Such a largest integer exists because G_d increases to infinity. Equation (81) gives the hard call cap on every outcome. If $j_m \leq j_{\text{stat}}$, maximality and integrality give $m + 1 \leq G_d(j_m + 1)$. Using (83), $\log(j_m + 1) \leq j_m$, and $(j_m + 1)\log 2 \leq j_m$, we get

$$s_m \leq s + \log K_d + \log(j_m + 1) + (j_m + 1)\log 2 \leq (\log K_d + 3) j_m.$$

Substitute in (82) to obtain (78). This branch includes $j_m = j_{\text{stat}}$. If $j_m > j_{\text{stat}}$, then $T/j_{\text{stat}} \leq \sqrt{dT}/64 \leq A_{d,T}/64$. The remainder in (82) is absorbed by the statistical term, proving the same bound. All resolution and saturation cases are covered. □

S10.3 Combining the information sources and proving sharpness

Proof of Theorem S10.1. The strategy of Lemma S10.5 ignores ν, while that of Lemma S10.3 requests no true-reference sample. Select the former if $s_m \geq s_\delta$ and the latter otherwise. The choice is made before play from supplied quantities, not from unknown risks. Enlarge C_d to cover both upper bounds. The selected bound is $C_d[A_{d,T} + T/\max\{s_m, s_\delta\}]$. Since

$$s_{m,\delta} = s_m + s_\delta - \log(eT) \leq 2\max\{s_m, s_\delta\},$$

this proves (69), with a factor two. The choice requires neither aggregation nor reconstruction of μ.

For the lower take $X = [0,1]$, $\nu = U$, and the fixed class of rational thresholds $h_\theta(x) = \mathbf{1}\{x \geq \theta\}$, including $\theta = 0, 1$. Equation (173) gives $31T/(2048 s_{m,\delta})$ for every hard-cap learner, even under the four simultaneous promises

$$\begin{aligned} \mu \geq (1-\delta)U, \quad & \mathrm{TV}(\mu, U) \leq \delta, \\ \mathrm{KL}(\mu \| U) \leq \delta, \quad & \mathrm{KL}(U \| \mu) \leq \delta. \end{aligned} \tag{84}$$

This invokes the complete hidden-path proof in Theorem S11.12, which keeps the reference and target fixed before play, has an autonomous input protocol, and verifies conditional KL at every actual full history, including after adaptive reference requests. Its starting endpoints are rational, successive restrictions take quarters, and its final midpoint target is rational. Thus the lower comparator belongs to this fixed countable class.

Independently, Lemma S11.11 gives $\sqrt{T\log(eT)}/256$ under the known uniform reference, even with unlimited iid reference draws. Restricting its comparator class to rational thresholds changes no finite-path best loss. If both sides of a real-threshold trace are present, choose a rational parameter between the largest zero-labeled coordinate and the smallest one-labeled coordinate. If the zero side is empty, use 0. If the one side is empty, realizability of that trace implies that the largest coordinate is below 1; choose a rational parameter above it and at most 1. On an empty list use 0. Repeated coordinates cannot lie on different prediction sides. This verifies trace equality with atoms and endpoints; the labels of the online game themselves need not be realizable.

The two references witnessing these lower terms may differ. Their maximum is at least half their sum. The rational class has VC dimension one and is admissible for every dimension bound $d \geq 1$. Finally $A_{d,T} \leq \sqrt{d}\sqrt{T\log(eT)}$. Absorbing this dimension factor into c_d proves the stated fixed-dimensional worst-case lower. The lower takes the worst case over supplied ν and classes. □

Corollary S10.6. *The same fixed-dimensional worst-case order in (69) holds for every real $m \geq 0$ if the expected number of true-reference requests is at most m uniformly over the announced admissible instance class. The lower remains valid if that class is restricted by (84).*

Proof. Run the hard-cap upper at $\lfloor m \rfloor$. Its actual count is at most m on every history, and $m + 1 < 2(\lfloor m \rfloor + 1)$ shows that its joint denominator decreases by at most $\log 2$. Both denominators are at least $\log(eT) > 1$, so only a universal factor is lost.

For the lower, apply the complete autonomous truncation argument of (174), with the announced δ unchanged. It gives $T/(512 s_{m,\delta})$ under the uniform expectation promise even on the four-promise class. The argument's hard reference remains in that same announced class; an expectation guarantee on a smaller, easier family is not substituted. The rational target and unrestricted post-input access persist as in the theorem's proof. Combine with Lemma S11.11, which already allows unlimited reference draws. The same maximum-of-witnesses and fixed-d comparisons prove the lower order. □

At exact accuracy $\delta = 0$, $\mu = \nu$ and Theorem S6.1 at budget T supplies the statistical upper $CA_{d,T}$. The uniform witness above supplies its fixed-dimensional worst-case lower. This extension concerns agnostic regret; the realizable reference-sample law remains specific to thresholds. The sample count measures access to the reference, while mathematical selection and probability operations are available in the statistical model. Implementation cost and empirical performance require separate analysis.

S11 Reference access for thresholds

First consider ordinary thresholds on $[0, 1]$ with one fixed realizable target and conditional KL at most one on every full preceding history. The unknown-reference learner receives no reference samples or other reference-access oracle.

Theorem S11.1. *For $T \geq 512$, the worst-case minimax expected mistakes are $\Theta(\sqrt{T})$ when the reference is known and $\Theta(T/\log(eT))$ when it is unknown. Both uppers use proper pre-input thresholds and permit adaptively chosen, pathwise realizable labels. Both lowers allow unrestricted post-input prediction. The known-reference lower uses the uniform law; the unknown-reference lower has an atomless full-support reference fixed before play.*

We prove Theorem S11.1 for

$$X = [0, 1], \qquad h_\theta(x) = \mathbf{1}\{x \geq \theta\}, \qquad \theta \in [0, 1].$$

A fixed Borel reference μ dominates the next-input law P_t conditional on the full preceding history, including earlier learner randomness and adversary state. Write $c_t = \mathrm{KL}(P_t \| \mu)$. Fresh current learner randomness is private. The lower bounds allow unrestricted prediction after observing the input and unlimited computation. The upper strategies choose a whole consistent threshold before that input.

The fixed-target game has one threshold chosen before play, with $y_t = h_{\theta^\star}(x_t)$ throughout. We also prove the uppers when labels are adaptively selected but remain realizable on every history. At a finite horizon, continued consistency guarantees a common final comparator. This extension is proved separately wherever the fixed-target argument does not apply directly.

In the unknown-reference game the learner is not supplied μ, reference samples, or another reference-access oracle. In the known- reference game the learner may use exact conditional medians of μ. This is statistical access; a finite-bit implementation for an arbitrary Borel measure requires additional representation assumptions.

For a target fixed before play, the per-round unknown-reference upper order already follows from Theorem 13 of Blanchard et al. (2026). The path lower below adapts the construction in Appendix C of Blanchard (2025). Records, version-space medians, and entropy variational inequalities are classical tools. We include complete proofs of the stated constants, the expected-total budget bounds, and the adaptive-realizability extension.

S11.1 Known reference: median shrinkage

Let V_t be the thresholds consistent with the first $t - 1$ observations, and let U_t be the set of input points where two members of V_t disagree. Define u_t to be the smallest preceding positive input, or 1 if there is none. If there is no preceding negative input then $V_t = [0, u_t]$. Otherwise, letting l_t be the largest negative input, we have $V_t = (l_t, u_t]$. Continued realizability makes this set nonempty. Its disagreement set is $[0, u_t)$ in the first case when $u_t > 0$, and (l_t, u_t) in the second. For $V_t = \{0\}$ it is empty.

When $s_t = \mu(U_t) > 0$, choose a median m_t of the probability measure $\mu(\cdot \cap U_t)/s_t$. It can be chosen in V_t, with

$$\mu(U_t \cap [0, m_t)) \leq s_t/2, \qquad \mu(U_t \cap (m_t, 1]) \leq s_t/2. \tag{85}$$

For completeness, take the left quantile at probability $1/2$. The restricted distribution function is right-continuous. This quantile cannot be an excluded lower endpoint, which has zero restricted cumulative mass, and it cannot exceed the upper version endpoint. The usual quantile inequalities give both strict-side bounds. This also handles an atom at m_t larger than half the mass. Evaluations of the restricted distribution function are measurable in the interval endpoints, so this quantile is a measurable selection for fixed μ. When $s_t = 0$, choose u_t, which is consistent. Absolute continuity implies that the next input lies outside U_t almost surely.

Predict with h_{m_t} when $s_t > 0$. A mistake on a negative label has $x_t \geq m_t$; consistency then requires $\theta > x_t$, so the new disagreement interval is contained in $U_t \cap (x_t, 1]$, hence in the strict right side of m_t. A mistake on a positive label has $x_t < m_t$; the new disagreement interval is contained in

$U_t \cap [0, x_t)$, hence in the strict left side. By (85), every mistake at least halves uncertain reference mass. Correct updates cannot increase it. If M_{t-1} counts preceding mistakes, then pathwise

$$\mu(U_t) \le 2^{-M_{t-1}}. \tag{86}$$

Let $r_t = P_t(U_t)$, and let p_t be the conditional probability of a mistake given the full preceding history. Every mistake has input in U_t, so $p_t \le r_t$. Binary data processing gives, for $s_t > 0$,

$$c_t \ge \mathrm{kl}(r_t, s_t) \ge r_t \log(1/s_t) - \log 2.$$

The last inequality follows by bounding binary entropy by $\log 2$ and discarding the nonnegative contribution $-(1 - r_t)\log(1 - s_t)$. Thus

$$M_{t-1} p_t \log 2 \le c_t + \log 2. \tag{87}$$

When $M_{t-1} = 0$ this is immediate; when $s_t = 0$ there is no mistake almost surely. A mistake increments M by one, so

$$\begin{aligned} \mathbb{E}[M_t^2 - M_{t-1}^2 \mid \text{full past}] &= (2M_{t-1} + 1)p_t \\ &\le \frac{2c_t}{\log 2} + 3. \end{aligned} \tag{88}$$

Summing and applying Jensen yields, whenever $\mathbb{E}\sum_{t=1}^{T} c_t \le B$,

$$\mathbb{E}M_T \le \min\left\{T, \sqrt{\mathbb{E}M_T^2}\right\} \le \min\left\{T, \sqrt{3T + \frac{2B}{\log 2}}\right\}. \tag{89}$$

At $B = T$ this is less than $3\sqrt{T}$. Every version-space inclusion requires only continued realizability, so the bound also holds for adaptive realizable labels.

S11.2 Uniform reference: a square-root lower

Let $\mu = \mathrm{Unif}[0, 1]$. For a lower-bound prior, choose independent fair bits $b_1, \dots, b_T$ before play and start from $J_1 = [3/4, 1]$. Let A_j be the open middle half of J_j. If $b_j = 1$, take J_{j+1} to be the closed left quarter of J_j; if $b_j = 0$, take its closed right quarter. Let $\theta^\star$ be the midpoint of J_{T+1}. All these endpoints are dyadic.

The bands A_j are pairwise disjoint and satisfy

$$|A_j| = 2^{-(2j+1)}. \tag{90}$$

Indeed $|J_1| = 1/4$ and each quarter step divides its length by four. The later intervals lie entirely on the chosen side of the current middle band. It follows that

$$h_{\theta^\star}(x) = b_j \quad (x \in A_j), \qquad h_{\theta^\star}(x) = 0 \quad (x \in [0, 1/2]). \tag{91}$$

Endpoint events have probability zero.

Let j be one plus the number of preceding active visits. Draw the next input from

$$P_t = (1 - p_j)\,\mathrm{Unif}[0, 1/2] + p_j\,\mathrm{Unif}(A_j), \qquad p_j = \frac{1}{64j}. \tag{92}$$

Use a fresh component coin and coordinate on every round, and increment j only upon an active visit. At most T active visits occur, so all needed bands have been defined. Conditional on the full fixed bit sequence and preceding history, the disjoint supports give

$$\begin{aligned} \mathrm{KL}(P_t \| \mu) &= (1 - p_j)\log(2(1 - p_j)) + p_j \log \frac{p_j}{|A_j|} \\ &\le \log 2 + p_j(2j + 1)\log 2 \le \frac{67}{64}\log 2 < 1. \end{aligned} \tag{93}$$

The current band A_j depends only on earlier bits, not on b_j. Under the prior, the current active label remains fair conditional on the learner's observations, current coordinate, and private randomness.

Inactive observations reveal no new bit. The learner need not be told whether a draw is active; the band lies above $3/4$ and the neutral interval below $1/2$, so that information is already in its coordinate. Even with it, each active visit has expected mistake probability $1/2$. This argument also applies to randomized or probabilistic predictions after observing the input.

The waiting times for successive active visits have geometric means $64j$. Their component coins are independent, and hence the expected time to m active visits is

$$\sum_{j=1}^{m} 64j = 32m(m+1).$$

For $T \geq 512$, set $m = \lfloor \sqrt{T/128} \rfloor$. Then $m \geq 2$, $m \geq \sqrt{T/128}/2$, and $32m(m+1) \leq 64m^2 \leq T/2$. Markov's inequality gives probability at least $1/2$ of m active visits by T. Consequently the prior expected mistake count is at least

$$\frac{m}{4} \geq \frac{\sqrt{T}}{8\sqrt{128}} \geq \frac{\sqrt{T}}{128}. \tag{94}$$

Averaging over the finite bit prior selects one fixed target and one input protocol with at least this expected loss for the given learner. The reference is uniform for all such selections and every horizon. After selection the protocol uses no learner actions; its fresh component coins are not part of the preceding adversary state.

S11.3 Unknown reference: a strict-record bound

Choose any threshold consistent with the previous labels. This requires no knowledge of μ, T, or a budget. A mistake on a negative input requires that its coordinate exceed every preceding negative coordinate. A mistake on a positive input requires that its coordinate be below every preceding positive coordinate. They are strict records: a repeated coordinate with the same label cannot produce a new mistake.

For fixed θ, let $R_\theta(x_1, \ldots, x_T)$ be the sum of upper records in the subsequence $x_t < \theta$ and lower records in the subsequence $x_t \geq \theta$, including each nonempty subsequence's first observation. If θ realizes the labels, then pathwise

$$M_T \leq R_\theta(x_1, \ldots, x_T). \tag{95}$$

Consider iid μ inputs for the moment. Conditional on their sequence of negative and positive indicators for this fixed θ, the two subsequences are independent iid samples from their conditional laws. To handle atoms, attach independent uniform tie keys and order coordinate/key pairs lexicographically. Every original strict record is a record in this ordering; the ranked pairs are distinct almost surely.

For n iid continuously ranked observations, their insertion ranks are independent and uniform on $\{1, \ldots, i\}$ at position i. To prove this, their order permutation is uniform, and its insertion-rank sequence is a bijective encoding. Each possible sequence has probability $1/n! = \prod_{i=1}^{n}(1/i)$. Upper-record indicators are therefore independent Bernoulli variables with probabilities $1/i$; the same holds for lower-record indicators. For $\lambda > 0$ the record-count moment generating function is

$$\prod_{i=1}^{n} \left(1 + \frac{e^\lambda - 1}{i}\right) \leq \exp\big((e^\lambda - 1)H_n\big), \qquad H_n = \sum_{i=1}^{n} \frac{1}{i}, \quad H_0 = 0. \tag{96}$$

Applying this to each sign subsequence and using $H_n \leq H_T$, then removing the conditioning, gives

$$\log \mathbb{E}_{\mu^T} e^{\lambda R_\theta} \leq 2(e^\lambda - 1)H_T. \tag{97}$$

The added keys only dominate the strict records, so the inequality holds for the original coordinate functional as written.

Let ν be the actual marginal input-path law for a fixed target and learner. The conditional law given past inputs alone is a mixture of the next-input laws conditional on the full past. Conditional convexity of KL followed by the entropy chain rule yields

$$\mathrm{KL}(\nu \| \mu^T) \leq \mathbb{E} \sum_{t=1}^{T} \mathrm{KL}(P_t \| \mu) \leq B. \tag{98}$$

For a bounded measurable path functional R and $\lambda > 0$, define Q_λ by density $e^{\lambda R}/\mathbb{E}_{\mu^T} e^{\lambda R}$ relative to μ^T. Nonnegativity of $\mathrm{KL}(\nu\|Q_\lambda)$ then proves

$$\lambda \mathbb{E}_\nu R \le \mathrm{KL}(\nu\|\mu^T) + \log \mathbb{E}_{\mu^T} e^{\lambda R}. \tag{99}$$

All record functionals here lie between zero and T, so the exponential normalizations are finite. Combining (95), (97), (98), and (99) gives

$$\mathbb{E}M_T \le \min\left\{T, \inf_{\lambda>0} \frac{B + 2(e^\lambda - 1)H_T}{\lambda}\right\}. \tag{100}$$

At $B = T$ and $T \ge 16$, choose $\lambda = (\log T)/2$. The inequalities $H_T \le 1 + \log T \le \sqrt{T}$ give

$$\mathbb{E}M_T \le \frac{6T}{\log T} \le \frac{9T}{\log(eT)}. \tag{101}$$

The last comparison uses $\log T \ge 2$. For smaller positive integer horizons, the trivial bound T gives the same final constant.

S11.4 Adaptive realizability

When labels are adaptive, a final realizing threshold may depend on the entire path and learner randomness. It would be invalid to substitute that random threshold into (97). Instead define the measurable path functional

$$R_{\max}(x_1, \dots, x_T) = \sup_{\theta \in [0,1]} R_\theta(x_1, \dots, x_T). \tag{102}$$

At most $T + 1$ divisions of the observed coordinates into negative and positive groups occur, so this supremum is a finite maximum. Endpoint restrictions may remove some divisions, which can only decrease it.

Under iid μ inputs, attach the independent tie keys again and condition on the unordered multiset of distinct coordinate/key pairs. Their arrival permutation is uniform. For each of the $T+1$ cuts between these ordered pairs, condition further on the arrival positions of its two groups. Their within-group permutations are independent and uniform. The insertion-rank argument for (96) therefore gives the conditional bound $\exp(2(e^\lambda - 1)H_T)$ for that cut's total records.

Every cut from an original coordinate threshold is a cut of the ordered pairs, without splitting an equal-coordinate group. Its strict records are bounded by the tie-broken records. Thus the exponential of the original maximum is bounded by the sum of the exponentials for the $T + 1$ rank cuts, pointwise for every set of tie keys. Summing the conditional bounds and integrating proves

$$\log \mathbb{E}_{\mu^T} e^{\lambda R_{\max}} \le \log(T + 1) + 2(e^\lambda - 1)H_T. \tag{103}$$

Continued realizability and consistency give $M_T \le R_{\max}$ pathwise. The marginal input-path inequality (98) does not require a fixed target or fixed labels. Applying (99) to $R_{\max}$ proves

$$\mathbb{E}M_T \le \min\left\{T, \inf_{\lambda>0} \frac{B + \log(T + 1) + 2(e^\lambda - 1)H_T}{\lambda}\right\}. \tag{104}$$

At $B = T$, the same choice $\lambda = (\log T)/2$ for $T \ge 16$, together with $\log(T + 1) \le T$, gives

$$\mathbb{E}M_T \le \frac{8T}{\log T} \le \frac{12T}{\log(eT)}.$$

The trivial bound handles smaller horizons. This proves the unknown- reference upper for pathwise adaptive realizability without conditioning on a hindsight-selected target.

S11.5 An atomless full-support unknown-reference lower

Use the bands and fixed-target construction in (90)–(91), but set one active probability throughout:

$$q = \frac{1}{8\log(eT)}. \tag{105}$$

For each fixed complete bit sequence b, define

$$\mu_b^0 = (1-q)\,\mathrm{Unif}[0,1/2] + \frac{q}{T}\sum_{j=1}^{T}\mathrm{Unif}(A_j), \qquad \mu_b = \frac{1}{2}\mu_b^0 + \frac{1}{2}\,\mathrm{Unif}[0,1]. \tag{106}$$

All components have densities, and the second mixture gives full support and density at least $1/2$ relative to Lebesgue measure. The reference is fixed before play. It is not revealed, and the learner receives no auxiliary reference samples.

At depth j, draw

$$P_t = (1-q)\,\mathrm{Unif}[0,1/2] + q\,\mathrm{Unif}(A_j), \tag{107}$$

with a fresh component coin and coordinate. Increment j only on an active visit. The neutral interval and active bands are disjoint, so

$$\mathrm{KL}(P_t\|\mu_b^0) = q\log T.$$

Since μ_b has density at least half that of μ_b^0,

$$\mathrm{KL}(P_t\|\mu_b) \le q\log T + \log 2 \le \frac{1}{8} + \log 2 < 1. \tag{108}$$

This holds at every full history, conditional on the complete fixed bit sequence. The reference does not track realized learner choices.

At depth j, the active coordinate law depends only on $b_1,\dots,b_{j-1}$. Under the prior, b_j remains fair conditional on everything observed by the learner before its prediction, including the current coordinate and its own randomness. The hidden measure may encode later bits, but it is not an observation in this game. Neutral draws reveal none of them. Fresh component coins make the active count exactly Binomial(T,q). Each active visit has expected mistake probability $1/2$, so

$$\mathbb{E}_{\text{prior}} M_T \ge \frac{qT}{2} = \frac{T}{16\log(eT)}. \tag{109}$$

For each learner, averaging selects one fixed b with at least this expected loss. Its target, reference μ_b, and input protocol are then fixed. The target has zero loss. The argument allows improper randomized post-input prediction with unlimited computation, even when T and the KL cap are known. It does not apply when the reference or reference samples are supplied, since those observations can reveal the path.

S11.6 Conclusion and limits

Proof of Theorem S11.1. Under $c_t \le 1$ on every full history, the expected total budget is at most T. The known-reference upper (89) is at most $3\sqrt{T}$. The uniform-reference lower (94) is at least $\sqrt{T}/128$ for $T \ge 512$. The unknown-reference upper (104) is at most $12T/\log(eT)$, while (109) is at least $T/(16\log(eT))$. Both lower bounds have fixed realizable targets. Both uppers allow pathwise adaptive realizability. Therefore the same orders hold in either version of the game. The unknown-reference lower has the atomless full-support density property in (106). □

This is a worst-case minimax comparison. The references and processes in the two lower bounds need not coincide. In the unknown-reference lower, the fixed hard reference may depend on the horizon and on the learner through minimax selection, but is chosen before realized play. The result above concerns realizable labels; Theorem S11.2 gives the corresponding unknown-reference agnostic order. Implementing medians in finite bit operations for an arbitrary reference requires additional access assumptions. The record upper's per-round order and the path lower's construction build on the results cited at the start of this appendix.

S11.7 Agnostic thresholds without reference samples

Let $h_\theta(x) = \mathbf{1}\{x \ge \theta\}$ for $x,\theta \in [0,1]$. The reference μ is any fixed Borel probability measure on $[0,1]$. It is unavailable to the learner and need not be atomless or have full support. The learner receives no reference samples, density, CDF, or reference-access oracle. Let P_t be the next-input law

conditional on the full preceding history, including earlier learner randomness and adversary state, and suppose

$$\mathbb{E}\sum_{t=1}^{T}\mathrm{KL}(P_t\|\mu)\le B. \tag{110}$$

Labels may depend on this history, the current input, and the announced prediction distribution, but not on the current private threshold sample. For the threshold $\widehat{\theta}_t$ sampled before x_t, write

$$\mathrm{Reg}_T=\sum_{t=1}^{T}\mathbf{1}\{h_{\widehat{\theta}_t}(x_t)\neq y_t\}-\min_{\theta\in[0,1]}\sum_{t=1}^{T}\mathbf{1}\{h_\theta(x_t)\neq y_t\}.$$

The minimum exists because only finitely many prediction vectors occur on a finite input path. The horizon is supplied; the learner does not use B.

The algorithm below is a threshold specialization of Cover, Algorithm 3 in Section 4.1 of Blanchard (2025). We retain its frozen all-past empirical cover and restarted exponential weights. The proof uses strict records in interleaved input sequences to bound the error of these covers under (110).

S11.7.1 The blockwise learner

Fix an integer $w\ge 1$ and a rate $\eta>0$. Divide the rounds into consecutive blocks of length w, with a possibly shorter final block. At the beginning of block k, let

$$G_k=\{0,1\}\cup\{x_t : t \text{ belongs to a strictly earlier block}\}.$$

Keep this finite set fixed during the block and assign every member initial weight one. Before observing the next input, privately sample a threshold with probability proportional to its current weight. After observing the input and binary label, multiply each threshold's weight by $\exp(-\eta\ell)$, where ℓ is its binary loss on that round. At the next block, discard these weights and restart on the enlarged grid. All preceding inputs are retained. Equal traces can remain as separate thresholds; they do not affect the bound $|G_k|\le T+2$.

Let $J=\lceil T/w\rceil$, $r_*=\min\{w,T\}$, and

$$X^{(r)}=(x_r,x_{r+w},x_{r+2w},\dots),\qquad n_r=\left\lfloor\frac{T-r}{w}\right\rfloor+1,\qquad 1\le r\le r_*. \tag{111}$$

Only indices at most T are included. Put $H_n=\sum_{i=1}^{n}1/i$.

Theorem S11.2. *For every integer $T\ge 1$, integer $w\ge 1$, and $\eta>0$, the proper pre-input learner above satisfies*

$$\mathbb{E}\,\mathrm{Reg}_T\le\min\left\{T,\ \frac{J\log(T+2)}{\eta}+\frac{\eta T}{8}\right.$$
$$\left.+\inf_{\lambda>0}\frac{B+\sum_{r=1}^{r_*}\log(n_r+1)+2(e^{\lambda}-1)\sum_{r=1}^{r_*}H_{n_r}}{\lambda}\right\}. \tag{112}$$

In particular, set

$$j=\lceil\log_2(T+2)\rceil,\qquad w=j^3,\qquad \beta=1-\frac{1}{j},\qquad \eta=-\log\beta. \tag{113}$$

Implement the update by the rational multiplier β on each mistaken expert. If $B\le T$, then

$$\mathbb{E}\,\mathrm{Reg}_T\le\frac{64T}{\log(eT)}. \tag{114}$$

For finitely encoded rational inputs, this specialization has an exact implementation with expected arithmetic bit cost

$$O(T^2\,\mathrm{polylog}(T+2)),$$

in addition to reading and comparing the input encodings. It does not have a deterministic worst-case fair-bit cap.

The condition $B\le T$ follows from conditional KL at most one on every full preceding history. The expected-total condition is sufficient for the theorem. The lower in (109) is at least $T/[16\log(eT)]$ and already uses realizable labels and an atomless full-support reference, while allowing unrestricted post-input prediction. Thus (114) also gives the zero-reference agnostic minimax order $T/\log(eT)$.

S11.7.2 Frozen-grid errors are delayed records

Fix a complete input-label path, including any repetitions and endpoint inputs, and fix any comparator $\theta \in [0,1]$. At the start of block k, define

$$b_k(\theta) = \min\{b \in G_k : b \geq \theta\}.$$

This minimum exists because $1 \in G_k$. The thresholds $b_k(\theta)$ and θ disagree only when $\theta \leq x < b_k(\theta)$. In particular, on a round in block k their loss difference is at most

$$\mathbf{1}\{\theta \leq x_t < b_k(\theta)\}. \tag{115}$$

If this indicator is one, then x_t is strictly smaller than every input at least θ in an earlier block: an earlier such input $z \leq x_t$ would give $b_k(\theta) \leq z \leq x_t$. If no earlier input is at least θ, the statement is vacuous. The strict upper endpoint in (115) ensures that a repeated grid point does not cause an approximation error, even when μ has atoms.

For a finite sequence $z = (z_1, \ldots, z_n)$, let $R_\theta(z)$ count strict upper records among the $z_i < \theta$ and strict lower records among the $z_i \geq \theta$, with the first member of either nonempty subsequence counted as a record. Define

$$R_{\max}(z) = \max_{\theta \in [0,1]} R_\theta(z).$$

There are at most $n+1$ distinct divisions into the two subsequences, so this is a measurable finite maximum and $0 \leq R_{\max}(z) \leq n$. This is the strict-record functional used in (103).

Let $D_w(\theta)$ be the sum of (115) over all rounds. If a round has offset r in its block, all earlier points in its stream $X^{(r)}$ belong to strictly earlier blocks. Every error counted by $D_w(\theta)$ is therefore a strict lower record at least θ in that stream. Consequently, pointwise and simultaneously for every comparator,

$$D_w(\theta) \leq \sum_{r=1}^{r_*} R_\theta(X^{(r)}) \leq \sum_{r=1}^{r_*} R_{\max}(X^{(r)}) =: S_w(x_1, \ldots, x_T). \tag{116}$$

The streamwise maxima may use different cuts, which only enlarges the bound. A shorter final block is already accounted for by the lengths n_r. We have $0 \leq S_w \leq \sum_r n_r = T$. No independence under the actual adaptive input law has been used.

S11.7.3 Finite-expert regret and private sampling

Consider one block of l rounds and K experts. Let p_t be its current distribution, ℓ_t its vector of binary losses, and W_t the sum of its unnormalized weights. For a random variable in $[0,1]$, the second derivative of its log moment-generating function is a variance under an exponential tilt and is at most $1/4$. Integrating this bound twice, with the value and derivative at zero, gives

$$\log \frac{W_{t+1}}{W_t} = \log \sum_a p_t(a) e^{-\eta \ell_t(a)} \leq -\eta \sum_a p_t(a)\ell_t(a) + \frac{\eta^2}{8}.$$

The block's initial total weight is K, and its final total weight is at least $e^{-\eta L_k(a)}$ for each expert a, where $L_k(a)$ is that expert's block loss. Summing the displayed inequality proves

$$\sum_{t \text{ in block } k} \sum_a p_t(a)\ell_t(a) - L_k(a) \leq \frac{\log K}{\eta} + \frac{\eta l}{8} \tag{117}$$

simultaneously for all its experts. This calculation is pathwise and allows adaptive loss vectors.

For each block, apply (117) to $b_k(\theta)$ and use (115) and (116). Since $K \leq T+2$, summing gives, simultaneously for all θ,

$$\sum_{t=1}^{T} \sum_a p_t(a)\ell_t(a) - \sum_{t=1}^{T} \mathbf{1}\{h_\theta(x_t) \neq y_t\} \leq \frac{J \log(T+2)}{\eta} + \frac{\eta T}{8} + S_w. \tag{118}$$

We may minimize the comparator loss on this same realized path. In particular, no random hindsight comparator is substituted into a fixed-comparator probability bound.

The current threshold is sampled privately before x_t. Conditional on the preceding history and then on the current input and label, its expected loss is the mixed loss in (118), because the input and label do not observe that private sample. Future input laws and labels may depend on all earlier sampled thresholds and predictions. This does not change the roundwise conditional equality. Its expectation, summed over rounds, converts the mixed loss in (118) to the expected realized learner loss, while leaving the same hindsight comparator on both sides.

S11.7.4 The reference-product moment and KL transfer

For iid $Z_1, \dots, Z_n$ from any Borel law μ, the all-cut record bound (103) states that

$$\log \mathbb{E}_{\mu^n} e^{\lambda R_{\max}(Z_1,\dots,Z_n)} \le \log(n+1) + 2(e^\lambda - 1)H_n, \qquad \lambda > 0. \tag{119}$$

Its proof includes atoms. Independent uniform tie keys make the ordered coordinate-key pairs distinct; conditional on their unordered set, their arrival permutation is uniform. For each rank cut the two within-group permutations have the record insertion-rank law, and a sum over at most $n+1$ cuts bounds the maximum. An original threshold does not divide a group of equal coordinates, and each strict original record remains a record after tie breaking. Thus the cited inequality applies to precisely the strict functional in (116), without an atomlessness assumption.

Under μ^T, the offset streams in (111) use disjoint coordinate sets and are independent. Therefore (119) gives

$$\log \mathbb{E}_{\mu^T} e^{\lambda S_w} \le \sum_{r=1}^{r_*} \log(n_r+1) + 2(e^\lambda - 1) \sum_{r=1}^{r_*} H_{n_r}. \tag{120}$$

This factorization is only under the product reference law. It makes no assertion about insertion ranks conditional on the actual labeled history.

Let ν be the actual marginal law of $(x_1, \dots, x_T)$, integrating out labels and learner randomness. Its input laws conditional on preceding inputs alone are mixtures of the laws conditional on the full preceding history. Convexity and the chain rule for relative entropy, as in (98), imply

$$\mathrm{KL}(\nu \| \mu^T) \le \mathbb{E} \sum_{t=1}^{T} \mathrm{KL}(P_t \| \mu) \le B.$$

This argument allows arbitrary admissible adaptive labels; it neither fixes the labels nor conditions on a selected target. Apply (99) to the bounded functional S_w:

$$\lambda \mathbb{E}_\nu S_w \le B + \log \mathbb{E}_{\mu^T} e^{\lambda S_w}. \tag{121}$$

For completeness, this inequality follows by normalizing $e^{\lambda S_w} \mu^T$ to a probability and expanding the nonnegative relative entropy from ν to that probability. Combining (118), (120), and (121), and then taking the infimum over $\lambda > 0$, proves (112). The separate bound $\mathrm{Reg}_T \le T$ supplies its minimum with T.

S11.7.5 A universal constant and an exact implementation

We finish the proof of Theorem S11.2. Use (113) and write $s = \log(eT)$. If $T < 2^{64}$, then $s < 46$, so the trivial bound T is at most $64T/s$. Suppose instead that $T \ge 2^{64}$. On this range,

$$s \le j \le 2s, \qquad w = j^3 \le 8s^3 \le T. \tag{122}$$

The first inequality follows from $\log_2 T \ge 1 + \log T$ on this range. The second follows from $j \le \log_2(T+2) + 1 \le 2\log(eT)$. For the last inequality, it holds at $T = 2^{64}$ and persists because $T/\log^3(eT)$ increases when $s > 3$.

Set $\lambda = (\log T)/2$. Since $n_r \le T$, $H_{n_r} \le s$, and $\log(n_r+1) \le s$, (120) yields

$$\log \mathbb{E}_{\mu^T} e^{\lambda S_w} \le ws(1 + 2\sqrt{T}) \le 24 s^4 \sqrt{T} \le T. \tag{123}$$

For the last inequality, at $T = 2^{64}$ we have $s < 46$ and $24 \cdot 46^4 < 2^{32}$. The logarithmic derivative of $\sqrt{T}/s^4$ with respect to $\log T$ is $1/2 - 4/s$, which is positive for $s > 8$. Thus the inequality persists on the stated range. If $B \le T$, (121) gives

$$\mathbb{E} S_w \le \frac{4T}{\log T} \le \frac{5T}{s}. \tag{124}$$

Since $j \geq 2$ and $x \leq -\log(1-x) \leq x/(1-x)$ for $0 \leq x < 1$, we have $1/j \leq \eta \leq 2/j$. Also $\log(T+2) \leq j$ and $J \leq T/j^3 + 1$. Hence

$$\frac{J\log(T+2)}{\eta} + \frac{\eta T}{8} \leq \left(\frac{T}{j^3}+1\right) j^2 + \frac{T}{4j} = \frac{5T}{4j} + j^2$$
$$\leq \frac{9T}{4j} \leq \frac{9T}{4s}. \tag{125}$$

Here $j^3 \leq T$ gives $j^2 \leq T/j$. Equations (124) and (125) sum to less than $8T/s$ in this range. Together with the small-horizon case, this proves (114).

The learner stores at most $T+2$ thresholds, all observed inputs or endpoints. A direct implementation uses $O(T^2)$ comparisons and scalar arithmetic operations. Sorting at block boundaries is optional; no reference integral or density computation is required. For rational inputs supplied in finite binary encodings, the sampled threshold is an input encoding or an endpoint. Reading and exact comparison costs depend on those encoding lengths and are charged separately.

The specialization is implemented exactly from integers. Compute $j = \lceil \log_2(T+2) \rceil$ from the integer $T+2$, then $w = j^3$ and $\beta = (j-1)/j$. No logarithm is evaluated to implement a weight update. If expert i has accumulated integer block loss $\ell_i \leq w$, assign it the positive integer weight

$$W_i = (j-1)^{\ell_i} j^{w-\ell_i} = j^w \beta^{\ell_i}. \tag{126}$$

The common factor cancels, so these weights give the exact distribution analyzed with $\eta = -\log\beta$. Each has $O(w \log j) = O(\log^3(T+2)\log\log(T+3))$ bits. Their sum has at most $\lceil \log_2(T+2) \rceil$ additional bits.

Compute the weights and their cumulative sums. To draw an integer uniformly from $\{0, \ldots, S-1\}$ for their sum $S > 1$, repeatedly draw $\lceil \log_2 S \rceil$ fair bits and reject the resulting integer if it is at least S. The accepted integer is uniform, and each attempt has acceptance probability greater than $1/2$. If $S = 1$, the choice is deterministic. Locate the accepted integer in the cumulative sums to sample the threshold. Every attempt has bit cost polynomial in $\log(T+2)$, and the expected number of attempts is less than two. Direct integer powering, summation, and lookup at each round therefore give the stated $O(T^2 \operatorname{polylog}(T+2))$ expected arithmetic bit bound, in addition to input reading and comparison. There is no numerical approximation error and no deterministic bound on the number of fair bits. For arbitrary real inputs, the same proper rule uses exact comparisons; the finite-bit bound applies to the finitely encoded rational inputs specified above.

S11.8 Reference-sample complexity for realizable thresholds

Consider the thresholds

$$h_\theta(x) = \mathbf{1}\{x \geq \theta\}, \qquad x, \theta \in [0,1].$$

A Borel probability measure μ is fixed before play and is not supplied to the learner. On each round, the next-input law P_t conditional on the full preceding history satisfies

$$\mathrm{KL}(P_t \| \mu) \leq 1 \qquad \text{almost surely.} \tag{127}$$

This history includes earlier learner randomness, already observed reference draws, and adversary state. It does not include fresh reference-sampling or input-generation coins. Current learner randomness is private. A target threshold is chosen before play and realizes every label. The upper bounds also permit adaptive labels provided that the entire realized history remains threshold-realizable.

The learner's only access to μ is an oracle returning an unlabeled sample with conditional law μ, independently of all preceding randomness, on every request. Samples do not count as online mistakes. The requests may be adaptive and interspersed with online observations. In the upper model, requests for a round are completed before its current input is generated. The lower bounds allow additional requests after that input and before prediction. They allow arbitrary randomized post-input predictions and unlimited computation. All procedures must produce their predictions; an infinite unfinished sequence of requests is not a prediction strategy.

There is no density, likelihood-ratio, CDF, quantile, or conditional- sampling oracle. No finite-bit representation of an arbitrary real sample is assumed. The upper algorithms are statistical sample-access procedures and select a whole proper threshold before the current input.

The exponential dependence of approximate rejection sampling on inverse accuracy under a KL constraint is established by Block and Polyanskiy (2023); we use the corrected version (Block and Polyanskiy, 2024) for this comparison. The statement below concerns prediction with online feedback, not selection of an output from the reference samples. Its hidden-path lower adapts Appendix C of Blanchard (2025). For a target fixed before play, the no-reference upper order already follows from Theorem 13 of Blanchard et al. (2026). Empirical CDF bounds, version-space medians, and sample-budget truncation are classical tools.

Theorem S11.3. *There are universal constants $c, C > 0$ such that, for every pair of integers $T \geq 512$ and $m \geq 0$, the following holds. In the game defined above with at most m reference-oracle calls on every history, the minimax expected number of mistakes has order*

$$\sqrt{T} + \frac{T}{\log(eT(m+1))}. \tag{128}$$

More precisely, one proper pre-input learner has expected mistakes at most C times (128) for every admissible reference and every adaptively realizable label protocol. Every unrestricted post-input learner has expected mistakes at least c times (128) on some fixed realizable target, fixed atomless full-support reference, and autonomous input protocol satisfying (127). Lower-bound instances may depend on T, m, and the learner, but are fixed before realized play.

S11.8.1 Uniform empirical interval control

The next lemma includes atoms and intervals selected after the reference sample has been observed. A quantile map is used only in its proof.

Lemma S11.4. *Let $m \geq 2^{18}$ be an integer and let $Z_1, \ldots, Z_m$ be iid with any Borel probability law μ on $[0,1]$. Write $\widehat{\mu}$ for their empirical measure. With probability at least $1 - 1/(4m^2)$, simultaneously for every interval $I \subseteq [0,1]$, including empty and singleton intervals and either endpoint convention,*

$$\mu(I) \leq \widehat{\mu}(I) + m^{-1/4}. \tag{129}$$

Proof. First let $U_1, \ldots, U_m$ be iid uniform variables on $[0,1]$, and let $G_m(u) = m^{-1} \sum_i \mathbf{1}\{U_i \leq u\}$. For a variable $X \in [0,1]$, the function $f(s) = \log \mathbb{E} e^{sX}$ has second derivative equal to a tilted variance. Under any such tilt, writing a for the mean, $X^2 \leq X$ gives variance at most $a - a^2 \leq 1/4$. Integrating twice gives

$$\log \mathbb{E} e^{s(X - \mathbb{E}X)} \leq s^2/8.$$

Independence and exponential Markov, with $s = 4\varepsilon$ for the sum of m indicators, imply

$$\Pr\{|G_m(u) - u| > \varepsilon\} \leq 2e^{-2m\varepsilon^2}. \tag{130}$$

Apply this at $u = j/m, 0 \leq j \leq m$. Monotonicity gives

$$\sup_{0 \leq u \leq 1} |G_m(u) - u| \leq \varepsilon + 1/m$$

whenever every grid inequality holds. For $\varepsilon = \sqrt{2\log(2m)/m}$ the failure probability is at most

$$2(m+1)e^{-4\log(2m)} = \frac{m+1}{8m^4} \leq \frac{1}{4m^3} \leq \frac{1}{4m^2}. \tag{131}$$

The same uniform bound holds for CDF left limits. Taking differences of CDF values or left limits therefore gives, simultaneously for every interval J, with any endpoint convention,

$$\mathrm{Leb}(J) \leq \frac{1}{m} \sum_i \mathbf{1}\{U_i \in J\} + 2\sqrt{\frac{2\log(2m)}{m}} + \frac{2}{m}.$$

For the stated range of m, the last error is at most $4\sqrt{\log(2m)/m}$, since $2\sqrt{2} + 2/\sqrt{m\log(2m)} \leq 4$. Also $16\log(2m) \leq \sqrt{m}$ for $m \geq 2^{18}$. To verify the latter inequality, at $m = 2^{18}$ its left side is $304\log 2 < 228 < 512$, and the ratio $\sqrt{m}/\log(2m)$ is increasing when $\log(2m) > 2$. Hence

$$4\sqrt{\frac{\log(2m)}{m}} \leq m^{-1/4}. \tag{132}$$

For general μ, couple $Z_i = F_\mu^{-1}(U_i)$ using the left quantile. Choose arbitrary definitions at the null endpoints of the uniform variable. The quantile is nondecreasing. The inverse image of each coordinate interval is an interval with a possible choice of open or closed endpoints. It has Lebesgue measure $\mu(I)$, and its sample count is the count of the Z_i in I. The simultaneous uniform bound just proved therefore implies (129) for every I, including singleton atoms. The learner does not evaluate this quantile map. □

S11.8.2 Empirical medians

Lemma S11.5. *For integers $T \geq 512$ and $m > T^2$, there is a proper pre-input learner using m reference samples before play such that, under (127) and adaptive pathwise realizability,*

$$\mathbb{E}M_T \leq \sqrt{12T} + \frac{24(\log 2)T}{\log m} + \frac{T}{4m^2}, \tag{133}$$

where M_T is its number of mistakes.

Proof. Request all m reference samples before online play. Let V_t be the thresholds consistent with the preceding online labels. Write u_t for the smallest preceding positive input, or 1 if there is none. If there has been no negative input then $V_t = [0, u_t]$. Otherwise, writing l_t for the largest preceding negative input, $V_t = (l_t, u_t]$. The disagreement interval U_t is respectively $[0, u_t)$ or (l_t, u_t); it is empty when $V_t = \{0\}$. Continued realizability ensures that V_t is nonempty.

If $\widehat{\mu}(U_t) > 0$, choose an empirical median of the reference points in U_t. An order statistic provides a measurable selection $a_t \in V_t$ satisfying

$$\widehat{\mu}(U_t \cap [0, a_t)) \leq \widehat{\mu}(U_t)/2, \qquad \widehat{\mu}(U_t \cap (a_t, 1]) \leq \widehat{\mu}(U_t)/2.$$

These are strict sides, so the inequalities allow repeated reference coordinates and an atom at the median. If $\widehat{\mu}(U_t) = 0$, choose $a_t = u_t$. Predict with h_{a_t}, then update the version space. The algorithm uses neither μ nor P_t nor any density.

A mistake with a negative label has $x_t \geq a_t$, so the new disagreement interval is contained in $U_t \cap (x_t, 1]$ and hence in the strict right side of a_t. A mistake with a positive label has $x_t < a_t$, so the new disagreement interval lies in $U_t \cap [0, x_t)$ and hence in its strict left side. Every mistake therefore halves uncertain empirical mass. Correct updates cannot increase it, and zero empirical mass stays zero. Pathwise,

$$\widehat{\mu}(U_t) \leq 2^{-M_{t-1}}. \tag{134}$$

Let G be the event in Lemma S11.4. It applies because $m > T^2 \geq 2^{18}$. On G,

$$\mu(U_t) \leq 2^{-M_{t-1}} + m^{-1/4}.$$

Put

$$L = \frac{\log m}{4 \log 2}, \qquad D_t = \min\{M_{t-1}, L\}. \tag{135}$$

Both terms in the mass bound are at most 2^{-D_t}, so

$$\mu(U_t) \leq 2^{1-D_t} \qquad \text{on } G. \tag{136}$$

Let p_t be the conditional probability of a mistake given the full preceding history, and let $r_t = P_t(U_t)$. Consistency gives $p_t \leq r_t$. If $\mu(U_t) = 0$ then absolute continuity gives $p_t = 0$. Otherwise, when $D_t \geq 2$, binary data processing and the binary entropy bound give

$$\begin{aligned} 1 &\geq \mathrm{KL}(P_t \| \mu) \geq \mathrm{kl}(r_t, \mu(U_t)) \\ &\geq r_t \log(1/\mu(U_t)) - \log 2 \geq p_t(D_t - 1)\log 2 - \log 2. \end{aligned}$$

Since $D_t/(D_t - 1) \leq 2$, this implies $D_t p_t \leq 2(1 + \log 2)/\log 2 < 5$. For $D_t < 2$, use $p_t \leq 1$. In every case, on G,

$$(\min\{M_{t-1}, L\} + 1)p_t \leq 6. \tag{137}$$

The event G is determined by the reference sample before play. Conditioning on that sample retains the full-history hypothesis (127). Thus conditioning on G in (137) does not condition on a future online event. The reference sample may be visible to the adversary: the interval bound is simultaneous over every subsequently chosen interval.

For integers $r \geq 0$, define

$$F(0) = 0, \qquad F(r) = \sum_{j=0}^{r-1} (\min\{j, L\} + 1).$$

A mistake increments $F(M)$ by $\min\{M, L\}+1$. Summing (137) and taking conditional expectations yields $\mathbb{E}[F(M_T) \mid G] \leq 6T$. For real $x \geq 0$, define

$$H(x) = \int_0^x \min\{u, L\}\, du = \begin{cases} x^2/2, & x \leq L, \\ Lx - L^2/2, & x > L. \end{cases}$$

On $u \in [j, j+1]$, $\min\{u, L\} \leq \min\{j, L\} + 1$, so $H(r) \leq F(r)$ for every integer r. Also

$$x \leq \sqrt{2H(x)} + H(x)/L.$$

For $x \leq L$ the square-root term is x. For $x > L$, $H(x)/L = x - L/2$ and $\sqrt{2H(x)} \geq L$, proving the inequality. Jensen's inequality therefore gives

$$\mathbb{E}[M_T \mid G] \leq \sqrt{12T} + 6T/L = \sqrt{12T} + \frac{24(\log 2)T}{\log m}.$$

On G^c there are at most T mistakes. Lemma S11.4 now gives (133). Continued realizability suffices throughout, including when the final consistent target depends on the observations. □

S11.8.3 The upper for every sample allowance

We use two bounds from Appendix S11. Equations (103) through (104) show that any proper consistent threshold learner, using no reference samples, satisfies

$$\mathbb{E}M_T \leq \min\left\{T, \inf_{\lambda > 0} \frac{B + \log(T+1) + 2(e^\lambda - 1)H_T}{\lambda}\right\}, \qquad H_T = \sum_{i=1}^{T} \frac{1}{i}, \tag{138}$$

under an expected total full-history input KL budget B. This bound allows atoms and adaptive pathwise realizability. Its proof uses the maximum over all $T+1$ rank cuts before applying the entropy inequality, not a fixed-cut bound conditioned on a hindsight target. In particular, at $B = T$ it gives $\mathbb{E}M_T \leq 12T/\log(eT)$.

