Minimax Tracking of Drifting Behavioural Strategies under Identifiable Partial Observation

§6.6 of a Larger Work

Michael Mannen

Abstract

We study high-probability tracking of a behavioural strategy that drifts over time in an imperfect-information extensive-form game, when the strategy is observed only through a known, possibly noisy emission channel at each recognized visit to an information set. We give a windowed estimator, projected onto the probability simplex, whose error decomposes into a drift-induced bias unaffected by the channel and a channel-conditioned sampling deviation controlled by a two-lemma supermartingale argument requiring no mixing assumption and no invalid conditioning on visit times; optimizing the window length yields a rate of $(\kappa_I^\circ)^{2/3} L^{1/3} (m \log(mT/\eta)/(\delta T))^{1/3}$ with probability $1-\eta$, uniformly over rounds, where κ_I° is the emission channel's tangent-space condition number and m its signal-alphabet size. We complement this with pointwise minimax lower bounds via the two-point (Le Cam) method, for both directly observed and noisy binary channels, matching the upper bound's polynomial dependence on T, L, δ, and, for binary channels, κ_I°; we do not claim these lower bounds resolve the necessity of the simultaneous-confidence logarithm, the alphabet-size factor $m^{1/3}$, or the general-channel geometry, each of which we flag explicitly as open. We also correct several errors present in earlier drafts of this material, detailed in the closing disclosure.

1 Classical Nonparametric Tracking Rate

We first record the fully observed benchmark, both because it motivates the window-length calculation used throughout and because we use it, later, as the standard against which the partially observed rate should honestly be compared.

Let $\theta : [0,1] \to \mathbb{R}$ be L-Lipschitz, and suppose $y_t = \theta(t/T) + \varepsilon_t$ for $t = 1, \dots, T$, with $\{\varepsilon_t\}$ independent and sub-Gaussian with parameter 1. For a *one-sided* sliding-window estimator using only past data,

$$\hat{\theta}_t = \frac{1}{h} \sum_{s=t-h+1}^{t} y_s, \qquad t \in [h, T],$$

the bias satisfies $|\mathbb{E}\hat{\theta}_t - \theta(t/T)| \le Lh/T$ (the window spans an interval of width at most h/T in the rescaled argument t/T), and, for any $\eta \in (0,1)$, setting $u = C\sqrt{\log(T/\eta)/h}$ gives $\mathbb{P}(|\hat{\theta}_t - \mathbb{E}\hat{\theta}_t| \ge u) \le 2\eta/T$ for a suitable absolute constant C; a union bound over the T values of t in range then gives, with probability at least $1-2\eta$,

$$\sup_{t \in [h,T]} |\hat{\theta}_t - \theta(t/T)| \;\le\; \frac{Lh}{T} + C\sqrt{\frac{\log(T/\eta)}{h}}.$$

Balancing the two terms gives, for $L > 0$,

$$h^\star = \Big[C\, T^{2/3} L^{-2/3} (\log(T/\eta))^{1/3} \Big]_{[1,T]}$$

(truncated to a legal window length between 1 and T; for $L = 0$ the drift bias vanishes identically and the correct choice is simply $h^\star = T$, giving the parametric rate $\sqrt{\log(T/\eta)/T}$ rather than an undefined expression), and, in the interior regime where the truncation is inactive,

$$\sup_{t\in[h^\star,T]} |\hat{\theta}_t - \theta(t/T)| = \Theta\big(L^{1/3}T^{-1/3}(\log(T/\eta))^{1/3}\big)$$

with probability at least $1 - 2\eta$. This is the same polynomial exponent identified, via the same balance, in the general nonparametric theory of estimation over Lipschitz and Hölder balls [4, 5]; those references establish the general minimax theory for that broader estimation problem, not this particular one-sided, sequential, simultaneous-coverage statement, which is the direct calculation above.

2 An Observation Model for Partial Observability

A tracking theorem for imperfect-information games should model an observation channel explicitly, rather than treat partial observability as mere intermittent visitation of an information set with the visited action revealed in full. We make this precise.

Let Γ be a finite extensive-form game of imperfect information with perfect recall, $\mathcal{I}_i$ the information sets of player i, and $A(I)$ the action set at $I \in \mathcal{I}_i$. Fix a single information set I under study, write $k := |A(I)|$ for its action-set size, and let $\mathcal{O}(I)$ denote the (possibly distinct) signal alphabet observed by the modeler, with $m := |\mathcal{O}(I)|$. At each round $s = 1, \dots, T$:

- $V_s(I) \in \{0, 1\}$ indicates whether I was reached *and recognized* at round s;
- if $V_s(I) = 1$, the party being modeled selects an action $A_s \in A(I)$ according to the (unobserved) behavioural strategy $\sigma_s^\star(I, \cdot) \in \Delta_k$;
- what the modeler actually observes is not A_s but a signal $O_s \in \mathcal{O}(I)$, drawn from a known emission distribution $O_s \sim Q_I(\cdot \mid A_s)$.

Write $B_I \in \mathbb{R}^{m\times k}$ for the (assumed known) emission matrix with entries $B_I(o, a) := Q_I(o \mid a) = \Pr(O = o \mid A = a)$; since $\sum_o B_I(o, a) = 1$ for each fixed action a, B_I is *column*-stochastic (a Markov kernel from actions to signals), and consequently $\|B_I p - B_I q\|_1 \le \|p - q\|_1$ for any $p, q \in \Delta_k$. Write $\kappa_I := \|B_I^+\|_2$ for the operator norm of its Moore–Penrose pseudoinverse, equivalently the reciprocal of B_I's smallest singular value. Two regimes are worth distinguishing explicitly, since they have been run together in informal treatments of this problem:

Directly logged actions. If B_I is the identity ($k = m$ and $O_s = A_s$ whenever $V_s(I) = 1$), the only source of partial observability is missing data from unrecognized visits, and the analysis reduces to a missing-data tracking problem, with $\kappa_I = 1$.

Genuinely hidden actions. If $B_I \neq \mathrm{Id}$, the observable distribution over signals at round s is $q_s^\star := B_I\, \sigma_s^\star(I, \cdot) \in \Delta_m$, a linear image of the true strategy. If B_I has full column rank k, then $B_I^+ B_I = I_k$, so $\sigma_s^\star(I, \cdot) = B_I^+ q_s^\star$ exactly, and an estimate of $q_s^\star$ can be deconvolved into an estimate of $\sigma_s^\star(I, \cdot)$, at a cost controlled not by the full operator norm $\kappa_I := \|B_I^+\|_2$ but by its restriction to the mean-zero tangent space that differences of probability vectors actually live in,

$$\kappa_I^\circ := \sup_{z\in\mathbf{1}_m^\perp,\, z\neq 0} \frac{\|B_I^+ z\|_2}{\|z\|_2} \;\le\; \kappa_I,$$

made precise in Section 4; for the binary guess/slip channel studied in Proposition 2 below, κ_I° takes the simple closed form given there. If B_I is rank-deficient, $\sigma_s^\star(I, \cdot)$ is not identifiable from O_s alone, and no estimator — however the visitation and drift structure is exploited — can recover it uniformly; this is a statement about the observation channel, not about any particular

estimation procedure. The one-step algebraic form of this channel matches the guess-and-slip emission layer used in Bayesian Knowledge Tracing: a correct response is observed with some probability even absent mastery (guessing) and an incorrect response is observed with some probability despite mastery (slipping), and B_I is full rank precisely when guess and slip rates are not so extreme as to make the two hidden states observationally indistinguishable. We emphasize that only this emission layer is captured here; a full knowledge-tracing model additionally specifies transition dynamics for a *persistent* latent mastery state, whereas A_s above is drawn afresh each visit from a drifting but otherwise memoryless distribution, and we do not claim to have encoded the former.

The remainder of this section develops the tracking theorem under the assumption that B_I is known and full column rank; we do not attempt a full treatment of the rank-deficient case beyond noting its impossibility.