Lemma S11.6. *For every pair of integers $T \geq 512$ and $m \geq 0$, one proper pre-input learner using at most m reference samples has expected mistakes at most a universal constant times (128), including under adaptive pathwise realizability.*

Proof. If $m \leq T^2$, ignore the reference oracle and use the consistent learner in (138). Since

$$\log(eT(m+1)) \leq \log(2eT^3) \leq 4\log(eT),$$

its expected mistakes are at most $48T/\log(eT(m+1))$. If $m > T^2$, use Lemma S11.5. Here

$$\log(eT(m+1)) \leq 1 + \log T + \log m + \log 2 \leq 2\log m.$$

The last inequality follows from $\log m > 2\log T$ and $T \geq 512$. Substitution in (133), with its last term bounded by 1, proves the stated universal-factor bound. □

S11.8.4 A nearly uniform reference whose path is hidden

Lemma S11.7. *For integers $T \geq 512$ and $m \geq 0$, every learner using at most m reference samples on every history has an instance with a fixed realizable threshold, an atomless full-support reference, and an autonomous input protocol satisfying (127) on which its expected mistakes are at least*

$$\frac{31T}{2048\log(eT(m+1))}. \tag{139}$$

Write $U = \mathrm{Unif}[0,1]$ and

$$\alpha = \frac{1}{64(m+1)(T+1)^2}, \qquad s = \log(eT(m+1)), \qquad q = \frac{1}{32s}. \tag{140}$$

The reference can satisfy all four promises

$$\mu \geq (1-\alpha)U, \qquad \mathrm{TV}(\mu, U) \leq \alpha,$$
$$\mathrm{KL}(\mu\|U) \leq \frac{1}{16(m+1)(T+1)}, \qquad \mathrm{KL}(U\|\mu) \leq 2\alpha. \tag{141}$$

The learner may know U and these promises, request samples adaptively including after the current input, and predict arbitrarily after it. Here total variation is the supremum of probability differences over measurable events.

Proof. Put $w = (T+1)^{-2}$ and fix the known interval $J_1 = [3/4, 3/4 + w]$. Choose independent fair bits $b_1, \ldots, b_T$ before play as a finite lower-bound prior. Let A_j be the open middle half of J_j. If $b_j = 1$, let J_{j+1} be its closed left quarter; if $b_j = 0$, let it be its closed right quarter. Let $\theta^\star$ be the midpoint of J_{T+1}. The bands are pairwise disjoint and

$$|A_j| = \frac{w}{2 \cdot 4^{j-1}}, \qquad \log(1/|A_j|) = 2\log(T+1) + (2j-1)\log 2.$$

Every later interval lies on the selected side of A_j, so

$$h_{\theta^\star}(x) = b_j \quad (x \in A_j). \tag{142}$$

The band's location uses earlier bits, not b_j. Outside the fixed J_1, every target in the prior labels points zero to its left and one to its right.

Write $Q_j = \mathrm{Unif}(A_j)$ and define the fixed reference

$$R_b = \frac{1}{T}\sum_{j=1}^{T} Q_j, \qquad \mu_b = (1-\alpha)U + \alpha R_b. \tag{143}$$

It is atomless, has full support, and dominates $(1-\alpha)U$. The mixture gives $\mathrm{TV}(\mu_b, U) \leq \alpha$. Convexity of KL in its first argument gives

$$\begin{aligned}\mathrm{KL}(\mu_b\|U) &\leq \frac{\alpha}{T}\sum_{j=1}^{T}\mathrm{KL}(Q_j\|U) \\ &= \alpha[2\log(T+1) + T\log 2] \leq 4\alpha(T+1).\end{aligned}$$

The equality uses $\sum_{j=1}^{T}(2j-1) = T^2$. The final inequality uses $\log(T+1) \leq T$ and $\log 2 < 1$. The density lower bound also gives

$$\mathrm{KL}(U\|\mu_b) \leq -\log(1-\alpha) \leq 2\alpha,$$

since $\alpha \leq 1/2$. This proves every promise (141) for each individual reference, not just its average under the prior.

Let j be one plus the number of preceding active visits. Generate the next input from

$$P_t = (1-q)U + qQ_j \tag{144}$$

using a fresh component coin and coordinate. Increment j only on an active visit, and label every input by $h_{\theta^\star}$. At most T active visits occur. The protocol ignores every learner action and reference-query time.

Conditional on the fixed bits and full preceding history, reference domination and convexity in the input argument give

$$\mathrm{KL}(P_t\|\mu_b) \leq (1-q)\log\frac{1}{1-\alpha} + q\log\frac{T}{\alpha}.$$

For $T \geq 512$, $s > 7$ and

$$\log(T/\alpha) = s - 1 + \log 64 + 2\log(T+1) \leq 4s.$$

Indeed $\log(T+1) \leq s - 1 + 1/T$, and $\log 64 - 3 + 2/T < 2 < s$. Consequently

$$\mathrm{KL}(P_t\|\mu_b) \leq 2\alpha + \frac{1}{8} \leq \frac{5}{32} < 1. \tag{145}$$

This holds after all preceding reference observations. The fresh input and reference coins are not preceding adversary state.

Pre-generate m potential reference-component coins. A call returns a draw from R_b with probability α, and otherwise an independent U draw. Reveal the next unused draw on each request. This implements iid μ_b sampling even under adaptive post-input requests. Also pre-generate T independent potential uniform online background coordinates, one for each round, independent of all active coins and bits. Use a round's coordinate only on a background round.

Let E be the event that every potential reference coin selects U and every potential online background coordinate lies outside the fixed J_1. This event is independent of the bits and active coins. It is defined on all potential draws, not on a realized stopping count. A union bound gives

$$\Pr(E) \geq 1 - m\alpha - Tw \geq 1 - \frac{1}{64(T+1)^2} - \frac{T}{(T+1)^2} \geq \frac{31}{32}. \tag{146}$$

For the final inequality each subtracted term is below $1/512$.

Conditional on E, the entire potential reference stream is iid U and independent of the bits. Its coordinates may geometrically fall inside path bands; their common unlabeled law still reveals no bit. The potential online background coordinates are uniform outside the fixed J_1, where all target labels agree. Conditioning leaves the active coins iid Bernoulli(q), independent of the bits, and leaves active within-band randomness unchanged.

At the active visit to depth j, earlier active feedback and the current band's location use only $b_1, \ldots, b_{j-1}$. On E, all reference and background interactions have bit-independent laws. Induction over successive online observations and oracle interactions therefore shows that b_j is fair given everything the learner observes before that prediction. This includes the current coordinate, all additional post-input reference requests and answers, and private learner randomness. Every active prediction has conditional expected error at least one half.

The active count remains Binomial(T, q) conditional on E. Mistakes off E are nonnegative, so

$$\mathbb{E}_{\text{prior}} M_T \geq \Pr(E) qT/2 \geq \frac{31T}{2048 \log(eT(m+1))}.$$

Finite-prior averaging selects one fixed target, reference, and autonomous protocol with this expected loss for the given learner. The selection precedes realized play. Realizability makes mistakes equal regret. The event E was used only to bound loss; it did not replace the actual online law in the KL verification. □

S11.8.5 Completion of the hard-cap law

The complete uniform-reference construction in (90) through (94) gives at least $\sqrt{T}/128$ mistakes for $T \geq 512$, even when the uniform law is disclosed. Its fair target bits are independent of any additional iid uniform reference stream. Adaptive requests into that stream, including requests after the current input, therefore reveal no unused target bit. Its lower proof continues to hold with unrestricted reference access.

Proof of Theorem S11.3. Lemma S11.6 proves the upper for every m. Every learner has worst-case expected mistakes at least both $\sqrt{T}/128$ by the preceding uniform-reference lower and (139) by Lemma S11.7. Their maximum is at least half their sum, proving the lower in (128) with a universal constant. The two witnesses need not be the same instance. Both lower references are atomless and have full support, and both lower targets are fixed before play. The upper permits adaptive pathwise realizability. □

S11.8.6 Uniformly bounded expected oracle calls

Corollary S11.8. *Let $T \geq 512$ be an integer and $m \geq 0$ a real number. Replace the hard cap by the requirement that the total number N of reference draws satisfy $\mathbb{E}N \leq m$ on every admissible reference, target, and input protocol. The minimax expected mistake count again has order (128), with universal constants. The upper is proper pre-input and uses a deterministic number of samples. The lower permits adaptive stopping and unrestricted post-input prediction, and uses an atomless full-support reference and an autonomous protocol.*

Proof. For the upper, use Theorem S11.3 with $n = \lfloor m \rfloor$. Its deterministic number of calls is at most $n \le m$. Since $n + 1 \ge (m + 1)/2$,

$$\log(eT(n+1)) \ge \log(eT(m+1)) - \log 2.$$

For $T \ge 512$ the two logarithms are within a universal factor, so the upper has the asserted order, including for adaptive labels.

For the lower, set

$$s = \log(eT(m+1)), \qquad K = 2^{20}, \qquad M = \lceil K(m+1)s \rceil.$$

Given a learner A with the uniform expected allowance, define A^M by running it until it attempts a $(M+1)$-st reference request. At that request, abandon its internal strategy and play h_1 on the current round if its prediction is still due and on all subsequent rounds, making no more reference calls. There is no need to simulate future requests by the abandoned strategy. This is a valid learner with at most M reference draws on every history.

Apply Lemma S11.7 to A^M. It supplies a fixed target, fixed reference, and autonomous input protocol with

$$\mathbb{E}M_T(A^M) \ge \frac{31T}{2048 \log(eT(M+1))}.$$

Couple A and A^M using identical reference draws, identical private randomness until truncation, and the same online component coins and coordinates. This is valid because (144) ignores learner actions and reference-request times; its depth depends only on earlier active visits. The target labels are fixed in both runs. Until truncation, the histories and predictions coincide.

On paths with no truncation, their mistake counts agree. On other paths, they differ by at most T. The resource guarantee holds for A on this particular admissible fixed instance, so Markov gives

$$\mathbb{E}M_T(A) \ge \frac{31T}{2048 \log(eT(M+1))} - T \Pr\{N > M\} \ge \frac{31T}{2048 \log(eT(M+1))} - \frac{Tm}{M}. \quad (147)$$

This couples two complete runs. It does not infer fair future labels by conditioning on an event about A's eventual stopping count.

Since $K(m+1)s \ge 2$, we have $M + 1 \le 2K(m+1)s$. Consequently

$$\log(eT(M+1)) \le s + \log(2K) + \log s \le 5s.$$

Indeed $s > 7$, $\log s \le s$, and $\log(2K) = 21 \log 2 < 63/4 < (9/4)s$, whose sum is at most $17s/4 < 5s$. Also $m/M \le 1/(Ks)$. Equation (147) now gives

$$\mathbb{E}M_T(A) \ge \left(\frac{31}{10240} - \frac{1}{2^{20}} \right) \frac{T}{s} \ge \frac{T}{512s}. \quad (148)$$

For the last inequality, $31/10240 - 1/512 = 11/10240 > 2^{-20}$. The uniform-reference lower $\sqrt{T}/128$ still holds with unrestricted reference sampling. Taking the maximum of the two worst cases, and then half their sum, proves the lower order. Their instances need not coincide. This truncation argument relies on the autonomy of the specified hard protocol; it does not couple arbitrary adversaries through divergent learner histories. □

S11.8.7 Accurate approximate reference knowledge

Corollary S11.9. *Restrict the references in Theorem S11.3 or Theorem S11.10 to the atomless full-support measures satisfying (141), with the T, m-dependent α in (140). Give the learner U and all four promises. The same respective minimax rates hold, with universal constants. They also hold when the expected number of calls is at most the real allowance m uniformly only over this restricted class.*

Proof. The previous upper bounds apply to every member of the restricted class. Lemma S11.7 gives the sample-dependent lower with its promises already known to the learner. The reference U itself satisfies all four promises, so the known-uniform statistical lowers used in the two theorems

also remain available with unrestricted reference samples. Taking the maximum of the two witnesses proves the hard-cap assertions.

For expected calls, use the truncation level M in Corollary S11.8. Since $M \geq m$, the hard reference constructed at allowance M satisfies tighter versions of all four promises than the reference class announced at allowance m. The truncated learner may retain its original announcement and ignore the stronger one. The original expected-resource guarantee therefore applies on this particular hard instance, even if required only on the restricted class. The same autonomous coupling and (147) through (148) give $T/[512\log(eT(m+1))]$. The known-uniform statistical lower and the deterministic $\lfloor m \rfloor$-sample upper complete both expected-call assertions. □

Thus small total variation, small KL in both directions, and even a pointwise density lower bound relative to a known uniform approximation do not remove the predictive access term. The uniform CDF error is also at most α. This does not bound the reference density from above or the input KL relative to U. Narrow path bands retain large density, and the true reference, not its approximation, defines the input budget.

S11.8.8 Consequences and scope

For every fixed $a \geq 0$, any allowance at most T^a leaves the worst-case order $\Theta_a(T/\log(eT))$. For a prescribed mistake bound $K_0\sqrt{T}$, the hard-cap lower requires

$$\log(eT(m+1)) \geq \frac{31}{2048K_0}\sqrt{T}, \qquad m+1 \geq \frac{\exp(31\sqrt{T}/(2048K_0))}{eT}.$$

For a uniformly expected allowance, (148) gives the same exponential order with the corresponding constant $1/(512K_0)$. Conversely, $m = \lceil e^{\sqrt{T}} \rceil$ gives $O(\sqrt{T})$ mistakes by the upper theorem; finitely many horizons where the empirical-median regime is not selected are absorbed by a universal constant.

These conclusions persist under the precise approximation promises above, but do not give sample requirements for a regular parametric family. The lower instance is selected separately for each horizon and sample allowance. These realizable results concern prediction from exact real samples; finite-bit implementation and density-estimation complexity require separate analyses. The agnostic and joint accuracy laws are in Theorems S11.10 and S11.12. The record bound (138) holds under an arbitrary expected-total input KL budget, while the full sample law here assumes the per-round constraint (127).

S11.9 Agnostic thresholds with a sampled reference

We combine the no-reference blockwise upper with the empirical-reference tree in Appendix S8. The latter uses only reference draws and finite class operations, not a numerical reference law. For ordinary thresholds we implement those class operations internally and count the reference draws separately. We also give a randomized-rank proof for atomless references and sample allowances at least the horizon. Real-coordinate comparisons and internal computation are unrestricted; no uniform bit bound for arbitrary real reference samples is asserted.

Theorem S11.10. *Let $T \geq 512$ and $m \geq 0$ be integers. For ordinary thresholds $h_\theta(x) = \mathbf{1}\{x \geq \theta\}$ on $[0,1]$, suppose the fixed Borel reference μ is unknown and $\mathrm{KL}(P_t\|\mu) \leq 1$ at every full preceding history, including reference samples and earlier learner randomness. Labels obey the current-round privacy in Section 2. With at most m iid unlabeled reference draws and no other reference access, the worst-case minimax expected regret over these references is*

$$\Theta\left(\sqrt{T\log(eT)} + \frac{T}{\log(eT(m+1))}\right). \tag{149}$$

The upper holds for every Borel reference, chooses proper thresholds before the current input, and uses a deterministic number of reference samples before play. The lower permits unrestricted post-input prediction and adaptive reference requests, including after the current input, with unlimited computation. Its references are atomless and have full support. The same rate holds for a real allowance $m \geq 0$ when the expected number of calls is at most m on every admissible instance. All implicit constants are universal.

S11.9.1 Countable thresholds and direct trace selection

Use the fixed countable class of thresholds whose parameters are rational numbers in $[0, 1]$, including 0 and 1. It has VC dimension one. On every finite list of inputs, each trace of a real threshold is also a trace of a rational threshold. Indeed, if both prediction sides are nonempty, the largest input assigned zero is strictly smaller than the smallest input assigned one. Choose a rational parameter strictly between them. If the zero side is empty, parameter 0 works. If the one side is empty, a realizing real parameter implies that the largest input is strictly below 1, so a rational parameter above it and at most 1 works. For an empty list, use 0. Repeated coordinates cannot be placed on different sides by a threshold and cause no exception.

This is a statement about all traces, not about realizability of the observed labels. Hence for arbitrary binary labels and on every realized finite input path, the best real and rational thresholds have equal loss. In particular the equality holds with atoms and at both endpoints.

Fix an enumeration of the rational parameters beginning with 0 and 1. For a nonempty sample of N inputs, sort its distinct coordinates as $a_1 < \cdots < a_k$, where $1 \le k \le N$. Every threshold trace has zeros on the first r distinct coordinates and ones on the rest, for some $0 \le r \le k$. For $r = 0$, choose parameter 0. For $1 \le r < k$, choose the first enumerated rational q with $a_r < q \le a_{r+1}$. The nonempty open interval between these inputs contains a rational, so this search terminates. The all-zero trace $r = k$ is feasible exactly when $a_k < 1$; choose parameter 1 in that case and omit this trace when $a_k = 1$. For an empty sample choose parameter 0. Repeated occurrences receive the same prediction. This supplies every feasible trace and at most $N + 1$ proper witnesses.

Each selected parameter is the first enumerated witness for its trace: the interior feasible set is exactly $(a_r, a_{r+1}]$, parameter 0 is first for the all-one trace, and parameter 1 is first for the feasible all-zero trace since parameter 0 cannot realize it. Ordering the witnesses by their original sample traces therefore gives the canonical list used in Appendix S8. This list is measurable. Sorting and equality tests partition the finite sample space into finitely many Borel ordering regions. On each region, the event that an enumeration index is the first satisfying an interior cut is its Borel inequality test intersected with the complements of the finitely many earlier tests. The endpoint cases are also Borel. Selection uses only the first sample; evaluation is the comparison $x \ge q$. Neither requires a reference density, CDF, conditional sample, label oracle, or further reference draw. Real-coordinate comparisons and internal computation remain unrestricted. There is no uniform bit bound for arbitrary real inputs or for finding a rational between arbitrarily close inputs.

S11.9.2 Reference draws at a prescribed tree resolution

Put

$$j_0 = \max\{16, \lceil \log_2(T + 2) \rceil\}, \qquad G(j) = 1072 j 2^j. \tag{150}$$

At dimension one and resolution $j \ge j_0$, the two sample sizes in (34) and (35) are

$$N = 1024 j 2^j, \qquad M = N + 1, \qquad R = 8 L_R 2^j, \qquad L_R = \min\{l \in \mathbb{N} : 2^l \ge 4M^2 2^j\}. \tag{151}$$

The first sample selects one rational threshold for every feasible trace by the construction above. The second independent sample estimates pairwise disagreement distances of those witnesses. There are exactly $N + R$ reference draws. Trace selection, tree construction, evaluations, and numerical prediction require no additional draws from μ. The number of witnesses is at most M.

Since $M \le 2N$, we have

$$L_R \le 25 + 2 \log_2 j + 3j \le 6j \qquad (j \ge 16).$$

For the last inequality use $\log_2 j \le j/4$ for $j \ge 16$ and $25 \le (25/16)j$. Thus the deterministic sample count satisfies

$$N + R \le 1072 j 2^j = G(j) \tag{152}$$

on every preprocessing outcome, including failed geometry events.

Apply the empirical-tree proof with $d = 1$, total budget $B = T$ and integer tuning input $b = T + 1$. The branch $b \ge T^2/d$ does not occur for $T \ge 512$. Its statistical resolution cap is

$$j_{\text{stat}} = \max\{j_0, \lceil 64\sqrt{T + 1} \rceil\}. \tag{153}$$

For every $j_0 \le j \le j_{\text{stat}}$, (55) gives

$$\mathbb{E}\,\mathrm{Reg}_T \le C\sqrt{T\log(eT)} + \frac{2T}{j\log 2}. \tag{154}$$

Here $\Phi_{1,T}(T) = \sqrt{T} + \sqrt{T\log(e+T)} = O(\sqrt{T\log(eT)})$. The first-sample net and second-sample disagreement estimates in that proof apply to the countable class just specified for every Borel μ; no change of reference is used.

The prediction distribution is computed from the acquired empirical tree. It selects a proper witness before the current input, with its choice and current randomness private through generation of the current labeled example. The local potential and the comparator transfer are proved at the actual conditional cost after preprocessing in (47)–(50). Their affine bounds are averaged in (51). This does not condition an unconditional budget on a good sample event. In the present per-round model the conditional cost is at most T after each fixed preprocessing outcome because the hypothesis holds at every full history. The finite probability approximation of Appendix S8 adds less than one expected loss and is already included in (154).

S11.9.3 An affordable resolution and the small-sample range

Suppose first that $m \ge G(j_0)$. Let j_m be the largest integer $j \ge j_0$ with $G(j) \le m$, and use $j = \min\{j_m, j_{\text{stat}}\}$. All samples are requested before play, and (152) gives the hard cap m.

If $j_m \le j_{\text{stat}}$, integrality and maximality imply $m+1 \le G(j_m+1)$. For $j_m \ge 16$,

$$\log G(j_m+1) = \log 1072 + \log(j_m+1) + (j_m+1)\log 2 \le 2j_m.$$

Indeed, $\log 1072 < 7$, $\log(j+1) \le j/4$ for $j \ge 16$, and $\log 2 < 3/4$, so the left side is at most $j + 31/4 \le 2j$. Also $m \ge G(j_0) \ge T$, whence

$$\log(eT(m+1)) \le 3\log(m+1) \le 6j_m.$$

It follows that

$$\frac{2T}{j_m \log 2} \le \frac{12}{\log 2}\,\frac{T}{\log(eT(m+1))}.$$

If instead $j_m > j_{\text{stat}}$, the residual term in (154) is at most $\sqrt{T}/(32\log 2)$ and is absorbed into the statistical term. This proves the upper in (149) when $m \ge G(j_0)$, including sample allowances above the statistical resolution cap.

For a simple range division, $m \ge T^4$ implies $m \ge G(j_0)$ for $T \ge 512$. In fact, $2^{j_0} \le 128T$, $j_0 \le 2\log_2 T$, and $\log_2 T \le T/8$ on this range. Therefore

$$G(j_0) \le 274432T\log_2 T \le 34304T^2 \le T^4.$$

The last inequality uses $T^2 \ge 512^2 > 34304$. For the remaining integer allowances $0 \le m < T^4$, ignore all reference samples and use Theorem S11.2. Since $m+1 \le T^4$,

$$\log(eT(m+1)) \le 1 + 5\log T \le 5\log(eT),$$

and hence

$$\mathbb{E}\,\mathrm{Reg}_T \le \frac{64T}{\log(eT)} \le \frac{320T}{\log(eT(m+1))}. \tag{155}$$

This completes the all-Borel hard-cap upper. The split at T^4 suffices for the two bounds to cover every sample allowance.

The next three subsubsections retain a direct alternative based on randomized ranks. Only in these subsubsections assume that μ is atomless and $m \ge T$. The preceding proof already covers all Borel references.

S11.9.4 A uniform event for reference gaps

Draw m iid reference points and sort them as $z_1 < \cdots < z_m$, with $z_0 = 0$, $z_n = 1$, and $n = m+1$. These inequalities, including the endpoints, are strict almost surely. Put $I_j = (z_j, z_{j+1})$ and $p_j =$

$\mu(I_j)$ for $0 \le j < n$. The omitted finite set has reference measure zero and hence zero conditional input probability.

The reference CDF is continuous, so its values at the sampled points are iid uniform on $[0,1]$. The vector (p_j) therefore has the law of uniform order-statistic spacings. Flat CDF intervals cause no exception: the transformed random value is uniform and CDF differences are the intervening reference masses.

The ordered uniform points have density $m!$ on their ordered simplex, by summing the $m!$ strict orderings of the unit-cube density. The determinant-one change to spacings gives density $(n-1)!$ on the simplex with last coordinate one minus the sum of the others. For nonnegative integers a_j, iterated integration gives

$$\mathbb{E}\prod_j p_j^{a_j} = \frac{(n-1)!\prod_j a_j!}{(n-1+\sum_j a_j)!}. \tag{156}$$

For completeness, integrating one coordinate and rescaling the others reduces this formula inductively to $\int_0^1 t^a(1-t)^b\,dt = a!b!/(a+b+1)!$. The latter follows by a integrations by parts, starting with $\int_0^1 (1-t)^b\,dt = 1/(b+1)$. Consequently, for $S = \sum_j p_j^2$,

$$\mathbb{E}S = \frac{2}{n+1}, \qquad \mathbb{E}S^2 = \frac{4(n+5)}{(n+1)(n+2)(n+3)}.$$

$$\mathrm{Var}(S) = \frac{4(n-1)}{(n+1)^2(n+2)(n+3)} \le \frac{4}{n^3}.$$

Since $4/n - \mathbb{E}S \ge 2/n$, Chebyshev gives $\Pr\{S > 4/n\} \le 1/n$. Translating one simplex coordinate by ε and rescaling gives $\Pr\{p_j > \varepsilon\} = (1-\varepsilon)^{n-1}$ for $0 \le \varepsilon \le 1$. A union bound therefore gives

$$\Pr\left\{\max_j p_j > \frac{3\log n}{m}\right\} \le n^{-2}.$$

For $m \ge 2^{18}$, $3\log(m+1) \le \sqrt{m}$. The inequality holds at 2^{18}, and $\sqrt{m}/\log(m+1)$ increases when $\log(m+1) > 2m/(m+1)$, which holds in this range. Thus the event

$$G = \left\{\sum_j p_j^2 \le \frac{4}{n}, \quad \max_j p_j \le m^{-1/2}\right\} \tag{157}$$

has probability at least $1 - 2/n$. The learner need not identify this event.

S11.9.5 The rank transformation and proper prediction

Fix a reference sample in G. For any input law P with $\mathrm{KL}(P\|\mu) \le 1$, write $q_j = P(I_j)$. Each p_j is positive almost surely. Data processing gives $\mathrm{KL}(q\|p) \le 1$. Nonnegativity of KL relative to the probability vector proportional to $(p_j \exp(\log(np_j)))_j$ gives

$$\sum_j q_j \log(np_j) \le \mathrm{KL}(q\|p) + \log\left(n\sum_j p_j^2\right).$$

It follows that

$$\mathrm{KL}(q\|\,\mathrm{Unif}\{0,\dots,n-1\}) \le 2\,\mathrm{KL}(q\|p) + \log\left(n\sum_j p_j^2\right) \le 2 + \log 4 < 4. \tag{158}$$

Let $r(x) = j$ for $x \in I_j$. If U is independently uniform on $(0,1)$, then $w = (r(x)+U)/n$ has density nq_j on $(j/n,(j+1)/n)$. Its KL from the continuous uniform law is exactly the left side of (158).

Run the uniform-reference algorithm in Theorem S18.1 with supplied budget $4T$ on virtual examples (w_t, y_t). Before the current original input, draw its threshold τ using its fresh private randomness and

independently draw U. Put $k = \lceil n\tau - U \rceil$, clipped to $\{0, \dots, n\}$, and select the original threshold z_k. After receiving the input and label, form w and update the virtual learner. For $x \in I_j$,

$$h_{z_k}(x) = \mathbf{1}\{j \geq k\} = \mathbf{1}\{j + U \geq n\tau\} = h_\tau(w). \tag{159}$$

The values $k = 0, n$ give the constant predictions one and zero outside the null endpoint set. Any fixed convention on that set completes the algorithm.

Current-round privacy makes τ independent of the current original example and of U, conditional on the full past. Thus the virtual example does not depend on the virtual learner's fresh threshold. The full past may include every previous U, the reference sample, the original adversary state, and all preceding learner randomness. Equation (158) holds conditional on each such history. Conditioning on a smaller virtual history forms a mixture and cannot increase KL. The fresh U is virtual input-generating noise, not information supplied to the original adversary before its current example. Drawing it before x has the same joint law as drawing it afterward; the virtual learner does not use it to choose τ. Conditional on any fixed good reference sample, the virtual learner therefore has expected regret at most $C\sqrt{T \log(eT)}$ against all virtual thresholds.

S11.9.6 Uniform transfer of the hindsight comparator

Round any original threshold to an adjacent reference threshold. This changes predictions only inside its one reference gap, or nowhere if it is already a sample endpoint. The virtual threshold k/n agrees with h_{z_k} outside the null set, since $\mathbf{1}\{j + U \geq k\} = \mathbf{1}\{j \geq k\}$. Writing $N_j = \sum_t \mathbf{1}\{x_t \in I_j\}$, we obtain pathwise

$$\mathrm{Reg}_T^{\text{original}} \leq \mathrm{Reg}_T^{\text{virtual}} + \max_j N_j. \tag{160}$$

The maximum is needed because the best comparator may depend on all inputs and labels.

Conditional on the fixed good reference sample, the input-path law Q has $\mathrm{KL}(Q\|\mu^T) \leq T$. Indeed, the input-only conditional laws are mixtures of the full-history laws; KL convexity and chain rule give this bound. Under μ^T, for every $\lambda > 0$,

$$\mathbb{E}_{\mu^T} e^{\lambda \max_j N_j} \leq \sum_j [1 + p_j(e^\lambda - 1)]^T \leq n \exp\{T \max_j p_j(e^\lambda - 1)\}.$$

The variational KL inequality follows by comparing Q with the exponential tilt of μ^T; applying it to $\max_j N_j$ with $\lambda = (\log m)/4$ gives

$$\mathbb{E}_Q \max_j N_j \leq \frac{T + \log n + Tm^{-1/4}}{(\log m)/4} \leq \frac{8T}{\log m} + 8. \tag{161}$$

Here we used (157) and $\log(m+1) \leq 2 \log m$. On G^c, original regret is at most T. Averaging (160) and (161) proves

$$\mathbb{E}\,\mathrm{Reg}_T \leq C\sqrt{T \log(eT)} + \frac{8T}{\log m} + 8 + \frac{2T}{m+1} \qquad (m \geq 2^{18}).$$

For $m \geq T \geq 512$, $\log(eT(m+1)) \leq 3 \log m$ and $2T/(m+1) \leq 2$. This gives the upper in (149). If $T \leq m < 2^{18}$, use any fixed threshold without reference calls: $\log(eT(m+1)) < 27$, so the bound T is within a universal constant of the required expression.

S11.9.7 The uniform-reference statistical lower

We verify an explicit conditional-per-round-KL lower for the statistical term in (149). This is the ordered-threshold specialization of the binary-search epoch construction used in Theorem 2.1, followed by the reference-quantile transfer in Appendix S19.8.

Lemma S11.11. *For every integer $T \geq 512$, ordinary thresholds under the uniform reference on $[0, 1]$ have minimax expected regret at least*

$$\frac{\sqrt{T \log(eT)}}{256}$$

under conditional KL at most one. The bound permits unrestricted post-input prediction and unlimited iid reference draws.

Proof. First use the atomic ordered reference that places mass $3/4$ at 0 and mass $[2^{l+1}l(l+1)]^{-1}$ at each point

$$x_{l,v} = (2v+1)/2^l, \qquad 0 \le v < 2^{l-1}, \quad l \ge 1.$$

The masses sum to one, since the total level-l mass is $1/[4l(l+1)]$. Maintain a nested binary-search interval starting from $(0,1)$. At epoch l, its midpoint is a level-l atom. With probability $q_l = 1/(80l)$ present that midpoint and a fresh fair label; otherwise present 0 with label zero. All mixture coins and informative labels are fresh. At the end of the epoch retain the half agreeing with its informative majority, with a fixed rule at a tie. Use open dyadic cells: retain the lower half for majority label one and the upper half for majority label zero. Every threshold in the retained half realizes that decision at the midpoint. After finitely many epochs the final cell is nonempty, so one threshold realizes all these majority decisions. It is positive and therefore also realizes the filler labels at zero.

At every full preceding history the current midpoint and epoch are fixed. The conditional KL is at most

$$\log(4/3) + q_l \log(2^{l+1}l(l+1)) \le \log(4/3) + 3lq_l = \log(4/3) + 3/80 < 1.$$

The first term bounds the filler contribution, and dropping $q_l \log q_l \le 0$ bounds the informative contribution. The inequality $\log(2^{l+1}l(l+1)) \le 3l$ holds for $l \ge 1$: the bound $l(l+1) \le 2^{2l-1}$ follows by induction, since its successive ratio is $(l+2)/l \le 3 < 4$. Padding rounds can all use input zero with label zero, with KL $\log(4/3) < 1$.

If epoch l has n_l rounds, let S_l be its informative label imbalance and $v_l = n_l q_l$. Its independent increments take values $1, -1, 0$ with probabilities $q_l/2, q_l/2, 1-q_l$, so $\mathbb{E}S_l^2 = v_l$ and $\mathbb{E}S_l^4 \le v_l + 3v_l^2$. Hölder's inequality yields

$$\mathbb{E}|S_l| \ge \frac{(\mathbb{E}S_l^2)^{3/2}}{(\mathbb{E}S_l^4)^{1/2}} \ge \frac{1}{2}\min\{v_l, \sqrt{v_l}\}.$$

Every learner makes an expected half mistake on each informative example. The final comparator saves $|S_l|/2$ there, and loses nothing on filler or padding rounds. Thus

$$\mathbb{E}\,\mathrm{Reg}_T \ge \frac{1}{4}\sum_l \min\{n_l q_l, \sqrt{n_l q_l}\}. \tag{162}$$

For $T \ge 2^{32}$, put

$$L = \lfloor T^{1/4} \rfloor, \qquad H = \sum_{l=1}^{L} \frac{1}{l}, \qquad n_l = \left\lfloor \frac{T}{2Hl} \right\rfloor.$$

The epochs use at most $T/2$ rounds. Every quantity inside a floor is at least two, and

$$n_l q_l \ge \frac{T}{320Hl^2} \ge \frac{\sqrt{T}}{320H} \ge 1.$$

For these claims it suffices to use $H \le 1 + \frac{1}{4}\log T$. They hold at 2^{32} and persist since $\sqrt{T}/(1 + \frac{1}{4}\log T)$ increases. The floor arguments are larger still, since $l \le T^{1/4}$. Also $H \ge \log(L+1) > \frac{1}{4}\log T$. Therefore (162) gives

$$\mathbb{E}\,\mathrm{Reg}_T \ge \frac{\sqrt{TH}}{4\sqrt{320}} \ge \frac{\sqrt{T\log T}}{8\sqrt{320}}. \tag{163}$$

Now replace each atom by its ordered reference-quantile interval of length equal to its mass, as proved in Appendix S19.8. Present a fresh uniform point in the interval of the selected atom. This makes the reference exactly uniform on $[0,1]$ and preserves the likelihood ratio and KL at every full preceding history. The pullback of each ordered comparator is an ordinary threshold, so its losses are unchanged. The extra within-interval coordinate is independent of the fresh fair informative label. Thus (163) continues to hold for every post-input learner under the uniform reference.

For $512 \le T < 2^{32}$, use iid uniform inputs and independent fair labels. Thresholds 0 and 1 give the two constant predictions outside a null endpoint set. Their better cumulative loss is $(T - |S|)/2$, where S is a sum of T fair signs. Since $\mathbb{E}S^2 = T$ and $\mathbb{E}S^4 \le 3T^2$, Hölder gives $\mathbb{E}|S| \ge \sqrt{T/3}$.

Expected regret is at least $\sqrt{T}/(2\sqrt{3})$. The two horizon ranges imply the stated constant $1/256$: in the small range $\log(eT) < 24$, while in the large range $\log(eT) \le 2\log T$ and $8\sqrt{640} < 256$.

Reference draws in both uniform constructions have a known fixed law independent of the current fair labels. Unlimited such draws, requested at any adaptive times, provide only additional independent randomness. The same fair-label lower therefore applies with unrestricted reference access. The conditional KL checks already hold on the stronger full histories containing these samples. □

S11.9.8 The sample lower and expected reference calls

Proof of Theorem S11.10. The preceding empirical-tree and blockwise constructions give the hard-cap upper for all integer $m \ge 0$ and all Borel references. Lemma S11.7 gives, for every learner with at most m calls,

$$\frac{31T}{2048\log(eT(m+1))}$$

expected regret on a fixed atomless full-support reference with a fixed realizable target. It is an agnostic lower because the comparator has zero loss. The lemma allows adaptive reference requests after the current input and unrestricted post-input predictions. Lemma S11.11 independently gives the statistical term under a uniform reference, even with unlimited reference access. The worst-case supremum includes both witnesses. Their maximum is at least half their sum; the instances need not agree.

For a uniformly expected real allowance $m \ge 0$, run the hard-cap upper at $k = \lfloor m \rfloor$. Its deterministic reference count is at most $k \le m$. Since $k + 1 \ge (m+1)/2$,

$$\log(eT(k+1)) \ge \log(eT(m+1)) - \log 2.$$

For $T \ge 512$ these logarithms differ by at most a universal factor, including $m < 1$. For the expected-budget lower, (148) gives $T/[512\log(eT(m+1))]$ under a resource promise uniform over all admissible instances. That proof uses an autonomous truncated learner and a full-run coupling; it does not condition a prediction bound on future nonoverflow. The uniform statistical lower is unchanged. Taking the maximum again proves the expected-budget version. □

For any fixed target constant in regret $O(\sqrt{T\log(eT)})$, the necessary reference allowance is at least $\exp(c\sqrt{T/\log(eT)})/(eT) - 1$ for a positive constant depending on that target. Conversely, an allowance $\lceil \exp(C\sqrt{T/\log(eT)}) \rceil$ suffices for all sufficiently large horizons. This exponential scale concerns the minimax prediction problem over unrestricted reference densities; a regular parametric restriction can change the sample requirement. Exponential KL-to-accuracy sample costs already occur in approximate rejection sampling (Block and Polyanskiy, 2024). The all-Borel upper above removes the sample-range and atomlessness restrictions from the minimax statement. The retained randomized-rank construction still has its explicitly stated atomless $m \ge T$ scope. These sample bounds use exact real reference observations.

S11.10 The joint price of reference accuracy and samples

The sample model is the one in Section S11.8, with conditional input KL at most one. In addition, the learner is given a number $0 < \delta \le 1$ and the promise $\mathrm{TV}(\mu, U) \le \delta$, where U is uniform on $[0,1]$ in the supplied threshold coordinate. Samples, when requested, are from the true reference μ, not from an estimated law.

Theorem S11.12. *For integers $T \ge 512$, $m \ge 0$, and $0 < \delta \le 1$, put $s_\delta = \log(eT(m+1)/\delta)$. The minimax rates over Borel references satisfying the supplied TV promise, with at most m reference calls, are*

$$\textit{realizable mistakes:} \quad \Theta\left(\sqrt{T} + \frac{T}{s_\delta}\right), \tag{164}$$

$$\textit{agnostic regret:} \quad \Theta\left(\sqrt{T\log(eT)} + \frac{T}{s_\delta}\right). \tag{165}$$

The constants are universal. Both uppers are proper pre-input and request all samples before play; the realizable upper also permits adaptive pathwise realizability. Both lowers permit unrestricted

post-input prediction, adaptive post-input reference requests, and unlimited computation. They already hold on the smaller atomless full-support class satisfying all four supplied promises

$$\mu \geq (1-\delta)U, \qquad \mathrm{TV}(\mu, U) \leq \delta, \qquad \mathrm{KL}(\mu\|U) \leq \delta, \qquad \mathrm{KL}(U\|\mu) \leq \delta. \tag{166}$$

The same rates hold for real $m \geq 0$ if the expected number of calls is at most m uniformly over the announced instance class, including when that class is restricted by (166).

At $\delta = 0$ the reference is exactly known uniform, and the statistical terms alone give the rates. The following proofs treat positive δ. Quantization, entropy transfer, and sample-budget truncation are classical tools; the proof gives the dependence on the two supplied resources explicitly.

S11.10.1 Agnostic prediction from a uniform approximation

First suppose $0 < \delta \leq T^{-2}$ and write $A_T = \sqrt{T \log(eT)}$, $L = \log(1/\delta)$. Set $k = \lfloor 1/\delta \rfloor \geq T^2$ and use the partition

$$I_j = [j/k, (j+1)/k) \quad (0 \leq j < k-1), \qquad I_{k-1} = [(k-1)/k, 1].$$

This includes all atoms and endpoints. If $p_j = \mu(I_j)$, then $p_j \leq 1/k + \delta \leq 2/k$. For a conditional input law P, put $q_j = P(I_j)$. Finite KL implies $q_j = 0$ whenever $p_j = 0$. Data processing gives

$$\mathrm{KL}(q\|\,\mathrm{Unif}\{0,\dots,k-1\}) = \mathrm{KL}(q\|p) + \sum_{j:q_j>0} q_j \log(kp_j) \leq 1 + \log 2. \tag{167}$$

No unknown cell mass is evaluated by the learner.

Run the known-uniform proper threshold learner of Theorem S18.1 with supplied integer budget $b = \lceil (1+\log 2)T \rceil \leq 2T$. Let τ_t be its fresh private threshold. Independently draw a fresh private V_t uniform on $(0,1)$ and choose, before the actual input,

$$\theta_t = \frac{1}{k}\,\mathrm{clip}_{[0,k]}(\lceil k\tau_t - V_t \rceil)\,. \tag{168}$$

When $x_t \in I_j$, give the virtual learner $z_t = (j + V_t)/k$ and the actual label y_t. For every $x_t < 1$, including all left grid endpoints,

$$h_{\theta_t}(x_t) = \mathbf{1}\{j \geq \lceil k\tau_t - V_t \rceil\} = \mathbf{1}\{j + V_t \geq k\tau_t\} = h_{\tau_t}(z_t).$$

Clipping preserves these predictions. At $x_t = 1$ this identity can fail, since every original threshold predicts one there. We charge this exception below.

Conditional on the actual full past, the virtual input has density kq_j in cell j, so its KL to uniform is (167). The current V_t is virtual input noise, not part of its preceding conditioning. Both V_t and τ_t remain private through generation of the actual input and label. In particular, the label mechanism cannot disclose τ_t through z_t. Past coins may be exposed; marginalizing surplus past state preserves the KL bound by convexity. Thus the virtual learner has expected regret at most CA_T under the required current-round privacy.

Let $N_1 = \sum_t \mathbf{1}\{x_t = 1\}$ and $Z = \max_j \sum_t \mathbf{1}\{x_t \in I_j\}$. Write L_o and L_v for original and virtual cumulative losses. Pathwise,

$$L_o(\text{learner}) \leq L_v(\text{learner}) + N_1.$$

For each fixed grid comparator r/k, its original and virtual predictions also agree except possibly at $x = 1$. Hence

$$\min_{r=0,\dots,k} L_o(h_{r/k}) \geq \min_{r=0,\dots,k} L_v(h_{r/k}) - N_1 \geq \inf_{\tau\in[0,1]} L_v(h_\tau) - N_1.$$

Rounding any original comparator θ upward to $\lceil k\theta \rceil/k$ changes predictions only within one cell, and changes nothing when θ is already a grid point. This is simultaneous in θ, so

$$\inf_\theta L_o(h_\theta) \geq \min_r L_o(h_{r/k}) - Z.$$

Combining all three inequalities and using $N_1 \leq Z$ gives

$$\mathrm{Reg}_o \leq \mathrm{Reg}_v + Z + 2N_1 \leq \mathrm{Reg}_v + 3Z. \tag{169}$$

Both endpoint charges are included. There is no conditioning on the cell of a hindsight-selected comparator.

Under μ^T, a fixed cell count is binomial, and $p_j \le 2/k$ implies, for $\lambda > 0$,

$$\log \mathbb{E}_{\mu^T} e^{\lambda Z} \le \log k + \frac{2T}{k}(e^\lambda - 1).$$

The full-history input KL assumption and the chain rule give $\mathrm{KL}(\mathcal{L}(x_{1:T})\|\mu^T) \le T$: first condition each input kernel on only the input past and use convexity, then sum. Exponential tilting therefore yields

$$\lambda \mathbb{E} Z \le T + \log k + \frac{2T}{k}(e^\lambda - 1).$$

This follows directly from nonnegativity of KL to the probability proportional to $e^{\lambda Z}\mu^T; 0 \le Z \le T$ makes all terms finite. With $\lambda = (\log k)/2$, we obtain

$$\mathbb{E} Z \le \frac{2T}{\log k} + 2 + \frac{4T}{\sqrt{k}\log k} \le \frac{2T}{\log k} + 3.$$

Since $\log k \ge L - \log 2 \ge L/2$, this proves a zero-reference proper agnostic upper for every Borel reference:

$$\mathbb{E}\,\mathrm{Reg}_T \le C A_T + \frac{12T}{L} + 9. \tag{170}$$

S11.10.2 Realizable prediction from the approximation

Maintain the interval of thresholds consistent with past labels and predict with its coordinate midpoint. The interval is $[0, u]$ before a negative input occurs and $(l, u]$ afterward; if it is $\{0\}$, play zero. On a mistake its length at least halves. Correct updates cannot enlarge it. If M_{t-1} is the number of past mistakes, the disagreement set D_t of the version interval has Lebesgue measure at most $2^{-M_{t-1}}$. These are the exact measurable disagreement sets, including their appropriate endpoints. Thus total variation gives

$$\mu(D_t) \le 2^{-M_{t-1}} + \delta.$$

Continued pathwise realizability ensures that a mistake can occur only in D_t; no preselected target is needed for this upper.

Put $L_2 = L/\log 2$, $d_t = \min(M_{t-1}, L_2)$, and let e_t be the conditional probability of a mistake. For $d_t \ge 2$, binary data processing and $\mu(D_t) \le 2^{1-d_t}$ give

$$1 \ge \mathrm{kl}(P_t(D_t), \mu(D_t)) \ge e_t(d_t - 1)\log 2 - \log 2.$$

When $\mu(D_t) = 0$, finite KL gives $e_t = 0$ directly. Since $d_t/(d_t - 1) \le 2$, we have $d_t e_t < 5$. For $d_t < 2$ use $e_t \le 1$. In both cases $(d_t + 1)e_t \le 6$.

For integers $r \ge 0$, define $F(r) = \sum_{i=0}^{r-1}(\min(i, L_2) + 1)$. Its conditional expected increment is at most six, so $\mathbb{E}F(M_T) \le 6T$. The function $H(x) = \int_0^x \min(u, L_2)\,du$ satisfies $H(r) \le F(r)$ for integer r, by comparison on each unit interval. Moreover

$$x \le \sqrt{2H(x)} + \frac{H(x)}{L_2} \qquad (x \ge 0).$$

For $x \le L_2$ the square-root term equals x. Otherwise $H(x) = L_2 x - L_2^2/2$ and the two terms exceed $L_2 + (x - L_2/2)$. Jensen's inequality now yields

$$\mathbb{E} M_T \le \sqrt{12T} + \frac{6(\log 2)T}{L}. \tag{171}$$

This algorithm uses no unknown-reference median or reference samples.

S11.10.3 Combining accuracy and samples

Let A denote $\sqrt{T}$ in the realizable case and A_T in the agnostic case. For $0 < \delta \le T^{-2}$, the preceding algorithms give $C(A + T/L)$. Theorems S11.3 and S11.10 give $C'(A + T/s_0)$ for every $m \ge 0$, where $s_0 = \log(eT(m+1))$. Choose the algorithm associated with the larger of the known denominators L, s_0. Since

$$\frac{1}{\max(L, s_0)} \le \frac{2}{L + s_0} = \frac{2}{s_\delta},$$

the claimed upper follows. If $T^{-2} < \delta \le 1$, then $0 \le L < 2\log T \le 2s_0$, and hence $s_\delta \le 3s_0$. Ignore the approximation and apply the sample theorem. This proves the upper throughout the stated interval, including $\delta = 1$ without dividing by $\log(1/\delta)$.

For a real expected allowance m, use the hard-cap upper at $\lfloor m \rfloor$. It satisfies the resource promise deterministically. Since $\lfloor m \rfloor + 1 \ge (m+1)/2$, its denominator decreases by at most $\log 2$, changing only a universal constant.