3 Assumptions

Assumption 1 (Conditional visitation)**.** *There exists $\delta > 0$ such that $\Pr(V_s(I) = 1 \mid \mathcal{F}_{s-1}) \geq \delta$ almost surely for every round s, where $\mathcal{F}_{s-1}$ is the filtration generated by the history through round $s-1$.*

Assumption 2 (Conditional action law)**.** *Given $\mathcal{F}_{s-1}$ and $V_s(I) = 1$, the action satisfies $\Pr(A_s = a \mid \mathcal{F}_{s-1}, V_s(I){=}1) = \sigma^\star_s(I, a)$, and the emission O_s is drawn from $Q_I(\cdot \mid A_s)$ conditionally independently of $\mathcal{F}_{s-1}$ given A_s.*

Assumption 3 (Lipschitz drift)**.** *$\|\sigma^\star_s(I,\cdot) - \sigma^\star_t(I,\cdot)\|_1 \leq L|s-t|/T$ for all rounds s, t.*

Assumption 4 (Identifiable channel)**.** *B_I is known and has full column rank k.*

Assumption 1 is deliberately an operational, filtration-conditional statement rather than a marginal one: a marginal bound $\Pr_{\sigma^\star(s)}(I) \geq \delta$ does not, on its own, imply any renewal structure, conditional lower bound, or concentration of visits within a sliding window. Assumptions 1–2 together provide exactly the conditional-mean structure a martingale argument needs; they do *not* imply that emissions are independent after conditioning on the complete visitation history, since adaptive future visitation may itself encode information about earlier emissions. The concentration argument of Section 4 accordingly works in the natural round-indexed filtration throughout, and does not condition on the realized visit times.

4 Main Theorem

For a window of w rounds, define $N_t := \sum_{s=t-w+1}^{t} V_s(I)$ and the empirical emission frequency among recognized visits,

$$\hat{q}_t(I,o) = \begin{cases} \dfrac{\sum_{s=t-w+1}^{t} V_s(I)\,\mathbf{1}\{O_s = o\}}{N_t}, & N_t > 0, \\ \text{uniform on } \mathcal{O}(I), & N_t = 0, \end{cases}$$

the second case simply fixing an arbitrary value so that $\hat{q}_t$ is a total function on the whole sample space; it is irrelevant on the high-probability event $N_t > 0$ that Lemma 1 below establishes. Define the estimated strategy $\hat{\sigma}_t(I,\cdot) = \Pi_{\Delta_k}\big(B_I^+\,\hat{q}_t(I,\cdot)\big)$, where Π_{Δ_k} denotes Euclidean projection onto the probability simplex Δ_k; since $B_I^+\,\hat{q}_t$ need not itself be a valid probability vector, this projection step ensures $\hat{\sigma}_t(I,\cdot)$ actually is one. Because Euclidean projection onto a closed convex set is non-expansive and $\sigma^\star_t(I,\cdot) \in \Delta_k$, this projection can only decrease the estimation error, so it suffices to bound $\|B_I^+\,\hat{q}_t - \sigma^\star_t\|_2$ throughout.

Theorem 1. *Under Assumptions 1–4, fix $\eta \in (0,1)$ and suppose the sample-size condition*

$$\delta T \geq C_0 \log(mT/\eta)$$

holds for a suitable absolute constant C_0; this is exactly what makes $w_{\min} := C_0\,\delta^{-1}\log(mT/\eta) \le T$, so that the admissible window range $[w_{\min}, T]$ used below is nonempty. (When $\delta T < C_0\log(mT/\eta)$, no window satisfies the theorem's hypotheses, and only the trivial simplex-diameter bound applies; setting $w = T$ does not repair this, since $T < w_{\min}$ in that regime.) Then, for every integer $w \in [w_{\min}, T]$ (ensuring every window contains at least one recognized visit with the required probability), with probability at least $1-\eta$, uniformly over admissible $t \in [w, T]$,

$$\|\hat{\sigma}_t(I) - \sigma_t^\star(I)\|_2 \;\le\; \frac{Lw}{T} + C\,\kappa_I^\circ\sqrt{\frac{m\log(mT/\eta)}{\delta w}}$$

for an absolute constant C, where $\kappa_I^\circ \le \kappa_I := \|B_I^+\|_2$ is the tangent-space-restricted condition number of Section 2 and $m = |\mathcal{O}(I)|$. Optimizing over admissible w gives

$$w^\star = \left[\left(\frac{(\kappa_I^\circ)^2\, m\, T^2\log(mT/\eta)}{\delta L^2}\right)^{1/3}\right]_{[w_{\min},\,T]}$$

(truncated to the admissible range), and, in the interior regime where this truncation is inactive,

$$\|\hat{\sigma}_t(I) - \sigma_t^\star(I)\|_2 \;\lesssim\; (\kappa_I^\circ)^{2/3}\,L^{1/3}\left(\frac{m\log(mT/\eta)}{\delta T}\right)^{1/3}$$

with probability at least $1-\eta$, uniformly over $t \in [w^\star, T]$.

Remark 1 (Generality of the argument)**.** *Nothing in the statement or proof of Theorem 1 uses extensive-form game structure beyond fixing a single information set I and reading $\sigma_t^\star(I,\cdot)$ as a probability vector. The theorem is, in substance, a general statement about tracking a drifting categorical distribution $p_t \in \Delta_k$, observed through a known linear channel B under adaptive (conditionally lower-bounded) missingness, via self-normalized martingale concentration; the game-theoretic reading (information sets, behavioural strategies) and the educational reading (knowledge components, guess/slip channels) of Section 7 are both instances of this same abstract result rather than requiring separate arguments.*

The factor $\sqrt{m}$ above comes from a coordinate-wise concentration bound converted to ℓ_2 by a union bound over the m signals in $\mathcal{O}(I)$; it is a correct consequence of the proof technique below, but we flag it as possibly not intrinsic to the ℓ_2 loss. For a multinomial empirical distribution, $\mathbb{E}\|\hat{q} - q\|_2^2 = (1 - \|q\|_2^2)/N \le 1/N$ regardless of alphabet size, which suggests a genuinely Hilbert-space-valued martingale concentration argument (rather than the per-coordinate argument used here) might remove the $\sqrt{m}$ factor entirely from the high-probability ℓ_2 bound. We have not derived such an argument and state the theorem with its coordinate-wise proof's actual dependence rather than claim a stronger bound we have not verified.

The truncation to $[w_{\min}, T]$ is not a technicality to be waved past: without it, the displayed optimizer can fail to be an admissible window at all — for instance when L is very small, the horizon T is short, or I is rarely visited — in which case the correct prescription is $w^\star = w_{\min}$ or $w^\star = T$ rather than the interior formula.

4.1 Proof

Bias. We first show the drift bias is controlled *without* any channel-conditioning factor, correcting a cruder bound that would multiply the entire error by κ_I or κ_I°. Let

$$\bar{q}_t := N_t^{-1}\sum_{s=t-w+1}^{t} V_s(I)\,q_s^\star$$

be the visit-weighted average of the true emission distributions inside the window (on the event $N_t > 0$). Since B_I is column-stochastic, $B_I^+ B_I = I_k$ under Assumption 4, so

$$B_I^+ \bar{q}_t = N_t^{-1} \sum_s V_s(I)\, B_I^+ q_s^\star = N_t^{-1} \sum_s V_s(I)\, \sigma_s^\star(I, \cdot),$$

a visit-weighted average of the true *strategies* in the window, with no channel factor anywhere in this identity. By Assumption 3, this average is within Lw/T of $\sigma_t^\star(I, \cdot)$ in ℓ_1 (hence in ℓ_2), exactly as in the classical bias calculation of Section 1. The channel-conditioning factor therefore enters only through the *estimation* error $\hat{q}_t - \bar{q}_t$; since this difference has coordinates summing to zero (both $\hat{q}_t$ and $\bar{q}_t$ are probability vectors), it lies in the tangent space $\mathbf{1}_m^\perp$ on which κ_I°, rather than the full operator norm κ_I, controls B_I^+:

$$\|B_I^+ \hat{q}_t - \sigma_t^\star\|_2 \le \|B_I^+ (\hat{q}_t - \bar{q}_t)\|_2 + \|B_I^+ \bar{q}_t - \sigma_t^\star\|_2 \le \kappa_I^\circ \|\hat{q}_t - \bar{q}_t\|_2 + \frac{Lw}{T}.$$