S11.10.4 The matching lower under all four promises

Use the entire fixed small-interval hidden-path construction from Lemma S11.7, changing only its parameters to

$$\alpha = \frac{\delta}{64(m+1)(T+1)^2}, \qquad q = \frac{1}{32 s_\delta}. \tag{172}$$

We verify the parameter-dependent steps. Start the quarter path in the fixed interval $J_1 = [3/4, 3/4 + w]$, $w = (T+1)^{-2}$. Its middle-half bands A_j have lengths $w/(2 \cdot 4^{j-1})$; Q_j is uniform on A_j. The T fair path bits are selected before play, and the final target labels band j by bit j. Band j uses only preceding bits in its location. Define the fixed reference

$$\mu_b = (1-\alpha)U + \frac{\alpha}{T}\sum_{j=1}^{T} Q_j.$$

It is atomless and has full support. Convexity gives, for each individual bit sequence,

$$\begin{aligned}
&\mathrm{TV}(\mu_b, U) \le \alpha \le \delta,\\
&\mathrm{KL}(\mu_b \| U) \le \alpha[2\log(T+1) + T\log 2] \le 4\alpha(T+1) \le \delta,\\
&\mathrm{KL}(U \| \mu_b) \le -\log(1-\alpha) \le 2\alpha \le \delta,\\
&\mu_b \ge (1-\alpha)U \ge (1-\delta)U.
\end{aligned}$$

The average band logarithm uses $T^{-1}\sum_{j=1}^{T}(2j-1) = T$.

At current depth j the autonomous online law is $P_t = (1-q)U + qQ_j$, incrementing depth only on an active draw. As in (145), at every actual full history for the fixed reference and target,

$$\mathrm{KL}(P_t \| \mu_b) \le 2\alpha + q\log(T/\alpha).$$

Here $s_\delta > 7$ and

$$\log(T/\alpha) = s_\delta - 1 + \log 64 + 2\log(T+1) \le 4s_\delta.$$

Indeed $\log(T+1) \le s_\delta - 1 + 1/T$, and $\log 64 - 3 + 2/T < s_\delta$. Thus the conditional KL is at most $2\alpha + 1/8 \le 5/32 < 1$, including after reference observations.

Pre-generate all m potential reference-component coins and all T potential uniform background coordinates as in the old proof. Let E say that all reference coins select U and all potential background coordinates avoid the fixed J_1. It is independent of the path bits and active coins, and

$$\Pr(E) \ge 1 - m\alpha - Tw \ge 31/32.$$

Given E, the whole potential reference stream has common law U^m, independent of the bits. Background labels outside J_1 are also bit-independent. The conditional-fairness induction in Lemma S11.7 therefore applies without change, even to additional requests after the current coordinate. The next active label remains fair. There are still Binomial(T, q) active rounds given E. Hence the prior-average expected mistakes are at least

$$\Pr(E)\frac{qT}{2} \ge \frac{31T}{2048 s_\delta}. \tag{173}$$

Finite-prior averaging selects a fixed hard reference, target, and autonomous process for each learner. The actual input KL calculation does not condition on E. Realizability makes this a regret lower as well as a mistake lower.

For a uniformly expected allowance, put $K = 2^{20}$ and $M = \lceil K(m+1)s_\delta \rceil$ and truncate at the attempted $(M+1)$-st query, using h_1 for each remaining prediction. Apply (173) at M and the same δ. The reference still belongs to the announced class, including its stronger four-promise version. The autonomous coupling from Corollary S11.8 and Markov's inequality give

$$\mathbb{E}M_T \geq \frac{31T}{2048\log(eT(M+1)/\delta)} - \frac{Tm}{M}.$$

Since $M+1 \leq 2K(m+1)s_\delta$, $s_\delta > 7$, $\log s_\delta \leq s_\delta$, and $\log(2K) = 21\log 2 < (9/4)s_\delta$, we have

$$\log(eT(M+1)/\delta) \leq 5s_\delta, \qquad \frac{m}{M} \leq \frac{1}{Ks_\delta}.$$

It follows that

$$\mathbb{E}M_T \geq \left(\frac{31}{10240} - 2^{-20}\right)\frac{T}{s_\delta} \geq \frac{T}{512s_\delta}. \tag{174}$$

Finally U itself satisfies all four promises for every δ. The known-uniform statistical lower bounds used in Theorems S11.3 and S11.10 remain valid with unrestricted independent uniform reference draws. The maximum of that witness and the corresponding hidden-reference witness is at least half their sum. They need not be the same instance. This proves both lower orders and completes Theorem S11.12.

S11.10.5 Interpretation and limits

At $\delta = 1$ the theorem recovers the sample-only law. At $m = 0$ it quantifies the accuracy needed to replace samples. Any fixed inverse-polynomial accuracy still leaves order $T/\log(eT)$. The logarithms of reciprocal accuracy and sample count substitute up to constant factors in this minimax statement.

The lower ranges over references with arbitrarily narrow density spikes and permits dependence on the horizon and allowance. Bounds under a fixed density-ratio ceiling, a regular parametric restriction, or a typical-data model require separate arguments. The approximation is uniform in the supplied coordinate. The randomized quantization upper uses exact real inputs; a finite-bit implementation requires specified encodings for those inputs and for δ.

S12 Persistent prediction loss under one fixed reference

We use a threshold path as in Appendix C of Blanchard (2025). An infinite nested path and its limiting threshold also appear in Appendix E.1 of Block and Polyanskiy (2024). Here polynomially decreasing reference masses conceal later path bits from accumulated reference samples. The reference, target and input generator remain fixed throughout one infinite run.

Let U be uniform on $[0,1]$ and $h_\theta(x) = \mathbf{1}\{x \geq \theta\}$ for $0 \leq \theta \leq 1$. The learner knows this ordinary threshold class, U, and the promise $\mathrm{TV}(\mu, U) \leq 1/4$. Its reference observations are iid unlabeled draws from μ. It receives no density, CDF, quantile, or conditional-sampling oracle. Write N_T for the number of reference answers through its Tth prediction, including preprocessing and interspersed requests. For predictions $v_t \in [0,1]$, put

$$L_T = \sum_{t=1}^{T} \ell(v_t, y_t), \qquad \ell(v,y) = y + (1-2y)v.$$

For binary predictions L_T is the actual number of mistakes. An autonomous input generator uses its fixed parameters, its own past state and fresh seeds, without using learner predictions or reference-request decisions.

Theorem S12.1. *Fix a finite real $p \geq 0$ and set $D = 2p + 10 + \log 64$. Consider one measurable anytime learner whose predictions are produced almost surely. The learner may know p. It may*

predict arbitrarily after observing the current input and may request reference samples before or after that input. Suppose that

$$\mathbb{E}N_T \le T^p \qquad \textit{for every integer } T \ge 1 \tag{175}$$

on every instance consisting of a fixed ordinary threshold realizing every label, a fixed atomless full-support reference μ with $\mathrm{TV}(\mu, U) \le 1/4$, and an input protocol satisfying $\mathrm{KL}(P_t\|\mu) \le 1$ at every full preceding history. The learner has no reference or target information beyond the announced promises, its online observations and these reference samples.

There exist one such reference μ, one ordinary threshold θ, and one autonomous infinite input generator, with $y_t = h_\theta(x_t)$, such that almost surely on this fixed instance,

$$L_T \ge \frac{T}{128D\log(eT)} \qquad \textit{for every sufficiently large integer } T. \tag{176}$$

In particular,

$$\liminf_{T\to\infty} \frac{\log(eT)}{T} L_T \ge \frac{1}{128D} \quad \textit{almost surely.}$$

For the same instance,

$$\mathbb{E}L_T \ge \frac{T}{256D\log(eT)} \qquad \textit{for every sufficiently large integer } T. \tag{177}$$

The reference and input laws can satisfy the stronger bounds

$$\begin{gathered} \mu \ge (31/32)U, \qquad \mathrm{TV}(\mu, U) \le 1/32, \\ \mathrm{KL}(\mu\|U) \le \frac{\log 4}{16} < \frac{1}{4}, \qquad \mathrm{KL}(U\|\mu) \le \frac{1}{16}, \\ \mathrm{KL}(P_t\|\mu) \le \frac{7}{8} \quad \textit{at every full preceding history.} \end{gathered} \tag{178}$$

The measure, target and generator may depend on the learner and p, but not on subsequently realized private randomness or a terminal horizon. The onset in (176) may be random. The onset in (177) is deterministic for the selected instance and learner. The target may be chosen irrational. The conclusions also hold under the stronger hard constraint $N_T \le \lfloor T^p \rfloor$ at every prefix on every outcome.

Proof. Fix the learner $\mathcal{A}$ and put $a = 2p + 10$, so $D = a + \log 64$. Give the infinite bit path $b = (b_j)_{j\ge1}$ independent fair coordinates for the proof prior. We first define the reference and generator for every fixed path.

The target and reference. Start with $V_0 = [3/4, 1]$. If $V_{j-1} = [v, v + w]$, let

$$A_j = (v + 3w/8, v + 5w/8).$$

For $b_j = 1$ retain $V_j = [v, v + w/4]$; for $b_j = 0$ retain $V_j = [v + 3w/4, v + w]$. The closed intervals are nested and their lengths tend to zero. Their intersection is the single point

$$\theta_b = \frac{3}{4} + 3\sum_{j\ge1}(1 - b_j)4^{-(j+1)}. \tag{179}$$

The bands are disjoint, $|A_j| = 4^{-(j+1)}$, and A_j depends only on $b_1, \dots, b_{j-1}$. Both retained quarters are separated from A_j. Consequently $h_{\theta_b}(x) = b_j$ for every $x \in A_j$. This includes the all-zero and all-one paths, whose targets are 1 and $3/4$, respectively, and all retained limit endpoints.

Let Q_j^b be uniform on A_j, and let Q_0 be uniform on $(0, 1/2)$, where every such target is zero. Define

$$\beta_j = \frac{j^{-a}}{64}, \qquad \beta = \sum_{j\ge1}\beta_j, \qquad \mu_b = (1 - \beta)U + \sum_{j\ge1}\beta_j Q_j^b. \tag{180}$$

Since $a \ge 10$ and $\sum_{j\ge1} j^{-2} \le 1 + \int_1^\infty x^{-2}\,dx = 2$, we have $\beta \le 1/32$. The density

$$r_b(x) = 1 - \beta + \sum_{j\ge1}\frac{\beta_j}{|A_j|}\mathbf{1}_{A_j}(x)$$

is jointly measurable in b, x. At each x at most one band term is nonzero. It integrates to one and is at least $31/32$. Thus μ_b is atomless, has full support and is a measurable probability kernel in b. Writing the spike part as β times a probability measure gives $\mathrm{TV}(\mu_b, U) \le \beta \le 1/32$.

To bound its forward KL, let $\phi(u) = u \log u - u + 1$, with $\phi(0) = 1$. This function is nonnegative and convex. Countable Jensen applied to the mixture of densities, followed by Tonelli, gives

$$\begin{aligned}\mathrm{KL}(\mu_b \| U) &\le \sum_{j \ge 1} \beta_j \, \mathrm{KL}(Q_j^b \| U) \\ &= \frac{\log 4}{64} \sum_{j \ge 1} (j+1) j^{-a} \le \frac{\log 4}{16}.\end{aligned}$$

Both sums $\sum j^{1-a}$ and $\sum j^{-a}$ in the last expression are at most two. The nonnegative ϕ integrand justifies the countable passage. Also,

$$\mathrm{KL}(U \| \mu_b) \le -\log(1-\beta) \le 2\beta \le \frac{1}{16}.$$

Here $-\log(1-\beta) = \int_0^\beta (1-s)^{-1} \, ds \le 2\beta$, since $\beta \le 1/2$. These bounds hold for every path before its selection.

An autonomous legal input process. At stage $j \ge 1$, let

$$q_j = \frac{1}{8(\log 64 + a \log(ej))}, \qquad P_j^b = (1 - q_j) Q_0 + q_j Q_j^b. \tag{181}$$

The probabilities q_j are positive, at most $1/(8D) < 1$, and nonincreasing. Start at stage one. At each round use a fresh activation coin of probability q_j. An inactive round draws from Q_0, has label zero, and leaves the stage unchanged. An active round draws from Q_j^b, has label b_j, and advances to stage $j+1$ after the prediction. All labels are realized by θ_b. The activation and within-component coordinate seeds are independent of the oracle and learner seeds. The generator uses neither predictions nor request decisions.

On the neutral interval $r_b = 1 - \beta$, so

$$\mathrm{KL}(Q_0 \| \mu_b) = \log \frac{2}{1-\beta} \le \log(64/31) < \frac{3}{4}.$$

The domination $\mu_b \ge \beta_j Q_j^b$ gives $\mathrm{KL}(Q_j^b \| \mu_b) \le \log 64 + a \log j$. By convexity in the input law,

$$\mathrm{KL}(P_j^b \| \mu_b) \le \frac{3}{4} + q_j(\log 64 + a \log j) \le \frac{3}{4} + \frac{1}{8} = \frac{7}{8}. \tag{182}$$

For fixed b, fresh input seeds give exactly this stage law conditional on the actual full preceding history, including the activation state, preceding learner randomness and reference answers. Additional reference calls before the input do not change its fresh law. Calls after the input do not change the law that generated it. If some past state is unobserved, averaging its conditional stage laws preserves (182) by convexity. This proves (178) in the original fixed-instance law.

One infinite oracle stream and finite capped procedures. Choose independent component indices C_i, independent of the entire bit path and all input and learner seeds, with

$$\Pr(C_i = 0) = 1 - \beta, \qquad \Pr(C_i = j) = \beta_j \quad (j \ge 1).$$

Give each index an independent uniform coordinate seed W_i. For index zero the answer is a U draw; for index $j \ge 1$ map W_i affinely into A_j. Null endpoint seeds may be mapped to the midpoint. For each fixed b this is an infinite iid μ_b list, independent of the online and learner seeds. Every reference request returns its next unused member. Thus it implements iid answers at arbitrary adaptive request times. All paths, input seeds, oracle seeds and learner seeds are chosen on one common product space. The maps defining a finite run are measurable, by the measurable interval endpoints, mixture kernel and learner decision kernels.

For analysis at a fixed integer $T \ge 2$, set

$$M = \lfloor T^{p+2} \rfloor, \qquad K = \lfloor \sqrt{T} \rfloor.$$

Define $\mathcal{A}^{[T]}$ by simulating $\mathcal{A}$ until an attempted $(M+1)$st reference request before its Tth prediction. At that attempt it abandons the simulation, makes no further requests, and predicts zero on the current prediction if it is still due and on every subsequent round. The capped procedure is almost surely completed and uses at most M answers on every execution. Pointwise termination on original nontermination null executions is not needed. They remain null for each constructed fixed instance and under the joint prior by integration.

The procedure $\mathcal{A}^{[T]}$ need not satisfy any smaller cap at earlier prefixes. We use only its total cap through this fixed horizon. Its dependence on T supplies no unknown path information. The original learner $\mathcal{A}$ remains one unchanged anytime policy.

Let

$$G_T = \{C_i \le K \text{ for all } 1 \le i \le M\}. \tag{183}$$

This event depends only on component indices and is independent of the entire bit path, all activations and all learner seeds. Moreover, $\Pr(G_T) \ge (1-\beta)^M > 0$. For $T \ge 4$, $K \ge \sqrt{T}/2$, so

$$\Pr(G_T^c) \le M \sum_{j>K} \beta_j \le \frac{T^{p+2} K^{1-a}}{64(a-1)} \le C_p T^{-5/2}, \tag{184}$$

$$C_p = \frac{2^{a-1}}{64(a-1)}.$$

The integral bound $\sum_{j>K} j^{-a} \le K^{1-a}/(a-1)$ gives the middle inequality; $a = 2p+10$ gives the last exponent. On G_T, every one of the first M potential reference positions depends only on $b_1, \dots, b_K$ and independent sample seeds. This holds for the entire list, including unrequested entries.

The finite-horizon information filtration. Fix T and condition on G_T for this paragraph. Reveal to a proof observer the first K bits, the component and coordinate seeds of all M potential answers, and the learner's independent private seed stream. Such a stream realizes its standard Borel decision kernels. This initial information determines every reference answer available to $\mathcal{A}^{[T]}$. The remaining bits are still iid fair and independent of the fresh online seeds. The original strategy is unchanged by this proof disclosure.

Let I_t indicate activation and $J_t = \sum_{s \le t} I_s$. The observer adds the activation and coordinate seeds used through round t, together with the path bits used by active labels. Its revealed prefix has length $\max\{K, J_t\}$. These variables form an increasing filtration. The actual history of $\mathcal{A}^{[T]}$ is measurable in it: each reference request returns the next predetermined list entry, and each preceding online label is a known prefix bit or zero. This follows by induction through the finite requests and predictions.

Before the current label, reveal the fresh activation and coordinate seed and then the learner's prediction. If the round is active at an index $j = 1 + J_{t-1} > K$, its current band uses only $b_1, \dots, b_{j-1}$, already known to the observer. The current input therefore reveals no b_j. All further post-input requests use the same finite list. The truncation decision and the default prediction, if needed, are also measurable in this information. Thus b_j remains fair immediately before its label is exposed, including after an adaptive number of post-input requests. Almost-sure completion suffices for these measurable predictions; conditioning on the positive-probability event G_T preserves the joint null sets.

This filtration does not reveal future mapped online locations, which can depend on unrevealed bits. It does not condition on the complete path or use oracle entries beyond M. It is used only through time T. The legality assertion (182) concerns the original fixed-instance law, separately from this observer law.

Loss and activation tails. Count only active indices greater than K. Let D_t be their number through t, and let B_t be their cumulative loss for $\mathcal{A}^{[T]}$. For any prediction $v \in [0,1]$, the preceding fairness gives

$$\mathbb{E}\Big[2^{-\ell(v,b_j)} \,\Big|\, \text{information before its label}\Big] = \frac{2^{-v} + 2^{-(1-v)}}{2} \le \frac{3}{4}.$$

The function on the right is convex in v, so it lies below the chord between its equal endpoint values $3/4$. Consequently

$$Z_t = 2^{D_t/3 - B_t}$$

is a nonnegative supermartingale through T, conditional on G_T. Its expected multiplier on a counted activation is at most $2^{1/3}(3/4) < 1$, because its cube is $54/64$. Otherwise its multiplier is one. Since $Z_0 = 1$, $\mathbb{E}[Z_T \mid G_T] \le 1$. For every real $z > 0$, Markov's inequality gives

$$\Pr\{B_T \le D_T/4,\ D_T \ge z \mid G_T\} \le 2^{-z/12}. \tag{185}$$

Indeed the event implies $Z_T \ge 2^{z/12}$.

Use independent uniform activation seeds ξ_t and put $u_T = Tq_T$. The stage at round t is $1 + J_{t-1} \le t \le T$. Since q_j decreases,

$$J_T \ge \sum_{t=1}^{T} \mathbf{1}\{\xi_t \le q_T\} \quad \text{on every path.}$$

The smaller count is binomial with mean u_T, also conditional on G_T, which is independent of its seeds. Exponential Markov at $\log 2$ yields

$$\begin{aligned}\Pr\{J_T < u_T/2 \mid G_T\} &\le e^{(\log 2)u_T/2}(1 - q_T/2)^T \\ &\le e^{-(1-\log 2)u_T/2}.\end{aligned} \tag{186}$$

No independence of the actual stage activations is required.

From (181),

$$q_T \ge \frac{1}{8D\log(eT)}.$$

For all $T \ge \lceil e^{64D} \rceil$, we have $\sqrt{T} \ge 32D\log(eT)$ and hence $K \le u_T/4$. At $T = e^{64D}$ the first inequality follows from $e^{32D} \ge (32D)^3/6 \ge 32D(1 + 64D)$ for $D \ge 1$. The ratio $\sqrt{T}/\log(eT)$ increases thereafter, since its derivative has the sign of $\log T - 1$. Rounding the threshold upward preserves the inequality.

On $J_T \ge u_T/2$, the number of counted activations is $D_T = (J_T - K)_+ \ge u_T/4$. Total loss $L_T(\mathcal{A}^{[T]})$ dominates B_T. If $L_T(\mathcal{A}^{[T]}) < u_T/16$, then $B_T < D_T/4$. Apply (185) with $z = u_T/4$, combine (186) and (184), and remove the conditioning:

$$\Pr\{L_T(\mathcal{A}^{[T]}) < u_T/16\} \le C_p T^{-5/2} + e^{-(1-\log 2)u_T/2} + 2^{-u_T/48}. \tag{187}$$

This calculation uses the procedure's total cap through the fixed horizon.

The original learner and one fixed path. Couple $\mathcal{A}$ and $\mathcal{A}^{[T]}$ with the same complete online input and label path, oracle stream and private seeds. The online path remains identical after divergence because its generator ignores predictions and reference-request decisions. If $N_T(\mathcal{A}) \le M$, truncation does not occur before prediction T, and all predictions and losses agree through that horizon. Therefore

$$\{L_T(\mathcal{A}) < u_T/16\} \subseteq \{L_T(\mathcal{A}^{[T]}) < u_T/16\} \cup \{N_T(\mathcal{A}) > M\}.$$

Every fixed μ_b, θ_b and its process is an admissible instance in (175). Integrating $\mathbb{E}[N_T(\mathcal{A}) \mid b] \le T^p$ over the fair-bit prior, then using $M \ge T^{p+2}/2$ for $T \ge 2$, gives

$$\Pr\{N_T(\mathcal{A}) > M\} \le \frac{\mathbb{E}N_T(\mathcal{A})}{M} \le \frac{2}{T^2}. \tag{188}$$

This count event is not a conditioning event in the fair-bit proof. Its dependence on labels and reference observations is immaterial to the union bound.

Combining (187) and (188) proves a summable bound for the failure probabilities of the original learner at every integer T. The polynomial terms are summable. For the exponential terms, use $u_T \ge T/[8D\log(eT)]$ and $T/\log(eT) \ge \sqrt{T}$ eventually. For any $c > 0$, grouping integers with $k^2 \le T < (k+1)^2$ bounds $\sum_T e^{-c\sqrt{T}}$ by a finite initial sum plus $\sum_k (2k+1)e^{-ck} < \infty$.

All original losses and counts are on the same infinite product space; the procedures $\mathcal{A}^{[T]}$ do not restart that run. The first Borel–Cantelli lemma therefore gives, jointly over the path and actual seeds, probability one that

$$L_T(\mathcal{A}) \geq u_T/16 \geq \frac{T}{128D\log(eT)}$$

for every sufficiently large integer T. Independence between the horizon-indexed events is not required. Fubini shows that this eventual event has conditional probability one for almost every fixed path b. Choose one such path. The choice depends on the learner's law and p, not on actual learner, oracle or input randomness. Its already defined measure, target and generator are then used at all horizons. This proves (176).

Different paths give different targets: at their first differing bit the retained closed quarters have a positive gap. Each individual bit path has prior probability zero. Thus the paths with rational targets form a null set, and the selection can avoid them.

For the selected path, the indicators of $\{L_T \geq T/[128D\log(eT)]\}$ converge to one almost surely. Bounded convergence makes their probabilities at least $1/2$ for all sufficiently large deterministic T. Nonnegativity of loss gives (177). Finally the hard constraint in the statement implies (175), so it is a special case. □

The target in Theorem S12.1 realizes the whole infinite run. Rational thresholds have the same hindsight comparison at every finite prefix, with a different quantifier on the comparator.

Lemma S12.2. *On every finite input list in $[0,1]$, ordinary and rational thresholds, including endpoint parameters, have identical prediction traces. Thus their hindsight loss infima agree for every finite labeled list. However, one fixed rational realizing target cannot replace the ordinary target in Theorem S12.1 for every learner.*

Proof. Fix an ordinary threshold trace. If both prediction sides are nonempty, let s be the largest input assigned zero and t the smallest input assigned one. Then $s < \theta \leq t$, so a rational parameter in (s,t) has the same trace. If the zero side is empty, parameter zero works. If the one side is empty on a nonempty list, its maximum s satisfies $s < \theta \leq 1$, and a rational parameter in $(s,1]$ works. For an empty list choose zero. Repeated inputs cannot receive different threshold predictions. These cases include inputs zero and one; an all-zero trace on a list containing one is impossible for both classes. The reverse trace inclusion follows because rational thresholds are ordinary thresholds. Equality of their prediction vectors gives equality of all finite hindsight loss infima, regardless of the labels.

For the obstruction, enumerate the rational thresholds as $h_1, h_2, \ldots$. After every finite labeled history, choose the least-index hypothesis consistent with the preceding labels and predict it on the next input. If none exists, use any fixed prediction. Consistency of a fixed index is a measurable condition on a finite history, and taking the least qualifying index is measurable. On a sequence realized by h_k, a qualifying index at most k always exists. Every mistake eliminates a previously chosen index smaller than k; h_k is never eliminated and never makes a mistake. Thus the learner makes at most $k-1$ mistakes on that entire run, using no reference samples. This disproves a diverging lower for every fixed rational realizing target. For the ordinary target in the theorem, a zero-loss rational comparator exists at each finite prefix by the first part, but it may change with the prefix. □

Proposition S12.3. *Every reference μ_b in (180) satisfies $D_{1+\eta}(\mu_b\|U) = \infty$ for each $\eta > 0$.*

Proof. Disjointness of the bands and $\mu_b \geq \beta_j Q_j^b$ give

$$\int r_b^{1+\eta}\, dU \geq \sum_{j\geq 1} \beta_j^{1+\eta} |A_j|^{-\eta} = 64^{-(1+\eta)} \sum_{j\geq 1} j^{-a(1+\eta)} 4^{\eta(j+1)} = \infty.$$

The summands do not tend to zero, since their exponential factor dominates their polynomial denominator. The definition $D_{1+\eta}(\mu_b\|U) = \eta^{-1}\log\int r_b^{1+\eta}\, dU$ proves the assertion. □

The persistent lower is sublinear. It concerns a fixed reference that may depend on the learner and the chosen polynomial exponent, and it has no positive forward moment promise. Its random almost-sure onset does not give a uniform finite-horizon bound over all such instances.

S13 Further reference-access laws

This section gives the joint sample-and-moment law and the persistence result for the sample/TV lower bound. The complete proofs are given in the indicated appendices.

S13.1 Persistence on one reference

Fix $p \geq 0$ and an anytime learner whose cumulative reference count satisfies $\mathbb{E}N_T \leq T^p$ at every prefix, uniformly over the announced TV family. At fixed accuracy $\delta = 1/4$, Theorem S12.1 selects one atomless full-support reference, one ordinary real target threshold, and one autonomous infinite process for this learner. The process obeys the same full-history KL bound, and the learner's cumulative mistake loss is at least $c_p T/\log(eT)$ eventually almost surely, for some $c_p > 0$; its expected loss has the same lower order. The instance may depend on the learner and p. The fixed target need not be rational, although rational thresholds have zero hindsight loss on every finite prefix.

This persistence statement fixes a positive TV accuracy: one fixed μ could satisfy $\mathrm{TV}(\mu, U) \leq \delta_T \to 0$ only if $\mu = U$. The joint sample-and-moment witnesses still depend on the horizon and allowance; the persistent reference has $D_{1+\eta}(\mu\|U) = \infty$ for every $\eta > 0$. In these information lower bounds, the reference encodes information about the target path. The additional penalty concerns recovery from samples and an approximation; with the reference known, Theorem 2.1 gives the statistical rate. These reference-information constructions permit arbitrarily fine density spikes; they do not establish the same information penalty for regular parametric reference families.

S13.2 Reference samples and forward moments

For $m \geq 0$ and $0 < \eta \leq 1$, put

$$S = \log(eT(m+1)), \qquad G_{T,m}(\eta) = \min\left\{\frac{T}{S}, \sqrt{\frac{T}{\eta}\log\left(e + \frac{T\eta}{S^2}\right)}\right\}. \tag{189}$$

Theorem S13.1. *There is a universal T_0 such that for every fixed positive integer d there are $c_d, C_d > 0$ with the following property. For all integers $T \geq \max\{d, T_0\}$ and $m \geq 0$, and every $0 < \eta \leq 1$: For every known nonempty countable measurable binary class $\mathcal{F}$ with $\mathrm{VC}(\mathcal{F}) \leq d$ and every supplied Borel probability ν, there is a measurable proper pre-input strategy depending only on $T, d, m, \eta, \mathcal{F}, \nu$ and observed history. For every fixed Borel reference μ satisfying $D_{1+\eta}(\mu\|\nu) \leq 1$ and every input-label protocol obeying $\mathrm{KL}(P_t\|\mu) \leq 1$ at every full preceding history and the current-round privacy convention, this strategy takes a deterministic number of at most m iid μ samples before play and satisfies*

$$\mathbb{E}\,\mathrm{Reg}_T \leq C_d\big(A_{d,T} + G_{T,m}(\eta)\big).$$

In the worst case over classes and supplied measures, every unrestricted learner has expected regret at least c_d times the same expression. The lower already uses rational thresholds, uniform ν, and a fixed atomless full-support μ. It permits adaptive post-input reference requests and unlimited computation. The same order holds for finite real uniformly expected call allowances $m \geq 0$.

The expected allowance is uniform over the announced moment family at the same T and η. The upper permits atoms and gives a measurable strategy. The rate is sharp for fixed d. The lower-bound reference is chosen before play and may depend on T, m, η and the learner. Theorem S17.7 gives the complete statement with the sufficient cutoff $\log T \geq 256$.

For every fixed $\eta > 0$, no samples are needed to recover the known-reference order $\Theta_{d,\eta}(A_{d,T})$. For each fixed $p \geq 0$, restricting $m \leq T^p$ leaves the same worst-case horizon order as $m = 0$, uniformly over $0 < \eta \leq 1$, with constants depending on d, p. Indeed, $\log(eT) \leq S \leq (p+2)\log(eT)$, and Lemma S17.5 compares the corresponding rates. An order improvement from using both sources therefore requires more than polynomially many samples. The following discussion gives one such regime and explains the transition as η tends to zero.

The information term depends on the dimensionless parameter $w = T\eta/S^2$:

$$G_{T,m}(\eta) = \frac{T}{S}H(w), \qquad H(w) = \min\{1, \sqrt{\log(e+w)/w}\}.$$

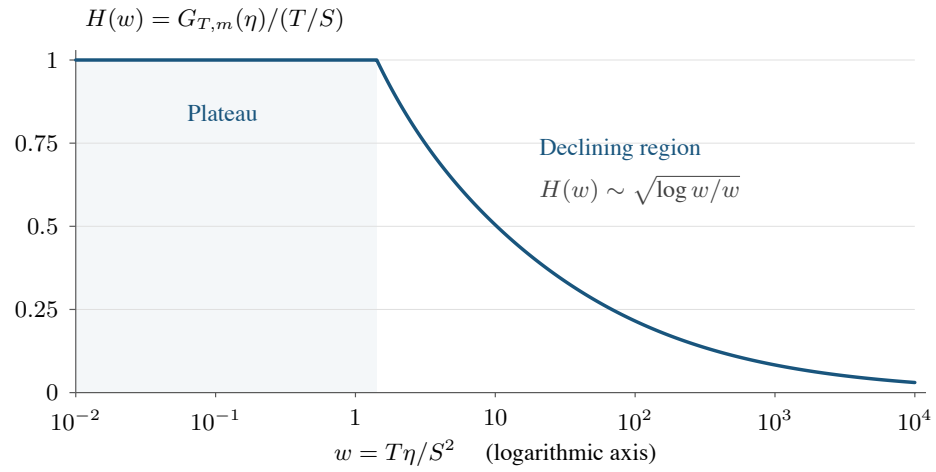

Figure 1: The normalized information term in Theorem S13.1. The full rate is $A_{d,T} + (T/S)H(w)$, up to constants depending on fixed d. The curve depicts the displayed rate formula, with the separate statistical floor omitted. Changing the sample allowance changes both w and the normalization T/S.

Figure 1 shows this normalized term. For $w \le 1$, the rate is $\Theta_d(A_{d,T} + T/S)$; as $w \to \infty$, $H(w) \sim \sqrt{\log w/w}$. With no samples, $S = \log(eT)$ and the statistical floor is absorbed into $G_{T,0}(\eta)$ at fixed d. The case $m = 0, d = 1$ of Corollary S17.7 gives the universal-constant threshold law $\Theta(G_{T,0}(\eta))$ for $\log T \ge 256$ and $0 < \eta \le 1$.

The order gain uses superpolynomial samples: as $n \to \infty$, take $R = 1, T = 2^n, m = 2^{n^2}, s = \log T$ and $\eta = S^2 \log s/T \to 0$. The joint rate is

$$\frac{T}{S}\sqrt{\frac{\log\log s}{\log s}},$$

while the sample-only and zero-sample moment guarantees are both $\Theta(T/S)$. Here $S \sim s^2/\log 2$ and the statistical floor is negligible.

When both sources are used, the learner first builds a finite tree of empirical disagreements at resolution $u \asymp S$. It then refines each cell using successive covers under the known measure ν, while retaining the sample trace of every function. This gives nested cells and proper representatives. The sample bound controls true-reference disagreement at the join by e^{-u}, hence input disagreement by $2/u$. The first ν-cover has resolution of order u/η. Joining the two parts costs $O_d(\sqrt{T/\eta})$, and the finer resolutions share total ν-KL at most $3T/\eta$. There are $O(\log(e + T\eta/u^2))$ such resolutions. Bounds on changes of the learning rates control the cost of joining the groups of levels. A uniform disagreement bound allows comparison with the best hypothesis in hindsight after the reference samples are observed. Appendix S16 proves the construction for every sample allowance and moment order in the theorem.

For the lower, a reference fixed before play places weight $e^{-4\eta j}/[64(m+1)j^3]$ on the band at depth j of one threshold path. With probability at least $31/32$, every potential reference request selects the uniform component and reveals no path bit. Noisy labels at several depths give the logarithm in (189). The complementary event has nonnegative expected excess loss relative to the fixed comparator, so it can be discarded. Appendix S17 proves the lower and extends it to expected sample allowances by truncating under the same announced moment constraint.

S14 Forward reference regularity

The reference-information lower permits narrow density spikes. This appendix gives a sufficient moment condition that removes its worst-case penalty and a quantitative lower when that condi-

tion weakens. The change-of-reference inequality is classical (Altschuler and Chewi, 2026, Appendix A.1). The lower adapts the fixed hidden-path construction in Appendix S11.10.

For $\alpha > 1$, define

$$D_\alpha(\mu\|\nu) = \frac{1}{\alpha - 1} \log \int \left(\frac{d\mu}{d\nu}\right)^\alpha d\nu,$$

with value infinity if $\mu \not\ll \nu$. Also set $D_\infty(\nu\|\mu) = \log \|d\nu/d\mu\|_{L^\infty(\mu)}$ when $\nu \ll \mu$.

Lemma S14.1. *For probability measures P, μ, ν and $\alpha > 1$,*

$$\mathrm{KL}(P\|\nu) \le \frac{\alpha}{\alpha - 1} \mathrm{KL}(P\|\mu) + D_\alpha(\mu\|\nu).$$

Proof. Suppose the right side is finite and put $a = \alpha/(\alpha - 1)$. For any bounded measurable real function f, the relative-entropy variational inequality and Hölder's inequality give

$$Pf \le a\,\mathrm{KL}(P\|\mu) + a \log \mu(e^{f/a}),$$
$$\mu(e^{f/a}) \le \left(\nu\left[(d\mu/d\nu)^\alpha\right]\right)^{1/\alpha} \left(\nu(e^f)\right)^{1/a}.$$

Hence $Pf - \log \nu(e^f) \le a\,\mathrm{KL}(P\|\mu) + D_\alpha(\mu\|\nu)$. The supremum on the left over bounded measurable f is $\mathrm{KL}(P\|\nu)$, proving the result. For infinite right side the inequality is immediate. This bounded-test argument does not decompose potentially infinite likelihood-ratio expectations. □

At $\alpha = 2$ this is the KL weak triangle inequality $\mathrm{KL}(P\|\nu) \le 2\,\mathrm{KL}(P\|\mu) + \log(1 + \chi^2(\mu\|\nu))$ of Altschuler and Chewi (2026, Corollary A.3).

Proposition S14.2. *In the full-history model of Section 2, suppose the learner knows a mathematical measure ν, $\alpha > 1$ and $R \ge 0$, and the fixed unknown reference satisfies $D_\alpha(\mu\|\nu) \le R$. For every known nonempty countable measurable binary class of VC dimension at most d, where $1 \le d \le T$, a proper pre-input strategy using no true-reference samples satisfies*

$$\mathbb{E}\,\mathrm{Reg}_T \le C \min\left\{T, \Phi_{d,T}\left(\frac{\alpha B}{\alpha - 1} + TR\right)\right\}$$

whenever $\mathbb{E}\sum_t \mathrm{KL}(P_t\|\mu) \le B$. The strategy does not depend on μ.

Proof. Apply Lemma S14.1 at every full history, sum its nonnegative costs, and take expectations. The total conditional KL relative to the known ν is at most $\alpha B/(\alpha - 1) + TR$. Theorem S6.1 supplies the proper strategy for that announced budget and reference, uniformly over the admissible μ. A fixed hypothesis gives the cap T. Choose between these announced guarantees before play. □

For conditional KL one and fixed $d, \alpha > 1, R$, this gives $O_{d,\alpha,R}(\sqrt{T \log(eT)})$ regret without reference samples. The worst-case lower of that order follows with $\mu = \nu = U$ from Lemma S11.11 and the rational-trace transfer in Appendix S10. This is a worst-case statement over classes and supplied references. If $\mu \le M\nu$ for known $M \ge 1$, the same bounded-test argument gives the sharper effective budget $B + T \log M$. Thus a fixed density ceiling is also sufficient.

The direction of the moment promise matters. For the existing accuracy-lower mixture $\mu = (1 - \beta)U + (\beta/T)\sum_j Q_j$ with $\beta = \delta/[64(m + 1)(T + 1)^2]$, its positive uniform component gives

$$D_\infty(U\|\mu) \le -\log(1 - \beta) \le 2\beta \le \delta.$$

That lower already has $\mathrm{KL}(\mu\|U) \le \delta$ and $\mathrm{TV}(\mu, U) \le \delta$. Reverse infinite-order control therefore does not substitute for forward order greater than one.

S14.1 A lower as the forward order approaches one

Theorem S14.3. *Let $T \ge 512$ and $m \ge 0$ be integers, $0 < \delta \le 1$, and $0 < \eta \le 1$. Put $s = \log(eT(m+1)/\delta)$. For rational thresholds on $[0, 1]$ with known supplied U, the worst-case realizable mistakes are at least*

$$c \min\{T/s, \sqrt{T/\eta}\}$$

under conditional KL one and the promises

$$D_{1+\eta}(\mu\|U) \le 1, \qquad \mathrm{TV}(\mu, U) \le \delta, \qquad \mathrm{KL}(\mu\|U) \le \delta, \qquad D_\infty(U\|\mu) \le \delta.$$

The reference is atomless, has full support and is fixed before play. The lower permits unrestricted post-input predictions, unlimited computation and at most m adaptive iid reference requests, including after a current input. The corresponding agnostic lower is

$$c' \left[\sqrt{T \log(eT)} + \min\{T/s, \sqrt{T/\eta}\} \right].$$

The constants are universal. The request allowance is a hard cap.

Proof. Set $M = \min\{T/s, \sqrt{T/\eta}\}$. First suppose $s \le T/2$, and put $n = \lfloor M \rfloor$. Then $M \ge 2$ and $M/2 \le n \le M$. Let $w = (T+1)^{-2}$ and $J_0 = [1/4, 1/4 + w]$. Choose independent fair bits $b_1, \dots, b_n$ before play. At depth j, let Q_j be uniform on the middle half of J_{j-1}. Retain the left quarter of J_{j-1} as J_j if $b_j = 0$ and its right quarter if $b_j = 1$. Choose the rational target θ at the midpoint of J_n. The active band depends only on preceding bits and has label $1 - b_j$. The bands are disjoint, and labels outside J_0 do not depend on the path.

For each fixed path define

$$S = \frac{1}{n} \sum_{j=1}^{n} Q_j, \qquad H = n \log 4 + 2 \log(T+1), \qquad z = \frac{\eta H}{1+\eta},$$

$$\beta = \frac{\delta e^{-z}}{64(m+1)(H+1)}, \qquad \mu = (1-\beta)U + \beta S.$$

The density $h = dS/dU$ satisfies $h \le e^H$ and $Uh = 1$, so $\|h\|_{1+\eta} \le e^z$. Minkowski's inequality gives

$$\begin{aligned} \|d\mu/dU\|_{1+\eta} &\le 1 - \beta + \beta e^z = 1 + \frac{\delta(1 - e^{-z})}{64(m+1)(H+1)} \\ &\le 1 + \frac{\delta \eta}{64(m+1)(1+\eta)}. \end{aligned}$$

Taking the logarithm and multiplying by $(1+\eta)/\eta$ yields $D_{1+\eta}(\mu\|U) \le \delta/[64(m+1)] \le 1$. Convexity gives $\mathrm{KL}(\mu\|U) \le \beta H \le \delta/64$. Also $\mathrm{TV}(\mu, U) \le \beta$ and $D_\infty(U\|\mu) \le 2\beta \le \delta$. The positive uniform component ensures full support, and the finite mixture is atomless. These conclusions hold for every fixed path.

Put $K = \log(n/\beta)$ and $q = 1/(8K)$. While fewer than n active draws have occurred, the next input law is $(1-q)U + qQ_j$, where j is one plus their count. Label every input by h_θ and advance only on an active component. After depth n use U. This autonomous protocol does not inspect predictions. The domination bounds $\mu \ge (1-\beta)U$ and $\mu \ge (\beta/n)Q_j$ give at every full history

$$\mathrm{KL}((1-q)U + qQ_j\|\mu) \le 2\beta + q \log(n/\beta) \le 1/32 + 1/8 < 1.$$

After exhaustion the conditional KL is at most 2β. All online and potential reference draws use fresh independent seeds conditional on the fixed path and preceding history.

Pregenerate m potential reference draws through independent mixture coins, T potential uniform background inputs, and T independent Bernoulli(q) active coins. Let E be the event that all reference mixture coins choose U and all online background locations avoid J_0. The unlabeled uniform reference locations need not avoid J_0. The event is independent of all hidden bits, active coins and active location seeds, and

$$\Pr(E) \ge 1 - m\beta - Tw \ge 1 - 1/64 - 1/(T+1) \ge 1/2.$$

On E, reference observations and background labels reveal no future bit. This remains true under adaptive requests, which reveal prefixes of the potential list. Even granting active indicators and preceding bits leaves the current active label fair after its input is observed. The realizable comparator has zero loss, so regret is nonnegative off E as well. It follows that

$$\mathbb{E}\,\mathrm{Reg}_T \ge \frac{\Pr(E)}{2} \mathbb{E} \min\{\mathrm{Binomial}(T, q), n\}. \tag{190}$$

Write $l = \log(eT)$ and $a = \log((m+1)/\delta)$, so $s = l + a$. Since $n \le T$ and $T \ge 512$, $H + 1 \le 5T$ and $\log 320 \le \log T$. Hence

$$\log \frac{64n(m+1)(H+1)}{\delta} \le a + 3\log T, \qquad z \le 2\eta n + 2l,$$

and $K \le 5s + 2\eta n$. The inequalities $ns \le T$ and $\eta n^2 \le T$ imply $nK \le 7T$ and $Tq/n \ge 1/56$. Divide $n\lfloor T/n \rfloor$ active coins into n equal groups. Their count of nonempty groups is at most the total success count and at most n. Therefore

$$\begin{aligned}\mathbb{E}\min\{\text{Binomial}(T,q), n\} &\ge n[1-(1-q)^{\lfloor T/n \rfloor}] \\ &\ge n[1 - e^{-qT/(2n)}] \ge n[1 - e^{-1/112}] \ge n/224.\end{aligned}$$

Equation (190) gives at least $M/1792$ regret, so $c = 1/2048$ suffices. Finite-prior averaging selects a deterministic path with this expected loss for any randomized learner; it does not select a path after seeing the learner's realized private randomness.

If $s > T/2$, then $M < 2$. Choose $\mu = U$ and a target uniformly from $\{1/4, 3/4\}$ before play. Use iid uniform inputs and its realizable labels. The first input is between the targets with probability one half, where its label is fair even after all permitted reference requests. The first-round expected loss is at least $1/4 > M/8$. All reference promises hold with divergence zero.

For the agnostic claim, separately use Lemma S11.11 and the rational-trace transfer in Appendix S10, with $\mu = U$. That witness permits unlimited reference draws. The maximum of the two lower bounds is at least half their sum. □

At $m = 0$, $\delta = 1$ and $\eta \le \log(eT)^2/T$, this lower retains order $T/\log(eT)$, matching the no-reference upper. For fixed positive η, Proposition S14.2 recovers the known-reference statistical order. The case $m = 0, d = 1$ of Corollary S17.7 gives the full agnostic transition for thresholds, with rate $\Theta(G_{T,0}(\eta))$ for $\log T \ge 256$ and $0 < \eta \le 1$.

S15 Gaussian and log-concave reference calibration

We prove Corollaries 5.1 and 5.2. The Gaussian moment calculations and parameter estimation are classical. The fixed-cost strategies of Lemmas S6.4 and S15.2 convert bounds on the estimated reference's KL cost into regret bounds. Quantizing the estimated parameters lets us choose the proper strategy from a countable family. When both Gaussian parameters are unknown, we can calibrate a Cauchy reference from online inputs alone. Its transfer bound uses logarithmic moments of the estimated scale. The same calibration extends to an unknown log-concave shape, with constants depending only on the ambient dimension.

S15.1 Known covariance

Proof with independent reference samples. First, for $\mu = N(a, \Sigma)$ and $\nu = N(b, \Sigma)$, put $v = \Sigma^{-1/2}(a - b)$. Under $x = b + \Sigma^{1/2}z$ the likelihood ratio is $\exp(v \cdot z - \|v\|^2/2)$. For $Z \sim N(0, I_k)$, completing the square in the Gaussian integral gives $\mathbb{E}e^{w \cdot Z} = e^{\|w\|^2/2}$, and hence

$$D_2(\mu\|\nu) = \log \mathbb{E}e^{2v \cdot Z - \|v\|^2} = \|v\|^2.$$

Let $Y_1, \dots, Y_n$ be the reference samples and put $\bar{Z} = \Sigma^{-1/2} n^{-1} \sum_i Y_i$ and $\theta = \Sigma^{-1/2}a$. Then $\sqrt{n}(\bar{Z} - \theta) \sim N(0, I_k)$ and, since $n = k$,

$$\mathbb{E}\|\bar{Z} - \theta\|^2 = k/n = 1.$$

Set $q = \lceil \sqrt{k} \rceil$ and round each coordinate of $\bar{Z}$ to the nearest multiple of $1/q$, with a fixed rule for ties. Call the result $\widehat{\theta}$ and put $\widehat{a} = \Sigma^{1/2}\widehat{\theta}$. The rounding error is at most $\sqrt{k}/(2q) \le 1/2$ in Euclidean norm. Hence

$$\begin{aligned}\mathbb{E}D_2(\mu\|N(\widehat{a}, \Sigma)) &= \mathbb{E}\|\theta - \widehat{\theta}\|^2 \\ &\le 2\mathbb{E}\|\theta - \bar{Z}\|^2 + 2\mathbb{E}\|\bar{Z} - \widehat{\theta}\|^2 \le \frac{5}{2}.\end{aligned}$$

For each $u \in q^{-1}\mathbb{Z}^k$, fix a measurable proper strategy from Lemma S6.4 for the known reference $\nu_u = N(\Sigma^{1/2}u, \Sigma)$. After preprocessing, run the strategy indexed by $u = \widehat{\theta}$. The lattice is countable and rounding is Borel measurable, so these fixed strategies combine into one measurable strategy. They depend on the known objects and the observed samples, not on a.

Let $S = (Y_1, \dots, Y_n)$ be the full preprocessing transcript. The spaces are standard Borel, so regular conditional laws exist. Fix a for the analysis and condition on almost every transcript $S = z$. The index $u = \widehat{\theta}(z)$ and its selected strategy are now fixed. The full-history hypothesis includes S, so the conditional input KL constraint and current-round privacy convention continue to hold in this conditional experiment. At each subsequent full history, Lemma S14.1 gives

$$\mathrm{KL}(P_t \| \nu_u) \le 2\,\mathrm{KL}(P_t \| \mu) + D_2(\mu \| \nu_u) \le 2 + \|\theta - u\|^2.$$

Thus the conditional expected cost

$$C_*(z) = \mathbb{E}\left[\sum_t \mathrm{KL}(P_t \| \nu_u) \mid S = z\right] \le T(2 + \|\theta - u\|^2)$$

is finite. The learner fixed for ν_u satisfies Lemma S6.4 for every such cost; it is not chosen or tuned using $C_*(z)$. The conditional adversary may still use the complete transcript and its other permitted state.