Lemma 1 (visit count). Under Assumption 1, write $p_s := \Pr(V_s(I){=}1 \mid \mathcal{F}_{s-1}) \ge \delta$. For any $\lambda > 0$, $\mathbb{E}[e^{-\lambda V_s(I)} \mid \mathcal{F}_{s-1}] = 1 - p_s + p_s e^{-\lambda} \le \exp(\delta(e^{-\lambda} - 1))$, using that the middle expression is decreasing in p_s and $1 - x \le e^{-x}$. Iterating over the w rounds in the window via the tower property gives $\mathbb{E}[e^{-\lambda N_t}] \le \exp(w\delta(e^{-\lambda} - 1))$, and optimizing λ in the resulting Chernoff bound yields the standard multiplicative lower-tail estimate

$$\Pr\big(N_t \le (1 - \varepsilon)\delta w\big) \le \exp\Big(-\frac{\varepsilon^2}{2}\delta w\Big);$$

taking $\varepsilon = 1/2$ gives $\Pr(N_t < \delta w/2) \le \exp(-\delta w/8)$, an elementary scalar Chernoff bound needing no external martingale-inequality citation. A union bound over the T window positions used by the estimator requires this probability to be at most η/T, which holds once $w \ge w_{\min} = C_0 \delta^{-1} \log(T/\eta)$ for suitable C_0; we fold the (larger) alphabet-dependent threshold from Lemma 2 into the single admissibility condition $w \ge w_{\min}$ stated in the theorem.

Lemma 2 (emission concentration). The argument works in the round-indexed filtration throughout and does not condition on the realized visit times. For a fixed signal $o \in \mathcal{O}(I)$, define $Y_{s,o} := V_s(I)\big(\mathbf{1}\{O_s = o\} - q_s^\star(o)\big)$. By Assumption 2, $\mathbb{E}\big[V_s(I)\, \mathbf{1}\{O_s = o\} \mid \mathcal{F}_{s-1}, V_s(I)\big] = V_s(I)\, q_s^\star(o)$ (keeping $V_s(I)$ inside the expectation on both sides avoids any need to define a special value for $\mathbf{1}\{O_s = o\}$ itself when $V_s(I) = 0$, since O_s is simply not defined on that event), so $\mathbb{E}[Y_{s,o} \mid \mathcal{F}_{s-1}] = 0$: $\{Y_{s,o}\}$ is a genuine martingale difference sequence with respect to $\{\mathcal{F}_s\}$, bounded in $[-1, 1]$, with no appeal to conditional i.i.d. structure or conditioning on N_t required. A conditional Hoeffding argument gives $\mathbb{E}[\exp(\lambda Y_{s,o} - \frac{\lambda^2}{8} V_s(I)) \mid \mathcal{F}_{s-1}, V_s(I)] \le 1$, so $\exp(\lambda M_{j,o} - \frac{\lambda^2}{8} \sum_{s \le j} V_s(I))$ is a nonnegative supermartingale in j, where $M_{j,o} := \sum_{s=t-w+1}^{j} Y_{s,o}$; applying Ville's maximal inequality for nonnegative supermartingales [1] with $\lambda = 4u$ gives, for any n_0,

$$\Pr\big(M_{t,o} \ge u N_t,\ N_t \ge n_0\big) \le \exp(-2u^2 n_0),$$

and the same bound applies to $-M_{t,o}$. Combined with Lemma 1's event $N_t \ge \delta w/2$ (taking $n_0 = \delta w/2$) and a union bound over the m signals and T window positions, this gives, with probability at least $1 - \eta$,

$$\|\hat{q}_t(I, \cdot) - \bar{q}_t\|_2 \ \le\ C_1 \sqrt{\frac{m \log(mT/\eta)}{\delta w}}$$

uniformly over admissible t, for an absolute constant C_1. We note, without proof, that this coordinate-wise argument may not be the tightest route to an ℓ_2 bound: since $\mathbb{E}\|\hat{q} - q\|_2^2 = (1 - \|q\|_2^2)/N \le 1/N$ for a multinomial empirical distribution regardless of alphabet size, a

genuinely vector-valued version of the argument above — applying a dimension-free vector Bernstein or Pinelis-type inequality directly to the Δ_m-valued martingale $\sum_s V_s(e_{O_s} - q^\star_s)$, whose increments have norm at most $\sqrt{2}$, rather than a per-coordinate bound combined with a union bound over m symbols — might remove the $\sqrt{m}$ factor entirely from the high-probability ℓ_2 bound, yielding $\|\hat{q}_t - \bar{q}_t\|_2 \lesssim \sqrt{\log(T/\eta)/(\delta w)}$ and correspondingly $(\kappa_I^\circ)^{2/3} L^{1/3} (\log(T/\eta)/(\delta T))^{1/3}$ in place of Theorem 1's rate. We have not carried out that stronger argument, state only what the coordinate-wise proof above actually establishes, and leave whether m-dependence is minimax-optimal for this ℓ_2 loss as an explicitly open question.