Apply the lemma conditionally and then average over S. The nonnegative costs are integrable by the mean squared error bound above, and regret is bounded between $-T$ and T. Therefore

$$\begin{aligned}\mathbb{E}\,\mathrm{Reg}_T &\le CA_{d,T} + C\frac{A_{d,T}}{T}\mathbb{E}C_*(S) \\ &\le CA_{d,T}\left(3 + \mathbb{E}\|\theta - \widehat{\theta}\|^2\right) \le C' A_{d,T}.\end{aligned}$$

□

Proof without independent reference samples. We first derive the moment bounds needed to calibrate from adaptive online inputs. Let P satisfy $\mathrm{KL}(P\|\gamma_k) \le 1$, where $\gamma_k = N(0, I_k)$. For bounded measurable g, tilting γ_k by $e^g/\gamma_k(e^g)$ and using nonnegativity of KL gives

$$Pg \le \mathrm{KL}(P\|\gamma_k) + \log \gamma_k(e^g).$$

Apply this to $g_M(z) = \min\{\|z\|^2/4, M\}$. The Gaussian integral $\gamma_k(e^{\|z\|^2/4}) = 2^{k/2}$ and monotone convergence give

$$P\|z\|^2 \le B_k := 4 + 2k\log 2.$$

In particular the first moments are finite. For a unit vector v and $t > 0$, the Gaussian tilt with density $e^{tv\cdot z - t^2/2}$ has an integrable log density under P. Nonnegativity of KL relative to this tilted measure yields

$$tv \cdot Pz \le \mathrm{KL}(P\|\gamma_k) + t^2/2 \le 1 + t^2/2.$$

Take $t = \sqrt{2}$ and then the supremum over unit v to obtain $\|Pz\| \le \sqrt{2}$.

Fix a deterministic integer $1 \le q \le T$. During the first q rounds, predict with a fixed $f_0 \in \mathcal{F}$. Let $\mathcal{H}_i$ be the full state immediately before input i, so the input and completed feedback of round i belong to $\mathcal{H}_{i+1}$. These sigma-fields include the preceding learner randomness and permitted adversary state. For the analysis write

$$Z_i = \Sigma^{-1/2}(X_i - a), \qquad m_i = \mathbb{E}[Z_i \mid \mathcal{H}_i], \qquad \xi_i = Z_i - m_i.$$

Whitening preserves relative entropy, so the conditional moment bounds just proved apply at each full state. They give

$$\begin{aligned}\|m_i\| &\le \sqrt{2}, \\ \mathbb{E}[\|\xi_i\|^2 \mid \mathcal{H}_i] &= \mathbb{E}[\|Z_i\|^2 \mid \mathcal{H}_i] - \|m_i\|^2 \le B_k.\end{aligned}$$

For $i < j$, ξ_i is $\mathcal{H}_j$-measurable and $\mathbb{E}[\xi_j \mid \mathcal{H}_j] = 0$. Their inner product is integrable by the second-moment bounds, so $\mathbb{E}\langle \xi_i, \xi_j\rangle = 0$. Consequently

$$\mathbb{E}\left\|\frac{1}{q}\sum_{i=1}^q \xi_i\right\|^2 \le B_k/q, \qquad \left\|\frac{1}{q}\sum_{i=1}^q m_i\right\|^2 \le 2.$$

Put

$$\bar{Z} = \Sigma^{-1/2}\frac{1}{q}\sum_{i=1}^{q} X_i, \qquad \theta = \Sigma^{-1/2}a.$$

The two-square inequality gives

$$\mathbb{E}\|\bar{Z} - \theta\|^2 \le 2B_k/q + 4.$$

The conditional means m_i need not vanish; only their uniform bound was used.

Set $h = \lceil\sqrt{k}\rceil$ and round each coordinate of $\bar{Z}$ to the nearest multiple of $1/h$ using a fixed Borel tie rule. Call the result $\widehat{\theta}$. Its rounding error has norm at most $1/2$, so

$$\mathbb{E}\|\widehat{\theta} - \theta\|^2 \le 4B_k/q + \frac{17}{2} \le C(1 + k/q), \tag{191}$$

where the last inequality uses $k \ge 1$.

Let $H = T - q$ be the number of remaining rounds. If $H = 0$, the whole regret is at most $q = T$. If $0 < H < d$, use any fixed member of $\mathcal{F}$ on the remaining rounds; its regret there is at most $H < d \le A_{d,T}$. Now suppose $H \ge d$. For every $u \in h^{-1}\mathbb{Z}^k$, fix the learner of Lemma S6.4 for reference $\nu_u = N(\Sigma^{1/2}u, \Sigma)$, class $\mathcal{F}$, dimension bound d and horizon H. These learners are fixed without knowledge of a or of the eventual estimation error. The countable lattice and Borel rounding make selection of the learner indexed by $u = \widehat{\theta}$ measurable and proper.

Condition on the full observed prefix transcript S, including its labels and learner randomness. For almost every transcript, the index u and chosen learner are fixed. The full-history input constraint continues to hold at almost every subsequent full state, since that state already includes S. The conditional adversary may use the prefix and its permitted state; fresh learner draws on the remaining rounds retain the privacy convention. The Gaussian divergence identity and Lemma S14.1 give

$$\mathrm{KL}(P_t\|\nu_u) \le 2\,\mathrm{KL}(P_t\|\mu) + D_2(\mu\|\nu_u) \le 2 + \|\theta - u\|^2.$$

Hence the conditional expected cost $C_*(S)$ on these H rounds is finite and at most $H(2 + \|\theta - u\|^2)$. Lemma S6.4, applied to the already chosen learner, bounds its conditional expected regret by

$$CA_{d,H}\left(1 + \frac{C_*(S)}{H}\right) \le CA_{d,H}(3 + \|\theta - u\|^2).$$

Averaging and using (191) gives expected regret at most $CA_{d,H}(1 + k/q)$ on the remaining rounds. The costs are integrable by that mean squared error bound. Finally, $A_{d,H} \le A_{d,T}$ because $x\log(ex/d)$ is increasing for $x \ge d$.

For a realized path, let $L_1(f)$ and $L_2(f)$ be the prefix and remaining comparator losses. Since

$$\inf_f[L_1(f) + L_2(f)] \ge \inf_f L_1(f) + \inf_f L_2(f),$$

whole-horizon regret is at most the sum of the two segment regrets. The prefix regret is at most q. The preceding bounds, including the cases $H = 0$ and $H < d$, therefore prove

$$\mathbb{E}\,\mathrm{Reg}_T \le \min\{T, q + CA_{d,T}(1 + k/q)\}. \tag{192}$$

The cap follows from the pathwise bound $\mathrm{Reg}_T \le T$ for this same learner. If $T > k$, take $q = k$ and absorb the factor two into C to obtain the corollary. If $T \le k$, a fixed hypothesis throughout gives the asserted bound directly. □

The prefix length can also be chosen for a smaller dimension cost when k is large. Put $A = A_{d,T}$ and $q = \min\{T, \lceil\sqrt{kA}\rceil\}$. Since $k, A \ge 1$, if $q < T$ then

$$\sqrt{kA} \le q \le \sqrt{kA} + 1 \le 2\sqrt{kA}, \qquad Ak/q \le \sqrt{kA}.$$

Equation (192) then gives $C(A + \sqrt{kA})$ regret. If $q = T$ and $T \ge 2$, the ceiling condition implies $\sqrt{kA} > T - 1 \ge T/2$, so the trivial bound T has the same guarantee up to a universal constant. At $T = 1$, $A = 1$ and the same conclusion holds. Thus a proper learner using no independent reference samples satisfies

$$\mathbb{E}\,\mathrm{Reg}_T \le C\min\{T, A_{d,T} + \sqrt{kA_{d,T}}\}.$$

These are sufficient calibration bounds; no optimal dependence on k is asserted.

S15.2 Unknown covariance

Proof with unknown mean and covariance. Fix $\mu = N(m, \Sigma)$ with $m \in \mathbb{R}^k$ and $\Sigma \succ 0$. Take

$$r = 128(k+1)^2, \qquad n = r + 1 \tag{193}$$

independent reference samples. For each known reference ν, Lemma S6.4 gives one strategy satisfying

$$\mathbb{E}\,\mathrm{Reg}_T \le C_0 A_{d,T} + C_0 \frac{A_{d,T}}{T} \mathbb{E} \sum_{t=1}^{T} \mathrm{KL}(P_t \| \nu) \tag{194}$$

whenever the expected total cost is finite. Its tuning does not depend on that cost. We construct a countably valued Gaussian estimate for which the expected transfer cost is bounded by a universal constant.

Uniform Gaussian KL transfer. Write $\gamma_k = N(0, I_k)$. We first prove that for any symmetric k by k matrix B and any P with $\mathrm{KL}(P\|\gamma_k) \le 1$,

$$\|\mathbb{E}_P Z\| \le \sqrt{2}, \qquad \mathbb{E}_P[Z^\top B Z] - \operatorname{tr} B \le 4\|B\|_F. \tag{195}$$

All moments appearing here are finite. For bounded measurable g, nonnegativity of KL relative to the tilted Gaussian gives

$$Pg \le \mathrm{KL}(P\|\gamma_k) + \log \gamma_k(e^g). \tag{196}$$

Apply this first to $g_M(z) = \min\{\|z\|^2/4, M\}$. The Gaussian integral is $\gamma_k(e^{\|z\|^2/4}) = 2^{k/2}$. Monotone convergence in (196) gives

$$P\|Z\|^2 \le 4 + 2k \log 2. \tag{197}$$

Thus all linear and quadratic forms are integrable. For unit v and $t > 0$, the Gaussian tilt with density $e^{tv^\top z - t^2/2}$ has integrable log density under P. Its KL nonnegativity gives $tv^\top PZ \le 1 + t^2/2$. Take $t = \sqrt{2}$ and the supremum over unit v to obtain the first assertion of (195).

For the second, let λ_i be the eigenvalues of B. For $0 < t < 1/(2\|B\|_{\mathrm{op}})$, with the restriction void if $B = 0$, Gaussian integration gives

$$\log \mathbb{E}_{\gamma_k} e^{t(Z^\top B Z - \operatorname{tr} B)} = \sum_i \left[-t\lambda_i - \frac{1}{2} \log(1 - 2t\lambda_i) \right].$$

Expanding the logarithm and using $1/j \le 1/2$ for $j \ge 2$ bounds each summand by $t^2\lambda_i^2/(1 - 2t\|B\|_{\mathrm{op}})$. This holds for either sign of λ_i, by bounding the absolute terms of the series. The entropy inequality, now justified by (197) and the integrable log density of the quadratic tilt, therefore gives

$$\mathbb{E}_P[Z^\top B Z] - \operatorname{tr} B \le \frac{1}{t} + \frac{t\|B\|_F^2}{1 - 2t\|B\|_{\mathrm{op}}}.$$

If $B \ne 0$, choose $t = (\|B\|_F + 2\|B\|_{\mathrm{op}})^{-1}$. The right side is $2\|B\|_F + 2\|B\|_{\mathrm{op}} \le 4\|B\|_F$. For $B = 0$ the assertion is immediate.

For a candidate Gaussian $\nu = N(q, Q)$ with $Q \succ 0$, put

$$S = \Sigma^{-1/2} Q \Sigma^{-1/2}, \qquad e = \Sigma^{-1/2}(q - m).$$

The following function is finite and nonnegative:

$$H(S,e) = 1 + \frac{1}{2}\{\log\det S + \operatorname{tr} S^{-1} - k\} + 2\|S^{-1} - I\|_F + \sqrt{2}\,\|S^{-1}e\| + \frac{1}{2}e^\top S^{-1}e. \tag{198}$$

Nonnegativity of the covariance term follows by summing $\log\lambda + \lambda^{-1} - 1 \geq 0$ over the eigenvalues of S. For every P with $\mathrm{KL}(P\|\mu) \leq 1$,

$$\mathrm{KL}(P\|\nu) \leq H(S,e). \tag{199}$$

To prove this, whiten $Z = \Sigma^{-1/2}(X - m)$. KL is unchanged under this invertible affine map. The Gaussian density identity is

$$\log\frac{d\mu}{d\nu}(X) = \frac{1}{2}\log\det S + \frac{1}{2}Z^\top(S^{-1} - I)Z - e^\top S^{-1}Z + \frac{1}{2}e^\top S^{-1}e. \tag{200}$$

By (197) its terms are integrable under P. Since $\mathrm{KL}(P\|\mu)$ is finite, the change of reference identity $\mathrm{KL}(P\|\nu) = \mathrm{KL}(P\|\mu) + P\log(d\mu/d\nu)$ is valid. For completeness, the negative part of $\log(dP/d\mu)$ is integrable because $x\log(1/x) \leq 1/e$ for $0 < x \leq 1$; its positive part is integrable by the assumed finite KL. Adding the integrable log-density ratio (200) therefore proves the identity and finiteness of the new KL. Apply (195) to $B = S^{-1} - I$ and bound the linear term by $\sqrt{2}\|S^{-1}e\|$. This proves (199).

The bound (199) is simultaneous for all such P. It is therefore valid even when P has been selected after seeing (q, Q) or the complete calibration transcript.

Sample mean and covariance moments. Let $Y_1, \ldots, Y_n$ be the reference observations. Put

$$\widehat{m} = \frac{1}{n}\sum_i Y_i, \qquad \widehat{\Sigma} = \frac{1}{r}\sum_i (Y_i - \widehat{m})(Y_i - \widehat{m})^\top. \tag{201}$$

For the analysis define

$$e_0 = \Sigma^{-1/2}(\widehat{m} - m), \qquad A = \Sigma^{-1/2}\widehat{\Sigma}\Sigma^{-1/2}.$$

Then $e_0 \sim N(0, I_k/n)$, independent of A, and

$$A = W/r, \qquad W = G^\top G, \tag{202}$$

where G has r independent standard Gaussian rows. To verify this without an independence assertion about estimated parameters, whiten the n original observations and apply an orthogonal transformation to their row coordinates whose first row is $n^{-1/2}(1, \ldots, 1)$. The transformed matrix still has independent standard Gaussian entries, by its joint Gaussian density and orthogonal invariance. Its first row is $\sqrt{n}\,e_0$. The remaining r rows form G, and centering removes exactly the first row from the quadratic sum. This proves (202) and independence.

The columns of G are almost surely linearly independent for $r \geq k$: inductively, the next independent Gaussian column lies in the span of its predecessors with probability zero. Thus $A \succ 0$ almost surely.

We prove the needed inverse moments. Write $V = W^{-1}$ and

$$a = r - k - 1, \qquad b = r - k - 3, \qquad c = r - k.$$

All three denominators are positive under (193). Partition the first column g of G from the other columns G_-. When $k \geq 2$, put

$$W_- = G_-^\top G_-, \quad v = W_-^{-1} G_-^\top g, \quad R = \|(I - \Pi_{G_-})g\|^2.$$

Conditional on G_-, the projection coefficients v are centered Gaussian with covariance W_-^{-1}. The projected and residual parts of g are independent Gaussian vectors. The residual has $r - k + 1$ independent standard Gaussian coordinates after a suitable orthogonal transformation. Its squared norm R thus has the χ^2_{r-k+1} law, which does not depend on G_-. It follows by conditioning that R is independent of the pair (v, G_-), not just conditionally independent of v. The block inverse formula gives

$$V_{11} = R^{-1}, \qquad V_{12} = -R^{-1} v_1. \tag{203}$$

For $k = 1$, the same diagonal formula holds with $R = \|g\|^2 \sim \chi^2_r$ and there is no off-diagonal entry. The density of a χ^2_ℓ variable is $x^{\ell/2-1} e^{-x/2} / (2^{\ell/2} \Gamma(\ell/2))$ for $x > 0$, as follows by polar coordinates in the standard Gaussian density. Integrating this density and using the Gamma recurrence gives, for $\ell > 4$,

$$\mathbb{E} R^{-1} = \frac{1}{\ell - 2}, \qquad \mathbb{E} R^{-2} = \frac{1}{(\ell - 2)(\ell - 4)}.$$

Thus, for all $k \geq 1$,

$$\mathbb{E} V_{11} = 1/a, \qquad \mathbb{E} V_{11}^2 = 1/(ab). \tag{204}$$

For $k \geq 2$, apply the already proved diagonal first-moment formula to W_-, which has $k-1$ columns and r rows. It gives $\mathbb{E}(W_-^{-1})_{11} = 1/(r-k) = 1/c$. By the independence in (203),

$$\mathbb{E} V_{12}^2 = \frac{1}{abc}. \tag{205}$$

Column permutation symmetry gives these moments at every diagonal or off-diagonal position. Off-diagonal means are zero, because changing the sign of one Gaussian column conjugates V by that sign change and preserves its law. Hence

$$\begin{aligned} \mathbb{E} \operatorname{tr} A^{-1} &= \frac{rk}{r - k - 1}, \\ \mathbb{E} \operatorname{tr} A^{-2} &= \frac{r^2 k (r-1)}{(r-k)(r-k-1)(r-k-3)}. \end{aligned} \tag{206}$$

The second formula sums k diagonal squares and $k(k-1)$ off-diagonal squares; for $k = 1$ the latter contribution is zero, and the displayed formula simplifies to the diagonal expression. In particular, since $a, b, c \geq r/2$,

$$\mathbb{E} \operatorname{tr} A^{-1} \leq 2k, \qquad \mathbb{E} \operatorname{tr} A^{-2} \leq 8k. \tag{207}$$

The same moments give the useful cancellation

$$\begin{aligned} \mathbb{E} \|A^{-1} - I\|_F^2 &= \frac{2kr^2}{a^2 b} + \frac{k(k-1)r^2}{abc} + \frac{k(k+1)^2}{a^2} \\ &\leq \frac{16k + 8k(k-1) + 4k}{r} \leq \frac{20k^2}{r}. \end{aligned} \tag{208}$$

The first line separates diagonal variance, off-diagonal squares, and the squared diagonal bias. For the second line use $a, b, c \geq r/2$ and $r \geq (k+1)^2$.

No log-determinant distribution is needed. The scalar inequalities $1 - \lambda^{-1} \le \log \lambda \le \lambda - 1$ imply

$$k - \operatorname{tr} A^{-1} \le \log\det A \le \operatorname{tr} A - k.$$

Both bounding variables are integrable, since $\mathbb{E}\operatorname{tr} A = k$ and (207) holds. Therefore $\log\det A$ is integrable and $\mathbb{E}\log\det A \le 0$. In particular,

$$\begin{aligned} 0 &\le \mathbb{E}[\log\det A + \operatorname{tr} A^{-1} - k] \\ &\le \frac{k(k+1)}{r-k-1} \le \frac{2k(k+1)}{r} \le \frac{1}{64}. \end{aligned} \tag{209}$$

Countable quantization. The learner cannot whiten by Σ. We quantize in its observed coordinates with mesh determined by the empirical covariance. Let $\lambda = \lambda_{\min}(\widehat{\Sigma}) > 0$ and set

$$h_Q = 2^{\lfloor \log_2(\lambda/(16k^2)) \rfloor}, \qquad h_m = 2^{\lfloor \log_2(\sqrt{\lambda}/(16k)) \rfloor}. \tag{210}$$

Round each entry on or above the diagonal of $\widehat{\Sigma}$ to its nearest multiple of h_Q, with a fixed Borel tie rule, and reflect these entries across the diagonal. Call the resulting symmetric matrix Q. Round each coordinate of $\widehat{m}$ to the nearest multiple of h_m with the same type of tie rule, giving q.

All coordinates of q and Q are dyadic rationals. These maps are Borel measurable: the sample mean and covariance are continuous, the least eigenvalue is continuous, the logarithm and floor in (210) are Borel on positive inputs, and coordinate rounding with fixed tie choices is Borel. On the null event where the empirical covariance is not positive definite, set $q = 0, Q = I_k$. This also gives Borel maps on all possible inputs.

Put $\delta = 1/(32k)$. Every covariance-entry error is at most $h_Q/2 \le \lambda/(32k^2)$. Thus

$$\|Q - \widehat{\Sigma}\|_{\mathrm{op}} \le \|Q - \widehat{\Sigma}\|_F \le k \max_{i,j} |Q_{ij} - \widehat{\Sigma}_{ij}| \le \delta\lambda.$$

It follows that

$$(1-\delta)\widehat{\Sigma} \preceq Q \preceq (1+\delta)\widehat{\Sigma}, \tag{211}$$

so Q is positive definite. Also

$$(q - \widehat{m})^\top \widehat{\Sigma}^{-1}(q - \widehat{m}) \le \frac{\|q - \widehat{m}\|^2}{\lambda} \le \frac{1}{1024k}. \tag{212}$$

The pair (q, Q) belongs to a fixed countable set of dyadic-rational vectors and symmetric positive-definite dyadic-rational matrices. This set, and its enumeration, are independent of m, Σ.

For the remaining analysis write

$$S = \Sigma^{-1/2} Q \Sigma^{-1/2}, \qquad d_0 = \Sigma^{-1/2}(q - \widehat{m}), \qquad e = e_0 + d_0.$$

Congruence in (211), and (212), give

$$(1-\delta)A \preceq S \preceq (1+\delta)A, \qquad d_0^\top A^{-1} d_0 \le \frac{1}{1024k}. \tag{213}$$

In particular,

$$\operatorname{tr} S^{-1} \le (1+2\delta)\operatorname{tr} A^{-1}, \qquad \log\det S \le \log\det A + k\delta, \tag{214}$$

using $(1-\delta)^{-1} \le 1+2\delta$ and $\log(1+\delta) \le \delta$.

We need a norm comparison that does not assume the matrices commute. Put $R = A^{-1/2} S A^{-1/2}$. Then $\|R^{-1} - I\|_{\mathrm{op}} \le 2\delta$, and

$$S^{-1} - A^{-1} = A^{-1/2}(R^{-1} - I)A^{-1/2}.$$

Diagonalize $R^{-1} - I = \sum_j \theta_j v_j v_j^\top$. The triangle inequality and $\|ww^\top\|_F = \|w\|^2$ show that

$$\|S^{-1} - A^{-1}\|_F \le \sum_j |\theta_j| \|A^{-1/2} v_j\|^2 \le 2\delta \operatorname{tr} A^{-1}. \tag{215}$$

Taking expectations in (214), (215), and using (207)–(209) gives

$$\begin{aligned} \mathbb{E}[\log\det S + \operatorname{tr} S^{-1} - k] &\le \frac{1}{64} + k\delta + 4k\delta < 1, \\ \mathbb{E}\|S^{-1} - I\|_F &\le \sqrt{20k^2/r} + 4k\delta < 1, \\ \mathbb{E}\operatorname{tr} S^{-1} &\le 3k. \end{aligned} \tag{216}$$

The first expression is nonnegative pointwise, as in (198). All its terms are integrable: (214) and its lower counterparts bound $\log\det S - \log\det A$, and the inverse trace is bounded by the inverse trace of A.

Also $S^{-1} \preceq 2A^{-1}$, so

$$\mathbb{E}\operatorname{tr} S^{-2} \le 4\mathbb{E}\operatorname{tr} A^{-2} \le 32k. \tag{217}$$

For clarity, if $0 \preceq L \preceq M$, then $\operatorname{tr}(M^2 - L^2) = \operatorname{tr}((M-L)(M+L)) \ge 0$. The last inequality follows from the nonnegative trace of a product of two positive-semidefinite matrices. Applying it to $L = S^{-1}$ and $M = 2A^{-1}$ proves (217) without commutativity.

Mean error after quantization. The matrix Q, and hence S, depends only on the empirical covariance. Therefore it is independent of e_0. Since e_0 has covariance I_k/n, (216), (217) give

$$\mathbb{E}[e_0^\top S^{-1} e_0] \le 3k/n, \qquad \mathbb{E}\|S^{-1} e_0\|^2 \le 32k/n. \tag{218}$$

No independence involving the rounding error d_0 is assumed. From (213) and $S^{-1} \preceq 2A^{-1}$,

$$d_0^\top S^{-1} d_0 \le \frac{1}{512k}. \tag{219}$$

The representation $S^{-1} = A^{-1/2} R^{-1} A^{-1/2}$, with $\|R^{-1}\|_{\mathrm{op}} \le 2$, also gives

$$\|S^{-1} d_0\|^2 \le 4\|A^{-1}\|_{\mathrm{op}}\, d_0^\top A^{-1} d_0 \le \frac{\operatorname{tr} A^{-1}}{256k}.$$

Taking expectations bounds this by $1/128$. Using $e = e_0 + d_0$ and the two-square inequality, we obtain

$$\begin{aligned} \mathbb{E}[e^\top S^{-1} e] &\le 6k/n + \frac{1}{256k} < 1, \\ \mathbb{E}\|S^{-1} e\|^2 &\le 64k/n + \frac{1}{64} < 1. \end{aligned} \tag{220}$$

Both strict bounds follow from $n \geq 128(k+1)^2$ and $k/(k+1)^2 \leq 1/4$. Cauchy–Schwarz also gives $\mathbb{E}\|S^{-1}e\| < 1$.

Equations (216) and (220) prove the desired universal expected transfer bound:

$$\mathbb{E}H(S,e) \leq 1 + \frac{1}{2} + 2 + \sqrt{2} + \frac{1}{2} < 6. \tag{221}$$

Here H is the explicit Borel function (198). For every realized positive-definite empirical covariance it simultaneously bounds the KL cost of every law in the original unit KL ball by (199). We do not need to form, or assume measurability of, an uncountable supremum over such laws.

Selecting the proper strategy. For every pair (q, Q) in the fixed countable set from the quantization construction above, fix once and for all a measurable proper strategy supplied by Lemma S6.4 for the known reference $N(q, Q)$, class $\mathcal{F}$, VC bound d, and horizon T. These choices depend on no unknown Gaussian parameter or eventual calibration error. Given the sample transcript, use its Borel quantization index to select the corresponding fixed strategy. A countable piecewise selection of measurable kernels is measurable, and every selected action remains in $\mathcal{F}$. Use fresh learner randomization with the current-round privacy convention.

Let $\mathcal{S} = (Y_1, \ldots, Y_n)$ be the full reference transcript. Fix m, Σ for the analysis. The underlying spaces are standard Borel, so regular conditional laws exist. For almost every transcript $\mathcal{S} = s$, the empirical covariance is positive definite and its selected reference $\nu_s = N(q(s), Q(s))$ and strategy are fixed. The subsequent full histories include $\mathcal{S}$. Consequently the assumed original KL constraint holds at almost every conditional full history as well. The adversary may use the complete transcript and any permitted state; this is covered by the simultaneous bound (199).

It follows at every such history that

$$\mathrm{KL}(P_t \| \nu_s) \leq H(S(s), e(s)).$$

The right side is finite for this fixed transcript, and does not depend on the adaptive choice of P_t. Therefore the conditional expected total cost obeys

$$C_*(s) := \mathbb{E}\left[\sum_{t=1}^{T} \mathrm{KL}(P_t \| \nu_s) \,\middle|\, \mathcal{S} = s\right] \leq TH(S(s), e(s)) < \infty. \tag{222}$$

The selected strategy in (194) is already fixed and works for every such actual cost; it is not chosen or tuned using $C_*(s)$. Conditionally applying (194) and then averaging gives

$$\begin{aligned} \mathbb{E}\,\mathrm{Reg}_T &\leq C_0 A_{d,T} + C_0 \frac{A_{d,T}}{T}\, \mathbb{E}C_*(\mathcal{S}) \\ &\leq C_0 A_{d,T}\big(1 + \mathbb{E}H(S,e)\big) \leq 7C_0 A_{d,T}. \end{aligned}$$

The nonnegative costs are integrable by (221), and regret is bounded between $-T$ and T, so these conditional-expectation steps are valid. This proves the first assertion of Corollary 5.1 with a universal constant.

All whitening in the proof is analytical. The algorithm uses the observed-coordinate sample mean, covariance, and dyadic rounding. Neither the construction nor the reference-family enumeration uses m, Σ, or a bound on their size or condition number.

□

S15.3 The statistical logarithm under a Gaussian reference

The following consequence transports the uniform-reference lower of Lemma S11.11 by the Gaussian distribution function.

Proposition S15.1. *Let $\gamma = N(0,1)$ be known and let*

$$\mathcal{F}_{\mathbb{Q}} = \{x \mapsto \mathbf{1}\{x \geq q\} : q \in \mathbb{Q}\}.$$

For every integer $T \geq 512$, this fixed countable class of VC dimension one has minimax expected regret at least $\sqrt{T \log(eT)}/256$ under the full-history constraint $\mathrm{KL}(P_t \| \gamma) \leq 1$. The lower permits arbitrary randomized post-input predictions, unlimited computation, and unlimited iid reference draws requested adaptively, including after the current input. The labels obey the current-round privacy convention.

Proof. Let λ be uniform probability on $[0,1]$, let Φ be the standard normal distribution function, and let $\psi = \Phi^{-1} : (0,1) \to \mathbb{R}$. These are strictly increasing continuous inverse maps, and $\Phi_{\#}\gamma = \lambda$, with the endpoints null.

Fix a learner for the Gaussian reference. Simulate it under the uniform reference by mapping each observed input u_t to $x_t = \psi(u_t)$, forwarding labels without change, and using the same randomized predictions. On each Gaussian reference request, return $\psi(V)$ for a fresh independent uniform reference draw V. This is a measurable simulation, including post-input requests. Use the same internal seed and announced prediction distributions under the coordinate change. Lemma S11.11 supplies a uniform-reference protocol against this learner with expected regret at least $\sqrt{T \log(eT)}/256$.

Transform that protocol to Gaussian coordinates. At a full Gaussian history h, apply Φ to every preceding input and reference draw, retaining labels, predictions, and seed coordinates. Let h' be the resulting uniform history, let $p_t^{h'}$ be its next-input law, and set $P_t^h = \psi_{\#} p_t^{h'}$. Finite conditional KL implies $p_t^{h'} \ll \lambda$. Writing $g_t^{h'} = dp_t^{h'}/d\lambda$, the density identity is

$$\frac{dP_t^h}{d\gamma}(x) = g_t^{h'}(\Phi(x)).$$

Changing variables by Φ gives

$$\begin{aligned}
\mathrm{KL}(P_t^h \| \gamma) &= \int_{\mathbb{R}} g_t^{h'}(\Phi(x)) \log g_t^{h'}(\Phi(x))\, d\gamma(x) \\
&= \int_0^1 g_t^{h'}(u) \log g_t^{h'}(u)\, du \\
&= \mathrm{KL}(p_t^{h'} \| \lambda) \leq 1.
\end{aligned}$$

The histories correspond by measurable inverses on the probability-one event where all uniform coordinates are interior. Set kernels at irrelevant null histories equal to the reference. This proves the bound conditional on the full history. The original lower uses fresh fair informative labels and the specified filler labels; the coordinate change introduces no dependence on a fresh private prediction draw, so current-round privacy is preserved.

Each transformed reference request is independent of the preceding transcript and has law γ. Adaptive request times cause no change: pre-generate a countable independent uniform sequence and reveal its successive entries. Even granting the whole sequence only adds independent randomness to the fair-label lower. Applying ψ coordinatewise simulates the Gaussian sequence and preserves the full-history argument.

Finally, the comparator losses agree pathwise. On a finite realized list $x_1, \ldots, x_T$, every trace of a real threshold is realized by a rational threshold. If both sides occur, let a be the largest coordinate assigned zero and b the smallest assigned one. Any threshold realizing the trace forces $a < b$; choose a rational $q \in (a,b)$. For an all-one trace choose q below the smallest coordinate, and for an all-zero trace choose q above the largest. Repeated coordinates cannot be assigned to different sides. The empty-list assertion is immediate.

Every rational real threshold maps to the uniform threshold $\Phi(q)$, and every interior uniform threshold maps to a real threshold whose finite trace is rationally realizable as above. Uniform thresholds at zero and one give the two constant traces because all $u_t = \Phi(x_t)$ are interior. Thus

$$\inf_{q \in \mathbb{Q}} \sum_{t=1}^{T} \mathbf{1}\{\mathbf{1}\{x_t \geq q\} \neq y_t\} = \inf_{\theta \in [0,1]} \sum_{t=1}^{T} \mathbf{1}\{\mathbf{1}\{u_t \geq \theta\} \neq y_t\}.$$

These are minima over finitely many feasible traces, and the countable infimum on the left is measurable. The simulated learner also has the same loss on every coupled round. The regrets therefore agree pathwise, and the uniform lower proves the claim for the arbitrary Gaussian learner. □

S15.4 Unknown covariance without reference samples

We first allow a known cost scale in the proper strategy. The tuning remains independent of the actual conditional KL expenditure.

Lemma S15.2. *Let $1 \le d \le H$ be integers, let $K \ge 1$ be known, let ν be a known reference, and let $\mathcal{F}$ be a known nonempty countable measurable binary class with $\mathrm{VC}(\mathcal{F}) \le d$ in the model of Section 2. There is a measurable proper pre-input strategy depending only on $H, d, K, \mathcal{F}, \nu$ such that every protocol with the current-round privacy convention and finite expected cost $C_* = \mathbb{E}\sum_{t=1}^{H} \mathrm{KL}(P_t \| \nu)$ satisfies*

$$\mathbb{E}\,\mathrm{Reg}_H \le C_1 \sqrt{K} A_{d,H} + C_1 \frac{A_{d,H}}{H\sqrt{K}} C_*$$

with universal C_1. The same strategy works for every C_* and does not observe the conditional laws or costs.

Proof. Put $R = H/d$ and $L = \log(eR)$. If $K \ge R$, use any fixed hypothesis. Its regret is at most $H \le \sqrt{KdHL} = \sqrt{K} A_{d,H}$, since $Kd \ge H$ and $L \ge 1$. This includes $d = H$.

Suppose $1 \le K < R$, and first take a finite input space and finite class with a full-support reference. Use the construction preceding (11) with deterministic tuning value $B_0 = KH$. Its scales satisfy

$$A_0 = 4K, \qquad 256\sqrt{KH/d} \le U < 512\sqrt{KH/d}.$$

The condition $KH < H^2/d$ puts this choice in the range of that construction. In particular, $64\sqrt{R/K} \le U/A_0 < 128\sqrt{R/K}$. There is at least one fine scale, and their number M satisfies

$$1 \le M \le 1 + \log_2(U/A_0) \le 8 + \tfrac{1}{2}\log_2 R \le 8\log(eR).$$

Set $\lambda = \sqrt{dM/(KH)}$ and $\tau = 8\lambda + 2/U$. Then

$$\tau \le C\sqrt{\frac{dL}{KH}} = C\frac{A_{d,H}}{H\sqrt{K}}. \tag{223}$$

The cost-free term in (11) is

$$A_* = C\left(\sqrt{dH} + \sqrt{dHA_0} + dA_0 + dU + \frac{dM}{\lambda} + d\right).$$

Its terms obey

$$\sqrt{dHA_0} = 2\sqrt{dKH}, \qquad dA_0 = 4dK \le 4\sqrt{dKH},$$
$$dU < 512\sqrt{dKH}, \qquad dM/\lambda = \sqrt{dMKH} \le \sqrt{8dKHL}.$$

Here $dK \le H$ gives the second bound, while $d \le \sqrt{dH}$ and $K, L \ge 1$ handle the remaining terms. Thus $A_* \le C\sqrt{K} A_{d,H}$. Also $U \ge 256\sqrt{R} \ge 4\log(eR)$, so the terminal-scale condition remains valid.

Equation (11) holds for every actual finite expected cost before imposing a bound by the tuning value. With this fixed τ, it gives $\mathcal{S}_\tau \le A_*$. All one-round costs on the finite full-support space are bounded. Lemma S6.2 therefore supplies one proper strategy with regret at most $2A_* + 2\tau C_*$. The bound on A_* and (223) prove the finite case independently of the actual cost.

For a countable class, fix a finite proper e^{-v}-cover $\mathcal{C} \subseteq \mathcal{F}$ under ν, where

$$v = \max\{4\log(eH/d), KH/d, 1\}.$$

For every $f \in \mathcal{F}$, fix $\pi f \in \mathcal{C}$ with $\nu\{f \ne \pi f\} \le e^{-v}$, and partition the space by the cover's evaluation patterns. Discard reference-null cells and assign any fixed continuation there. Finite

expected KL cost implies $P_t \ll \nu$ at almost every full history, so null cells are visited with probability zero. The induced reference has full support and its class has VC dimension at most d: representatives of shattered cells would shatter the same number of original points.

Apply the finite strategy to this partition and select its hypotheses from $\mathcal{C}$. Data processing and conditional averaging bound the induced expected KL cost by C_*. Pathwise regret against $\mathcal{F}$ is at most regret against $\mathcal{C}$ plus

$$\sup_{f\in\mathcal{F}} \sum_{t=1}^{H} \mathbf{1}\{f(X_t) \neq \pi f(X_t)\}.$$

Lemma S6.3 bounds the expectation of this countable disagreement supremum by $Cd + 2C_*/v$. These terms are absorbed because

$$d \le \sqrt{K} A_{d,H}, \qquad \frac{1}{v} \le \frac{d}{KH} \le \frac{A_{d,H}}{H\sqrt{K}}.$$

The evaluation partition is measurable, and the finite strategy uses finitely many comparator-loss states. The lifted strategy is consequently measurable and proper. Its cover, rates and strategy were fixed independently of C_*. □

We next control the scatter of an adaptive prefix by changing measure from independent reference inputs. Write $\log^+ x = \max\{0, \log x\}$.

Lemma S15.3. *Let $k \ge 1$ be an integer, and let ν be a probability measure on $\mathbb{R}^k$. Suppose every unit projection under ν has a Lebesgue density bounded by $L > 0$, and*

$$\mathbb{E}_\nu \exp\{\|Z\|/(B_0\sqrt{k})\} \le M, \qquad B_0 > 0, \quad M \ge 1.$$

Define

$$\begin{aligned} m &= \lceil 16k \log(ek) \rceil, \qquad n = 2m, \\ R &= \max\{1, B_0(2\log M + 8)\}, \qquad K_L = \max\{1, L\sqrt{8\pi e}\}, \\ A &= \log K_L + \frac{\log(3R) + \frac{1}{2}\log 17 + \frac{3}{2}}{16}, \qquad C' = A + \log 3, \\ C_{\mathrm{inv}}(L, B_0, M) &= 16 + 8C' + \log 4. \end{aligned}$$

For $z_1, \dots, z_n \in \mathbb{R}^k$, let

$$\bar{z} = \frac{1}{n}\sum_i z_i, \qquad S = \frac{1}{n}\sum_i (z_i - \bar{z})(z_i - \bar{z})^\top,$$
$$F_S = \log^+(1/\lambda_{\min}(S)),$$

where $F_S = +\infty$ if S is singular. Then

$$\mathbb{E}_{\nu^n} e^{(m/8)F_S} \le \exp\{(C' + \log 4/8)m\}. \tag{224}$$

For any joint law P_n of the n inputs with $\mathrm{KL}(P_n\|\nu^n) \le n$, the scatter is positive definite almost surely and

$$\mathbb{E}_{P_n} \log^+(1/\lambda_{\min}(S)) \le C_{\mathrm{inv}}(L, B_0, M). \tag{225}$$

This conclusion applies if every input law, conditional on the full preceding history containing the earlier inputs, has KL at most one from ν. The inputs need not be independent.

Proof. First take $Z_1, \dots, Z_n$ independent with law ν. Put $D_j = Z_{2j} - Z_{2j-1}$ and let D be the matrix whose m rows are $D_j^\top$. Define $G = m^{-1}\mathsf{D}^\top\mathsf{D}$ and $F_G = \log^+(1/\lambda_{\min}(G))$, with the same convention at a zero eigenvalue. The differences are independent. By the triangle inequality and independence within each pair,

$$\mathbb{E} e^{\|D_j\|/(B_0\sqrt{k})} \le M^2.$$

Writing $W = \sum_j \|D_j\|$, exponential Markov gives, for $t \ge 0$,

$$\begin{aligned} \Pr\{W > Rm\sqrt{k}\, e^{t/4}\} &\le \exp\{2m\log M - (R/B_0)me^{t/4}\} \\ &\le e^{-8m-2mt}. \end{aligned} \tag{226}$$

The last inequality uses $R/B_0 \ge 2\log M + 8$ and $e^{t/4} \ge 1 + t/4$.

Let $\delta = e^{-3t/4}/(R\sqrt{mk}) \le 1$. A δ-net of the unit sphere has cardinality at most $(1 + 2/\delta)^k \le (3/\delta)^k$. Indeed, take a maximal δ-separated set. The disjoint balls of radius $\delta/2$ about its points lie in the ball of radius $1 + \delta/2$; volume comparison bounds its cardinality, and maximality gives the net property. If $\lambda_{\min}(G) \le e^{-t}$, there is a unit v with $\|\mathsf{D}v\| \le \sqrt{m}e^{-t/2}$. On the complement of the event in (226), $\|\mathsf{D}\|_{\mathrm{op}} \le \|\mathsf{D}\|_F \le W$. A net point w at distance at most δ from v then has

$$\|\mathsf{D}w\| \le \sqrt{m}e^{-t/2} + Rm\sqrt{k}\, e^{t/4}\delta = 2\sqrt{m}e^{-t/2}.$$

For fixed unit w, each $w^\top D_j$ has a density bounded by L: its density is the convolution of a density bounded by L with a reflected probability density. For $q = e^t/8$,

$$\mathbb{E}e^{-q(w^\top D_j)^2} \le L\int_{\mathbb{R}} e^{-qz^2}\, dz = L\sqrt{\pi/q}.$$

Independence across j and exponential Markov therefore give

$$\Pr\{\|\mathsf{D}w\|^2 \le 4me^{-t}\} \le e^{4qme^{-t}}(L\sqrt{\pi/q})^m \le K_L^m e^{-mt/2}.$$

The union bound over the net and (226) imply

$$\Pr\{F_G > t\} \le e^{-8m-2mt} + (3R\sqrt{mk})^k K_L^m e^{-(m/2-3k/4)t}.$$

Since $16k\log(ek) \le m \le 17k\log(ek)$ and $\log(ek) \le k$,

$$\begin{aligned} k\log(3R\sqrt{mk}) &\le k\log(3R) + \tfrac{k}{2}\log 17 + \tfrac{3k}{2}\log k \\ &\le \frac{m}{16}\{\log(3R) + \tfrac{1}{2}\log 17 + \tfrac{3}{2}\}. \end{aligned}$$

Also $m \ge 16k$ gives $m/2 - 3k/4 \ge m/4$. It follows that $\Pr\{F_G > t\} \le 2e^{Am - mt/4}$ for $t \ge 0$. With $r = m/8$, Tonelli's tail formula yields

$$\mathbb{E}e^{rF_G} = 1 + r\int_0^\infty e^{rt}\Pr\{F_G > t\}\, dt \le 1 + 2e^{Am} \le e^{C'm}.$$

This argument permits $F_G = +\infty$ initially; the finite bound proves that G is positive definite almost surely. For any $x, y, z \in \mathbb{R}^k$,

$$(x-z)(x-z)^\top + (y-z)(y-z)^\top = \tfrac{1}{2}(x-y)(x-y)^\top + 2\left(\tfrac{x+y}{2} - z\right)\left(\tfrac{x+y}{2} - z\right)^\top.$$

Taking $z = \bar{Z}$ and summing gives $S \succeq G/4$. Thus $F_S \le \log 4 + F_G$, proving (224), also when $k = 1$.

For a joint law P_n with $\mathrm{KL}(P_n\|\nu^n) \le n$, apply the entropy variational inequality to $r(F_S \wedge a)$, where a is finite. For bounded g, that inequality follows from nonnegativity of KL relative to $e^g\nu^n/\int e^g\, d\nu^n$. Using (224) gives

$$r\mathbb{E}_{P_n}(F_S \wedge a) \le 2m + \log \mathbb{E}_{\nu^n} e^{rF_S} \le 2m + C'm + r\log 4.$$

Monotone convergence as $a \to \infty$ proves (225) and almost-sure positive definiteness.

Finally, let $\mathcal{H}_i$ be the full history before input i. The conditional law given $Z_{<i}$ is a mixture of the laws given $\mathcal{H}_i$, since the latter contains the earlier inputs. Conditional convexity of KL and the finite-product chain rule give

$$\begin{aligned} \mathrm{KL}(P_n\|\nu^n) &= \sum_{i=1}^n \mathbb{E}\,\mathrm{KL}(\mathcal{L}(Z_i \mid Z_{<i})\|\nu) \\ &\le \sum_{i=1}^n \mathbb{E}\,\mathrm{KL}(\mathcal{L}(Z_i \mid \mathcal{H}_i)\|\nu) \le n. \end{aligned}$$

These conditional laws exist on the standard Borel input space. Finite conditional KL gives absolute continuity at every required step. No entropy of labels or other history variables is charged. This verifies the final assertion without independence assumptions. □

Proof without reference samples and with unknown covariance. Fix $\mu = N(a, \Sigma)$, where $\Sigma \succ 0$. Use the deterministic prefix length

$$m = \lceil 16k \log(ek) \rceil, \qquad n = 2m \le C_n k \log(ek), \qquad C_n = 34. \tag{227}$$

The bound follows from $m \le 17k \log(ek)$. If $T \le n$, use any fixed member of $\mathcal{F}$ throughout. The horizon comparison at the end bounds its regret. Suppose $T > n$. Use a fixed member of $\mathcal{F}$ during the first n rounds and collect their inputs. We construct a countably valued Cauchy reference $\widehat{\nu}$ such that

$$\mathrm{KL}(P\|\widehat{\nu}) \le 1 + \widehat{H} \quad \text{whenever } \mathrm{KL}(P\|\mu) \le 1, \qquad \mathbb{E}\widehat{H} \le C_2 k, \tag{228}$$

where C_2 is a specified universal constant below. Neither the construction nor its tuning uses a or Σ.

Conditional moments and the scatter. For the analysis write $Z_i = \Sigma^{-1/2}(X_i - a)$. Let $\mathcal{H}_i$ be the full state immediately before input i; round-i observations and feedback belong to $\mathcal{H}_{i+1}$. The conditional law of Z_i has KL at most one from $\gamma_k = N(0, I_k)$. Equation (197), applied to this conditional law, gives

$$\mathbb{E}[\|Z_i\|^2 \mid \mathcal{H}_i] \le B_k := 4 + 2k \log 2 \le 6k. \tag{229}$$

The conditional mean $b_i = \mathbb{E}[Z_i \mid \mathcal{H}_i]$ obeys $\|b_i\|^2 \le 2$ by (195). For clarity, the Gaussian linear tilt gives $v^\top b_i \le 1 + \|v\|^2/2$ for every v, and taking $v = b_i$ proves the bound. Put $\xi_i = Z_i - b_i$. For $i < j$, ξ_i is $\mathcal{H}_j$-measurable and $\mathbb{E}[\xi_j \mid \mathcal{H}_j] = 0$, so $\mathbb{E}\langle \xi_i, \xi_j \rangle = 0$. The conditional variance identity and (229) give $\mathbb{E}\|\xi_i\|^2 \le 6k$. Therefore, for the pooled whitened mean and scatter,

$$e = \frac{1}{n}\sum_{i=1}^{n} Z_i, \qquad S = \frac{1}{n}\sum_{i=1}^{n} (Z_i - e)(Z_i - e)^\top,$$

Jensen's inequality and the martingale-difference orthogonality yield

$$\begin{aligned} \mathbb{E}\|e\|^2 &\le 2\mathbb{E}\left\|\frac{1}{n}\sum_i b_i\right\|^2 + 2\mathbb{E}\left\|\frac{1}{n}\sum_i \xi_i\right\|^2 \le 4 + \frac{12k}{n} \le 16, \\ \mathbb{E}\,\mathrm{tr}\, S &\le \frac{1}{n}\sum_i \mathbb{E}\|Z_i\|^2 \le 6k. \end{aligned} \tag{230}$$

These bounds use full-history conditional moments, not independence or zero conditional means of the online inputs.

The reference γ_k satisfies Lemma S15.3 with $L = 1$, $B_0 = 2$, and $M = 2$. Its unit projections have density at most $(2\pi)^{-1/2} \le 1$. For $x \ge 0$, $x/2 \le x^2/4 + 1/4$, so the Gaussian square-norm integral gives

$$\mathbb{E}_{\gamma_k} e^{\|Z\|/(2\sqrt{k})} \le e^{1/4}(1 - 1/(2k))^{-k/2} \le e^{1/4}\sqrt{2} < 2.$$

The second inequality follows from Bernoulli's inequality $(1 - 1/(2k))^k \ge 1/2$. The conditional KL hypotheses were verified above. The lemma therefore gives $S \succ 0$ almost surely and, with $\lambda = \lambda_{\min}(S)$,

$$\mathbb{E}[\log(1/\lambda)]_+ \le L_0 := C_{\mathrm{inv}}(1, 2, 2).$$

This is a known universal constant. Neither the independent reference inputs in the lemma's proof nor the whitening transformation is used by the learner. Together with (230), Jensen's inequality and AM–GM give

$$\begin{aligned} \mathbb{E}\log(1 + 2/\lambda) &\le \log 3 + L_0, \\ \mathbb{E}[\log\det S]_+ &\le k \log 7, \\ \mathbb{E}\log(1 + \|e\|^2) &\le \log 17. \end{aligned} \tag{231}$$

Here $\log(1 + 2/\lambda) \le \log 3 + [\log(1/\lambda)]_+$, and $[\log\det S]_+ \le k \log(1 + \mathrm{tr}\, S/k)$.

A Cauchy reference and a uniform density bound. For any $e \in \mathbb{R}^k$ and $S \succ 0$, define

$$q_{e,S}(z) = \frac{\Gamma((k+1)/2)}{\pi^{(k+1)/2}\sqrt{\det S}}\big(1 + (z-e)^\top S^{-1}(z-e)\big)^{-(k+1)/2}. \tag{232}$$

This is a probability density. To see the normalization explicitly, let $U \sim N(0, I_k)$ and $V \sim N(0,1)$ be independent. The density of $U/|V|$ at w is

$$\int_0^\infty (2\pi)^{-k/2} v^k e^{-v^2\|w\|^2/2}\sqrt{2/\pi}\, e^{-v^2/2}\, dv = \frac{\Gamma((k+1)/2)}{\pi^{(k+1)/2}}(1+\|w\|^2)^{-(k+1)/2}.$$

The equality follows by substituting $t = v^2(1+\|w\|^2)/2$ into the gamma integral. The affine map $w \mapsto e + S^{1/2}w$ gives (232). Its original-coordinate version uses the observed sample mean and scatter, without knowing a or Σ.

For every z,

$$1 + (z-e)^\top S^{-1}(z-e) \le (1+2/\lambda)(1+\|e\|^2)(1+\|z\|^2).$$

Write ϕ_k for the standard Gaussian density and put

$$c_k = \tfrac{1}{2}\log\pi - \tfrac{k}{2}\log 2 - \log\Gamma((k+1)/2), \qquad x = (k+1)/2 \ge 1.$$

The maximum over $v \ge 0$ of $-v/2 + (k+1)\log(1+v)/2$ is attained at $v = k$. The gamma integral restricted to $[x, x+1]$ gives $\Gamma(x) \ge x^{x-1}e^{-x-1}$. Consequently the normalization constant and radial maximum satisfy

$$\begin{aligned} c_k - k/2 + \tfrac{k+1}{2}\log(k+1) &\le \tfrac{1}{2}\log(2\pi) + \tfrac{3}{2} + \log x \\ &\le \log(k+1) + 3. \end{aligned}$$

It follows pointwise in z that

$$\log\frac{\phi_k(z)}{q_{e,S}(z)} \le H(e,S), \tag{233}$$

where the finite nonnegative function

$$\begin{aligned} H(e,S) = {} & \log(k+1) + 3 + \tfrac{1}{2}[\log\det S]_+ \\ & + \tfrac{k+1}{2}\{\log(1+2/\lambda) + \log(1+\|e\|^2)\} \end{aligned}$$

has, by (231),

$$\mathbb{E}H(e,S) \le k + 3 + \tfrac{k}{2}\log 7 + k(\log 3 + L_0 + \log 17) \le (L_0 + 9)k.$$

We used $\log(k+1) \le k$, $(k+1)/2 \le k$, $\frac{1}{2}\log 7 < 1$, and $\log 51 < 4$.

For any P with $\mathrm{KL}(P\|\gamma_k) \le 1$, its second moment is finite by (197). For each fixed e, S, the absolute log-density ratio in (233) is bounded by a finite constant times $1 + \|z\|^2$. It is consequently integrable under P, and the change-of-reference identity gives

$$\mathrm{KL}(P\|q_{e,S}) \le 1 + H(e,S). \tag{234}$$

This bound holds simultaneously for every admissible P.

Countable selection of the reference. Let $\bar{X}, V$ be the observed-coordinate sample mean and scatter divided by n. Choose the largest dyadic power $h_V \le \lambda_{\min}(V)/(4k)$. Round the upper-triangular entries of V to nearest multiples of h_V using a fixed tie rule, and copy them symmetrically to obtain Q. The operator norm of the error is at most $kh_V/2 \le \lambda_{\min}(V)/8$, so

$$(7/8)V \preceq Q \preceq (9/8)V.$$

Choose the largest dyadic power $h_b \le \sqrt{\lambda_{\min}(V)}/(2\sqrt{k})$, and round $\bar{X}$ coordinatewise to obtain b. Then

$$(b-\bar{X})^\top V^{-1}(b-\bar{X}) \le 1/16.$$

The choices are Borel: eigenvalues, the floor of the dyadic logarithm, and fixed-rule rounding are Borel maps. The pair (b, Q) ranges over a countable set of rational vectors and rational positive-definite matrices. On the null event $V \not\succ 0$, set $(b, Q) = (0, I_k)$.

For every x, put $v = (x - \bar{X})^\top V^{-1}(x - \bar{X})$. The preceding estimates give

$$1 + (x - b)^\top Q^{-1}(x - b) \le 1 + (16/7)(v + 1/16) \le 3(1 + v).$$

Also $\det Q \le (9/8)^k \det V$. Substitution into (232) proves the pointwise bound

$$\log \frac{q_{\bar{X},V}(x)}{q_{b,Q}(x)} \le \tfrac{k}{2} \log(9/8) + \tfrac{k+1}{2} \log 3 \le 2k. \tag{235}$$

Take $\widehat{\nu}$ to have density $q_{b,Q}$ and put $\widehat{H} = H(e, S) + 2k$. Affine invariance of KL, (234) and (235) prove (228), since $\mathbb{E}\widehat{H} \le (L_0 + 11)k =: C_2 k$. The logarithmic density ratios remain integrable under every admissible P because $Q \succ 0$ and P has finite second moments.

The proper strategy on the remaining rounds. Put $h = T - n > 0$. If $h < d$, use the fixed hypothesis also on the remaining rounds. Their regret is at most $h < d \le A_{d,T}$, and the prefix is bounded in the horizon comparison below. Suppose $h \ge d$ and put $K = (L_0 + 12)k$. For every rational pair (b, Q) with $Q \succ 0$, fix the proper strategy of Lemma S15.2 for reference $q_{b,Q}$, class $\mathcal{F}$, dimension bound d, horizon h, and this same deterministic K. Its tuning is independent of the eventual cost; if $K \ge h/d$, the lemma uses a fixed hypothesis. The countable family and Borel choice of (b, Q) define a measurable proper learner independent of a, Σ.

Condition on the complete observed prefix transcript, including labels and learner randomness. For almost every transcript the selected reference is fixed and $\widehat{H}$ is finite. Every subsequent full state contains that transcript, so the original conditional KL constraint continues to hold at almost every subsequent full state. The conditional adversary may use the prefix and its permitted state; fresh learner draws retain the privacy convention. By (228), the conditional expected KL expenditure on the remaining rounds is at most $h(1 + \widehat{H})$. Lemma S15.2 therefore gives

$$\mathbb{E}[\mathrm{Reg}_{\mathrm{suffix}} \mid \mathrm{prefix}] \le C_1 \sqrt{K} A_{d,h} + C_1 \frac{A_{d,h}}{\sqrt{K}} (1 + \widehat{H}).$$

This affine guarantee holds for the actual conditional cost, even if that cost exceeds Kh on a particular transcript. Since $1 + \mathbb{E}\widehat{H} \le (L_0 + 12)k = K$, averaging gives

$$\mathbb{E}\,\mathrm{Reg}_{\mathrm{suffix}} \le 2C_1 \sqrt{K} A_{d,h} \le 2\sqrt{L_0 + 12}\, C_1 \sqrt{k} A_{d,T}.$$

The last inequality uses the monotonicity of $x \log(ex/d)$ for $x \ge d$. All costs are integrable by (228). No tuning or cover depends on the unknown expenditure.

Prefix regret is at most n. If $L_1(f), L_2(f)$ are the two segments' comparator losses, then $\inf_f (L_1(f) + L_2(f)) \ge \inf_f L_1(f) + \inf_f L_2(f)$. Whole-horizon regret is therefore at most the sum of the two segment regrets.

It remains to bound the prefix, including short horizons. For $1 \le d \le T$, the function $d \log(eT/d)$ is increasing in d, because its derivative is $\log(T/d) \ge 0$. Hence $A_{d,T}^2 \ge T \log(eT)$. If $T \le n$ and $T \le k$, then $T \le \sqrt{k} A_{d,T}$. If $k < T \le n$, (227) gives

$$\frac{T}{k \log(eT)} \le \frac{n}{k \log(ek)} \le C_n,$$

so the fixed learner's regret T is at most $\sqrt{C_n k} A_{d,T}$. If $T > n$, then $n \ge k$ and $n \le C_n k \log(en)$. Monotonicity of $x \log(ex)$ gives

$$n^2 \le C_n k n \log(en) \le C_n k T \log(eT) \le C_n k A_{d,T}^2.$$

This bounds the prefix by $\sqrt{C_n k} A_{d,T}$. For $0 < h < d$, the suffix bound $h < d \le A_{d,T}$ is absorbed as well. Combining all cases proves

$$\mathbb{E}\,\mathrm{Reg}_T \le C \sqrt{k} A_{d,T}$$

with universal C. The same learner has pathwise regret at most T, giving the stated minimum. The construction uses only the observed mean and scatter, rational rounding, and deterministic parameters depending on k, d, T. It imposes no bound on a, the covariance eigenvalues, or the condition number, and permits every adaptive protocol satisfying the full-history KL constraint. □

S15.5 A longer Gaussian calibration prefix

Density-comparison sets are used in minimum-distance estimation by Yatracos (1985); see also Ashtiani et al. (2020, Definition 2.6). We fit a Gaussian from adaptively chosen inputs with failure probability at most $1/T$, then use a known-reference strategy.

Theorem S15.4. *Let k, T, d be positive integers with $d \le T$, and let $\mathcal{F}$ be a known nonempty countable class of measurable binary functions on $\mathbb{R}^k$ with VC dimension at most d. There is a measurable proper pre-input learner, using no independent reference samples and receiving no mean or covariance, whose expected regret is at most*

$$\min\{T,\ C\sqrt{k}\,A_{d,T},\ C(k^2 + A_{d,T})\}$$

for every fixed $\mu = N(a, \Sigma)$ with $\Sigma \succ 0$ and every protocol satisfying the current-round privacy convention and $\mathrm{KL}(P_t\|\mu) \le 1$ conditional on the full preceding history. Here C is universal and $A_{d,T} = \sqrt{dT\log(eT/d)}$.

We first give the required uniform prefix estimate. Write $\mathcal{Q}_k$ for the countable class of sets $\{x : q(x) > 0\}$ and $\{x : q(x) \ge 0\}$, where q ranges over polynomials of degree at most two with rational coefficients. This class contains complements. Fix the universal constants

$$M = e^{128}, \qquad \delta = (10^4 M)^{-1}, \qquad v = \frac{(k+1)(k+2)}{2}.$$

Lemma S15.5. *For every integer $n \ge 2^{16}v$, every nonsingular Gaussian μ and every n-round process with the stated full-history conditional KL bound, with probability at least $1 - 3e^{-n/16}$ one has*

$$\sup_{A \in \mathcal{Q}_k :\, \mu(A) \le 2\delta} \sum_{t=1}^{n} \mathbf{1}\{X_t \in A\} \le n/4.$$

The same assertion holds conditionally on any preceding full history from which such a process is continued.

Proof. The feature vector of monomials of degrees zero, one and two has dimension v. Strict quadratic signs are homogeneous linear signs of these vectors, whose VC dimension is at most v: among $v+1$ vectors choose a nonzero linear dependence, reverse its sign if necessary, and prescribe positive signs at positive coefficients and nonpositive signs at negative coefficients. A realizing linear functional would make the dependence strictly positive, a contradiction. On any finite set, a closed quadratic sign is realized by a strict one after adding a sufficiently small positive rational constant. Thus the same trace bound applies to $\mathcal{Q}_k$. The Sauer bound used in Lemma S2.1 gives

$$\Pi_{\mathcal{Q}_k}(s) \le (es/v)^v \qquad (s \ge v).$$

Let $\mathcal{H}_{t-1}$ be the full history before input t. As in the truncation in Appendix S9, choose jointly measurable versions $r_t = dP_t/d\mu$, and put

$$C_t = \mathbf{1}\{r_t(X_t) > M\}, \quad D_n = \sum_{t=1}^{n} C_t, \qquad Z_t = X_t \text{ if } C_t = 0, \quad Z_t = * \text{ otherwise.}$$

Since $u \log u \ge -1/e$,

$$\Pr(C_t = 1 \mid \mathcal{H}_{t-1}) \le \frac{\mathrm{KL}(P_t\|\mu) + 1/e}{\log M} < \frac{1}{64}.$$

Iteration of the conditional bound $\mathbb{E}[e^{C_t} \mid \mathcal{H}_{t-1}] \le \exp((e-1)/64)$ and Markov's inequality give

$$\Pr(D_n > n/8) \le \exp\{-n/8 + (e-1)n/64\} \le e^{-n/16}. \tag{236}$$

Put $\rho_0 = (\mu + \delta_*)/2$. Conditional on the full original history, Z_t has a density $g_t \le 2M$ relative to ρ_0. Couple these variables to one infinite sequence of iid ρ_0 proposals, with independent uniforms. At each trial in round t, accept a proposal V when its uniform is at most $g_t(V)/(2M)$, and continue until acceptance. Given the entire previously revealed construction, each trial accepts with probability

$\sigma = 1/(2M)$; its accepted value has the required law of Z_t. After acceptance, reconstruct the original X_t and the remaining round variables from their regular conditional law given that value and the full original past. These kernels exist by the standard-Borel model. They preserve the full original process, including labels, learner randomness and adversary state. Fresh independent seeds are used for this reconstruction. In particular future densities may depend on the reconstructed history, rather than only on the Z_t. This is an analysis coupling and supplies no proposals to the learner.

Let τ_n be the number of proposals used for n accepted values and $L = \lceil 8Mn \rceil$. Continue trials with density one after the nth acceptance if necessary. If S_L counts acceptances in the first L trials, its conditional success probability at each trial remains σ. Iterating its exponential moment therefore gives

$$\Pr(\tau_n > L) = \Pr(S_L < n) \le e^n \mathbb{E} e^{-S_L} = e^n (1 - \sigma + \sigma/e)^L \le \exp\{n - (1 - e^{-1}) L\sigma\} \le e^{-n}.$$

All of the first L proposals are iid ρ_0, without conditioning on this stopping event. Off that event every Z_t is among them.

Extend the indicators of $A \subseteq \mathbb{R}^k$ by zero at $*$; adding this common-valued point does not increase their trace count. The nonempty countable class with $\mu(A) \le 2\delta$ has ρ_0-probabilities at most δ. Lemma S2.2, with $N = L$ and exponential parameter one, bounds its iid-pool maximum W by

$$\log \mathbb{E} e^W \le \log 2 + v \log(3eL/(\delta v)) + 5L\delta(e-1).$$

Indeed its ghost-pool size is $L + \lceil L/\delta \rceil \le 3L/\delta$ and is at least v. Since $L \le 9Mn$, put $K_0 = 27eM/\delta$, for which $\log K_0 < 300$. For $x \ge 2^{16}$,

$$\frac{d}{dx} \frac{\log(K_0 x)}{x} = \frac{1 - \log(K_0 x)}{x^2} < 0, \qquad \frac{\log(K_0 2^{16})}{2^{16}} < \frac{300 + 16 \log 2}{2^{16}} < \frac{1}{64}.$$

Applying this at $x = n/v$ gives

$$v \log(K_0 n/v) \le n/64, \qquad \log 2 \le n/64, \qquad 5L\delta(e-1) \le 90n/10^4 \le n/64.$$

Consequently $\Pr(W > n/8) \le e^{-n/16}$ by Markov's inequality. On the complement of these three exceptional events, every such A has at most $D_n + W \le n/4$ original observations. Their total probability is at most $3e^{-n/16}$. Every argument applies to a regular conditional continuation with the same fixed μ and KL bound, proving the last assertion. □

Lemma S15.6. *For every integer $n \ge 2^{16} v$, there is a countably valued measurable Gaussian fit $\widehat{\rho}$ from these n inputs, independent of μ, such that*

$$\Pr\{D_{\mathrm{B}}(\mu, \widehat{\rho}) > c_0\} \le 3e^{-n/16}, \qquad c_0 = \log(4/\delta), \qquad D_{\mathrm{B}}(\mu, \rho) = -\log \int \sqrt{d\mu \, d\rho}.$$

Proof. Enumerate Gaussians with rational mean and rational positive-definite covariance. Call a candidate ρ feasible if, for every $A \in \mathcal{Q}_k$ with $\rho(A) \ge 1 - \delta$,

$$n^{-1} \sum_{t=1}^{n} \mathbf{1}\{X_t \in A\} \ge 3/4.$$

Choose the first feasible candidate, or $N(0, I_k)$ if none exists. Each feasibility event is a countable intersection of Borel events; first selection and the default are therefore Borel and countably valued.

On the event of Lemma S15.5, any rational Gaussian at TV distance less than δ from μ is feasible: the complement of each tested set belongs to $\mathcal{Q}_k$ and has μ-mass at most 2δ. Such a Gaussian exists without bounds on the unknown parameters. Approximate the mean and covariance by rational parameters, keeping positive definiteness. The resulting densities converge pointwise to that of μ, and in L^1: the integrals of their minima with the density of μ tend to one by dominated convergence.

Suppose a feasible ρ has $D_{\mathrm{B}}(\mu, \rho) > c_0$. The set $A = \{x : \rho(x) \ge \mu(x)\}$ satisfies

$$\mu(A) \le e^{-D_{\mathrm{B}}(\mu,\rho)} < \delta/4, \qquad \rho(A^c) < \delta/4,$$

because the smaller of two densities is bounded by their geometric mean on each of the two sets. The polynomial $\log \rho - \log \mu$ is a nonzero quadratic, possibly of lower degree. Its zero set has Lebesgue

measure zero. To justify this fact, regard a nonzero polynomial as a polynomial in one coordinate with polynomial coefficients in the others: outside the null set where all its nonzero coefficients vanish, it has finitely many roots; induction on dimension and Fubini complete the argument. Both Gaussians therefore give its zero set measure zero. Rational approximation of its coefficients gives sets $A_j \in \mathcal{Q}_k$ whose indicators converge off that zero set. Dominated convergence supplies a j with $\mu(A_j) < \delta/2$ and $\rho(A_j) > 1 - \delta/2$. Feasibility requires empirical mass at least $3/4$, while the rare-set event bounds it by $1/4$, a contradiction. The asserted failure probability follows. No effective procedure for deciding all feasibility tests is claimed. □

Lemma S15.7. *There are universal constants $\eta > 0$ and $R < \infty$, depending only on the fixed c_0 above, such that nonsingular Gaussians with $D_{\mathrm{B}}(\mu, \rho) \le c_0$ satisfy $D_{1+\eta}(\mu\|\rho) \le R$ in every dimension. Consequently every P with $\mathrm{KL}(P\|\mu) \le 1$ has $\mathrm{KL}(P\|\rho) \le K_1 := (1+\eta)/\eta + R$.*

Proof. Whiten μ and rotate to write $\mu = N(0, I_k)$ and $\rho = N(b, \mathrm{diag}(\lambda_1, \dots, \lambda_k))$. The identity $\int_{\mathbb{R}} e^{-Az^2/2+tz}\, dz = \sqrt{2\pi/A}\, e^{t^2/(2A)}$ for $A > 0$, obtained by completing the square, gives

$$D_{\mathrm{B}}(\mu, \rho) = \sum_i h(\lambda_i) + \sum_i \frac{b_i^2}{4(1+\lambda_i)}, \qquad h(\lambda) = \frac{1}{2} \log \frac{1+\lambda}{2\sqrt{\lambda}}.$$

All summands are nonnegative. With $l = e^{-4c_0}/4$, the inequality $h(\lambda_i) \le c_0$ implies $l \le \lambda_i \le 1/l$: use $(1+\lambda)/(2\sqrt{\lambda}) \ge \sqrt{\lambda}/2$ and $\ge 1/(2\sqrt{\lambda})$. Set $\eta = l/2$, $\alpha = 1 + \eta$ and $B(\lambda) = \alpha\lambda - \eta \ge \lambda/2$ on this interval.

For one coordinate the integrand in $\int \mu^\alpha \rho^{-\eta}$ is

$$(2\pi)^{-1/2} \lambda^{\eta/2} \exp\left\{ -\frac{B(\lambda)}{2\lambda} z^2 - \frac{\eta b}{\lambda} z + \frac{\eta b^2}{2\lambda} \right\}.$$

The same integral identity and division of its logarithm by η give the dimension-additive formula

$$D_\alpha(\mu\|\rho) = \sum_i g(\lambda_i) + \sum_i \frac{\alpha b_i^2}{2B(\lambda_i)}, \qquad g(\lambda) = \frac{\alpha \log \lambda - \log B(\lambda)}{2\eta}.$$

Both g and h vanish at one, and

$$h'(\lambda) = \frac{\lambda - 1}{4\lambda(1+\lambda)}, \qquad g'(\lambda) = \frac{\alpha(\lambda - 1)}{2\lambda B(\lambda)}.$$

Their derivative ratio away from one is $2\alpha(1+\lambda)/B(\lambda) \le R_0 := 4\alpha(1 + 1/l)$. Integrating from one to λ proves $g(\lambda) \le R_0 h(\lambda)$; for $\lambda < 1$ both derivatives are negative and the integral is taken in the reverse direction. The ratio of the two mean coefficients is the same expression and has the same bound. Thus $D_\alpha(\mu\|\rho) \le R_0 c_0 =: R$.

Lemma S14.1, with $\alpha = 1 + \eta$, now gives

$$\mathrm{KL}(P\|\rho) \le \frac{1+\eta}{\eta} \mathrm{KL}(P\|\mu) + D_{1+\eta}(\mu\|\rho) \le K_1.$$

Its hypotheses are exactly the finite KL and Renyi bounds just stated. □

Proof of Theorem S15.4. Use a fixed member of $\mathcal{F}$ for an initial prefix of length

$$n = \left\lceil \max\{2^{16} v,\ 16 \log(3T)\} \right\rceil.$$

If $T \le n$, use this hypothesis throughout. Otherwise fit $\widehat{\rho}$ from the n observed inputs as above and put $h = T - n$. Let

$$\mathsf{G} = \{D_{\mathrm{B}}(\mu, \widehat{\rho}) \le c_0\}.$$

This event is determined by the observed prefix for each fixed μ and is used only in the analysis. Lemma S15.6 gives $\Pr(\mathsf{G}^c) \le 3e^{-n/16} \le 1/T$.

If $0 < h < d$, continue the fixed hypothesis. Its suffix regret is at most $h < d \le A_{d,T}$. Suppose $h \ge d$. For every rational Gaussian candidate ρ, including the default, fix the proper strategy of Lemma S15.2 with reference ρ, horizon h, class $\mathcal{F}$, dimension bound d and the known universal

tuning K_1 from Lemma S15.7. Select this strategy using $\widehat{\rho}$ on every prefix, without testing G. Countability of the candidates and the Borel selection make this one measurable learner, independent of μ. The suffix uses fresh private hypothesis draws under the same current-round privacy convention.

Condition on the full prefix history, including labels, past learner draws and the permitted adversary state. For almost every such history in G, the fitted reference is fixed and Lemma S15.7 gives

$$\mathrm{KL}(P\|\widehat{\rho}) \le K_1 \quad \text{for every } P \text{ with } \mathrm{KL}(P\|\mu) \le 1.$$

Every subsequent full history extends this prefix, so the original conditional KL constraint still holds almost surely. Thus the conditional expected total KL cost on the suffix is at most hK_1. The scaled-cost strategy gives

$$\mathbb{E}[\mathrm{Reg}_{\text{suffix}} \mid \text{prefix}] \le C_1\sqrt{K_1}A_{d,h} + C_1\frac{A_{d,h}}{h\sqrt{K_1}}hK_1 = 2C_1\sqrt{K_1}A_{d,h} \quad \text{on } \mathsf{G}.$$

Averaging over any unobserved prefix state preserves this bound. The strategy uses only observed history; the conditional argument allows every permitted adversary continuation.

On G^c use only the pathwise bound $\mathrm{Reg}_{\text{suffix}} \le h$. No bound on its transferred KL cost is needed. Splitting the expectation by G gives

$$\mathbb{E}\mathrm{Reg}_{\text{suffix}} \le 2C_1\sqrt{K_1}A_{d,h} + h\Pr(\mathsf{G}^c) \le 2C_1\sqrt{K_1}A_{d,T} + 1.$$

Here $A_{d,h} \le A_{d,T}$ because $\frac{d}{dx}[x\log(ex/d)] = \log(ex/d) + 1 > 0$ for $x \ge d$. The prefix is at most

$$n \le 3\cdot 2^{16}k^2 + 32\log(eT) + 1 \le C_0\big(k^2 + \log(eT)\big), \qquad C_0 = 3\cdot 2^{16} + 33,$$

since $v \le 3k^2$, $\log(3T) \le 2\log(eT)$ and $k, T \ge 1$. For $1 \le d \le T$, the function $d\log(eT/d)$ is increasing in d, as its derivative is $\log(T/d) \ge 0$. Consequently

$$A_{d,T}^2 \ge T\log(eT) \ge \log^2(eT), \qquad \log(eT) \le A_{d,T},$$

where $\log(eT) \le T$ follows from $\log T \le T - 1$. Also $d \le A_{d,T}$, since $A_{d,T}^2 \ge dT \ge d^2$. Prefix regret is at most n. Full regret is at most the sum of segment regrets, since the infimum of the sum of comparator losses is at least the sum of their infima. The preceding estimates prove

$$\mathbb{E}\mathrm{Reg}_T \le C'(k^2 + A_{d,T})$$

with universal C', including $T \le n$, when regret is at most $T \le n$, and the short-suffix case $0 < h < d$. The prefix length is $O(k^2 + \log T)$; the learner receives only $T, d, k, \mathcal{F}$ and the observed history. It does not use G or the conditional laws in the analysis coupling.

The online-only Gaussian construction in Appendix S15.4 already has bound $C''\sqrt{k}A_{d,T}$. Choose between these two learners before play by comparing the known quantities $k^2 + A_{d,T}$ and $\sqrt{k}A_{d,T}$. Their larger universal constant bounds the smaller quantity. Every learner's regret is also at most T, giving the three-way minimum in the statement. All choices precede the current input. This construction gives no efficient search or finite-bit guarantee and asserts no optimality of the prefix's dimension dependence. □

S15.6 Unknown log-concave references

We prove Corollary 5.2 using the Cauchy reference from (232). The log-concave facts below are Theorem 3.1, Lemma 3.4(a), and Theorems 3.9(e) and 3.17 of the Newton preprint NI02055 of Lovász and Vempala (2007); these numbers refer to that version. We account for the dimension restriction in its Section 3.2. Determinant anti-concentration for adaptive inputs under density-ratio smoothing relative to log-concave measures appears in Block and Simchowitz (2022, supplement, Proposition 41). Here the prefix is controlled only in conditional KL, and we bound the expected KL cost of the fitted reference.

Proof of Corollary 5.2. Choose the upper-semicontinuous version of the log-concave density of μ. Its mean a and covariance Σ are finite. To justify this, each coordinate marginal is an integrable one-dimensional log-concave density g, by the cited marginal theorem. If its support is unbounded to the right, choose interior points $x_0 < x_1$ with $0 < g(x_1) < g(x_0) < \infty$. Such an x_1 exists, since

otherwise g would be bounded below by $g(x_0)$ on the right half-line. Concavity gives, for $x \geq x_1$ in the support,

$$\log g(x) \leq \log g(x_0) + \frac{\log g(x_1) - \log g(x_0)}{x_1 - x_0}(x - x_0).$$

This bounds that tail by a decaying exponential. If the support is bounded to the right, that tail is zero instead. Apply the same argument to $g(-x)$ for the left tail. Integrability on compact intervals then gives finite second moments for every coordinate. If $v^\top \Sigma v = 0$ for a nonzero v, the reference would be supported on a hyperplane, contrary to full dimensionality. Thus $\Sigma \succ 0$.

For analysis alone, put $Z = \Sigma^{-1/2}(X - a)$ and let f be its reference density. Then f is isotropic and log-concave, and affine changes preserve KL. No whitening parameter is supplied to the learner. Write $\ell_k = \log(ek)$.

Reference density and input first moments. Section 3.2 of NI02055 assumes its ambient dimension exceeds five. Let $W = (Z, G)$ under the reference, where $G \sim N(0, I_6)$ is independent. Its product density is isotropic and log-concave in dimension $k+6 \geq 7$. Theorem 3.9(e), with its bound weakened from $2^{8(k+6)}$ to $2^{9(k+6)}$, gives, by evaluating at $(z, 0)$,

$$f(z) \leq (2\pi)^3\, 2^{9(k+6)}(k+6)^{(k+6)/2} \leq e^{60k\ell_k}. \tag{237}$$

Indeed, $k + 6 \leq 7k$ and $\log(k+6) \leq \log 7 + \log k$ bound the logarithm by $56k + (7/2)k \log k \leq 60k\ell_k$. The same product extension permits Theorem 3.17: for every integer $q \geq 1$,

$$(\mathbb{E}_f \|Z\|^q)^{1/q} \leq (\mathbb{E}\|W\|^q)^{1/q} \leq 2q\, \mathbb{E}\|W\| \leq 2q\sqrt{k+6} \leq 2q\sqrt{7k}.$$

The penultimate step is Cauchy–Schwarz and isotropy. The elementary bound $q! \geq (q/e)^q$ follows by integrating $\log x$ between 1 and q. Tonelli's theorem applied to the nonnegative exponential series therefore gives

$$\mathbb{E}_f \exp\left(\frac{\|Z\|}{8e\sqrt{7k}}\right) \leq \sum_{q=0}^{\infty} 4^{-q} = \frac{4}{3}.$$

For bounded g, nonnegativity of KL relative to the density $e^g f / \int e^g f$ gives $Pg \leq \mathrm{KL}(P\|f) + \log \int e^g f$. Apply this to $g_R(z) = \min\{\|z\|, R\}/(8e\sqrt{7k})$ and use monotone convergence. Thus, with a universal constant

$$B = 8e\sqrt{7}(1 + \log(4/3)), \qquad \mathbb{E}_P\|Z\| \leq B\sqrt{k} \quad \text{if } \mathrm{KL}(P\|f) \leq 1. \tag{238}$$

Neither the auxiliary Gaussian coordinates nor the whitening transformation is part of the learner. No second moment of P is asserted or used.

We also use a sharper bound for the isotropic density. Write $h(f) = -\int f \log f$. Equation (6) of the published Klartag and Lehec (2025), with their definition (3), gives $e^{-h(f)/k} \leq C_s$ for a universal $C_s \geq 1$. This statement includes all log-concave probability densities, without a symmetry assumption, in every dimension. Proposition I.2 of Bobkov and Madiman (2011), arXiv version 2, gives $h(f) \leq k - \log \|f\|_\infty$. Consequently

$$\|f\|_\infty \leq C_0^k, \qquad C_0 = eC_s. \tag{239}$$

An essential-supremum bound suffices below. The coarser (237) and the explicit first-moment bound remain available for the other calibration consequences.

The prefix mean and minimum-determinant scatter. By Theorem 3.1 and Lemma 3.4(a), every unit projection of the isotropic reference has density at most one. These one-dimensional statements precede the dimension restriction. Together with the exponential moment proved above, this verifies Lemma S15.3 with $L = 1$, $B_0 = 8e\sqrt{7}$, and $M = 4/3$. Take the deterministic prefix sizes

$$m = \lceil 16k\ell_k \rceil, \qquad n = 4m \leq C_n k\ell_k, \qquad C_n = 68.$$

If $T \leq n$, use a fixed hypothesis throughout; its regret is bounded in the horizon comparison below. For $T > n$, use a fixed hypothesis during the first n rounds and collect their inputs. Let $\mathcal{H}_i$ be the

full state immediately before input i; its observations and feedback belong to $\mathcal{H}_{i+1}$. For analysis put $Z_i = \Sigma^{-1/2}(X_i - a)$ and define

$$e = \frac{1}{n}\sum_{i=1}^{n} Z_i, \qquad e_J = \frac{1}{2m}\sum_{i\in J} Z_i,$$
$$S_J = \frac{1}{2m}\sum_{i\in J}(Z_i - e_J)(Z_i - e_J)^\top \qquad (|J| = 2m).$$

Select J_* minimizing $\det S_J$, with lexicographic ties, and put $S = S_{J_*}$. This is the scatter part of the minimum covariance determinant construction; see Rousseeuw and Van Driessen (1999, p. 213). Our Cauchy center is the ordinary mean e of all n inputs, not the selected subset's mean. The learner performs the minimization in observed coordinates, as verified below.

Let P_n be the marginal law of this standardized prefix. The conditional law given $Z_{<i}$ is a mixture of those given $\mathcal{H}_i$. Conditional convexity and the KL chain rule give

$$\begin{aligned}\mathrm{KL}(P_n \| f^{\otimes n}) &= \sum_{i=1}^{n} \mathbb{E}\,\mathrm{KL}(\mathcal{L}(Z_i \mid Z_{<i}) \| f) \\ &\le \sum_{i=1}^{n} \mathbb{E}\,\mathrm{KL}(\mathcal{L}(Z_i \mid \mathcal{H}_i) \| f) \le n. \end{aligned} \tag{240}$$

This is a bound on the input prefix without conditioning on its labels. For every nonnegative Borel function V and $r > 0$ with finite reference exponential moment, the entropy inequality applied first to $r(V \wedge b)$ and then monotone convergence give

$$r\mathbb{E}_{P_n} V \le n + \log \mathbb{E}_{f^{\otimes n}} e^{rV}. \tag{241}$$

We use this bound for two functions of the prefix.

First let $F_J = \log^+(1/\lambda_{\min}(S_J))$, with value $+\infty$ at a singular scatter. Under $f^{\otimes n}$, each fixed $2m$-element subset has the iid law required in (224). Denote its universal constant C' for $(L, B_0, M) = (1, 8e\sqrt{7}, 4/3)$ by C_0'. Since the exponential of a maximum is at most the sum,

$$\mathbb{E}_{f^{\otimes n}} \exp\{(m/8)\max_J F_J\} \le \binom{n}{2m} \exp\{(C_0' + \log 4/8)m\} \le \exp\{n\log 2 + (C_0' + \log 4/8)m\}.$$

No independence between different subsets is needed. Equation (241) and $n = 4m$ imply

$$\mathbb{E}_{P_n} \max_J F_J \le L_0 := 32(1 + \log 2) + 8C_0' + \log 4.$$

In particular every subset scatter is positive definite almost surely under the actual prefix law. Thus $S \succ 0$ almost surely and, with $\lambda = \lambda_{\min}(S)$, $\mathbb{E}\log^+(1/\lambda) \le L_0$. All scatters are finite for every finite real prefix; their expected traces need not be finite.

Next, under the iid reference law, let R be the $2m$-th smallest of the n numbers $\|Z_i\|/\sqrt{k}$. For the subset J of the $2m$ smallest norms, $\operatorname{tr} S_J = (2m)^{-1}\sum_{i\in J}\|Z_i\|^2 - \|e_J\|^2 \le kR^2$. AM–GM and determinant minimization give

$$(\det S)^{1/k} \le R^2, \qquad D_+ := [\log\det S]_+ \le 2k\log^+ R.$$

For $r \ge 1$, isotropy and Markov's inequality give $\Pr_f\{\|Z\| > r\sqrt{k}\} \le r^{-2}$. If $R > r$, at least $2m + 1$ input norms exceed $r\sqrt{k}$. Requiring only $2m$ such inputs and using their independence therefore gives $\Pr_{f^{\otimes n}}\{R > r\} \le 2^n r^{-4m} = 2^n r^{-n}$. Tonelli's tail formula now yields

$$\begin{aligned}\mathbb{E}_{f^{\otimes n}} e^{nD_+/(8k)} &\le \mathbb{E}_{f^{\otimes n}} e^{(n/4)\log^+ R} \\ &\le 1 + \frac{n}{4}\int_0^\infty e^{nt/4} 2^n e^{-nt}\,dt = 1 + 2^n/3 \le 2^n.\end{aligned}$$

Applying (241) gives $\mathbb{E}D_+ \le C_D k$, where $C_D = 8(1 + \log 2)$. This controls the determinant directly, without an estimate of $\mathbb{E}\operatorname{tr} S$.

Applying (238) conditionally on each $\mathcal{H}_i$, then averaging and using the triangle inequality, gives $\mathbb{E}\|e\| \le n^{-1}\sum_{i=1}^n \mathbb{E}\|Z_i\| \le B\sqrt{k}$. Since $1+x^2 \le (1+x)^2$, Jensen's inequality then yields

$$\mathbb{E}\log(1+\|e\|^2/k) \le 2\mathbb{E}\log(1+\|e\|/\sqrt{k}) \le L_e := 2\log(1+B). \tag{242}$$

The preceding scatter estimates give

$$\mathbb{E}\log(1+2/\lambda) \le \log 3 + L_0, \qquad \mathbb{E}[\log\det S]_+ \le C_D k. \tag{243}$$

These estimates require only first moments under the actual adaptive input law, and hold also when $k=1$.

Integrated KL transfer to a Cauchy reference. Use $q_{e,S}$ from (232), whose probability normalization was proved there, and put $c_k = \Gamma((k+1)/2)/\pi^{(k+1)/2}$. The gamma integral on $[1,2]$ gives $\Gamma((k+1)/2) \ge e^{-2}$, so $-\log c_k \le 4k$. For the sharper bound needed here, put $x = (k+1)/2 \ge 1$. Integration on $[x, x+1]$ gives $\Gamma(x) \ge x^{x-1}e^{-x-1}$. Since $k/x \le 2$, $\log x \le x-1$, and $x \le k$,

$$-\log c_k + \frac{k+1}{2}\log k \le x\log(2\pi) + x + \log x + 1 \le (\log(2\pi)+2)k < 4k. \tag{244}$$

For every z,

$$\begin{aligned} 1 + (z-e)^\top S^{-1}(z-e) &\le 1 + (2/\lambda)(\|z\|^2 + \|e\|^2) \\ &\le k(1+2/\lambda)(1+\|z\|^2/k)(1+\|e\|^2/k). \end{aligned}$$

The second inequality follows by expanding the right side and using $k \ge 1$. Equations (239) and (244) bound $\log(f(z)/q_{e,S}(z))$, for almost every z with $f(z) > 0$, by the nonnegative function

$$\begin{aligned} g_{e,S}(z) = k(\log C_0 + 4) &+ \tfrac{1}{2}[\log\det S]_+ \\ &+ \tfrac{k+1}{2}\{\log(1+2/\lambda) + \log(1+\|e\|^2/k) + \log(1+\|z\|^2/k)\}. \end{aligned}$$

For every P with $\mathrm{KL}(P\|f) \le 1$, (238) and Jensen's inequality give

$$P\log(1+\|Z\|^2/k) \le 2P\log(1+\|Z\|/\sqrt{k}) \le 2\log(1+B) =: L_B.$$

Thus $L_B = L_e$. Hence $Pg_{e,S} \le H_k(e,S)$ simultaneously for all such P, where

$$\begin{aligned} H_k(e,S) = k(\log C_0 + 4) &+ \tfrac{1}{2}[\log\det S]_+ \\ &+ \tfrac{k+1}{2}\{\log(1+2/\lambda) + \log(1+\|e\|^2/k) + L_B\}, \\ \mathbb{E}H_k(e,S) \le C_H k, \qquad C_H &= \log C_0 + 4 + \tfrac{1}{2}C_D + \log 3 + L_0 + L_e + L_B. \end{aligned}$$

The expectation uses (242) and (243), with $(k+1)/2 \le k$.

We justify the change of reference before integrating logarithms. The measure P has a Lebesgue density p, since $P \ll f$. At their P-almost-sure finite values put $Y = \log(p/f)$ and $X = \log(p/q_{e,S})$. For any probability reference ρ with $P \ll \rho$, writing $r = dP/d\rho$ gives

$$\int \left[\log\frac{dP}{d\rho}\right]_- dP = \int_{\{r<1\}} r\log(1/r)\,d\rho \le 1/e.$$

Thus finite KL makes Y integrable. The pointwise chain rule and the preceding bound give $X \le Y + g_{e,S}$, so $X_+ \le Y_+ + g_{e,S}$ is integrable under P. The displayed inequality with $\rho = q_{e,S}$ also makes X_- integrable. Now $\log(f/q_{e,S}) = X - Y$ is integrable, and the KL identity is legitimate. It yields

$$\mathrm{KL}(P\|q_{e,S}) = \mathrm{KL}(P\|f) + P\log(f/q_{e,S}) \le 1 + H_k(e,S). \tag{245}$$

This is an integrated bound, not a pointwise density domination by $e^{H_k(e,S)}$. It allows the log-concave density to vanish.

Observed parameters, countable selection and the suffix. The learner computes the ordinary observed mean $\bar{X}$ and, for each $2m$-element subset J, its centered scatter V_J divided by $2m$. It chooses the least determinant with the same lexicographic rule and writes $V = V_{J_*}$. Under whitening, $V_J = \Sigma^{1/2} S_J \Sigma^{1/2}$, so every determinant is multiplied by the same positive $\det \Sigma$. The selected index is therefore unchanged, and $\bar{X} = a + \Sigma^{1/2} e$, $V = \Sigma^{1/2} S \Sigma^{1/2}$. Finite minimization of continuous determinants with fixed ties is Borel. This exact subset selection is used only as a statistical construction; no efficient subset-search algorithm is asserted. Apply the dyadic construction preceding (235) to this observed pair; its proof requires only a center and a positive-definite scatter, not that they use the same mean. Round symmetric entries of V at the largest dyadic mesh at most $\lambda_{\min}(V)/(4k)$ to obtain Q, and round $\bar{X}$ at the largest dyadic mesh at most $\sqrt{\lambda_{\min}(V)}/(2\sqrt{k})$ to obtain b. Use the fixed tie rules and set $(b, Q) = (0, I_k)$ if $V \not\succ 0$, an event of probability zero here. The deterministic bounds proved there give

$$(7/8)V \preceq Q \preceq (9/8)V, \qquad (b - \bar{X})^\top V^{-1} (b - \bar{X}) \le 1/16, \qquad q_{\bar{X},V} \le e^{2k} q_{b,Q}.$$

Eigenvalues, dyadic logarithms, and fixed-rule rounding are Borel; the index (b, Q) lies in a countable set of rational vectors and rational positive-definite matrices. Affine invariance and (245) give the unrounded bound in observed coordinates. The last pointwise inequality adds at most $2k$ to KL. Integrability follows from the same positive-part and negative-part argument just given. With $\widehat{\nu} = q_{b,Q}$ and $\widehat{H} = H_k(e, S) + 2k$, we have simultaneously for every admissible P,

$$\mathrm{KL}(P \| \widehat{\nu}) \le 1 + \widehat{H}, \qquad \mathbb{E}\widehat{H} \le (C_H + 2)k. \tag{246}$$

The measure P may be selected after observing the prefix.

Put $h = T - n > 0$. If $h < d$, continue with the fixed hypothesis; its suffix regret is at most $h < d \le A_{d,T}$. For $h \ge d$, set the known deterministic cost scale $K = (C_H + 3)k$. For every rational pair (b, Q), fix the proper pre-input strategy of Lemma S15.2 for reference $q_{b,Q}$, class $\mathcal{F}$, dimension bound d, horizon h and this same K. Countability and Borel selection make the resulting learner measurable, independently of the unknown shape, mean and covariance.

Condition on the complete observed prefix transcript, including labels and learner randomness. The reference index is then fixed and $\widehat{H}$ is finite almost surely. Every subsequent full state includes that transcript, so the original conditional KL constraint persists. The adversary may use its permitted state and the prefix; fresh learner draws retain the current-round privacy convention. Equation (246) bounds the actual conditional expected suffix cost by $h(1 + \widehat{H})$. The selected strategy therefore satisfies

$$\mathbb{E}[\mathrm{Reg}_{\mathrm{suffix}} \mid \mathrm{prefix}] \le C_1 \sqrt{K} A_{d,h} + C_1 \frac{A_{d,h}}{\sqrt{K}} (1 + \widehat{H}).$$

This is its affine actual-cost guarantee; no transcriptwise bound by Kh or random tuning is used. Since $1 + \mathbb{E}\widehat{H} \le K$, averaging and monotonicity of $x \log(ex/d)$ for $x \ge d$ give

$$\mathbb{E}\,\mathrm{Reg}_{\mathrm{suffix}} \le 2C_1 \sqrt{K} A_{d,h} \le 2C_1 \sqrt{C_H + 3}\sqrt{k}\, A_{d,T}.$$

All expected costs are finite by (246).

Whole-horizon regret is at most the sum of the segment regrets: $\inf_f (L_1(f) + L_2(f)) \ge \inf_f L_1(f) + \inf_f L_2(f)$. The prefix costs at most $n \le C_n k \ell_k$. For completeness, $A_{d,T}^2 \ge T \log(eT)$, since $d \log(eT/d)$ is increasing for $1 \le d \le T$. If $T \le k$, then $T \le \sqrt{k} A_{d,T}$. If $k < T \le n$, then $T/(k \log(eT)) \le C_n$. If $T > n$, then $n \ge k$ and

$$n^2 \le C_n k n \log(en) \le C_n k T \log(eT) \le C_n k A_{d,T}^2.$$

These bounds absorb the fixed learner for $T \le n$ and the prefix for $T > n$ into $\sqrt{C_n k} A_{d,T}$. The case $0 < h < d$ adds at most $A_{d,T}$. Combining all cases proves $\mathbb{E}\,\mathrm{Reg}_T \le C\sqrt{k} A_{d,T}$ with universal C. The same learner has pathwise regret at most T, proving the cap. It uses no independent reference samples or unknown parameters. Only the fixed reference is assumed log-concave. □

S16 Joint reference samples and forward moments: the general upper

Let X and all round-state spaces be standard Borel, and let $\mathcal{F} = (f_i)_{i \ge 1}$ be a known nonempty finite or countable class of measurable binary functions with $\mathrm{VC}(\mathcal{F}) \le d$, where $d \ge 1$ is fixed. A

finite enumeration may be repeated. The learner knows a mathematical Borel probability ν. Finite proper covers under ν and their probabilities are available as mathematical objects, as in the statistical known-reference model. One unknown Borel reference μ is fixed before play and satisfies $D_{1+\eta}(\mu\|\nu) \le 1$. Every input law conditional on the full preceding history satisfies $\mathrm{KL}(P_t\|\mu) \le 1$ almost surely. The history includes preprocessing, preceding learner choices, initial state, and preceding adversary variables. The current input-label kernel is independent of the fresh private hypothesis choice for that round. Labels are otherwise adaptive and agnostic. A reference request returns a fresh independent μ draw.

For $0 < \eta \le 1$ and $m \ge 0$, write

$$S = \log(eT(m+1)), \qquad A_{d,T} = \sqrt{dT\log(eT/d)},$$
$$G_{T,m}(\eta) = \min\left\{\frac{T}{S}, \sqrt{\frac{T}{\eta}\log\left(e + \frac{T\eta}{S^2}\right)}\right\}. \tag{247}$$

Theorem S16.1. *For each fixed positive integer d, there is a finite constant C_d such that the following holds for every integer $T \ge d$ with $\log T \ge 256$, every integer $m \ge 0$, and every $0 < \eta \le 1$. For every class and supplied measure described above, there is a measurable proper pre-input strategy such that the following bound holds for every admissible unknown reference and input-label protocol:*

$$\mathbb{E}\,\mathrm{Reg}_T \le C_d\big(A_{d,T} + G_{T,m}(\eta)\big).$$

The strategy may depend on $T, d, m, \eta, \mathcal{F}, \nu$ and observed history. Its number of reference samples is deterministic, at most m, and all samples are taken before play. The strategy has no other access to μ or to the conditional input laws.

These are measurable statistical strategies for arbitrary class and measure encodings. Computing the selected representatives would require deciding which countable subclasses are nonempty. The bounds are for fixed d, with dimension-dependent constants. The matching worst-case lower and the uniformly expected allowance are treated in Appendix S17.