Combining. Substituting Lemma 2's bound into the bias/deviation decomposition above gives the stated theorem, with $C = C_1$.

Optimization. Setting Lw/T equal to $\kappa_I^\circ \sqrt{m \log(mT/\eta)/(\delta w)}$ and solving for w gives $w^\star = ((\kappa_I^\circ)^2 m T^2 \log(mT/\eta)/(\delta L^2))^{1/3}$ before truncation; substituting back, both terms are of the same order, giving the stated rate. Truncating to $[w_{\min}, T]$ handles the boundary cases noted above. ■

5 Minimax Lower Bound

Theorem 1 is an upper bound. We now show the polynomial exponent 1/3, the dependence on δ, and (in a refined version) the dependence on the channel conditioning are not artifacts of this particular estimator.

Proposition 1 (Lower bound, directly observed case)**.** *Consider a single information set I with $|A(I)| = 2$, visited according to i.i.d.* Bernoulli(δ) *indicators V_s independent of the action draws and identical under both candidate paths below, with the action directly observed whenever $V_s = 1$. There is an absolute constant $c > 0$ such that, for any estimator $\hat{\sigma}_T$ of $\sigma_T^\star(I)$ at the final round,*

$$\inf_{\hat{\sigma}_T} \sup_{\sigma^\star \in \mathrm{Lip}(L)} \mathbb{E}\,|\hat{\sigma}_T(I) - \sigma_T^\star(I)| \;\geq\; c \min\Big\{1,\; \big(L/(\delta T)\big)^{1/3}\Big\}.$$

Proof sketch. Compare the constant path $p_0(s) \equiv 1/2$ (probability of action 1) against a path p_1 that agrees with p_0 except for a one-sided linear ramp of height $\Delta := \min\{1/4,\, c'(L/(\delta T))^{1/3}\}$ built up over the final $w \asymp \Delta T/L$ rounds, the largest ramp consistent with Assumption 3's slope bound L/T per round; the truncation at $\Delta \leq 1/4$ is exactly what makes the proposition's $\min\{1, \cdot\}$ form necessary, since for small $\delta T/L$ the unconstrained optimum would exceed the largest ramp height the Bernoulli parameter space and the quadratic KL approximation below can support. Restricting $\Delta \leq 1/4$ keeps both paths' Bernoulli parameters in a compact subset of $(0,1)$ on which the quadratic approximation to KL divergence below is uniform. Since visitation is independent Bernoulli(δ) and each recognized round contributes an independent Bernoulli action draw, the total Kullback–Leibler divergence between the two induced observation processes over the ramp is

$$\begin{aligned}\mathrm{KL}(P_1 \| P_0) &= \delta \sum_s \mathrm{KL}\big(\mathrm{Bernoulli}(p_{1,s}) \,\|\, \mathrm{Bernoulli}(p_{0,s})\big) \\ &= \Theta(\delta w \Delta^2) = \Theta\Big(\delta \cdot \frac{\Delta T}{L} \cdot \Delta^2\Big) = \Theta\Big(\frac{\delta T \Delta^3}{L}\Big).\end{aligned}$$

For Δ as defined above with c' chosen small enough, this divergence is $O(1)$, so the two processes are statistically indistinguishable by Le Cam's two-point method, while $p_0(T)$ and $p_1(T)$ differ by $\Theta(\Delta)$. Any estimator therefore incurs expected error $\Omega(\Delta)$ against at least one of the two paths, giving the stated bound. □

This matches Theorem 1's polynomial exponent and δ-dependence, in the interior regime $\delta T \gtrsim L$, but — using directly observed actions — says nothing about the channel-conditioning factor. The following sharper construction closes that gap for a binary emission channel.