S16.1 Change of reference and simultaneous charges

Throughout the proof put $s = \log T$ and $K = 1/\eta \ge 1$. The moment promise implies

$$\mathrm{KL}(\mu\|\nu) \le 1, \qquad \mathrm{KL}(P_t\|\nu) \le K + 2 \le 3K. \tag{248}$$

For the first assertion, with $r = d\mu/d\nu$, Jensen gives $\mathbb{E}_\mu \log r \le \eta^{-1}\log \mathbb{E}_\mu r^\eta \le 1$; bounded truncations justify the inequality when needed. For a bounded test h, Hölder gives

$$\log \mathbb{E}_\mu e^{\eta h/(1+\eta)} \le \frac{\eta}{1+\eta} + \frac{\eta}{1+\eta}\log \mathbb{E}_\nu e^h.$$

Combining this with the entropy inequality relative to μ shows

$$\mathbb{E}_P h - \log \mathbb{E}_\nu e^h \le \frac{1+\eta}{\eta}\,\mathrm{KL}(P\|\mu) + 1.$$

Taking the supremum over bounded h proves the second assertion at every actual full preceding history.

We shall use two consequences for fractional disagreement functions. If $0 \le q \le 1$, $\nu q \le 2e^{-a}$, and $a \ge 4$, then

$$Pq \le \frac{8}{\log(e+a)} \qquad \big(\mathrm{KL}(\mu\|\nu) \le 1,\ \mathrm{KL}(P\|\mu) \le 1\big). \tag{249}$$

Indeed, convexity of the exponential on $[0,1]$ and entropy at test $(a/2)q$ give

$$\mu q \le \frac{2}{a}\left[1 + \log(1 + 2e^{-a/2})\right] \le \frac{4}{a}.$$

For $a \ge 16$, apply entropy relative to μ at test $\log(a/4)q$. Its exponential moment is at most two, and hence

$$Pq \le \frac{1+\log 2}{\log(a/4)} \le \frac{8}{\log(e+a)}.$$

For $4 \le a < 16$ the last upper bound exceeds one.

For the second consequence, let ξ be any reference and let the scales $a \ge 4$ double successively. Suppose $0 \le q_a \le 1$ and $\xi q_a \le c_0 e^{-a}$. For $r = dP/d\xi$, split

$$Pq_a \le c_0 e^{-a/2} + H_a, \qquad H_a = \int_{\{r > e^{a/2}\}} q_a r \, d\xi.$$

Then

$$\sum_a a H_a \le 8 \, \mathrm{KL}(P\|\xi). \tag{250}$$

At a fixed point the contributing dyadic scales have $a < 2 \log r$, so their sum is at most $4 \log r$. They require $r > e^2$, where $r \log r \le 2(r \log r - r + 1)$. The latter integrand is nonnegative everywhere and integrates to $\mathrm{KL}(P\|\xi)$, proving the claim. All these statements hold for predictable q_a at each actual history.

S16.2 A finite-tree potential with bounded positive rate variation

We give the variation estimate needed to join the two reference trees. For a finite rooted tree with leaf values L_f, positive internal rates ρ_v, and fixed positive child weights $w_{c|v}$, set

$$V_f(L) = L_f, \qquad V_v(L) = -\rho_v^{-1} \log \sum_{c \text{ child of } v} w_{c|v} e^{-\rho_v V_c(L)}.$$

The weights may be normalized or unnormalized. Let $p(c \mid v; L)$ be the corresponding normalized branch probabilities and $p_f(L)$ their product along the path to f.

Lemma S16.2. *For an internal root and a path with internal rates $\rho_0, \dots, \rho_J$, define*

$$B_f = \rho_0 + \sum_{j=1}^{J} (\rho_j - \rho_{j-1})_+, \qquad B = \max_f B_f.$$

For every leaf loss vector $\ell \in [0,1]^{\text{leaves}}$ and $0 \le z \le 1$,

$$p_f(L + z\ell) \le e^{zB} p_f(L).$$

If each path consists of at most b consecutive nondecreasing rate blocks and every rate is at most r, then $B \le br$. The Hessian comparison with descendant squared loss differences has Taylor factor

$$I(B) = \int_0^1 (1-z) e^{zB} \, dz = \frac{e^B - 1 - B}{B^2}, \qquad I(0) = \frac{1}{2}.$$

In particular $I(B)$ is universally bounded for three blocks with rates at most $1/2$.

Proof. If the root is a leaf, take $B = 0$ and $p_f = 1$; all claims are immediate. Otherwise differentiate along $L + z\ell$ and put $g_v = dV_v/dz$. We have $g_f = \ell_f$ and $g_v = \sum_c p(c \mid v) g_c$, so every g_v is in $[0,1]$. All functions involved are finite positive smooth sums. One branch contributes $\rho_v(g_v - g_c)$ to the logarithmic derivative of a leaf probability. Summing along its path gives

$$\frac{d}{dz} \log p_f(L + z\ell) = \rho_0 g_{v_0} + \sum_{j=1}^{J} (\rho_j - \rho_{j-1}) g_{v_j} - \rho_J \ell_f. \tag{251}$$

The positive coefficients have total mass B_f. The absolute mass of the negative coefficients is $\sum_j (\rho_{j-1} - \rho_j)_+ + \rho_J = B_f$, by telescoping. Thus the derivative lies in $[-B_f, B_f]$, and integration proves the probability bound.

In the first nondecreasing block, the initial rate and its increases sum to its final rate, at most r. A later block contributes at most its final rate as well: if its first rate exceeds the previous block's last rate, add that positive interface increment to its internal increases; otherwise only the internal increases contribute. Each contribution is at most r, proving $B \le br$, including at increasing interfaces.

Put $H_v = -d^2V_v/dz^2$. Direct differentiation yields

$$H_v = \sum_c p(c \mid v) H_c + \rho_v \operatorname{Var}_{c|v}(g_c), \qquad H_{\text{root}} = \sum_{v \text{ internal}} p_v \rho_v \operatorname{Var}_{c|v}(g_c).$$

Here p_v is the probability of reaching v. This identity uses no monotonicity of the rates. For a nonroot node with fixed representative loss ℓ_{h_v}, conditional total variance gives

$$\operatorname{Var}_{c|v}(g_c) \le \operatorname{Var}_{f|v}(\ell_f) \le \mathbb{E}_{f|v}(\ell_f - \ell_{h_v})^2.$$

The root variance is at most one. Multiplying by reach probabilities turns the descendant expectations into sums of leaf probabilities. The already proved tilt bound compares each perturbed probability with its value at $z = 0$. Taylor's integral remainder consequently has factor $I(B)$; the root term is covered too since $I(B) \ge 1/2$. Integrating the exponential gives the displayed formula, with continuous value $1/2$ at zero and $I(B) \le e^B/2$. □

In the learning constructions we use unnormalized weights $w_{c|v} = 1$. The potential is then at most the best leaf loss, and its initial entropy penalty is at most the sum of the log cardinalities of child levels divided by their split rates, by (19)–(20). For binary losses, $(\ell_f(x,y) - \ell_{h_v}(x,y))^2 = \mathbf{1}\{f(x) \ne h_v(x)\}$. At a level with ancestor representatives h_v, the pre-round function

$$q_t(x) = \sum_f p_f(L_{t-1}) \mathbf{1}\{f(x) \ne h_{\text{anc}(f)}(x)\}$$

is predictable and lies in $[0,1]$. Combining the lemma with the potential telescoping argument in (27) gives

$$\mathbb{E}\operatorname{Reg}_T(\text{leaves}) \le -V(0) + I(B) \left[T\rho_{\text{root}} + \mathbb{E} \sum_{t,\text{nonroot splits}} \rho_{\text{split}} q_t(X_t) \right]. \tag{252}$$

Here the expected current loss is the mixture loss because the fresh private selection is independent of the current example, conditional on the preceding history. Every representative is a member of $\mathcal{F}$. The proof applies with fixed preprocessing and its actual conditional online process.

S16.3 Known-reference cells and terminal disagreement

Fix deterministic scales $v_l = 2^l v_0$, $v_0 \ge 4$. By Lemma S2.1, choose a finite proper ν-cover Q_l of radius $\delta_l = e^{-2v_l}$ and fix an ordering of its centers. For each class index i, let $\pi_l(i)$ be its first center within δ_l. These maps are deterministic. Partition the class by the whole tuple $(\pi_0(i), \dots, \pi_l(i))$ and choose the least-index representative of every nonempty cell. Then

$$\log(\text{number of cells at level } l) \le \sum_{r=0}^{l} \log |Q_r| \le C d v_l. \tag{253}$$

Deleting the last component gives the containing parent cell, so the partitions are nested. Every member of a cell shares π_l with its representative; their disagreement has ν mass at most

$$2e^{-2v_l} \le e^{-v_l}. \tag{254}$$

Every cell has descendants at every subsequent finite level, and every countable comparator has a terminal representative. All these objects are fixed before play; no effective search for empty cells is required.

We use the following terminal bound from Lemma S6.3. For any fixed nonempty countable $\mathcal{Q} \subseteq \mathcal{F} \triangle \mathcal{F}$ with $\nu q \le e^{-v}$ for all q, if $v \ge 4\log(eT/d)$ and the actual expected total full-history KL cost relative to ν is $B_\nu < \infty$, then

$$\mathbb{E} \sup_{q \in \mathcal{Q}} \sum_{t=1}^{T} q(X_t) \le Cd + \frac{2B_\nu}{v}. \tag{255}$$

The proof combines the countable rare-class MGF (1) with KL convexity from full histories to input histories, the path chain rule, and bounded entropy transfer. Its unsimplified bound is

$$\frac{2B_\nu}{v} + 4d + \frac{2\log 2 + 4d\log(3eT/d) + 10Te^{-v/2}}{v}.$$

Lemma S2.1 supplies the uniform growth bound $\Pi_{\mathcal{Q}}(n) \le (en/d)^{2d}$ for $n \ge d$. These results allow atoms and repeated observations. Later we apply them separately to the fixed class resulting from each good preprocessing outcome, with the actual conditional budget.

S16.4 A zero-sample upper for the general class

We first prove the following fallback for every class and supplied measure in Theorem S16.1:

$$\mathbb{E}\,\mathrm{Reg}_T \le C_d \min\left\{\frac{T}{s},\ \sqrt{\frac{T}{\eta}\log\left(e+\frac{T\eta}{s^2}\right)}\right\}, \tag{256}$$

using no reference samples. Put $z = T\eta$. If $z < s^2$, use Theorem S9.1 under total cost at most T. Its upper is $C_d T/\log(eT) \le C_d T/s$. Since $\log(e + z/s^2) > z/s^2$, the minimum in (256) equals T/s in this range.

Suppose henceforth $z \ge s^2$, so $K \le T/s^2$. Use the nested cells above, starting at $v_0 = 4$ and ending at the first $V = v_J \ge 256\sqrt{TK}$. Then $V < 512\sqrt{TK}$. Give the root above the first partition rate $1/(2\sqrt{T})$. The nonroot split scales are $a = v_{l-1}, 1 \le l \le J$. Put $\ell = \log(eK)$ and $A_0 = 4K\ell$. At coarse scales $a \le A_0$ choose

$$\rho(a) = \min\left\{\frac{1}{2},\ \sqrt{\frac{a\log(e+a)}{8T}}\right\}.$$

Let M_f be the number of scales $a > A_0$, and set $\lambda = \sqrt{M_f/(TK)}$. At those fine scales choose

$$\gamma(a) = 2e^{-a/2}, \qquad \rho(a) = \min\left\{\frac{1}{2},\ \lambda a,\ \sqrt{\frac{a}{T\gamma(a)}}\right\}. \tag{257}$$

The first scale is coarse, and $M_f \ge 1$. Indeed,

$$\frac{V}{A_0} \ge \frac{64\sqrt{z}}{\ell} \ge \frac{64s}{s+1} > 32,$$

and the last split scale is $V/2$. All rates are nondecreasing. The first coarse rate exceeds the root rate, and each candidate increases within its block. For completeness, let the adjacent boundary scales be $a_f = 2a_c > A_0$ and put $h_c = \log(e + a_c)$. Since $h_c \le \log(e + A_0) \le 3\ell$, the squared ratio of the fine λa_f candidate to the preceding coarse candidate is

$$\frac{16M_f a_f}{Kh_c} > \frac{64M_f\ell}{h_c} \ge \frac{64M_f}{3}.$$

The corresponding background-candidate ratio is $16/(\gamma(a_f)h_c) > 1$, because $\gamma(a_f)h_c \le 2a_c e^{-a_c} < 1$. The common cap $1/2$ preserves the ordering. Thus Lemma S16.2 applies with one block.

Each final leaf and its ancestor satisfies (254). The coarse predictable disagreements therefore obey (249). By (253) and the chosen rate, entropy plus expected variance at scale a is at most

$$C_d\left[a + \sqrt{\frac{Ta}{\log(e+a)}}\right].$$

The square-root terms sum geometrically: since $\log(e+2a) \le \log(e+a) + \log 2$ and $\log(e+a) \ge \log(e+4)$, the ratio of successive $a/\log(e+a)$ is bounded below by a universal number greater than one. Moreover

$$A_0 \le 8\sqrt{TK}, \qquad \frac{A_0}{\log(e+A_0)} \le 4K.$$

For the first inequality use $\ell \le s+1$ and $\sqrt{z} \ge s$; the second uses $A_0 \ge eK$. The complete coarse contribution, including the root, is therefore at most $C_d\sqrt{TK}$.

For the fine block, the entropy sum is bounded by

$$C_d\left[V + \frac{M_f}{\lambda} + \sqrt{T}\sum_a \sqrt{a\gamma(a)}\right].$$

Here $a/\rho(a)$ is at most a constant times $a + 1/\lambda + \sqrt{Ta\gamma(a)}$. The expected background charge is at most $\sqrt{T}\sum_a \sqrt{a\gamma(a)}$, and the high-density charge is at most $8\lambda \sum_t \mathrm{KL}(P_t\|\nu)$ by (250). The background series is universally convergent. By (248), the total fine cost is thus at most

$$C_d\left[V + \frac{M_f}{\lambda} + \sqrt{T} + \lambda TK\right] \le C_d\sqrt{TKM_f}. \tag{258}$$

The disagreements between original comparators and their final representatives form a fixed countable subclass of $\mathcal{F} \triangle \mathcal{F}$, of ν mass at most e^{-V}. Their supremum controls the full hindsight approximation, because each comparator has its own final representative. The condition $V \ge 256\sqrt{TK} \ge 4\log(eT/d)$ permits (255) with $B_\nu \le 3TK$. The terminal contribution is at most $Cd + 6TK/V \le C_d\sqrt{TK}$.

It remains to count fine scales. If $K \ge \sqrt{T}$, then $\ell \ge s/2$ and

$$V/A_0 < 256\sqrt{z}/s, \qquad M_f \le C\log(e + z/s^2).$$

If $K < \sqrt{T}$, then $\log z > s/2$ and $\log(z/s^2) \ge s/4$ for $s \ge 256$. The whole tree has $O(s)$ levels since $V < 512T$, giving the same bound on M_f. Finally, for $w = z/s^2 \ge 1$, $\log(e + w) \le 2w$. Hence the resulting square-root bound is at most $\sqrt{2}T/s$, and is within a universal factor of the minimum in (256). This proves that bound for arbitrary countable $\mathcal{F}$ and arbitrary supplied ν with a finite proper tree.

S16.5 Empirical acquisition and deterministic parameter branches

The sample-only strategy in Lemma S10.5 gives $C_d(A_{d,T}+T/S)$ with a deterministic preprocessing count at most m. We shall use it in several parameter branches. For $1 \le d \le T$,

$$A_{1,T} \le A_{d,T} \le \sqrt{d}\, A_{1,T}. \tag{259}$$

The first inequality follows from $(d-1)(1+\log T) - d\log d \ge d - 1 - \log d \ge 0$; the second follows from $\log(eT/d) \le \log(eT)$.

Set $j_0 = \max\{16, \lceil \log_2(T+2)\rceil\}$. For $j \ge j_0$, the measurable trace acquisition of Lemma S10.4, followed by an independent distance sample, has the exact sizes

$$\begin{aligned} &\rho = 2^{-j}, && u = j\log 2, && N = 256(d+1)^2 j 2^j, \\ &M = (N+1)^d, && L_R = \left\lceil \log_2 \frac{4M^2}{\rho} \right\rceil, && R = 8L_R 2^j. \end{aligned} \tag{260}$$

The deterministic envelope (81) gives

$$N + R \le F_d(j) := 512(d+1)^2 j 2^j. \tag{261}$$

It uses the deterministic witness bound M, including repeated sample points. The proof bounds $\log_2(N+1) \le 5j$ and $L_R \le 12(d+1)j$, whence $N + R \le 352(d+1)^2 j 2^j \le F_d(j)$.

If $S \ge T/A_{d,T}$, the sample-only upper is at most $2C_dA_{d,T}$, which proves the desired result. Suppose $S < T/A_{d,T}$. If $m < F_d(j_0)$, use (256). In this branch $S \le 8s$: the terms $s+1$, $\log[512(d+1)^2]$, $\log j_0$, and $j_0 \log 2$ are at most $2s, 3s, s, 2s$, respectively, since $d \le T$ and $s \ge 256$. The expressions with denominators s and S are within a universal factor. Indeed, for $c \ge 1$,

$$\log(e + c^2 x) \le (1 + 2\log c)\log(e + x), \qquad x \ge 0, \tag{262}$$

and $1 \le S/s \le 8$ bounds both terms in their respective minima.

Otherwise choose the largest $j \ge j_0$ with $F_d(j) \le m$, and use (260). Put $u = j\log 2$. Then

$$s < u \le S \le 8u, \qquad j < 2S < \frac{2T}{A_{d,T}} \le \frac{2T}{A_{1,T}} < \sqrt{T}. \tag{263}$$

For $S \le 8u$, maximality and integrality imply $m + 1 \le F_d(j+1)$. Bound the four terms of $s+1+\log F_d(j+1)$ by $2u, 3u, u, 2u$ respectively; use $d \le T$, $\log(j+1) \le j\log 2$, and $(j+1)\log 2 \le 2u$. The other inequalities use $j \ge j_0$, $u = j\log 2$, and $S < T/A_{d,T}$.

Write $w_u = T\eta/u^2$. If $w_u < 64$, use the sample-only upper: $w_S = T\eta/S^2 \le w_u$ and $G_{T,m}(\eta) \ge T/(8S)$ since $\log(e + w_S) \ge 1$. It remains to treat (263) with $w_u \ge 64$.

S16.6 Joint cells and measurable representatives

Run the empirical construction at this j. Let C contain the least-index witness for each realized first-sample trace, and retain exactly the empirical pruned center tree ending in the full witness level C from Appendix S8. For a realized trace α let c_α denote its witness. The simultaneous first-sample event (36) and pair-distance event (38) have conjunction E with $\Pr(E^c) \le 2\rho$. This event depends only on preprocessing for each fixed μ; the learner need not recognize it. On E, every pair of class members agreeing on the first sample has μ disagreement at most ρ, and the acquired pair distances satisfy

$$\mu\{f \ne g\} \le 2\widehat{D}(f,g) + \rho, \qquad \widehat{D}(f,g) \le 2\mu\{f \ne g\} + \rho \quad (f,g \in C).$$

To specify the imported tree geometry, put $b_l = 4 \cdot 2^l$ and let H be the largest $l \ge 0$ with $2b_l \le j-1$. Its empirical radii are 2^{-2b_l}. Nonroot split scales are $a_r = b_{r-1}$, $1 \le r \le H+1$, the last split ending at all of C. By (41)–(43), the root child log cardinality is at most Cd, each nonroot child log cardinality is at most Cda_r, and a witness in C differs from its scale-a_r ancestor on μ mass at most $2e^{-a_r}$. Pruning keeps every witness as a leaf and one node at each retained level on each path. These are simultaneous geometry bounds; we shall recompute the potential charges for the enlarged final leaf set.

Let v_0 be the smallest number $4 \cdot 2^q \ge u/\eta$ with integer $q \ge 0$. Set $v_l = 2^l v_0$ and let J be least with $V = v_J \ge 256\sqrt{TK}$. The remaining parameter range gives

$$uK \le v_0 < 2uK \le \frac{\sqrt{TK}}{4}, \qquad 256\sqrt{TK} \le V < 512\sqrt{TK}, \qquad J \ge 1. \tag{264}$$

Choose the deterministic covers Q_l and assignments π_l from Appendix S16.3. At level l, partition the class by the joint identifier

$$\big(\text{first-sample trace of } f_i, \pi_0(i), \dots, \pi_l(i)\big).$$

Let H_l contain the least-index member of every nonempty joint cell. The parent in H_{l-1} is the containing cell obtained by dropping π_l, and the parent of an H_0 cell is its c_α in C. The empirical tree above C is unchanged. Each nonempty class cell has descendants at every later level, and every C witness gets descendants, so the empirical pruning property is preserved. All final leaves belong to $\mathcal{F}$.

There are at most $(N+1)^d$ realized sample traces. The cover products therefore give

$$\log|H_l| \le d\log(N+1) + \sum_{r=0}^{l} \log|Q_r| \le Cd(u+v_l) \le Cdv_l. \tag{265}$$

Here $\log(N+1) \le Cu$ follows from the explicit N, $d \le T$, $u > s$, and $j \ge j_0$. In particular the bridge log cardinality is at most Cdv_0, and the fine split into H_l has log cardinality at most Cda_l for $a_l = v_{l-1}$. Joint refinement needs only this product bound.

All selections are measurable. For each proposed finite identifier, its nonemptiness is a countable union over i of measurable trace events intersected with fixed π conditions. Its least-index member is measurable on that event. At every used level there are finitely many possible trace bit strings and cover tuples. Omit empty cells and order the rest by fixed finite orders; this makes the finite tree shape, parent assignments, and evaluations measurable in preprocessing.

For a more explicit terminal selection, fix a comparator index i. At preprocessing outcome z, let $k(z,i) \le i$ be the first index matching its terminal identifier. For a proposed k, equality of identifiers and failure for indices below k are finitely many measurable tests. Thus $k(z,i)$ is measurable, and

$$(z,x) \longmapsto \mathbf{1}\{f_i(x) \ne f_{k(z,i)}(x)\}$$

is jointly measurable by its countably many index cases. The supremum over i of the corresponding finite-path counts is measurable too.

On E, each final leaf and its C ancestor c_α agree on the first sample, so their μ disagreement is at most $\rho = e^{-u}$. The final leaf and an empirical ancestor at scale a consequently disagree on μ mass at most

$$\rho + 2e^{-a} \le 3e^{-a}, \tag{266}$$

since $a \le j/2 < u$. A final leaf and its H_{l-1} ancestor share π_{l-1}, giving ν disagreement at most $2e^{-2v_{l-1}} \le e^{-a_l}$. The bridge thus uses a μ discrepancy, and fine levels use their own ν discrepancies; no comparison between these metrics is needed. For every preprocessing outcome, even outside E, each f_i and its terminal representative share π_J and have ν disagreement at most e^{-V}.

S16.7 The three rate blocks and their conditional costs

Use the unnormalized finite-tree strategy on the empirical center tree, its level C, and the appended levels $H_0, \dots, H_J$. The empirical root rate is $1/\sqrt{T}$; at its first nonroot scale $a = 4$ choose $\min\{1/2, \sqrt{a/T}\}$. Let M_μ be the number of remaining empirical scales $a \ge 8$. It is positive for $j \ge j_0$. Put

$$\lambda_\mu = \sqrt{M_\mu/T}, \qquad \gamma_\mu(a) = 4e^{-a/2}, \qquad \rho_\mu(a) = \min\left\{\frac{1}{2}, \lambda_\mu a, \sqrt{\frac{a}{T\gamma_\mu(a)}}\right\}. \tag{267}$$

The bridge split from C to H_0 has rate

$$\tau = \min\left\{\frac{1}{2}, \sqrt{\frac{v_0 u}{T}}\right\}. \tag{268}$$

For the $M_\nu = J$ fine splits, use

$$\lambda_\nu = \sqrt{M_\nu/(TK)}, \qquad \gamma_\nu(a) = 2e^{-a/2}, \qquad \rho_\nu(a) = \min\left\{\frac{1}{2}, \lambda_\nu a, \sqrt{\frac{a}{T\gamma_\nu(a)}}\right\}. \tag{269}$$

Dimension factors in the entropy estimates are absorbed into C_d.

The empirical sequence is nondecreasing. At the transition from 4 to 8, the two squared candidate ratios are $16M_\mu$ and $2/\gamma_\mu(8) > 1$; thereafter both candidates increase. The root rate is smaller than the first coarse rate. The ν fine block also increases. Including the positive singleton bridge gives at most three nondecreasing blocks, with all rates at most $1/2$. Lemma S16.2 therefore gives a universal Taylor factor at both junctions.

Fix a good full preprocessing outcome z outside a null exceptional set. The per-round μ cap holds given this outcome, so the actual conditional expected total μ cost is at most T; by (248) the corresponding ν cost is at most $3TK$. The per-round full-history promise includes preprocessing and therefore applies to this conditional process. All representatives are now fixed; their charging functions are predictable from the actual past, and the fresh private selection is independent of the current example.

For the empirical block, (266) gives $\mu q_a \le 3e^{-a}$. Its first scale and root cost $C_d\sqrt{T}$. At the remaining scales, (250) relative to μ controls the high-density sum, and γ_μ dominates the background. As in the explicit fine-cost calculation above, entropy and variance together are at most

$$\begin{aligned} C_d\left[j + \sqrt{T} + \frac{M_\mu}{\lambda_\mu} + \lambda_\mu T\right] &\le C_d\left[j + \sqrt{T\log(ej)}\right] \\ &\le C_d A_{d,T}. \end{aligned} \tag{270}$$

The dyadic scale sum is $O(j)$, the background series $\sum_a \sqrt{a\gamma_\mu(a)}$ is bounded, and $M_\mu = O(\log(ej))$. The last inequality uses $j < \sqrt{T}$ from (263) and (259). This charges disagreements of the enlarged final leaves, including their additional ρ distance through each empirical ancestor.

At the bridge, $\mu q \le e^{-u}$. Convexity gives $\mathbb{E}_\mu e^{uq} \le 2$, so entropy with $\mathrm{KL}(P_t\|\mu) \le 1$ implies $P_t q \le 2/u$. The bridge entropy and variance cost is at most

$$C_d\left[\frac{v_0}{\tau} + \frac{T\tau}{u}\right] \le C_d\left[v_0 + \sqrt{\frac{Tv_0}{u}}\right] \le C_d\sqrt{TK}, \tag{271}$$

by (264). In the capped case, $v_0 u/T \ge 1/4$ also bounds T/u by $4v_0$, verifying the first inequality in that case.

For the ν fine block, the discrepancy at scale a has ν mean at most e^{-a}. Use (250), the actual conditional cost at most $3TK$, and the entropy bound (265). They give

$$C_d\left[V+\sqrt{T}+\frac{M_\nu}{\lambda_\nu}+\lambda_\nu TK\right]\le C_d\sqrt{TKM_\nu}. \tag{272}$$

The background sums are universally bounded. Moreover

$$2^J=V/v_0<512\sqrt{T\eta}/u,\qquad M_\nu\le C\log(e+T\eta/u^2). \tag{273}$$

All entropy bounds are deterministic and uniform over good outcomes. The learner observes neither KL costs nor the indicator of E.

S16.8 Countable hindsight control and completion

For a fixed preprocessing outcome z, define

$$\mathcal{Q}_z=\{\mathbf{1}\{f_i\neq f_{k(z,i)}\}:i\ge 1\}\subseteq\mathcal{F}\triangle\mathcal{F}.$$

Every member has ν mean at most e^{-V}, with the same growth bound $(en/d)^{2d}$ independently of z. The joint measurability of the indexed functions and their countable supremum was proved in Appendix S16.6. For every realized input-label path,

$$\min_{h\in H_J}L_h-\inf_{f\in\mathcal{F}}L_f\le\sup_{q\in\mathcal{Q}_z}\sum_{t=1}^{T}q(X_t). \tag{274}$$

Indeed, the leaf representing any comparator differs from its loss by at most its own disagreement count, and then one takes the supremum.

For almost every fixed good z, apply (255) to this fixed countable class and the actual conditional online process. Its total ν KL is at most $3TK$ and $V\ge 256\sqrt{TK}\ge 4\log(eT/d)$, so

$$\mathbb{E}\left[\sup_{q\in\mathcal{Q}_z}\sum_t q(X_t)\,\middle|\,z\right]\le Cd+\frac{6TK}{V}\le C_d\sqrt{TK}. \tag{275}$$

This bounds the realized hindsight maximum over all comparators. The rare-class comparison uses the product law ν^T only as a comparison measure; the actual conditional inputs can be adaptive. Atoms, repeated observations, and auxiliary history variables are permitted by the terminal bound.

Combining (252) and (270)–(275) gives, conditional on each good outcome outside a null set,

$$\mathbb{E}[\mathrm{Reg}_T\mid z]\le C_d\left[A_{d,T}+\sqrt{\frac{T}{\eta}\log\left(e+\frac{T\eta}{u^2}\right)}\right].$$

On E^c regret is at most T. Since $\Pr(E^c)\le 2e^{-u}$ and $u\ge\log(T+2)$, this contributes at most two after averaging. Finally $u\le S\le 8u$ and $w_u\ge 64$ imply $w_S=T\eta/S^2\ge 1$ and

$$\log(e+w_u)\le\log(e+64w_S)\le(1+\log 64)\log(e+w_S).$$

For $w_S\ge 1$, $\log(e+w_S)\le 2w_S$. Thus the square-root contribution is within a universal factor of $G_{T,m}(\eta)$, including when T/S is the smaller branch. The parameter branches in Appendix S16.5 cover every other case. This proves Theorem S16.1.

S17 The joint reference-information lower

We use the notation $A_{d,T}$, $S=\log(eT(m+1))$, and $G_{T,m}(\eta)$ from (247). Throughout this section, U is uniform on $[0,1]$ and $h_\theta(x)=\mathbf{1}\{x\ge\theta\}$, with $0\le\theta\le 1$. The rational subclass includes both endpoint parameters. The next observation transfers every hindsight comparison between the two threshold classes.

Lemma S17.1. *On every finite input list in $[0,1]$, the prediction traces of ordinary thresholds equal those of rational thresholds in $[0,1]$. Consequently their comparator infima have equal loss for every finite labeled path, including repeated inputs and endpoint inputs.*

Proof. Fix the trace of a real parameter θ. If both prediction sides are nonempty, let a be the largest input assigned zero and b the smallest input assigned one. Then $a < \theta \le b$, so $a < b$. A rational parameter strictly between a and b has exactly this trace. If the zero side is empty, parameter zero works. If the one side is empty and the list is nonempty, its maximum a satisfies $a < \theta \le 1$; a rational parameter in $(a, 1]$ works. For an empty list use zero. Repeated inputs cannot lie on different sides of a threshold split. These arguments include inputs zero and one; in particular, an all-zero trace on a list containing one is unrealizable for both classes.

The rational class is a subclass of the real class, proving the reverse trace inclusion. Their possible prediction vectors on the list are therefore identical, so their loss infima agree for arbitrary labels. The statement is pathwise and requires neither realizability of the labels nor independence of the inputs. □

An autonomous input-label protocol uses its fixed instance parameters, its own preceding inputs, labels and activation state, and fresh seeds. It does not use learner predictions or reference-request decisions. Its input laws need not form an independent deterministic sequence. This property will permit comparison of complete runs with different reference-call rules.

Theorem S17.2. *Let T, m be integers with $\log T \ge 256$ and $m \ge 0$, and let $0 < \eta \le 1$. For every learner using at most m iid reference requests on every outcome, some atomless full-support reference μ, fixed before play, satisfies $D_{1+\eta}(\mu\|U) \le 1$ and admits an autonomous input-label protocol such that*

$$\mathrm{KL}(P_t\|\mu) \le 1 \quad \textit{at every full preceding history}, \qquad \mathbb{E}\,\mathrm{Reg}_T \ge 2^{-15} G_{T,m}(\eta).$$

For this G-only assertion, one rational comparator is also fixed before play. The full worst-case lower is

$$\mathbb{E}\,\mathrm{Reg}_T \ge 2^{-16}\big(A_{1,T} + G_{T,m}(\eta)\big).$$

Both assertions permit unrestricted post-input predictions, unlimited computation, and adaptive requests, including after a current input and before its prediction. They hold for ordinary and rational thresholds.

S17.1 A reference with sample-hidden bands

Set $V_0 = (3/4, 1)$. If $V_{j-1} = (v, v + w)$, let

$$A_j = (v + 3w/8, v + 5w/8).$$

A bit $b_j = 0$ retains the right outer quarter of V_{j-1} as V_j; a bit $b_j = 1$ retains the left outer quarter. For a finite depth N, choose all bits before play and let θ_b be the midpoint of V_N. The bands are disjoint, $|A_j| = 4^{-(j+1)}$, and $h_{\theta_b} = b_j$ on A_j. Moreover, A_j depends only on the preceding bits. All endpoints and θ_b are rational. Let Q_j be uniform on A_j, and let Q_0 be uniform on $(0, 1/2)$, where every such comparator predicts zero.

For a deterministic retained set $I \subseteq \{1, \dots, N\}$, put

$$\begin{aligned} \beta_j &= \frac{e^{-4\eta j}}{64(m+1)j^3}, \qquad \beta = \sum_{j \in I} \beta_j, \\ \mu_b &= (1-\beta)U + \sum_{j \in I} \beta_j Q_j. \end{aligned} \tag{276}$$

The weights do not depend on the bits. Since $\sum_{j \ge 1} j^{-3} \le 2$, we have $\beta \le 1/[32(m+1)]$. In particular, μ_b is atomless, has full support, and its density relative to U is at least $31/32$.

Lemma S17.3. *For every fixed bit path, the measure in (276) satisfies*

$$D_{1+\eta}(\mu_b\|U) \le \frac{5}{32(m+1)}.$$

For $j \in I$, define

$$q_j = \frac{1}{32[\log(ej) + \log(m+1) + \eta j]}, \qquad P_j = (1-q_j)Q_0 + q_j Q_j. \tag{277}$$

Then $\mathrm{KL}(P_j\|\mu_b) \le 29/32$, and $\mathrm{KL}(Q_0\|\mu_b) < 3/4$. These bounds remain true after any finite number of independent reference observations.

Proof. Write $r = d\mu_b/dU$, and on A_j put $u_j = \beta_j/|A_j|$. The bands are disjoint, so $r = 1-\beta+u_j \le 1+u_j$ there; outside them, $(\log r)_+ = 0$. Since $u_j \le 4^{j+1}$ and $j \ge 1$,

$$\log(1+u_j) \le \log(1+4^{j+1}) \le 4j.$$

Monotonicity of $x \log_+ x$ and $\log(1+u) \le u$ give

$$\begin{aligned}\int_{A_j} r(\log r)_+ \, dU &\le |A_j|(1+u_j)\log(1+u_j) \\ &\le \beta_j + 4j\beta_j.\end{aligned}$$

For $z \ge 0$, $e^{\eta z} \le 1 + \eta z e^{\eta z}$. Replacing a negative logarithm by zero and integrating against μ_b therefore gives

$$\mathbb{E}_{\mu_b} r^\eta \le 1 + \eta \sum_{j \in I} e^{4\eta j}\beta_j(1+4j).$$

Taking the logarithm and dividing by η yields

$$\begin{aligned}D_{1+\eta}(\mu_b \| U) &\le \frac{1}{64(m+1)} \sum_{j \ge 1} (j^{-3} + 4j^{-2}) \\ &\le \frac{5}{32(m+1)}.\end{aligned}$$

For the last inequality, each series is at most two by a first-term plus integral comparison.

The density lower bound gives

$$\mathrm{KL}(Q_0 \| \mu_b) \le \log(64/31) < 3/4.$$

Also $\mu_b \ge \beta_j Q_j$, so

$$\begin{aligned}\mathrm{KL}(Q_j \| \mu_b) &\le \log 64 + \log(m+1) + 3\log j + 4\eta j \\ &\le 5[\log(ej) + \log(m+1) + \eta j].\end{aligned}$$

Convexity of KL in its first argument and (277) now give $\mathrm{KL}(P_j \| \mu_b) \le 3/4 + 5/32 = 29/32$. Each assertion holds for the fixed measures themselves. An online generator using fresh seeds for the prescribed stage law, independently of reference seeds, has that same law conditional on the full preceding history, including additional reference observations. □

Pregenerate a potential list of m reference observations using independent component coins for (276). Let E be the event that all these coins select U. Then

$$\Pr(E) \ge 1 - m\beta \ge 31/32. \tag{278}$$

For $m = 0$ the event is certain. The event is independent of the path bits, online input coins and label noise because the mixture weights are independent of the path. Conditional on E, the full potential list consists of independent U draws, independent of the path and online process. Adaptive requests reveal successive entries of this list. Their transcript, including requests made after a current input, is a function of this independent list and the information already available to the learner.

S17.2 Testing with adaptive reference requests

Lemma S17.4. *Fix a stage j of deterministic integer length t_j with $t_j q_j \ge 1$. Use P_j for its input law. On an active input from Q_j, give label b_j with probability $(1+\gamma_j)/2$, where*

$$\gamma_j = \frac{1}{8\sqrt{t_j q_j}}.$$

Use fresh label noise and give label zero on the neutral component. Conditional on E and all earlier bits, if the current bit is fair before its stage, its contribution to expected loss minus the loss of h_{θ_b} is at least

$$\frac{3}{64}\sqrt{t_j q_j}.$$

Proof. Grant the learner the earlier bits, the current activation indicators, and the full potential reference list conditional on E. Proving the claim with this additional information suffices. The active band depends only on earlier bits. All earlier-stage observations, the potential reference list and the current input therefore have the same law for either value of b_j. Only the past active labels in this stage distinguish those values.

The KL divergence between their two Bernoulli label laws is

$$\gamma_j \log \frac{1+\gamma_j}{1-\gamma_j} \le 4\gamma_j^2,$$

since $\gamma_j \le 1/8$. The chain rule bounds the KL of any available stage transcript by $4t_jq_j\gamma_j^2 = 1/16$. Adaptive calls and prediction decisions are functions of the granted independent list, past observations and private randomness, and add no term to this bound. Pinsker's inequality bounds total variation by $\sqrt{1/32} < 1/4$. With a fair prior on b_j, every decision therefore disagrees with b_j with prior-averaged probability at least $3/8$. This remains true after the current input, whose conditional law is identical for the two bits, and after any permitted current requests.

The comparator gives the majority label on an active input. A prediction that disagrees with b_j has expected excess loss γ_j over that comparator; neutral excess is nonnegative. Summing the active contributions yields $(3/8)t_jq_j\gamma_j = (3/64)\sqrt{t_jq_j}$. Randomized binary decisions, or their probabilities in $[0,1]$, satisfy the same bound by averaging. □

The complementary event also needs a loss comparison. Fix all path bits and condition on the reference-component coins, the full potential reference list and online input locations. Whether E holds does not depend on the fresh noisy labels. Conditional on the past at an active prediction, the current label still has majority b_j; on a neutral input the comparator is correct. Thus the expected excess over this fixed comparator is nonnegative also on E^c, even if reference observations there reveal the path. Consequently the unconditional expected comparator excess is at least $\Pr(E)$ times the sum of the preceding stage bounds. This argument conditions on the reference common event only for the testing calculation; the actual input KL laws remain those already verified for every full history.

S17.3 Allocating rounds and completing the hard-cap lower

Proof of Theorem S17.2.

Put $s = \log T$, $S = \log(eT(m+1))$, and $w = T\eta/S^2$. First suppose $w \ge e^{64}$, and put

$$\begin{aligned} y &= \log w, \qquad a = \left\lceil \frac{2S}{\eta} \right\rceil, \quad N = \left\lfloor \sqrt{\frac{T}{1024\eta y}} \right\rfloor, \\ I &= \{a, \dots, N\}, \quad Q = \sum_{j=a}^{N} q_j, \quad t_j = \left\lfloor \frac{Tq_j}{2Q} \right\rfloor. \end{aligned} \tag{279}$$

The estimates below show that I is nonempty and the stage lengths are positive. Set all bits before a to zero and choose the bits in I independently and fairly before play. Use the reference and stage laws above, with stages ordered by j.

Since $S \ge s \ge 256, 64 \le y \le s$. The quantity inside the floor defining N is at least $\sqrt{T}/(32\sqrt{s}) > 2$. Also $\eta \ge e^{64}S^2/T$, so that quantity is at most $T/\sqrt{1024e^{64}S^2y} < T$. Thus $N \le T$, while the ceiling gives $a \le 4S/\eta$. Using $N+1 > \sqrt{T/(1024\eta y)}$ gives

$$\log \frac{N+1}{a} \ge \frac{y}{2} - \log 128 - \frac{1}{2}\log y \ge \frac{y}{4}. \tag{280}$$

For the last inequality, the difference is positive at $y = 64$ and has derivative $1/4 - 1/(2y) > 0$ thereafter. In particular $N \ge a$. For retained j, the bounds $j \le T$ and $\eta j \ge 2S$ imply

$$\log(ej) + \log(m+1) \le S \le \eta j/2, \qquad \frac{1}{64\eta j} \le q_j \le \frac{1}{32\eta j}.$$

Integral comparisons with the harmonic sum yield

$$\frac{y}{256\eta} \le Q \le \frac{y}{32\eta}. \tag{281}$$

Indeed, the lower uses (280); for the upper, $N/a \le \sqrt{w}/(64\sqrt{y})$ and $\sum_{j=a}^{N} 1/j \le 1+\log(N/a) \le 1 + y/2 \le y$. Moreover,

$$q_j \ge q_N \ge \frac{1}{64\eta N} \ge \frac{1}{2}\sqrt{\frac{y}{T\eta}}.$$

It follows that $Tq_j^2/(2Q) \ge 4$. Since $q_j \le 1$, the quantity inside each stage-length floor is at least four, and flooring loses at most a factor two. Consequently

$$t_j \ge \frac{Tq_j}{4Q}, \qquad t_j q_j \ge 2, \qquad \sum_{j=a}^{N} t_j \le T/2. \tag{282}$$

Use Q_0 with label zero for the remaining rounds.

Each stage has the noisy labels of Lemma S17.4. Earlier bits may be revealed at the start of later stages; no later bit is used to choose an earlier band or law. The reference list conditional on E is independent of every unrevealed bit, so the testing argument holds for every stage. The nonnegative complementary-event comparison and (278) give

$$\begin{aligned}\mathbb{E}\,\mathrm{Reg}_T &\ge \frac{31}{32}\frac{3}{64}\sum_{j=a}^{N}\sqrt{t_j q_j} \\ &\ge \frac{31}{32}\frac{3}{128}\sqrt{TQ} \ge \frac{31}{32}\frac{3}{2048}\sqrt{\frac{T}{\eta}}\,y.\end{aligned}$$

Regret to the hindsight optimum is at least excess over the fixed θ_b on every path. Every bit-path instance obeys the moment and full-history KL promises. Averaging over the finite prior therefore selects a deterministic reference and rational comparator attaining at least the displayed expected regret for the given randomized learner. The reference is selected before the realized play, not after the learner's private randomness is observed. For $w \ge e^{64}$, $\log(e+w) \le 2y$ and $\log(e+w) \le w$. Thus the coefficient of $G_{T,m}(\eta)$ is at least $(31/32)(3/2048)/\sqrt{2} > 2^{-15}$. Conditional on its fixed instance, this multiscale protocol uses deterministic stage lengths and fresh independent input and label seeds.

If $0 < w < e^{64}$, apply Theorem S14.3 with the same T, m, η and $\delta = 1$. Its complete proof gives the explicit constant $c_0 = 1/2048$ in

$$c_0 \min\left\{\frac{T}{S}, \sqrt{\frac{T}{\eta}}\right\}.$$

Here $1 \le \log(e+w) \le 65$. Since $\min(x, rb) \le r\min(x, b)$ for $r \ge 1$, this bound is at least $G_{T,m}(\eta)/(2048\sqrt{65}) \ge 2^{-15}G_{T,m}(\eta)$. The imported construction fixes an atomless full-support reference and rational target before play and verifies the same moment order. Its next active band is determined by its own number of preceding active component draws. Its active coins, within-band coordinates and potential reference observations use independent seeds conditional on the fixed path. It does not inspect predictions or request decisions. For $S > T/2$, that proof instead uses iid uniform inputs and one target selected from $\{1/4, 3/4\}$ before play, with the same autonomy. Thus the G-only witness is autonomous in every parameter range, although the active-count branch need not have independent inputs.

Finally, Lemma S11.11 gives the separate lower $A_{1,T}/256$ under $\mu = U$, even with unlimited adaptive reference requests. Its online generator uses its own fresh fair labels and preceding generated majorities, not learner decisions or requests. The reference moment is zero and its full-history input KL is at most one. Lemma S17.1 transfers the complete hindsight comparison to the rational class as well. Taking the maximum of this witness and the G-only witness gives at least $2^{-16}(A_{1,T} + G_{T,m}(\eta))$. The two terms need not be attained by the same reference or comparator. □

S17.4 Uniformly expected reference calls

We first record the denominator comparison needed for rounding and for a larger truncation cap.

Lemma S17.5. *Fix $T, \eta > 0$, and for $r > 0$ define*

$$g(r) = \min\left\{\frac{T}{r}, \sqrt{\frac{T}{\eta}\log\left(e + \frac{T\eta}{r^2}\right)}\right\}.$$

For every $a \geq 1$,

$$\frac{g(r)}{a} \leq g(ar) \leq g(r).$$

Consequently, if $r \leq r' \leq ar$, then $g(r)/a \leq g(r') \leq g(r)$.

Proof. Both terms defining g decrease with r. For the lower bound put $x = T\eta/r^2$ and $b = a^{-2} \leq 1$. Concavity of the logarithm gives

$$\begin{aligned}\log(e + bx) &= \log\big((1-b)e + b(e+x)\big)\\ &\geq (1-b)\log e + b\log(e+x) \geq b\log(e+x).\end{aligned}$$

Thus the square-root term at ar is at least $1/a$ times that term at r, and the other term is exactly $1/a$ times its value. Taking their minimum proves the assertion, including changes of the minimizing branch. The intermediate-r' statement follows by monotonicity. □

For a finite real allowance $m \geq 0$, let $N_{\mathcal{L}}$ be the total number of reference requests made by a learner through the horizon. Its uniformly expected allowance means

$$\mathbb{E}_I N_{\mathcal{L}} \leq m \quad \text{for every admissible fixed instance } I \tag{283}$$

in the announced family at the same T and η. An instance includes the reference and the legal input-label protocol. The expectation includes learner, oracle and online randomness. The promise holds for every reference and legal input-label protocol in that family. A strategy must finish its reference requests and produce each prediction.

We first prove the expected-call lower for thresholds.

Proposition S17.6. *Let T be an integer with $\log T \geq 256$, let $m \geq 0$ be finite and real, and let $0 < \eta \leq 1$. Every threshold learner satisfying (283), even with unrestricted post-input predictions and adaptive requests, has expected regret at least*

$$2^{-19}\big(A_{1,T} + G_{T,m}(\eta)\big)$$

on some admissible atomless full-support reference fixed before play. The online protocol can be autonomous.

Proof. Put $A = A_{1,T}$, $G = G_{T,m}(\eta)$ and $s = \log T$. If $G \leq A$, the uniform-reference statistical floor gives $A/256 \geq (A+G)/512$ even with unlimited reference requests. Assume $G > A$, fix a learner $\mathcal{L}$ satisfying (283), and put

$$M = \left\lceil 2^{20}(m+1)T^2 \right\rceil. \tag{284}$$

Run $\mathcal{L}$ with its original announced T, m, η until its first attempted $(M+1)$st request. At that attempted request abandon its internal strategy, make no further requests, and use h_1 for every prediction still due. If the current prediction already occurred, keep it and switch at the next prediction. Denote the resulting learner by $\mathcal{L}^M$. This is a valid hard-cap learner with at most M requests. It may retain the original m as auxiliary information independent of the unknown instance. The hard-cap lower quantifies over all such policies and does not require any simulation after the switch.

Apply the G-only assertion of Theorem S17.2 to $\mathcal{L}^M$ at cap M and the same T, η. It supplies a fixed admissible reference μ_M and autonomous protocol I_M with

$$\mathbb{E}_{I_M} \operatorname{Reg}_T(\mathcal{L}^M) \geq 2^{-15} G_{T,M}(\eta). \tag{285}$$

This instance belongs to the original announced family for $\mathcal{L}$. That family is specified by the moment promise at η and the unit input KL cap, not by how many samples a particular hard witness was designed to hide. Therefore the uniform resource promise gives

$$\mathbb{E}_{I_M} N_{\mathcal{L}} \le m. \tag{286}$$

Uniformity makes this guarantee available at the selected hard instance, as well as at U.

Realize I_M with its fixed parameters, its own online seeds, an infinite iid μ_M reference list, and learner randomness. Couple the two learners using the same reference list and the same private randomness until truncation. Give them the same entire online input-label path. This is possible because the protocol ignores learner actions and request decisions. In the multiscale witness, stages and their laws are fixed before play. In the single-depth witness, use the same active-component seeds and within-band coordinates; its active state depends only on its own preceding active draws and remains identical after the learner histories separate. The elementary endpoint is iid. Placing future seeds on the common probability space does not reveal them in the preceding history.

Allowing the original learner to reveal further entries beyond M does not invalidate the input constraint. Conditional on the fixed instance, all reference seeds are independent of fresh online seeds. The moment and domination bounds in the lower constructions hold for the same current stage or active-count state after any additional finite reference observations. In particular, no actual input law is conditioned on the lower proof's common event.

Until the first attempted $(M+1)$st request, histories and predictions are identical. That stopping event is exactly $H = \{N_{\mathcal{L}} > M\}$ in the complete original run. Off H, predictions agree through the horizon. On H, cumulative prediction losses can differ by at most T. The input-label path is common, so its full hindsight comparator loss cancels in the difference of the two regrets. Pathwise,

$$\left|\mathrm{Reg}_T(\mathcal{L}) - \mathrm{Reg}_T(\mathcal{L}^M)\right| \le T\mathbf{1}_H. \tag{287}$$

For randomized binary decisions use the same sampling seeds before separation. Predictions represented by probabilities in $[0,1]$ also have binary expected losses in $[0,1]$, giving the same comparison. Requests within a round or after its current input are included in H. This comparison concerns complete runs. The call-count event enters only through the pathwise loss comparison above.

Markov's inequality for the integer-valued count and (286) give

$$\mathbb{E}_{I_M} \mathrm{Reg}_T(\mathcal{L}) \ge 2^{-15} G_{T,M}(\eta) - \frac{Tm}{M+1}. \tag{288}$$

Let $x = 2^{20}(m+1)T^2$. Then $x \ge 2, M \ge x$ and $M+1 \le x+2 \le 2x$. With $S_M = \log(eT(M+1))$, this gives

$$S \le S_M \le S + 2s + 21\log 2 \le 4S, \qquad \frac{Tm}{M+1} \le \frac{2^{-20}}{T}.$$

The last denominator bound uses $S \ge s+1$ and $s \ge 256$. Lemma S17.5 therefore implies $G_{T,M}(\eta) \ge G/4$. Substitution in (288) yields

$$\mathbb{E}_{I_M} \mathrm{Reg}_T(\mathcal{L}) \ge 2^{-17} G - 2^{-20}/T \ge 2^{-18} G \ge 2^{-19}(A+G),$$

where $G > A \ge 1$ was used. The uniform-reference branch and this branch prove the assertion for every finite real m, without changing η. □

We combine this deduction with the general upper.

Corollary S17.7. *Fix a positive integer d, and let $T \ge d$ be an integer with $\log T \ge 256$ and $0 < \eta \le 1$. In the model of Theorem S16.1, the worst-case minimax expected regret over classes of VC dimension at most d and supplied measures ν is*

$$\Theta_d\big(A_{d,T} + G_{T,m}(\eta)\big).$$

This holds both for integer hard call caps $m \ge 0$ and for finite real uniformly expected allowances $m \ge 0$ in (283). For every supplied countable measurable class and Borel ν, the upper is attained by a measurable proper pre-input learner using a deterministic number of calls at most $\lfloor m \rfloor$, all before play. The lower already holds for rational thresholds with $\nu = U$, atomless full-support fixed references, arbitrary post-input prediction and adaptive requests. The same order holds for ordinary thresholds. The constants may depend on d; a lower for every class or every supplied ν is not asserted.

Proof. For an integer cap use Theorem S16.1. For a finite real allowance let $n = \lfloor m \rfloor$ and apply that theorem at n with the original $T, d, \eta, \mathcal{F}, \nu$. The count is deterministically at most $n \le m$, so it satisfies the uniform expected promise on the entire announced family. Since $n + 1 \ge (m + 1)/2$, its denominator S_n obeys

$$S - \log 2 \le S_n \le S, \qquad S/2 \le S_n \le S.$$

Lemma S17.5 gives $G_{T,n}(\eta) \le 2G_{T,m}(\eta)$. Thus its expected regret is at most $2C_d(A_{d,T} + G_{T,m}(\eta))$. This includes $m = 0$ and $0 < m < 1$, where no samples are requested. The current private selection and its independence from the current input-label kernel are unchanged. The floor is a mathematical parameter choice in this statistical theorem; a finite encoding of an arbitrary real allowance is not assumed.

For the lower use Proposition S17.6 on the fixed rational-threshold class, whose VC dimension is one and which is therefore allowed for every bound $d \ge 1$. Because $A_{d,T} \le \sqrt{d}\, A_{1,T}$,

$$2^{-19}(A_{1,T} + G_{T,m}(\eta)) \ge \frac{2^{-19}}{\sqrt{d}}\big(A_{d,T} + G_{T,m}(\eta)\big).$$

A strategy with hard cap m also satisfies the uniformly expected allowance, so this lower covers that case as well. This gives the stated worst case over classes and supplied references. For ordinary thresholds apply the generic upper to their rational subclass. Its selections are proper ordinary thresholds, and Lemma S17.1 preserves the exact hindsight infimum on each realized path. The lower transfers by the same lemma, which proves the claimed ordinary-threshold consequence. □

S18 Finite-bit known-reference threshold guarantee

For ordinary thresholds on $[0, 1]$ under known uniform μ, the sharp statistical rate has a direct finite-bit realization. The comparator includes every real threshold; the learner uses a finite dyadic grid.

Theorem S18.1. *There are universal constants $C, c > 0$ with the following property. Given integers $T, b \ge 1$, a proper learner for ordinary thresholds satisfies*

$$\mathbb{E}\,\mathrm{Reg}_T \le C \min\{T, \Phi_{1,T}(b)\}$$

whenever $\mathbb{E}\sum_{t=1}^{T} \mathrm{KL}(P_t\|\mu) \le b$, under the current-round privacy in Section 2. For $T \ge 2$ and $b < T^2$, it reads a prefix of

$$K < 768\sqrt{\max\{T, b\}}$$

binary digits per input, stores at most $1 + 2TK$ tree nodes, and uses $O(TK[\log(TK + 2)]^c)$ bit operations. Its storage and fair-bit use are $O(TK \log(TK + 2))$. It uses no evaluation, learning, reference-sampling, or integration oracle. For $T = 1$ or $b \ge T^2$, a fixed threshold suffices.

The bit bound uses random access with charged input and output bits and logarithmic-size addresses. Each input encoding must supply the stated prefix of binary digits. Choosing $b = \max\{1, \lceil B \rceil\}$ changes $\Phi_{1,T}(B)$ only by a constant factor, including at $B = 0$. The matching statistical lower transfers from ordered atoms to uniform intervals.

Appendix S19 gives the complete proof, including rates, proper sampling, precision and bit accounting.

S19 An explicit threshold implementation

Proof of Theorem S18.1. Let μ be Lebesgue probability measure on $[0, 1]$, and define

$$\mathcal{F}_{\text{thr}} = \{h_a : a \in [0, 1]\}, \qquad h_a(x) = \mathbf{1}\{x \ge a\}.$$

In this section the order and the reference are given explicitly. The learner selects h_a before observing the current input. Its current selection bits remain private until the input and label have been generated. The adversary may use the complete preceding history, including previous selections and labels, and its current label may depend on its current input. Write

$$B_* = \mathbb{E}\sum_{t=1}^{T} \mathrm{KL}(P_t\|\mu),$$

where P_t is conditional on that complete preceding history. The conditional laws and their divergences are not revealed to the learner.

The rate below is the dimension-one specialization of (8):

$$\Phi_T(B) = \begin{cases} \sqrt{T} + \sqrt{B \log(e+T)}, & 0 \le B \le T, \\ \sqrt{T} + \sqrt{B \log(eT/\sqrt{B})}, & T < B \le T^2, \\ T, & B > T^2. \end{cases} \tag{289}$$

We prove the bound for supplied positive integers T, b and $B_* \le b$, including its finite implementation. The operation count uses random-access addressing with its bit cost, and streamed input and output bit strings. The learner calls no ERM, evaluation, reference-sampling, integration, or conditional-simulation oracle.

The supplied input digits determine all losses on the algorithm's grid. Absolute continuity implies that dyadic inputs occur with probability zero. On such an exceptional input we use its terminating expansion. The endpoint $x = 1$ is treated separately: all thresholds used by the algorithm predict one there. The integer budget is supplied before play. The guarantee concerns expected regret at the announced horizon and budget under the known reference.

S19.1 The conceptual tree and its rates

Assume $T \ge 2$ and $1 \le b < T^2$. Put

$$u_j = 4 \cdot 2^j, \qquad J = \min\{j : u_j \ge 256\sqrt{\max\{T, b\}}\}, \qquad U = u_J,$$

and set

$$k_j = 6 \cdot 2^j, \qquad K = k_J = 3U/2, \qquad a_j = u_{j-1} \quad (1 \le j \le J), \qquad A_0 = 4\max\{1, b/T\}.$$

Thus $U < 512\sqrt{\max\{T, b\}}$. At level j, use the representatives

$$\mathcal{C}_j = \{i2^{-k_j} : 0 \le i < 2^{k_j}\}.$$

The representative a denotes threshold h_a. Its parent is obtained by truncating to k_{j-1} binary digits. Connect $\mathcal{C}_0$ to one root. A leaf and its level-$(j-1)$ ancestor disagree on an interval of reference mass at most

$$2^{-k_{j-1}} \le e^{-a_j} \le 2e^{-a_j}. \tag{290}$$

The root has 2^6 children. Every split to level $j \ge 1$ has $2^{k_j - k_{j-1}}$ children, so its log branching number is at most Ca_j. This complete regular tree defines the distribution; the implementation will not enumerate it.

A scale j is coarse if $a_j \le A_0$, and fine otherwise. Let M be the number of fine scales. If $b \le T$, then $U/A_0 \ge 64$; if $b > T$, then $U/A_0 \ge 64T/\sqrt{b} > 64$. Since the largest a_j is $U/2$, in either case $M \ge 1$. Define

$$\begin{aligned} \eta_0 &= \min\{1/2, 1/\sqrt{T}\}, \\ \eta_j &= \min\{1/2, \sqrt{a_j/T}\} && \text{if } j \text{ is coarse}, \\ \lambda &= \sqrt{M/b}, \qquad \gamma_j = 2^{1-a_j/2}, \\ \eta_j &= \min\{1/2, \lambda a_j, \sqrt{a_j/(T\gamma_j)}\} && \text{if } j \text{ is fine}. \end{aligned} \tag{291}$$

These rates satisfy

$$\frac{1}{2\sqrt{T}} \le \eta_0 \le \eta_1 \le \cdots \le \eta_J \le \frac{1}{2}. \tag{292}$$

The coarse terms increase. Both uncapped fine terms increase because a_j increases and γ_j decreases. At their transition write $a_f = 2a_c$. Since $\gamma_f \le 2$ and $a_f > A_0 \ge 4b/T$,

$$\sqrt{\frac{a_f}{T\gamma_f}} \ge \sqrt{\frac{a_c}{T}}, \qquad \frac{(\lambda a_f)^2}{a_c/T} = \frac{2MTa_f}{b} > 8M \ge 1.$$

Taking minima with $1/2$ preserves the comparisons. The first scale is coarse because $a_1 = 4 \le A_0$, and its rate is at least η_0.

S19.2 A LOCAL BOUND FOR THE POTENTIAL

At a leaf f let $V_f = L_f$, its cumulative binary loss. At an internal node v with rate $\eta(v)$ define

$$V_v = -\frac{1}{\eta(v)} \log \sum_{c \text{ child of } v} e^{-\eta(v) V_c}. \tag{293}$$

The root uses η_0; the split from level $j-1$ to level j uses η_j. Choose a child with probability proportional to $e^{-\eta(v)V_c}$, and let $p_f(L)$ be the product of these probabilities along a leaf path. Differentiation gives $\nabla V_{\text{root}} = p(L)$. Repeatedly comparing a soft minimum with its smallest argument gives

$$V_{\text{root}}(L) \le \min_f L_f, \qquad -V_{\text{root}}(0) \le \frac{C}{\eta_0} + C \sum_{j=1}^{J} \frac{a_j}{\eta_j}. \tag{294}$$

Indeed, each split contributes at most its log branching number divided by its rate to the difference from a minimum along a path.

Fix a binary loss vector ℓ and differentiate along $L + z\ell$, $0 \le z \le 1$. Put $g_v = dV_v/dz$ and $H_v = -d^2 V_v/dz^2$. At a leaf, $g_f = \ell_f$ and $H_f = 0$. Differentiating (293) gives

$$g_v = \sum_c p(c \mid v) g_c, \qquad H_v = \sum_c p(c \mid v) H_c + \eta(v) \operatorname{Var}_{c|v}(g_c).$$

Consequently $0 \le g_v \le 1$ and, writing p_v for the probability of reaching v,

$$H_{\text{root}} = \sum_{v \text{ internal}} p_v \eta(v) \operatorname{Var}_{c|v}(g_c). \tag{295}$$

Conditional total variance bounds the local variance by the variance of the descendant leaf loss. At a nonroot node with representative h_v, the latter is at most

$$\mathbb{E}_{f|v}(\ell_f - \ell_{h_v})^2 = \mathbb{E}_{f|v} \mathbf{1}\{f(x) \ne h_v(x)\}. \tag{296}$$

At the root it is at most one.

Differentiating the logarithm of a leaf probability gives the telescoping identity

$$\frac{d}{dz} \log p_f = \eta_0 g_{\text{root}} + \sum_{\substack{v \text{ nonroot internal} \\ v \text{ on the path to } f}} (\eta(v) - \eta(\text{parent}(v))) g_v - \eta_J \ell_f.$$

By (292), the absolute value is at most $\eta_J \le 1/2$. Thus $p_f(L+z\ell) \le e^{z/2} p_f(L)$. From the pre-round losses define

$$q_{t,j}(x) = \sum_f p_f(L_{t-1}) \mathbf{1}\{f(x) \ne h_{\text{ancestor}(f,j-1)}(x)\}.$$

This function is predictable, belongs to $[0,1]$, and has reference mean at most $2e^{-a_j}$ by (290). Taylor's formula with integral remainder, (295), and (296) give

$$p(L) \cdot \ell - \left[V_{\text{root}}(L+\ell) - V_{\text{root}}(L)\right] \le \eta_0 + \sum_{j=1}^{J} \eta_j q_{t,j}(x). \tag{297}$$

Here the integral factor is $\int_0^1 (1-z) e^{z/2}\, dz < 1$. Conditional on the preceding history and current labeled example, fresh private selection gives expected loss $p(L) \cdot \ell$. Summing (297) and using (294) proves

$$\mathbb{E}\,\text{Reg}_T(\mathcal{C}_J) \le \frac{C}{\eta_0} + C \sum_j \frac{a_j}{\eta_j} + T\eta_0 + \mathbb{E} \sum_{t,j} \eta_j q_{t,j}(X_t). \tag{298}$$

In particular, this step uses the declared pre-input timing, not a distribution chosen after observing X_t.

S19.3 The total divergence charge

Fix any full preceding history and write $r = dP/d\mu$. At a fine scale, split Pq_j into integrals over $r < e^{a_j/2}$ and $r \geq e^{a_j/2}$. The first integral is at most $2e^{-a_j/2} \leq \gamma_j$. Call the second H_j. At any point with $r \geq e^2$, the dyadic sum satisfies

$$\sum_{a_j \leq 2\log r} a_j \leq 4 \log r.$$

Also $r \log r \leq 2(r \log r - r + 1)$ on this set, and $r \log r - r + 1 \geq 0$ everywhere. Therefore

$$\sum_{j \text{ fine}} a_j H_j \leq 8\,\mathrm{KL}(P\|\mu). \tag{299}$$

This calculation does not add a constant at each round.

The coarse contributions in (298) are at most

$$C \sum_{j \text{ coarse}} (a_j + \sqrt{Ta_j}) \leq C(A_0 + \sqrt{TA_0}).$$

For fine scales, taking reciprocals of the minimum in (291) gives

$$\sum_{j \text{ fine}} \frac{a_j}{\eta_j} \leq C \left(U + \frac{M}{\lambda} + \sqrt{T} \sum_j \sqrt{a_j \gamma_j} \right).$$

The last sum is bounded by a numerical constant: it is at most $\sum_{n \geq 4} \sqrt{2(n+1)}\, 2^{-n/4} < \infty$. Using $\eta_j \leq \lambda a_j$ for the high-density terms and $\eta_j \leq \sqrt{a_j/(T\gamma_j)}$ for the low-density terms gives

$$\mathbb{E} \sum_t \sum_{j \text{ fine}} \eta_j q_{t,j}(X_t) \leq T \sum_{j \text{ fine}} \eta_j \gamma_j + 8\lambda B_* \leq C\sqrt{T} + 8\lambda B_*.$$

Together with the root terms, this proves the affine bound

$$\mathbb{E}\,\mathrm{Reg}_T(\mathcal{C}_J) \leq C \left(\sqrt{T} + \sqrt{TA_0} + A_0 + U + \frac{M}{\lambda} \right) + 8\lambda B_*. \tag{300}$$

It holds for any actual expected budget, before substituting $B_* \leq b$. There is no preprocessing event on which a conditional budget must be controlled.

S19.4 Retrospective threshold approximation

Put $n = 2^K$, partition $[0,1)$ into its n half-open grid cells, and let Z be the maximum number of inputs in one cell. A threshold $a < 1$ can be rounded down to the left endpoint of its cell; replace $a = 1$ by $1 - 1/n$. The original and rounded predictions differ only in that cell. Thus, on every path outside the zero-probability endpoint exceptions,

$$\min_{f \in \mathcal{C}_J} L_f - \inf_{a \in [0,1]} L_{h_a} \leq Z. \tag{301}$$

This is simultaneous over all retrospective thresholds.

Under the product reference law, for each integer $r \geq 1$,

$$\Pr_{\mu^T}(Z \geq r) \leq \binom{T}{r} n^{1-r}.$$

Indeed, choose r time indices and a cell containing their independent inputs, and sum over these choices. With $s = (\log n)/2$, this gives

$$\begin{aligned}
\mathbb{E}_{\mu^T} e^{sZ} &\leq e^s + \sum_{r=2}^{T} e^{sr} \binom{T}{r} n^{1-r} \\
&\leq \sqrt{n} + n\left[(1 + n^{-1/2})^T - 1 - Tn^{-1/2}\right] \\
&\leq \sqrt{n} + \frac{T^2}{2} e^{T/\sqrt{n}}.
\end{aligned}$$

The last inequality uses $\binom{T}{r} \le T^r/r!$ and $e^z - 1 - z \le z^2 e^z/2$. Our choice satisfies $n \ge T^4$: $\log T \le \sqrt{T}$ and $K \log 2 \ge 384\sqrt{T} \log 2 \ge 4 \log T$. Since $T \ge 2$, it follows that

$$\mathbb{E}_{\mu^T} e^{sZ} \le 3\sqrt{n}. \tag{302}$$

Let P be the actual input-path law. Conditioning only on preceding inputs averages the full-history conditional input laws. Convexity of KL and then its chain rule yield $\mathrm{KL}(\mathsf{P}\|\mu^T) \le B_*$. For a bounded random variable Y, nonnegativity of divergence relative to the probability measure with density $e^Y/\mathbb{E}_Q e^Y$ gives $\mathbb{E}_P Y \le \mathrm{KL}(P\|Q) + \log \mathbb{E}_Q e^Y$. Apply this to $Y = sZ$ and use (302) to obtain

$$\mathbb{E}Z \le 1 + \frac{2(B_* + \log 3)}{K \log 2} \le C + \frac{2B_*}{U}. \tag{303}$$

The statistic Z is measurable on a finite partition, so this step does not require a measurability convention for an uncountable supremum.

S19.5 Evaluation of the rate

Suppose $B_* \le b < T^2$. The parameter definitions give

$$U + \frac{b}{U} + A_0 + \sqrt{TA_0} \le C\sqrt{\max\{T, b\}}, \qquad \frac{M}{\lambda} = \lambda b = \sqrt{bM}.$$

If $b \le T$, then $M \le C \log(e + T)$. If $T < b < T^2$, put $r = T/\sqrt{b} > 1$. Since $U/A_0 < 128r$, the number of fine scales is at most $C \log(er)$. Substitution into (300), (301), and (303) gives $C\Phi_T(b)$ in both ranges. Regret is also at most T, so the minimum with T is valid after changing the constant.

If $T = 1$ or $b \ge T^2$, the trivial bound proves the theorem. For a numerical bound $B \ge 0$, take $b = \max\{1, \lceil B \rceil\}$. For $B < 1$, all the added terms are $O(\sqrt{T})$; for $B \ge 1$, $b \le 2B$. Within either nonsaturated range the square-root term changes by at most a numerical factor. The two formulas have comparable values at their common boundary. If rounding instead gives $b \ge T^2$, then for $T \ge 2$, $B > T^2 - 1 \ge T^2/2$, and $\Phi_T(B) \ge cT$. This proves the zero-budget, small-budget, and saturation assertions.

S19.6 The sparse binary implementation

Use a complete binary tree of depth K, with leaves indexed in order by $0, \dots, 2^K - 1$. A node at depth r has rate

$$\beta_r = \begin{cases} \eta_0, & 0 \le r < k_0, \\ \eta_j, & k_{j-1} \le r < k_j, \quad 1 \le j \le J. \end{cases}$$

Use the normalized binary soft minimum

$$S_\beta(a, b) = -\frac{1}{\beta} \log \frac{e^{-\beta a} + e^{-\beta b}}{2}. \tag{304}$$

An induction on d shows that repeating this operation at a fixed rate in a complete depth-d subtree gives $-\beta^{-1} \log(2^{-d} \sum_i e^{-\beta v_i})$. Each repeated-rate block therefore collapses to the corresponding conceptual split with uniform normalization. At a given depth the normalization changes the unnormalized value by a constant that is the same at every node. To verify this assertion, start at the leaves, where the constant is zero, and induct upward: equal child constants pass through a soft minimum as a common additive shift, and the normalizing shift depends only on the branching number and rate. Hence normalization does not change any conditional probability. The binary distribution is exactly the conceptual distribution, with the initial penalty in (294) unchanged. In particular, the binary refinement does not introduce entropy penalties at additional rates.

An unexpanded node represents a homogeneous subtree whose leaf losses all equal one integer w. Its normalized value is exactly w at every depth. A stored internal node has two children, one value, and an integer lazy tag representing an increase common to its descendant losses. A whole-subtree increment c increases its value and tag by c, using

$$S_\beta(a + c, b + c) = c + S_\beta(a, b).$$

Before a partial update, push a tag to both children and reset it. Expand a homogeneous node only when a partial update reaches it, giving both children its current integer value. Recompute the values of updated ancestors using (304).

After reading the first K digits of $x < 1$, let $z = \lfloor 2^K x \rfloor$. Exactly indices $0, \dots, z$ predict one. For label zero, increment that prefix; for label one, increment its complementary suffix. There is at most one partially intersected node at each depth. Completely intersected siblings are updated through their tags, without traversing their descendants. Thus an example expands at most K nodes, creates at most $2K$ children, and uses $O(K)$ operations. A prefix of size 2^K is a whole-root update, and an empty suffix requires no update. Repeated inputs and inputs at a grid boundary obey the same rule.

Before the next input, sample at each stored internal node using probabilities proportional to the exponentials of its child values. The parent's common tag cancels. Upon reaching an unexpanded node, sample its remaining binary suffix uniformly: all descendant losses and all remaining branching structures are equal. This step allocates no nodes. The result is one explicit K-bit dyadic threshold chosen before the current input. Its later prediction is obtained by a lexicographic comparison of the input and threshold strings.

Induction on updates shows that stored values, including inherited tags, equal the conceptual full-tree values. Induction on branch choices then proves equality of leaf distributions. There are at most $1+2TK$ stored nodes and $O(TK)$ update and sampling operations. For the bit count, traverse input strings instead of comparing K-bit interval endpoints at every node. Form $z+1$ by a single $O(K)$ carry operation and follow its bits. Store child pointers and depths, not K-bit coordinates at every node. Write the sampled threshold as a string during traversal. These operations use $O(K)$ input and output bit work per round, without another factor K.

S19.7 Finite precision

Exact normalized values lie in $[0, T]$, because a normalized soft minimum lies between its arguments. Put

$$\zeta = \frac{1}{64(T+1)^2(K+1)^2}.$$

At each recomputation approximate (304) to error at most ζ relative to the exact rate and the supplied approximate children. Keep integer tags and whole-subtree integer additions exact. Since the gradient of a soft minimum is a probability vector, it is one-Lipschitz in the supremum norm of its arguments. Induction on depth therefore bounds the maintained value error at depth r by $(K-r)\zeta$. This error does not accumulate with the number of rounds: recomputation replaces the old parent value by a new function of its children. Expanding a homogeneous node starts from exact integers. A heterogeneous subtree is never collapsed merely because two approximate values agree.

For a fixed $\beta \le 1/2$, softmax is 2β-Lipschitz from the supremum norm of its arguments to the ℓ_1 distance of its probabilities. In a direction w, its derivative has coordinates

$$-\beta p_i \left(w_i - \sum_j p_j w_j \right),$$

whose absolute sum is at most $2\beta\|w\|_\infty$. Approximate child values thus change a conditional distribution by at most $K\zeta$ in ℓ_1. Evaluate its first probability, clip to $[0,1]$, and round to a dyadic, with at most ζ further ℓ_1 error. Its complement is exact. A dyadic denominator $2^q \ge 16/\zeta$ suffices when the preliminary evaluation error is at most $\zeta/16$. Exactly q fair bits implement this choice; homogeneous suffix bits remain fair.

Couple the exact and implemented branch samplers until they first choose different branches. Their leaf laws have total variation at most

$$K(K+1)\zeta \le \frac{1}{64(T+1)^2}.$$

Hence the sum of their additional expected losses is less than one. This comparison uses the exact conceptual gradient at the same actual history and loss vector as the implementation. The potential inequalities and the full-history KL calculation continue to apply at the implemented history with that gradient.

We specify bounded computations for the indicated approximations. Select rate minima by comparing squared positive rationals. If $\gamma_j \le 4a_j/T$, the background candidate $\sqrt{a_j/(T\gamma_j)}$ is at least $1/2$ and can be discarded. Decide this comparison from the exponent in γ_j and the bit length of $T/(4a_j)$ before constructing a power-of-two denominator. A denominator is materialized only if this exponent is $O(\log(T+2))$. Every nondiscarded squared candidate rate thus has $O(\log(TK+2))$ bits. Bisection gives square roots at the required precision.

For fixed child values of magnitude at most $T+1$, differentiation in the rate gives

$$\left|\frac{\partial S_\beta}{\partial \beta}\right| = \frac{\mathrm{KL}(p\|\,\mathrm{Unif}\{0,1\})}{\beta^2} \le \frac{\log 2}{\beta^2}.$$

On $\beta \ge 1/(4\sqrt{T})$ this is at most $16T\log 2$. The ℓ_1 rate derivative of the child probabilities is at most $2(T+1)$. Approximating each exact rate to accuracy $\zeta/[256(T+1)]$ preserves this lower bound on the rate and makes its contributions smaller than the allocated fractions of ζ.

Evaluate the soft minimum in the stable form

$$\min\{a,b\} + \frac{\log 2 - \log(1+e^{-\beta|a-b|})}{\beta}.$$

For a sufficiently large computable integer $N = O(\log((T+1)/\zeta))$, an exponential with argument at least N can be replaced by zero within its error allowance. This follows from $e \ge 2$ and the lower bound on β. On the remaining bounded argument range, range reduction and Taylor expansion give certified exponential enclosures in polynomially many operations at $O(\log(TK+2))$ working precision. For logarithms on $[1,2]$ use

$$\log z = 2\sum_{j=0}^{\infty} \frac{w^{2j+1}}{2j+1}, \qquad w = \frac{z-1}{z+1}, \quad |w| \le \frac{1}{3}.$$

Its geometric remainder gives the required precision. The stable logistic expression supplies binary probabilities without overflow. Use rounded fixed-precision arithmetic with guard bits and interval enclosures, rather than retaining exact Taylor denominators. The number of terms is polynomial in $\log(TK+2)$; increasing the working precision by the logarithm of that number controls the sum of rounding errors. Choose enclosures within the allocated error bounds and round each final value to a fixed dyadic grid of mesh at most $\zeta/16$. Integer tags preserve this grid.

These procedures terminate by explicit Taylor, geometric-series, and bisection bounds. They test no transcendental equality. There are $O(TK)$ operations of this kind. Values, tags, pointers, depths, and rounded rates have $O(\log(TK+2))$ bits. This proves the bit, storage, and bounded fair-bit assertions in Theorem S18.1. □

S19.8 The matching statistical obstruction and the access scope

The lower construction for Theorem S6.1 at $d = 1$ can be placed on the same uniform reference used here. We detail the atom-to-interval transfer from Appendix S4 to specify what information is preserved. Let μ_0 be the fixed atomic reference on the ordered threshold domain in that construction. For each positive-mass atom z, assign the interval

$$I_z = \left(\sum_{w<z} \mu_0(w),\ \sum_{w\le z} \mu_0(w)\right).$$

These intervals are disjoint, respect the order, and their lengths sum to one. The complement has Lebesgue measure zero. Replace a draw of atom z by a uniform point in I_z. If its conditional atomic law is P, the new likelihood ratio on I_z is $P(z)/\mu_0(z)$, and therefore

$$\mathrm{KL}(P^{\mathrm{new}}\|\mu) = \sum_z P(z)\log\frac{P(z)}{\mu_0(z)} = \mathrm{KL}(P\|\mu_0).$$

A threshold with boundary θ becomes the ordinary threshold with boundary $\sum_{z<\theta}\mu_0(z)$. Its predictions agree at every interval outside endpoints. The additional within-interval coordinate is independent of the current fair label conditional on the atomic construction. It gives an unrestricted

learner additional randomness, but not information about that label. This verifies the same loss lower bound after the transfer, including the fixed-reference epochs in Appendix S6. The inherited lower proof, together with Theorem S18.1, consequently gives the statistical rate $\Theta(\min\{T, \Phi_T(B)\})$ for this explicit threshold class and uniform reference.

At $B = T$, the upper is $O(\sqrt{T \log(eT)})$ with deterministic polynomial bit work. For supplied budget sequences $B_T = o(T^2)$ it is sublinear. The learner receives the input order and constructs ordinary thresholds directly. These computational bounds apply to the specified threshold class and uniform reference.

A known nonatomic reference on the line has a mathematical reduction through its continuous CDF. To use that reduction computationally one must supply the first K bits of a CDF rank, or an equivalent quantile bin representation, together with executable resulting thresholds. An ordinary sample oracle for an unknown reference does not supply that access. General CDF evaluation, exact quantile computation, atom splitting, and transcendental boundary tests lie outside the stated computational model. The uniform-reference algorithm uses none of these operations and has the polynomial work bound proved above.

S20 Quantitative generalized-smoothness comparisons

We give the prior-bound substitutions used in the main text. The theorem numbering for Block and Polyanskiy (2024) refers to arXiv version 3, and that for Blanchard et al. (2026) refers to the published COLT version. All logarithms below are natural. Assume $\log T \geq 256$. The class is nonempty, finite or countable, measurable, and has VC dimension at most d, where $1 \leq d \leq T$. The input and causal round-state spaces are standard Borel. Input constraints hold conditional on the full preceding history of Section 2. The current private hypothesis draw is hidden from the current input-label kernel. The comparisons are statistical existence statements with mathematical access to the supplied reference.

S20.1 Optimizing the KL specialization of approximate rejection

For $\mathrm{KL}(P_t \| \mu) \leq K$ with $K \geq 1$, set $\varphi(u) = u \log u - u + 1$ and $\sigma = 1/K$ in Theorem 10, equation (17), of Block and Polyanskiy (2024). Since $(\varphi')^{-1}(v) = e^v$, the displayed statistical bound becomes, up to universal constants,

$$\sqrt{dT \log T} + \inf_{0<\epsilon<1} \left\{ \epsilon T + \sqrt{dT(\log T + K/\epsilon)} \right\}.$$

The inequalities $(\sqrt{u} + \sqrt{v})/\sqrt{2} \leq \sqrt{u+v} \leq \sqrt{u} + \sqrt{v}$ reduce its optimization to that of $\epsilon T + \sqrt{dTK/\epsilon}$, apart from the displayed baseline. On the positive axis this function has unique minimizer

$$\epsilon_* = (dK/(4T))^{1/3}, \qquad \epsilon_* T + \sqrt{dTK/\epsilon_*} = 3\,2^{-2/3} T^{2/3} (dK)^{1/3}.$$

Indeed its derivative is $T - \sqrt{dTK}\,\epsilon^{-3/2}/2$, which changes sign once. If $dK \leq T$, this minimizer belongs to $(0, 1)$. If $dK > T$, every permitted ϵ gives $\sqrt{dTK/\epsilon} > T$, so use the trivial regret bound T. Thus the optimized, capped expression has order

$$\min \left\{ T, \sqrt{dT \log(eT)} + T^{2/3} (dK)^{1/3} \right\}. \tag{305}$$

In particular, its unit-KL order at fixed d is $T^{2/3}$. Only the unrestricted statistical guarantee is invoked from that theorem. An order statement about its optimized upper-bound expression does not lower-bound the algorithm's regret.

S20.2 Known families and fragmentation

For a family $\mathcal{V}$ of input laws, let $N_{\mathcal{V}}(a)$ be the largest number of disjoint measurable sets A_i satisfying $\sup_{P \in \mathcal{V}} P(A_i) \geq a$ for every i. This is Definition 15 of Blanchard et al. (2026). Their Theorem 18 gives the expression

$$\mathcal{B}_T(\mathcal{V}) = \min \left\{ T, \inf_{0<a\leq 1} \left[\sqrt{dT \log(TN_{\mathcal{V}}(a))} + aT \right] \right\}. \tag{306}$$

The two supplied families used here are

$$\mathcal{K}_\nu = \{P : \mathrm{KL}(P\|\nu) \le 1\}, \qquad \mathcal{C}_\eta(\nu) = \{P : \exists \mu,\ \mathrm{KL}(P\|\mu) \le 1,\ D_{1+\eta}(\mu\|\nu) \le 1\},$$

where $0 < \eta \le 1$. The learner knows these sets as mathematical objects from ν and the announced parameters. The second family permits a different witness μ for each P. It contains every actual conditional input law of our moment model, so an upper for that family applies using zero samples from the hidden reference. A lower for the larger family need not hold under one common μ.

The proof of their Theorem 18 invokes the cover argument of Haghtalab et al. (2020, Theorem 3.1). To make the full-history comparison explicit, we use the adaptive disagreement estimate of Haghtalab et al. (2024, Lemma 3.3). The following deduction gives the required proper learner and explicit tuning.

S20.3 The adaptive cover deduction

We state the two imported cover facts used below (Haghtalab et al., 2024, Sections 3.2–3.3). For a binary VC class of dimension at most d and a fixed probability measure ξ, a proper $L^1(\xi)$ cover $\mathcal{F}_\epsilon \subset \mathcal{F}$ can be chosen with

$$\log |\mathcal{F}_\epsilon| \le C d \log(C/\epsilon).$$

Choose for each $f \in \mathcal{F}$ a cover element g_f with $\xi\{f \ne g_f\} \le \epsilon$. If the successive conditional densities of Z_t relative to ξ are at most $1/\sigma$, the adaptive disagreement estimate is

$$\mathbb{E} \sup_{f \in \mathcal{F}} \sum_{t=1}^{T} \mathbf{1}\{f(Z_t) \ne g_f(Z_t)\} \le C \left[\sqrt{\frac{\epsilon}{\sigma} T \log T \, d \log(1/\epsilon)} + \frac{\epsilon}{\sigma} T \log T \right]. \tag{307}$$

We use it only for $0 < \epsilon \le 1/4$ satisfying the stronger tuning restriction printed in their Section 3.3,

$$\epsilon > \frac{120 d \sigma \log(4e^2/\epsilon)}{T \log T}. \tag{308}$$

The reference may be arbitrary, as allowed by their model and general-measure coupling. The independent proposals in that coupling are indexed samples; repeated values at atoms are distinct draws.

Lemma S20.1. *Suppose* $\log T \ge 256$ *and* $a, b \in (0,1)$ *are known. Assume that every allowed conditional input law is absolutely continuous relative to a known* ν *and satisfies*

$$\nu(A) \le b \quad \Longrightarrow \quad P(A) \le a$$

for every measurable A*. There is a proper pre-input learner with*

$$\mathbb{E}\,\mathrm{Reg}_T \le C \min\{T, \sqrt{dT \log(eT/b)} + aT\}. \tag{309}$$

It uses zero true-reference calls and requires no observed input laws.

Proof. At the actual preceding history, put $r_t = dP_t/d\nu$. Replace X_t by a new dummy point only if $r_t(X_t) > 1/b$, and denote the resulting variable by $\widetilde{X}_t$. The clipped set has ν-mass at most b because $\int r_t \, d\nu = 1$. Its conditional P_t-mass is therefore at most a. With $\widetilde{\nu} = (\nu + \delta_\emptyset)/2$, the retained density is at most $2/b$, and the dummy density is at most two. Thus the clipped process is σ-smooth for $\sigma = b/2$. This construction is for analysis; the learner does not perform the clipping.

Give every hypothesis value zero at the dummy and choose the proper cover under $\widetilde{\nu}$. This extension does not increase VC dimension. Run randomized Hedge on the cover using the original observed losses. For every realized path,

$$\sup_f \sum_t \mathbf{1}\{f(X_t) \ne g_f(X_t)\} \le \sup_f \sum_t \mathbf{1}\{f(\widetilde{X}_t) \ne g_f(\widetilde{X}_t)\} + \sum_t \mathbf{1}\{X_t \ne \widetilde{X}_t\}.$$

The difference between the two comparator losses is at most their disagreement, for either label. Hedge's bound against its finite cover and this inequality therefore compare with the full hindsight infimum. The private pre-input draw has conditional expected loss equal to the Hedge mixture's

loss, since the current input-label kernel does not observe it. Consequently expected regret is at most $C\sqrt{T \log |\mathcal{F}_\epsilon|}$, the maximum disagreement in (307), and aT. The finite-expert inequality is the Hedge bound used in the cited cover argument, valid for every realized loss-vector sequence.

Put $L = \log(eT/\sigma)$. If $dL \le T/16$, choose

$$\epsilon = \frac{1024 d\sigma L}{T \log T}.$$

Then $0 < \epsilon \le 64\sigma/\log T \le 1/4$ and

$$\log(1/\epsilon) = \log(T/\sigma) + \log\log T - \log 1024 - \log d - \log L \le L.$$

Here $L \ge \log T$, $d \ge 1$, and $\log\log T \le \log L$. Since $L \ge 256$, also $\log(4e^2/\epsilon) \le 2L$. The coefficient $1024 > 240$ gives the strict restriction (308). The two disagreement terms are $O(dL)$, and the Hedge term is $O(\sqrt{dTL})$. Since $dL \le T$, the former is absorbed by the latter. If $dL > T/16$, use any fixed hypothesis: its regret is at most $T \le 4\sqrt{dTL}$. Finally $L = \log(eT/b) + \log 2$, and capping regret at T proves (309).

All clipping density bounds hold conditional on the full actual preceding state. If the imported lemma uses less history, conditional averaging preserves the density bound. In particular, the original inputs may remain in the causal state although the auxiliary process records clipped inputs. Hedge is always run on the actual history, with its actual labels and state updates. Countability makes the disagreement supremum measurable; standard Borel kernels supply measurable density versions. The finite cover and all strategy parameters are fixed from known objects before play. □

S20.4 The known-KL and moment upper bounds

Put $c_0 = 1 + \log 2$. Binary data processing gives, for any event with $P(A) \ge a$ and $\mathrm{KL}(P\|\nu) \le K$,

$$a \log(1/\nu(A)) \le K + h(P(A)) \le K + \log 2,$$

where h is binary entropy. Thus $\log N_{\mathcal{K}_\nu}(a) \le c_0/a$. The same inequality supplies the mass implication in Lemma S20.1 with $b = e^{-c_0/a}/2$. For $d < T$, choose $a = (d/T)^{1/3}$; for $d = T$, use the trivial bound. The resulting proper guarantee is

$$C \min\left\{T, \sqrt{dT \log(eT)} + d^{1/3}T^{2/3}\right\}. \tag{310}$$

It has order $O_d(T^{2/3})$ at fixed d, as does the corresponding substitution in (306).

The following proof of reference transfer does not require logarithm integrability under P. Let $\alpha = \eta/(1+\eta)$ and let g be bounded and measurable. Entropy variation and Hölder give

$$\mathbb{E}_P g \le \frac{\mathrm{KL}(P\|\mu)}{\alpha} + \frac{\log \mathbb{E}_\mu e^{\alpha g}}{\alpha},$$
$$\mathbb{E}_\mu e^{\alpha g} \le \left(\int (d\mu/d\nu)^{1+\eta}\, d\nu\right)^{1/(1+\eta)} (\mathbb{E}_\nu e^g)^\alpha.$$

Subtracting $\log \mathbb{E}_\nu e^g$ and taking the entropy variational supremum over bounded g yields

$$\mathrm{KL}(P\|\nu) \le \frac{1+\eta}{\eta}\,\mathrm{KL}(P\|\mu) + D_{1+\eta}(\mu\|\nu). \tag{311}$$

Under our radius-one promise this is at most $2 + 1/\eta \le 3/\eta$. Substitution into (305) gives $O_d(\min\{T, T^{2/3}\eta^{-1/3}\})$ for sufficiently large T, uniformly in $0 < \eta \le 1$. This is an input KL promise, not an input Rényi moment.

There is a second fragmentation bound. Jensen gives $\mathrm{KL}(\mu\|\nu) \le D_{1+\eta}(\mu\|\nu) \le 1$. Writing $r = d\mu/d\nu$, the positive part is integrable by the moment bound, and the negative part of $r\log r$ is bounded in absolute integral, so this use of Jensen is legitimate. Applying binary KL twice, $P(A) \ge a$ implies

$$\mu(A) \ge e^{-c_0/a}, \qquad \log(1/\nu(A)) \le \frac{c_0}{\mu(A)} \le c_0 e^{c_0/a}.$$

Together with (311), disjointness gives

$$\log N_{\mathcal{C}_\eta(\nu)}(a) \le \min\left\{c_0 e^{c_0/a}, \frac{3+\log 2}{\eta a}\right\}. \tag{312}$$

If an event supremum is not attained, apply these implications at levels below a and pass to the limit. For the proper learner, either event bound gives the hypothesis of Lemma S20.1 by taking b to be half the resulting lower bound on $\nu(A)$.

Choose $a = 2c_0/\log T$ in the first bound. The square-root term in that lemma is $O_d(T^{3/4} + \sqrt{T\log T})$, and $aT = O(T/\log T)$. For fixed d and sufficiently large T depending only on d, this gives $O_d(T/\log T)$. For the second bound choose $a = (d/(\eta T))^{1/3}$ if this is below one; otherwise use the trivial cap. It gives $O(\sqrt{dT\log(eT)} + T^{2/3}(d/\eta)^{1/3})$ when nontrivial. Choose between the two strategies before play. At fixed d, the statistical term is absorbed into either displayed branch for sufficiently large T, uniformly in η. Thus there is a proper learner using zero true-reference calls with

$$\mathbb{E}\,\mathrm{Reg}_T \le C_d \min\left\{\frac{T}{\log(eT)}, T^{2/3}\eta^{-1/3}\right\}, \qquad 0 < \eta \le 1. \tag{313}$$

The threshold for T depends only on fixed d. The same tuning upper-bounds the fragmentation expression (306). This derivation retains the actual-history and properness requirements through Lemma S20.1.

S20.5 Order of the displayed expression on nonatomic measures

We now assume that ν is nonatomic. The following lower calculations concern (306), not regret. For the known KL ball and $0 < a \le 1/8$, let $n = \lfloor e^{1/(2a)} \rfloor$, partition into n equal-ν-mass cells A_i, and put $Q_i = \nu(\cdot \mid A_i)$ and $P_i = (1-a)\nu + aQ_i$. Convexity gives $\mathrm{KL}(P_i\|\nu) \le a\log n \le 1/2$, while $P_i(A_i) \ge a$. Hence

$$\log N_{\mathcal{K}_\nu}(a) \ge \log n \ge \frac{1}{2a} - \log 2 \ge \frac{1}{4a}.$$

Together with the preceding upper bound, this is of order $1/a$ near zero. Minimizing $\sqrt{dT/(4a)} + aT$ gives a constant times $d^{1/3}T^{2/3}$. The remaining scales $a > 1/8$ already cost at least $T/8$. Therefore, for fixed d and large T, $\mathcal{B}_T(\mathcal{K}_\nu) = \Theta_d(T^{2/3})$.

For the moment family and $0 < a \le 1/8$, put

$$\gamma = e^{-1/(2a)}, \qquad n = \left\lfloor \left(1 + \frac{e^\eta - 1}{\gamma}\right)^{1/\eta} \right\rfloor.$$

Take n equal-mass cells, define Q_i as above, and set

$$\mu_i = (1-\gamma)\nu + \gamma Q_i, \qquad P_i = (1-a)\mu_i + aQ_i.$$

Convexity of the power function yields

$$\int (d\mu_i/d\nu)^{1+\eta}\, d\nu \le 1 - \gamma + \gamma n^\eta \le e^\eta.$$

Also $\mu_i \ge \gamma Q_i$, so

$$\mathrm{KL}(P_i\|\mu_i) \le a\,\mathrm{KL}(Q_i\|\mu_i) \le a\log(1/\gamma) = 1/2, \qquad P_i(A_i) \ge a.$$

It follows that $N_{\mathcal{C}_\eta(\nu)}(a) \ge n$. Write

$$F(a,\eta) = \eta^{-1}\log(1 + \eta e^{1/(2a)}).$$

The function $\log(1+u\eta)/\eta$ decreases in $\eta > 0$, since $\log(1+x) \ge x/(1+x)$. Thus $F(a,\eta) \ge \log(1+e^4) > 4$ throughout the stated range. Since $e^\eta - 1 \ge \eta$, taking the integer floor loses at most $\log 2$, and

$$\log N_{\mathcal{C}_\eta(\nu)}(a) \ge F(a,\eta)/2. \tag{314}$$

To lower-bound every tuning in (306), first consider $a \geq 1/(8\log T)$. Its approximation term is at least $T/(8\log T)$. Otherwise $a < 1/8$ and we may set $\eta_0 = e^{-1/(4a)}$. If $\eta \geq \eta_0$, then

$$\log(1+\eta e^{1/(2a)}) \geq \frac{1}{4a}, \qquad \log N_{\mathcal{C}_\eta(\nu)}(a) \geq \frac{1}{8\eta a}.$$

The sum $\sqrt{dT/(8\eta a)} + aT$ is at least a universal constant times $T^{2/3}(d/\eta)^{1/3}$, by the same positive-axis minimization used above. If $\eta < \eta_0$, monotonicity gives

$$F(a,\eta) \geq F(a,\eta_0) \geq \frac{e^{1/(4a)}}{4a}.$$

Here $1/(4a) > 2\log T$, so (314) gives $\log N_{\mathcal{C}_\eta(\nu)}(a) > T^2 \log T$; its square-root cost exceeds T. These cases cover every $a \in (0,1]$. The outer cap at T preserves the lower bound because $T/\log T \leq T$. Combining with (313)'s expression calculation proves

$$\mathcal{B}_T(\mathcal{C}_\eta(\nu)) = \Theta_d\left(\min\left\{\frac{T}{\log(eT)}, T^{2/3}\eta^{-1/3}\right\}\right) \tag{315}$$

uniformly for $0 < \eta \leq 1$ and sufficiently large T depending only on fixed d. The mixtures μ_i differ across cells. Neither this construction nor the order of (306) gives a common-reference regret lower. For measures with atoms the upper bounds remain valid, but fragmentation and the optimized expression may be smaller.

S20.6 The prior lower expression and the remaining comparison

Theorem 16 of Blanchard et al. (2026) states that, when $N_{\mathcal{V}}(a) \geq 3d$, some class of VC dimension at most d has a regret lower bound proportional to

$$\sqrt{adT\log(N_{\mathcal{V}}(a)/d)} \wedge aT. \tag{316}$$

For the known KL ball, $\log N(a) \leq c_0/a$ bounds this expression by $C_d\sqrt{T}$ for every permitted scale. For a nonatomic known reference, choose a fixed sufficiently small $a > 0$ so that the construction above gives $N(a) \geq 3d$. Then its square-root term has order $\sqrt{T}$ and aT eventually exceeds it. Thus the largest displayed lower has order $\sqrt{T}$ for fixed d and a nonatomic reference. It does not directly give the extra $\sqrt{\log T}$ in our matching lower.

For example, write $L = \log(eT)$ and take $\eta = L^3/T$ with T sufficiently large. On a nonatomic supplied measure, (315) has order T/L. Our zero-sample law instead has information term

$$\sqrt{\frac{T}{\eta}\log\left(e + \frac{T\eta}{L^2}\right)} = \frac{T\sqrt{\log(e+L)}}{L^{3/2}}.$$

This is below T/L and dominates the statistical floor $\sqrt{TL}$ for large T. The ratio of the two displayed guarantees grows as $\sqrt{L/\log L}$. At $\eta = 1/T$, both expressions have order T/L.

The family-only upper uses no hidden-reference samples and supplies no sample-allowance dependence. However, the union relaxation alone does not explain the zero-sample gap: our zero-sample upper can also use a per-law family. The multiscale analysis improves the displayed single-scale balance in that setting. Common-reference structure is additionally used in our sample-dependent geometry and lower constructions.

S21 A full-history consequence of the simulator potential

The exponential potential of Voitovych et al. (2026, Lemmas 4.2–4.3 and Proposition 4.4) also gives a bound for our full-history budget. We give the deduction because the source's stated trajectory-KL guarantee and its adaptive-label assumption do not directly identify that budget.

Proposition S21.1. *Let $1 \leq d \leq T$ be integers, let $0 \leq B < \infty$, and let $\mathcal{F}$ be a known nonempty finite or countable class of measurable binary functions with $\mathrm{VC}(\mathcal{F}) \leq d$. Let μ be a known Borel*

probability measure. Use the standard Borel round states, measurable causal kernels, and full preceding histories of Section 2. There is a measurable unrestricted post-input learner such that

$$\mathbb{E}\,\mathrm{Reg}_T \le C\min\{T, \sqrt{dT} + \sqrt{TB}\} \tag{317}$$

whenever $\mathbb{E}\sum_{t=1}^{T}\mathrm{KL}(P_t(\cdot \mid H_{t-1})\|\mu) \le B$*, where* C *is universal. The bound holds for fractional loss and for expected binary loss when the label does not observe the fresh current prediction coin. The learner may use* $T, B, \mathcal{F}, \mu$ *as mathematical objects and integrate exactly; no properness, computational, or sample-access guarantee is asserted.*

Proof. Fix $\lambda > 0$ and put $g_f = (2f-1)/2$. For a labeled list $s_{1:t} = ((x_1, y_1), \dots, (x_t, y_t))$, write $L_t(f) = \sum_{j=1}^{t} \ell_f(x_j, y_j)$ and define

$$A_t(s_{1:t}) = \mathbb{E}_{Z,\varepsilon} \exp\left(\lambda \sup_{f\in\mathcal{F}} \left\{\sum_{j=t+1}^{T} \varepsilon_j g_f(Z_j) - L_t(f)\right\}\right). \tag{318}$$

Here $Z_{t+1:T} \sim \mu^{T-t}$ and the ε_j are independent fair signs, independent of Z. They are integration variables, not revealed future randomness. The bounded exponent makes A_t finite and strictly positive. Countability makes the supremum measurable, and integration against the fixed product law preserves measurability in $s_{1:t}$.

After observing x at time t, put $A_y = A_t(s_{1:t-1}, (x, y))$ and choose

$$q_t(s_{1:t-1}, x) = \frac{1}{2} + \frac{\log A_1 - \log A_0}{2\lambda}. \tag{319}$$

Changing y changes each comparator loss by at most one, so $e^{-\lambda} \le A_1/A_0 \le e^{\lambda}$. Thus q_t is measurable and belongs to $[0, 1]$. It uses the observed labeled list and current input, without access to the adversary's state or conditional laws. For either $y \in \{0, 1\}$,

$$e^{\lambda\ell(q_t,y)} A_y = e^{\lambda/2}\sqrt{A_0 A_1} \le e^{\lambda/2}\frac{A_0 + A_1}{2}.$$

The binary-loss identity $\ell_f(x, (1+\varepsilon)/2) = 1/2 - \varepsilon g_f(x)$ shows that integrating the right side over $x \sim \mu$ gives $A_{t-1}(s_{1:t-1})$. Consequently, for every labeled list,

$$\mathbb{E}_{x\sim\mu} \sup_{y\in\{0,1\}} e^{\lambda\ell(q_t(s_{1:t-1},x),y)} A_t(s_{1:t-1}, (x, y)) \le A_{t-1}(s_{1:t-1}). \tag{320}$$

Fix an admissible causal environment and let P denote its complete interaction law with this learner. Construct Q in the same temporal order, with the same initial state law and the same label, prediction, and auxiliary-state kernels, replacing only each input kernel by a fresh draw from μ. The kernels of the environment are used only in this comparison, and are not supplied to the learner.

We specify this construction also on null histories. Choose regular measurable kernel versions on the standard Borel history spaces and fix their extensions once. Finite expected cost implies finite conditional input KL at P-almost every preceding history at each of the finitely many rounds. Replace an input kernel by μ wherever its KL is infinite. This is a measurable modification on a P-null set and does not change P. All input kernels now admit measurable densities with respect to μ. Use the same fixed extensions of the other kernels under both laws. Histories reached only under Q need not be typical under P; inequality (320) holds at those histories as well.

The product of the conditional input density ratios gives $\mathsf{P} \ll \mathsf{Q}$. The chain rule in the actual temporal order yields

$$\mathrm{KL}(\mathsf{P}\|\mathsf{Q}) = \mathbb{E}_{\mathsf{P}} \sum_{t=1}^{T} \mathrm{KL}(P_t(\cdot \mid H_{t-1})\|\mu) \le B. \tag{321}$$

Identical initial, label, learner, and state-transition kernels contribute zero conditional divergence. The negative parts of the input log density ratios are integrable, since $\int_{\{r<1\}} r\log(1/r)\,d\mu \le 1/e$ for each conditional density. Thus the displayed chain rule is also valid under the expected-cost

assumption. Only the announced full preceding causal history is used at each input stage; no future seed stream or future input coordinates are added.

Under Q, the next input has law μ given that full history. Since (320) is uniform over the label, it applies to every allowed causal label and state rule. Hence

$$M_t = \exp\left(\lambda \sum_{j=1}^{t} \ell(q_j, Y_j)\right) A_t(S_{1:t})$$

is a nonnegative Q-supermartingale. At time T the potential equals $A_T(S_{1:T}) = \exp(-\lambda \inf_f L_T(f))$, so

$$\mathbb{E}_{\mathsf{Q}} e^{\lambda \operatorname{Reg}_T} \le A_0. \tag{322}$$

The regret here is fractional regret, with the hindsight infimum kept inside the expectation.

We use the classical uniform VC Rademacher inequality, in the deterministic-input form used in the proof of Voitovych et al. (2026, Lemma 4.3): for every fixed $x_{1:T}$,

$$\mathbb{E}_\varepsilon W_x(\varepsilon) \le C\sqrt{dT}, \qquad W_x(\varepsilon) = \sup_{f \in \mathcal{F}} \sum_{t=1}^{T} \varepsilon_t g_f(x_t). \tag{323}$$

This classical complexity bound is the only VC estimate imported here. Changing one sign changes W_x by at most one. Expose the independent signs successively: each conditional-expectation martingale difference has conditional range of length at most one. Hoeffding's exponential bound for these differences, iterated over the T signs, gives

$$\mathbb{E}_\varepsilon e^{\lambda W_x(\varepsilon)} \le \exp\left(C\lambda\sqrt{dT} + \frac{\lambda^2 T}{8}\right).$$

Integrating $x_{1:T} \sim \mu^T$ preserves this uniform bound, and therefore $\log A_0 \le C\lambda\sqrt{dT} + \lambda^2 T/8$.

The entropy variational inequality for the full laws P, Q and the bounded variable Reg_T now gives

$$\begin{aligned}\mathbb{E}_{\mathsf{P}} \operatorname{Reg}_T &\le \frac{\mathrm{KL}(\mathsf{P}\|\mathsf{Q}) + \log \mathbb{E}_{\mathsf{Q}} e^{\lambda \operatorname{Reg}_T}}{\lambda} \\ &\le C\sqrt{dT} + \frac{\lambda T}{8} + \frac{B}{\lambda}.\end{aligned}$$

The inequality follows by tilting Q by $e^{\lambda \operatorname{Reg}_T}/\mathbb{E}_{\mathsf{Q}} e^{\lambda \operatorname{Reg}_T}$ and using nonnegativity of KL. For $B > 0$, take $\lambda = \sqrt{8B/T}$. For $B = 0$, take $\lambda = 1/T$; the additional $1/8$ is absorbed by $\sqrt{dT}$. Since regret is at most T pathwise, this proves (317) for fractional predictions.

For binary predictions, draw a fresh private Bernoulli variable of mean q_t after observing X_t. Include its kernel and preceding realized coins in the two interaction laws above. The same argument bounds fractional regret on this actual interaction. The current label cannot observe the fresh coin, so conditioning on the current history, input, and label gives expected binary loss $\ell(q_t, Y_t)$. Summing these identities proves equality of the expected binary and fractional regrets with the same hindsight comparator. Past realized coins may affect later inputs and labels. □

At per-round conditional KL one, the substitution $B = T$ in (317) gives only $O(T)$. Applying Lemma S14.1 to a supplied forward reference moment gives a larger per-round KL bound, so this deduction is again capped at T and gives no dependence on a reference-sample allowance. The proof uses the complete interaction laws, rather than equating the input-trajectory divergence with the full-history cost. It supplies an unrestricted post-input comparison; the proper upper and the sharper budget dependence are provided by Theorem S6.1.

S22 Finite-bit thresholds with an unknown reference

An input x is supplied as follows. Write $j = \lfloor x \rfloor$. Encode its sign, then the length $\ell \ge 1$ of the binary expansion of $|j|$ as ℓ ones followed by a zero, then those ℓ bits, then the binary digits of $x - j$.

Zero has length one and positive sign; dyadic fractions use their terminating expansion followed by zeros. Each input is a retained read-only forward stream. Reading and advancing by one bit costs one bit operation. Only the first sixteen streams need to be retained between rounds. Output integers and positive denominators have the same self-delimiting encoding. We charge memory addresses and all arithmetic; no operation on an unread real number is available.

Theorem S22.1. *There are universal constants $C, c > 0$ such that, for each integer $T \geq 1$, one proper learner independent of the reference has the following property. For every nondegenerate log-concave probability measure μ on $\mathbb{R}$, with mean a and standard deviation σ, suppose every input law, conditional on the full preceding history, has KL divergence at most one from μ, with the current-round privacy of Section 2. The learner selects finite rational thresholds before the current input and satisfies*

$$\mathbb{E}\,\mathrm{Reg}_T \leq C\sqrt{T\log(eT)}$$

against all real thresholds. Under the raw binary-stream encoding specified in Appendix S22, its computation terminates almost surely and its expected bit work, including input, output, addressing and arithmetic, is at most $C(T + L_\mu)^c$, where

$$L_\mu = 1 + \log_2(1 + |a| + \sigma) + \log_2(1 + \sigma^{-1}).$$

The learner receives only T and the online inputs and labels. The reference's shape, mean and scale are not supplied. It takes no external reference samples. It uses no evaluation, learning, integration, CDF or quantile oracle.

Proof of Theorem S22.1. The learner below depends on T but not on the reference. Put $h_r(x) = \mathbf{1}\{x \geq r\}$. If $T \leq 5016$, use h_0 throughout. Its prediction is determined by the first sign bit, and its regret is at most $T \leq \sqrt{5016}\sqrt{T\log(eT)}$. Henceforth $T > 5016$. Use h_0 for the first $n = 16$ rounds.

Finite-read calibration. Read the integer parts of these inputs. At fractional precision $r = 0, 1, \ldots,$ maintain the signed integers $m_i(r) = \lfloor 2^r X_i \rfloor$, updating by $m_i(r+1) = 2m_i(r) + \text{next bit}$. Stop at the first $r = r_*$ for which $d_r = \max_i m_i(r) - \min_i m_i(r) \geq 8$. Set

$$b = 2^{-r_*} m_1(r_*), \qquad s = 2^{\lceil \log_2 d_{r_*} \rceil - r_*}. \tag{*}$$

The ceiling is computed from an integer's leading bit and a power-of-two test. Both parameters are rational, and $s > 0$. If $R_x = \max_i X_i - \min_i X_i$, the dyadic enclosures give $|2^{-r} d_r - R_x| \leq 2^{-r}$. Consequently

$$\begin{gathered} R_x/2 \leq s \leq 3R_x, \qquad |b - X_1| \leq R_x/7, \\ r_* \leq \max\{0, \lceil \log_2(16/R_x) \rceil\}. \end{gathered} \tag{324}$$

The last inequality follows because $2^{-r} \leq R_x/16$ implies $d_r \geq 15$. The first two follow from $d_{r_*} \geq 8$: $2^{-r_*} \leq R_x/7$ and $6R_x/7 \leq 2^{-r_*} d_{r_*} \leq 8R_x/7$.

For analysis alone let $Z_i = (X_i - a)/\sigma$, let $M = \max_i |Z_i|$, and let $R = \max_i Z_i - \min_i Z_i$. The one-dimensional reference facts proved in Appendix S15.6 give density at most one and, by (238), conditional $\mathbb{E}|Z_i| \leq 144$. Thus $\mathbb{E}M \leq 2304$. An interval of length e^{-u} has reference mass at most e^{-u}. If its actual probability is p and reference probability is $q > 0$, binary KL is at least $p\log(1/q) - \log 2$, since binary entropy is at most $\log 2$ and $-(1-p)\log(1-q) \geq 0$. If $q = 0$, absolute continuity gives $p = 0$. Thus, writing $c_0 = 1 + \log 2$, conditional KL at most one gives

$$\Pr\{|Z_{2i} - Z_{2i-1}| < e^{-u}/2 \mid \mathcal{H}_{2i}\} \leq \min\{1, c_0/u\}. \tag{**}$$

Here $\mathcal{H}_{2i}$ is the full state before input $2i$, including input $2i - 1$. Apply conditional expectation successively to the eight pairs. The event $R < e^{-u}/2$ requires all eight displayed events, so, with $J = [\log(1/(2R))]_+$ and $V = \log(1 + M)$,

$$\begin{gathered} \Pr\{J > u\} \leq \min\{1, (c_0/u)^8\}, \\ \mathbb{E}J^j \leq \frac{8c_0^j}{8 - j}, \quad \mathbb{E}V^j \leq 2304\, j! \quad (1 \leq j \leq 4). \end{gathered} \tag{325}$$

The middle bound integrates the tail at c_0. The last uses $V^j \leq j!(e^V - 1) = j!M$. In particular $R > 0$ almost surely, and calibration terminates.

Let $A = 1 + r_* + \max_i \ell(\lfloor X_i \rfloor)$. The encoded sizes of b, s and every calibration register are $O(A)$: the exponent of s lies between $-r_*$ and the largest integer-part length plus a constant. Moreover, the preceding range bound gives $r_* \le C + \log_2(1 + \sigma^{-1}) + J/\log 2$, and $|X_i| \le |a| + \sigma M$ gives $\ell(\lfloor X_i \rfloor) \le C + \log_2(1 + |a| + \sigma) + V/\log 2$. Hence

$$A \le C(L_\mu + 1 + J + V). \tag{326}$$

Calibration uses at most $r_* + 1$ passes over sixteen integers of $O(A)$ bits. Schoolbook signed arithmetic, its counters and charged addresses therefore cost at most CA^4 bits. All stream reads in these passes are forward; no digit is reread.

A rational reference and its cost. Use the density and increasing bijection

$$q(x) = \frac{s}{2(s + |x - b|)^2}, \qquad F(x) = \frac{1}{2} + \frac{x - b}{2(s + |x - b|)}. \tag{327}$$

Integration on either side of b gives mass $1/2$, and differentiation gives $F' = q$. For $0 < u < 1$ its inverse is

$$F^{-1}(u) = \begin{cases} b - s(1 - 2u)/(2u), & u < 1/2, \\ b + s(2u - 1)/(2(1 - u)), & u \ge 1/2. \end{cases} \tag{328}$$

These are rational operations at dyadic u.

In standardized coordinates put $e = (b - a)/\sigma$ and $\tau = s/\sigma$. Calibration gives $R/2 \le \tau \le 3R$ and $|e| \le 2M$. If f is the standardized reference density, then $f \le 1$, and, wherever $f(z) > 0$,

$$\log \frac{f(z)}{\tau/[2(\tau + |z - e|)^2]} \le 9 \log 2 + J + 2V + 2 \log(1 + |z|). \tag{329}$$

Indeed, $-\log \tau \le 2 \log 2 + J$ and $\tau + |z - e| \le 8(1 + M)(1 + |z|)$. For every P with $\mathrm{KL}(P \| f) \le 1$, its first-moment bound and Jensen's inequality imply $P \log(1 + |Z|) \le \log 145$. Since $9 \log 2 + 2 \log 145 < 32$, define

$$H = 32 + J + 2V, \qquad \mathrm{KL}(P_{\text{raw}} \| q) \le 1 + H, \qquad 1 + \mathbb{E}H < 5000. \tag{330}$$

This holds simultaneously over all admissible laws, including laws selected after seeing the prefix. The KL identity is legitimate: the negative part of a log probability-density ratio has integral at most $1/e$; its positive part here is bounded by the integrable positive part of $\log(dP/df)$ plus the nonnegative right side of (329). Thus both new log-ratio parts are integrable before the identity is used. The expectation bound follows from $\mathbb{E}J \le 8c_0/7$ and $\mathbb{E}V \le 2304$. Equations (325) and (326) also give

$$\mathbb{E}A^j \le C_j L_\mu^j \ (1 \le j \le 4), \qquad \mathbb{E}[A^j(1 + H)] \le C_j L_\mu^j \ (0 \le j \le 3). \tag{331}$$

For the second assertion, expand $(L_\mu + 1 + J + V)^j (33 + J + 2V)$ and use moments through $j + 1$.

The midpoint tree. Set $h = T - 16$ and use the rates and depths of Appendix S19 with horizon h and deterministic integer budget $b_h = 5000h < h^2$. At each depth k_j replace its representatives by $(i + 1/2)2^{-k_j}$, with the same index i and parent index $\lfloor i/2^{k_j - k_{j-1}} \rfloor$. A leaf and its ancestor lie in the same ancestor dyadic cell. Their uniform disagreement mass is at most $2^{-k_{j-1}}$, which is exactly the bound used in (290). Branching numbers and rates are unchanged. The potential (293) depends only on the leaf loss vector. Its derivative identities and (296) therefore apply to these representatives; the predictable disagreements have the same reference means. This verifies every hypothesis of (297)–(300). For hindsight approximation, send a real rank in $[0, 1)$ to its terminal cell's midpoint, and send rank one to the last midpoint. Disagreement is confined to that cell, so (301) and the unchanged occupancy calculation (303) apply.

For the implementation, leaves remain ordered by their K-bit indices, where $K = k_J$. The binary soft minimum (304) has the same repeated-rate blocks and equal normalization shifts. At any input exactly an initial segment of leaves predicts one. Given its size $w \in \{0, \dots, 2^K\}$, the prefix/suffix lazy updates in Appendix S19.6 visit one partial node per depth; use the $K + 1$ bits of w in place of the former value $z + 1$. The homogeneous-subtree representation and branch sampler consequently maintain the same conceptual distribution. Values still lie in $[0, h]$, and the number of recomputations and branches is unchanged. The error, Lipschitz and dyadic-sampling bounds in Appendix S19.7

therefore apply without a changed precision parameter. They add less than one to expected regret and use $O(hK[\log(hK+2)]^{c_1})$ bit work.

For a sampled leaf output the finite rational threshold $F^{-1}((i+1/2)2^{-K})$ before the current input. The full calibration transcript is initial side information. Conditional on the entire subsequent history, KL invariance gives the transformed input divergence to uniform as $\mathrm{KL}(P_t\|q) \le 1+H$. Its unconditional total expectation is at most $5000h$. The proved midpoint version of Theorem S18.1, with the fixed budget b_h, therefore gives $C\sqrt{h\log(eh)}$ expected suffix regret. Its proof uses precisely these full-history costs, so it permits this initial side information. The input and label are still generated without the fresh private leaf draw. The bijection F preserves every real threshold's losses on the path and maps their parameter set onto $(0,1)$. The midpoint approximation was simultaneous over all ranks, so the suffix bound is against the infimum over all real thresholds. Adding sixteen bounds the full regret as claimed.

Reading a bin with finite expected work. It remains to compute w from raw bits. Let E be the total encoded length of b, s, and put $D = K+E+10$. Each leaf boundary is a rational of numerator and denominator length $O(D)$ by (328); in particular $|b|, s, s^{-1} \le 2^E$ and every boundary has magnitude at most $|b| + s2^K < 2^{2D}$. Read the input's sign and unary integer-length header. If it exceeds $2D+5$, stop: the magnitude of its integer part is at least 2^{2D+5}, so its sign determines whether all or none of the leaves predict one. A saturating counter of $O(\log D)$ bits suffices, and the unread stream is discarded. Otherwise read that integer part and the first $r_0 = 2D+5$ fractional digits. They give an interval $[l, l+2^{-r_0})$ containing the input.

Since $q \le 1/(2s)$, successive quantiles have spacing at least $2s\,2^{-K}$. Our interval is shorter than this. Binary search over the leaf indices, using exact rational comparisons with l, finds the number of boundaries at most l. Test whether the next boundary, if present, is strictly below $l+2^{-r_0}$. If not, this is w; if it is, this single boundary θ is the only unresolved comparison. All searches and cross multiplications involve $O(D)$-bit integers and cost at most CD^4, including addresses. Their intermediate binary encodings have polynomial size in D.

For that last comparison generate the binary expansion of $\theta - \lfloor\theta\rfloor$ by ordinary division. The remainder always has at most the denominator's $O(D)$ bits: double it, compare with the denominator, and subtract if needed. Generate its first r_0 digits using bounded work, then compare each further digit with the next raw input bit until they differ. This last loop retains one bounded-size remainder and a current bit, with no growing prefix or position counter. It costs $C(D+\log(hK+2))^2$ per further digit, including local and tree addresses, and reads the stream only forward. Thus it has work linear in its random length.

Let Z be that further length, zero if no last comparison is needed. For integers $u \ge 0$, under the fitted density q, the event $Z > u$ requires the input to be within 2^{-r_0-u} of one of the 2^K boundaries. The density bound and $s^{-1} \le 2^E$ give

$$\begin{aligned} \Pr_q\{Z > u\} &\le 2^{K+E-r_0-u} \le 2^{-u}, \\ \mathbb{E}_q 2^{Z/2} &= 1 + (\sqrt{2}-1)\sum_{u=0}^{\infty} 2^{u/2}\Pr_q\{Z>u\} \le 1+\sqrt{2} < 3. \end{aligned} \tag{332}$$

Apply the entropy inequality to truncated Z and then monotone convergence. At every actual suffix history,

$$\mathbb{E}[Z \mid \text{history}] \le \frac{2}{\log 2}\{\mathrm{KL}(P_t\|q) + \log 3\} \le C(1+H). \tag{333}$$

Equality to a boundary is null because $P_t \ll \mu$ has a density. The threshold convention predicts one on equality; the streaming comparison may fail to halt only on that null event. Dyadic inputs and the two-expansion ambiguity are also null, with the terminating convention fixed above.

Outputting a sampled rational threshold costs CD^4 by integer arithmetic; no transcendental evaluation occurs. Since $E = O(A)$, combining the calibration cost, the tree cost, the deterministic bin work, and the last-loop bound gives

$$\begin{aligned} \mathbb{E}[\text{bit work}] \le{}& CL_\mu^4 + Ch(K + L_\mu + \log(h+2))^4 \\ &+ ChK[\log(hK+2)]^{c_1}. \end{aligned} \tag{334}$$

Here the random-length term uses $\mathbb{E}[(K+A+\log(h+2))^2(1+H)] \le C(K+L_\mu+\log(h+2))^2$, exactly the mixed moments in (331). No higher moment of the suffix comparison length was used. Finally $K < 768\sqrt{5000h}$, so the display is at most $C(T+L_\mu)^c$ for universal constants. All accesses, storage addresses, rational outputs and fair bits are included in these computations. This completes the proof. □

S23 Approximate log-concavity without reference access

Blanchard et al. (2026, Lemma 12 and Theorem 14) use clipping to a latent sequence and control its disagreements. We use a dummy point in a penalized proper game, so that unobserved corrupted rounds have a linear regret cost. No clipping flags are supplied to the learner.

Put $A_T = \sqrt{T\log(eT)}$ and define

$$R_T(0) = A_T, \qquad R_T(\delta) = A_T + \frac{T}{\log(eT/\delta)} \quad (0 < \delta \le 1).$$

Theorem S23.1. *For each positive integer d there is a constant $C_d > 0$ with the following property, uniformly over positive integers k, T. For every known nonempty countable measurable binary class $\mathcal{F}$ on $\mathbb{R}^k$ with $\mathrm{VC}(\mathcal{F}) \le d$, there is one measurable proper pre-input learner, depending only on $T, d, k, \mathcal{F}$ and the observed history, such that the following holds. For every $0 \le \delta \le 1$, every fixed Borel probability μ, and every full-dimensional log-concave probability ρ with $\mathrm{TV}(\mu,\rho) \le \delta$, its expected regret is at most*

$$C_d\left(\sqrt{k}\,A_T + \frac{T}{\log(eT/\delta)}\right) \quad (\delta > 0), \qquad C_d\sqrt{k}\,A_T \quad (\delta = 0),$$

under the full-history conditional constraint $\mathrm{KL}(P_t\|\mu) \le 1$ and current-round privacy in Section 2. It receives neither μ, ρ, δ nor external reference samples. For each fixed d, k there is $c_{d,k} > 0$ such that, for $T \ge 512$, the supremum of the minimax expected regret over these classes and reference pairs is at least $c_{d,k}R_T(\delta)$ for unrestricted learners, even when ρ and δ are supplied.

The lower matches the horizon and radius dependence at fixed d, k; no optimal dimension dependence is asserted. The upper is a statistical strategy, without a runtime claim. We first prove the reference-use statement needed after calibration.

Lemma S23.2. *Let $1 \le d \le h$, let $K \ge 1$ be known, let ν be a known Borel probability, and let $\mathcal{F}$ be a known nonempty countable measurable binary class of VC dimension at most d on a standard Borel space. There is one measurable proper pre-input strategy, depending only on $h, d, K, \mathcal{F}, \nu$, with the following property. For every predictable decomposition $P_t = G_t + B_t$ into nonnegative subprobability kernels, put $b_t = B_t(X)$ and*

$$c_t = (1-b_t)\,\mathrm{KL}(G_t/(1-b_t)\|\nu), \qquad C_* = \mathbb{E}\sum_{t=1}^{h} c_t, \qquad B_* = \mathbb{E}\sum_{t=1}^{h} b_t.$$

The expression defining c_t is zero when $b_t = 1$. If $C_* < \infty$, the strategy satisfies

$$\mathbb{E}\,\mathrm{Reg}_h \le C\sqrt{K}A_{d,h} + C\frac{A_{d,h}}{h\sqrt{K}}C_* + CB_*.$$

The constant C is universal. The decomposition, its costs and its latent flags are not observed.

Proof. Put $R = h/d$. If $K \ge R$, a fixed hypothesis has regret at most $h \le \sqrt{K}A_{d,h}$, so assume $1 \le K < R$. First let the space and class be finite. Let $X_+ = \{x : \nu(x) > 0\}$, add a point $*$, and extend every classifier by $f(*) = 0$. Restricting to $X_+ \cup \{*\}$ does not increase VC dimension: the new point cannot be shattered. Write $\bar\nu = (\nu + \delta_*)/2$. Use the signed derivation in the proof of Lemma S15.2, before its proper-game step, with $H = h$, deterministic tuning $B_0 = Kh$, and full-support reference $\bar\nu$. Its range $1 \le K < h/d$ holds here. Equations (11) and (223) give fixed parameters

$$A_* \le C\sqrt{K}A_{d,h}, \qquad \tau \le C\frac{A_{d,h}}{h\sqrt{K}},$$

and the signed inequality

$$\mathbb{E}\sup_f \sum_t \varepsilon_t f(\widetilde{X}_t) - \tau\mathbb{E}\sum_t \mathrm{KL}(\widetilde{P}_t\|\bar{\nu}) \le A_* \tag{335}$$

for every adapted signed process on $X_+ \cup \{*\}$. This is (11), before imposing an actual-cost bound, for the full-support reference $\bar{\nu}$.

For a probability vector p on the original space, define

$$\psi(p) = \min_{\substack{0\le g_x\le p_x \\ g_x=0\ (x\notin X_+)}} \left\{ \tau\,\mathrm{KL}(g + (1-\textstyle\sum_x g_x)\delta_* \,\|\, \bar{\nu}) + 1 - \displaystyle\sum_x g_x \right\}. \tag{336}$$

The objective is continuous and strictly convex on these compact boxes, so its minimizer is unique. The value is convex: convex combinations of feasible pairs (p, g) remain feasible, and the objective is convex in g. It is continuous as well. If $p^{(j)} \to p$, any limit of minimizers is feasible at p, giving the lower limit inequality. For a minimizer g at p, the vectors $g_x^{(j)} = \min\{g_x, p_x^{(j)}\}$ are feasible and converge to g, giving the upper limit inequality. The same compactness argument and uniqueness show that the minimizer is a continuous function of p. Finally $0 \le \psi(p) \le 1 + \tau\log 2$, by choosing $g = 0$.

For the auxiliary payoff $\mathrm{Reg}_h - 2\sum_t \psi(P_t)$, write L for the comparator-loss vector and set $W_{h+1}(L) = -\min_f L(f)$. Its backward recursion is

$$W_t(L) = \min_{q\in\Delta(\mathcal{F})} \max_D \{qD\ell - 2\psi(D_X) + \mathbb{E}_{S\sim D} W_{t+1}(L + \ell_S)\}.$$

The objective is affine in q and concave and continuous in the joint input-label law D, by convexity and continuity of ψ. Minimax on these compact simplexes interchanges the extrema. The minimizing proper strategy depends only on L, which has finitely many reachable values. Iterating the recursion and using $\min_g D_t\ell_g \le D_t\ell_f$ gives

$$W_1(0) \le \sup_{D_t} \mathbb{E}\left[\sup_f \sum_t \{D_t\ell_f - \ell_f(S_t)\} - 2\sum_t \psi((D_t)_X)\right].$$

The tangent-sample construction in the proof of Lemma S6.2 now applies to each fixed law. Its conditional Jensen step leaves the penalty unchanged. Split that penalty equally between the two signed branches. Each branch has input marginal $(D_t)_X$ given its preceding filtration, including after the label-dependent sign change; its signed label-only term has expectation zero. The two branch expectations are therefore bounded by the same signed supremum, giving

$$\mathbb{E}\,\mathrm{Reg}_h \le 2\mathcal{S}_\psi + 2\mathbb{E}\sum_t \psi(P_t), \qquad \mathcal{S}_\psi = \sup \mathbb{E}\left[\sup_f \sum_t \varepsilon_t f(X_t) - \sum_t \psi(P_t)\right].$$

All histories may contain predictable auxiliary state. The strategy does not observe the penalty.

For a signed process in this supremum, take the minimizing g_t in (336). Mark an input clean with subprobability law g_t, and otherwise replace it by $*$. This defines a joint kernel with original marginal P_t; the marking is drawn before the independent fair sign. Let N_b count the marked bad rounds. Pathwise,

$$\sup_f \sum_t \varepsilon_t f(X_t) \le \sup_f \sum_t \varepsilon_t f(\widetilde{X}_t) + N_b.$$

The conditional censored law is $g_t + (1 - g_t(X))\delta_*$ and $\mathbb{E}N_b = \mathbb{E}\sum_t(1 - g_t(X))$. The bad-count terms cancel the corresponding part of ψ. Equation (335) then gives $\mathcal{S}_\psi \le A_*$. For any specified decomposition from the lemma, finite c_t forces G_t to vanish on ν-null cells, and

$$\mathrm{KL}(G_t + b_t\delta_*\|\bar{\nu}) = c_t + \log 2 + (1-b_t)\log(1-b_t) + b_t\log b_t \le c_t + \log 2. \tag{*}$$

The usual convention $0\log 0 = 0$ includes both endpoints. Hence $\psi(P_t) \le \tau(c_t + \log 2) + b_t$. The finite bound follows, since $\tau h \le CA_{d,h}/\sqrt{K} \le C\sqrt{K}A_{d,h}$.

For a countable class, choose a proper e^{-v}-cover $\mathcal{C}$ under ν, with $v = \max\{4\log(eh/d), h\sqrt{K}/d, 1\}$, and fixed approximants πf. Partition the original space by the evaluation patterns of $\mathcal{C}$, retaining

every nonempty cell, including ν-null cells. Its finite class has VC dimension at most d. Apply the finite strategy on these cells. Data processing bounds the induced clean KL cost by C_*, while the bad mass is unchanged. The finite guarantee allows the original full history as auxiliary state, so no pathwise budget is inferred.

The difference from the full hindsight comparator is at most $\sup_f \sum_t \mathbf{1}\{f(X_t) \neq \pi f(X_t)\}$. Censor using the specified G_t, B_t, and extend every function by zero at $*$. The displayed difference is bounded by its censored counterpart plus N_b. Under $\bar{\nu}$ each censored disagreement has mass at most $e^{-v}/2$; by $(*)$ the censored process has expected KL cost at most $C_* + h \log 2$. Lemma S6.3 bounds the expected difference by

$$Cd + 2(C_* + h \log 2)/v + B_*.$$

Since $h/v \leq d/\sqrt{K} \leq \sqrt{K} A_{d,h}$, $d \leq \sqrt{K} A_{d,h}$, and $1/v \leq d/(h\sqrt{K}) \leq A_{d,h}/(h\sqrt{K})$, these terms are absorbed. The evaluation partition is measurable, and the finite learner uses finitely many loss states and selects members of $\mathcal{C} \subseteq \mathcal{F}$. This proves measurability and properness without revealing any censoring flag. □

Proof of Theorem S23.1. Put $m = \lceil 16k \log(ek) \rceil$, $n = 8m \leq 136k \log(ek)$. If $T \leq n + d$, use a fixed hypothesis. For larger T, the calibration expert uses a fixed hypothesis for n rounds, then selects a $4m$-element subset J_* of the observed inputs minimizing the determinant of its centered scatter, divided by $4m$, with lexicographic ties. It uses that subset's mean $\bar{X}$ and scatter V. This is the minimum covariance determinant location and scatter construction (Rousseeuw and Van Driessen, 1999, p. 213); no efficient subset-search claim is made. If $V \succ 0$, apply the rational rounding preceding (235) to $(\bar{X}, V)$; otherwise use $(b, Q) = (0, I_k)$. The mesh for symmetric scatter entries is the largest dyadic number at most $\lambda_{\min}(V)/(4k)$, and that for mean coordinates is the largest at most $\sqrt{\lambda_{\min}(V)}/(2\sqrt{k})$, with the fixed ties there. For each rational index (b, Q) fix the strategy of Lemma S23.2 with $\nu = q_{b,Q}$, $h = T - n$ and known scale $K = C_{\text{cal}} k$. The universal constant $C_{\text{cal}} \geq 1$ is specified below. Finite determinant minimization, subset means and rounding are Borel, and the reference index is countably valued. This expert uses none of μ, ρ, δ.

Virtual clean observations. Fix a reference pair for analysis. Put $\epsilon = \mathrm{TV}(\mu, \rho)$ and $\xi = \mu \wedge \rho$, defined by taking the smaller densities relative to $\mu + \rho$. For $\epsilon < 1$, mark each actual input clean with probability $(d\xi/d\mu)(X_i)$, using independent analytical coins. On a clean flag put $W_i = X_i$; otherwise draw W_i freshly from ρ. The actual protocol ignores these variables. Constructing this joint process sequentially with the original transition kernels ensures that, given its augmented preceding state, the actual input still has its original conditional law P_i. No future flag is conditioned upon.

Let b_i be the bad probability and Q_i the normalized clean law at that state. The marked reference laws are $(\xi, \mu - \xi)$ and have the same likelihood ratio as P_i relative to μ. Their joint KL is at most one. Binary entropy is at most $\log 2$, so for $0 < \epsilon \leq \delta < 1$,

$$b_i \leq \beta := \frac{1 + \log 2}{\log(1/\delta)}, \qquad (1 - b_i)\,\mathrm{KL}(Q_i \| \rho) \leq 1 + \log \frac{1}{1 - \epsilon}. \tag{337}$$

For the second inequality, the clean reference $\rho_0 = \xi/(1 - \epsilon)$ has density at most $(1 - \epsilon)^{-1}$ relative to ρ; apply the conditional KL chain rule, then change reference. If $\epsilon = 0$, all flags are clean and $Q_i = P_i$.

Suppose first that $\delta = 0$, with $\beta = 0$, or $D_\delta := \log(1/\delta) \geq D_0 := 4(1 + \log 2)$. Then $b_i \leq \beta \leq 1/4$, the weighted clean KL is at most two, and $\mathrm{KL}(Q_i \| \rho) \leq 4$. The virtual conditional law is $(1 - b_i)Q_i + b_i \rho$. Convexity bounds its KL from ρ by two. Conditioning only on the preceding virtual observations mixes these laws. The KL chain rule therefore gives

$$\mathrm{KL}(\mathcal{L}(W_{1:n}) \| \rho^n) \leq 2n. \tag{338}$$

Let G be the event that at most $2m$ prefix flags are bad. Markov's inequality gives $\Pr(G^c) \leq n\beta/(2m) = 4\beta$. This bound does not grow with dimension.

Geometry of the selected subset. Write a, Σ for the mean and positive-definite covariance of ρ, justified in Appendix S15.6. For analysis standardize the virtual points as $Z_i = \Sigma^{-1/2}(W_i - a)$; their reference density f is isotropic and log-concave. Its unit projection densities are at most one,

$\|f\|_\infty \le C_0^k$, and $\mathbb{E}_f e^{\|Z\|/(B_0\sqrt{k})} \le 4/3$, where $B_0 = 8e\sqrt{7}$, by (239) and its preceding facts. The constants C_0, B_0 are universal. Let e, S be the standardized mean and scatter of the actual selected subset. Every scatter determinant is multiplied by $\det\Sigma$ under this change, so the selected index is unchanged; its mean is affine equivariant as well.

For each $2m$-element subset I, let S_I be its virtual centered scatter divided by $2m$, and put $F_I = \log^+(1/\lambda_{\min}(S_I))$, with value $+\infty$ at singularity. Equation (224), a sum over at most 2^n subsets, and (338) give

$$\log \mathbb{E}_{f^n} e^{(m/8)\max_I F_I} \le n\log 2 + (C_0' + \log 4/8)m,$$
$$\mathbb{E}\max_I F_I \le 128 + 64\log 2 + 8C_0' + \log 4.$$

Here C_0' is the universal constant in that lemma for $(L, B_0, M) = (1, 8e\sqrt{7}, 4/3)$. The entropy inequality is applied to bounded truncations and then monotone convergence; no conditioning on G is used. It also shows that every virtual subset scatter is nonsingular almost surely.

On G, the selected $4m$-subset contains at least $2m$ clean indices. Let I be the first $2m$ such indices and write $\widetilde{e} = (2m)^{-1}\sum_{i\in I} Z_i$. Expanding the selected scatter about this clean subset's mean gives

$$S \succeq \tfrac{1}{2}S_I + \tfrac{1}{2}(\widetilde{e} - e)(\widetilde{e} - e)^\top.$$

Thus $S \succ 0$ on G, and, with $\lambda = \lambda_{\min}(S)$,

$$\mathbb{E}[\mathbf{1}_G \log^+(1/\lambda)] \le L_0, \qquad \|\widetilde{e} - e\|_{S^{-1}}^2 \le 2,$$
$$L_0 = 128 + 65\log 2 + 8C_0' + \log 4.$$

Products with $\mathbf{1}_G$ here and below are defined to be zero outside G; set $\widetilde{e} = 0$ there.

Let R be the $6m$-th smallest of the virtual norms $\|Z_i\|/\sqrt{k}$. On G, at least $4m$ of these $6m$ points are clean actual points. Their scatter has trace at most kR^2. Determinant minimization and AM–GM consequently imply $\mathbf{1}_G[\log\det S]_+ \le 2k\log^+ R$. Under f^n, isotropy and Markov's inequality bound each norm tail by r^{-2} for $r \ge 1$. The event $R > r$ requires at least $2m+1$ large norms. Independence and a union over $2m$-subsets bound its probability by $2^n r^{-4m} = 2^n r^{-n/2}$. Hence

$$\mathbb{E}_{f^n} e^{(n/8)\log^+ R} \le 1 + \frac{n}{8}\int_0^\infty e^{nt/8} 2^n e^{-nt/2}\, dt = 1 + 2^n/3 \le 2^n.$$

Entropy transfer at cost $2n$ gives

$$\mathbb{E}[\mathbf{1}_G(\log\det S)_+] \le C_D k, \qquad C_D = 16(2 + \log 2).$$

For $j = 2, 4$, any law with KL at most j from f has first moment at most $B_j\sqrt{k}$, where $B_j = B_0(j + \log(4/3))$, by the same truncated entropy inequality used in (238). The virtual conditional laws have KL at most two, so $\mathbb{E}\sum_i \|Z_i\| \le nB_2\sqrt{k}$. Since $\mathbf{1}_G\|\widetilde{e}\| \le (2m)^{-1}\sum_i \|Z_i\|$, the inequality $1 + x^2 \le (1+x)^2$ and Jensen give

$$\mathbb{E}[\mathbf{1}_G \log(1 + \|\widetilde{e}\|^2/k)] \le L_e := 2\log(1 + 4B_2).$$

All selected quantities were bounded by virtual-prefix functions or its total norm. No selected subset was treated as iid, and no actual-input second moment was assumed.

Transfer and the suffix. On G, for every standardized point z,

$$1 + \|z - e\|_{S^{-1}}^2 \le 5 + (4/\lambda)(\|z\|^2 + \|\widetilde{e}\|^2)$$
$$\le k(5 + 4/\lambda)(1 + \|z\|^2/k)(1 + \|\widetilde{e}\|^2/k).$$

The first step uses the preceding Mahalanobis bound and the squared triangle inequality. Expansion proves the second for $k \ge 1$. For each Q with $\mathrm{KL}(Q\|\rho) \le 4$, the standardized radial log has expectation at most $L_B = 2\log(1 + B_4)$ by the first-moment bound and Jensen. The density bound $f \le C_0^k$, the gamma cancellation (244), and rounding domination (235) therefore give, simultaneously for all these Q, on G,

$$\mathrm{KL}(Q\|\widehat{\nu}) \le \Psi(e, S, \widetilde{e}),$$
$$\Psi = 4 + k(\log C_0 + 6) + \tfrac{1}{2}[\log\det S]_+$$
$$+ \tfrac{k+1}{2}\{\log(5 + 4/\lambda) + \log(1 + \|\widetilde{e}\|^2/k) + L_B\}. \tag{339}$$

Indeed $\log(5+4/\lambda) \le \log 9 + \log^+(1/\lambda)$, so the preceding bounds give

$$\mathbb{E}[\mathbf{1}_G \Psi] \le C_{\text{cal}} k,$$
$$C_{\text{cal}} = 10 + \log C_0 + C_D/2 + \log 9 + L_0 + L_e + L_B.$$

This specifies the fixed strategy scale used above. The change-of-reference integrations are legitimate even if f vanishes. Write Q' and q' for the standardized input law and fitted density. Let $g(z)$ be the expression for Ψ with its initial 4 removed and L_B replaced by $\log(1+\|z\|^2/k)$. The pointwise bounds give $\log(f/q') \le g$ on $\{f>0\}$, and $Q'g \le \Psi - 4 < \infty$. Finite $\mathrm{KL}(Q'\|f)$ makes $\log(dQ'/df)$ integrable, since the negative part of a log probability ratio has integral at most $1/e$. The new log ratio is bounded above by that old ratio plus g. Its positive part is therefore integrable; its negative part again has integral at most $1/e$. The integrated chain identity gives the claimed bound. The $2k$ rounding cost is included in g. Affine invariance returns the result to observed coordinates.

Condition on the complete actual prefix state and its auxiliary variables. All auxiliaries were generated sequentially from past inputs and fresh independent randomness and never affect the protocol. Future actual input kernels are therefore unchanged, and (337) continues to hold. On G, the conditional weighted clean suffix cost is at most $h\Psi$, and the bad mass at most $h\beta$. The scaled robust strategy consequently has conditional regret at most

$$C\sqrt{K}A_{d,h} + C\frac{A_{d,h}}{\sqrt{K}}\Psi + C\beta h.$$

On G^c its regret is at most h. Average, use $K = C_{\text{cal}}k$ and $\Pr(G^c) \le 4\beta$, and add the prefix cost to obtain

$$\mathbb{E}\,\mathrm{Reg}_T \le C\sqrt{k}\,A_{d,T} + n + C\beta T \le C_d\sqrt{k}\,A_T + C\beta T. \tag{340}$$

For the prefix inequality when $T > n$, use $n^2 \le 136kn\log(en) \le 136kT\log(eT)$. For the initial case $T \le n+d$, if $T \le 2d$ then $T \le \sqrt{2d}A_T$; otherwise $T \le 2n$. If also $T \le k$, then $T \le \sqrt{k}A_T$; if $T > k$, use $T \le 272k\log(ek) \le 272k\log(eT)$. Thus the fixed learner in every short-horizon case has the same $C_d\sqrt{k}A_T$ bound.

One learner for every radius. For $T > n+d$, combine this expert with the reference-free Cover expert of Theorem S9.1. At each round let their proper distributions be $q_{t,1}, q_{t,2}$ and let $\ell_{t,i} = q_{t,i}\ell(X_t, Y_t) \in [0,1]$. Run exponential weights on these two losses, starting uniformly with rate $\eta = \min\{1/2, \sqrt{\log 2/T}\}$, and draw the actual hypothesis from $\sum_i w_{t,i}q_{t,i}$ before the input. Both experts are updated on the actual input and label; their states are included in the full history. Lemma S3.2, applied to rewards $-\ell_{t,i}$, gives on every path

$$\sum_{t=1}^{T}\sum_{i=1}^{2} w_{t,i}\ell_{t,i} \le \min_{i=1,2}\sum_t \ell_{t,i} + \log 2/\eta + \eta T.$$

Current-round privacy makes the actual expected loss equal to the mixture's expected loss. The regret guarantees for each expert permit all the other predictable state and the master's announced distribution; only the current private draw is hidden. Thus the master has expected regret at most $C\sqrt{T}$ plus the smaller of the two expert bounds on its actual history.

Write $L = \log(eT)$ and $D = \log(1/\delta)$ when $\delta > 0$. For $D \ge D_0$, combine (340) with Cover's C_dT/L bound. The result is at most $C_d\{\sqrt{k}A_T + T/(L+D)\}$, since $\min\{T/L, T/D\} \le 2T/(L+D)$. For $D < D_0$, Cover suffices because $(L+D)/L \le 1 + D_0$, with universal D_0. At $\delta = 0$, use (340) with $\beta = 0$. Neither the master nor either expert tests or receives δ. The short-horizon fixed strategy satisfies the same bound, so this proves the upper for every T. Its pathwise regret is also at most T.

Matching lower bound. For $\delta > 0$, take $m = 0$ in the agnostic lower of Theorem S11.12. Its supplied uniform reference on $[0,1]$ is log-concave, and its fixed true reference lies within TV δ of it. That lower is $cR_T(\delta)$ for $T \ge 512$, even for unrestricted post-input prediction. For $k > 1$, take products of both references with uniform measure on $[0,1]^{k-1}$, and use thresholds in the first coordinate only. TV and conditional KL are unchanged. The nuisance coordinates are independent fresh draws; a one-dimensional learner can generate them privately and simulate any proposed k-dimensional learner. Thus they cannot invalidate the lower. The product approximating reference

is full-dimensional and log-concave, and the countable threshold class has VC dimension one, at most d. At $\delta = 0$, Proposition S15.1 gives cA_T under a known standard Gaussian; independent Gaussian nuisance coordinates give the same embedding. Revealing ρ and δ only strengthens these lower-bound learners. The reference may depend on the horizon as in the imported lower, but is fixed before play. This completes the proof. □