Proposition 2 (Lower bound, noisy binary channel)**.** *In the setting of Proposition 1, suppose instead that the action is observed through a binary channel with guess parameter* $g := \Pr(\textit{observed correct} \mid \textit{action } 0) \in [0,1)$ *and slip parameter* $s := \Pr(\textit{observed incorrect} \mid \textit{action } 1) \in [0,1)$*. Writing* p *for the true (hidden) action-*1 *probability, the observed correct-response probability is* $q = g + \rho p$ *where* $\rho := 1 - g - s \neq 0$*, so that* $\kappa := |\rho|^{-1}$ *plays the role of the channel condition number (this is* κ_I° *of Section 2 for the binary channel, exactly, rather than merely an upper bound on it). Assume further that the baseline observed probability* $q_0 := g + \rho/2$ *stays in* $[\varepsilon, 1-\varepsilon]$ *for some fixed* $\varepsilon > 0$*, so that the local quadratic approximation to Bernoulli KL divergence is uniform rather than degenerating near the boundary. Then there is a constant* $c_\varepsilon > 0$ *such that*

$$\inf_{\hat{\sigma}_T} \sup_{\sigma^\star} \mathbb{E}\,|\hat{\sigma}_T(I) - \sigma_T^\star(I)| \;\geq\; c_\varepsilon \min\left\{1,\; \kappa^{2/3}\big(L/(\delta T)\big)^{1/3}\right\}.$$

Proof sketch. Two hidden paths separated by Δ at the final round induce *observed* paths separated by $|\rho|\Delta$. Repeating the construction of Proposition 1 with this attenuated separation, and with $q_0 \in [\varepsilon, 1-\varepsilon]$ ensuring the constant in the quadratic KL approximation is uniform, gives $\mathrm{KL}(P_1 \| P_0) = \Theta(\delta w(\rho\Delta)^2) = \Theta(\delta T \rho^2 \Delta^3 / L)$ over a ramp of the same width $w \asymp \Delta T/L$. Setting this to $O(1)$, and truncating exactly as in Proposition 1, gives $\Delta \asymp \min\{1/4, (L/(\delta\rho^2 T))^{1/3}\} = \min\{1/4, \kappa^{2/3}(L/(\delta T))^{1/3}\}$, and Le Cam's method again lower-bounds the estimation error by this Δ. □

Since $\kappa = \kappa_I^\circ$ exactly for this binary channel, this matches Theorem 1's $(\kappa_I^\circ)^{2/3}$ dependence exactly, in the interior regime, using κ_I° rather than the full operator norm κ_I (a distinction that matters: an earlier draft's claim of an exact match used κ_I, which need not equal κ_I° for a general channel, even though the two coincide here). For a general emission matrix B_I, the corresponding lower bound is more naturally phrased in terms of local Fisher information along a least-informative tangent direction, since even κ_I° need not control the relevant KL divergence unless emission probabilities are also bounded away from 0 and 1, as the $q_0 \in [\varepsilon, 1-\varepsilon]$ condition above makes explicit for the binary case; we leave that generalization open, along with the alphabet-size factor $m^{1/3}$, which these binary ($m = 2$) propositions do not address. In both propositions, the logarithmic factor present in Theorem 1 does not appear, because the lower bounds concern pointwise risk at a single fixed round, whereas the theorem controls risk simultaneously over all T rounds; the logarithm arises specifically in our simultaneous upper bound (via the union bound over window positions), and the pointwise lower bounds above do not determine whether it is actually necessary.

The propositions are stated for a scalar Bernoulli parameter p, while Theorem 1 bounds a vector ℓ_2 error; the two are directly comparable via the embedding $\sigma(p) = (1-p, p) \in \Delta_2$, under which $\|\sigma(\hat{p}) - \sigma(p)\|_2 = \sqrt{2}\,|\hat{p} - p|$, so the scalar lower bounds above translate into ℓ_2 lower bounds of the same order.

6 Extensions and Remarks

Hölder generalization. For a drift assumption $\|\sigma_s^\star - \sigma_t^\star\|_1 \leq L(|s-t|/T)^\alpha$ with $\alpha \in (0,1]$, the same balance — bias $L(w/T)^\alpha$ against deviation $\kappa_I^\circ\sqrt{m\log(mT/\eta)/(\delta w)}$ — gives

$$w^\star \asymp T^{\frac{2\alpha}{2\alpha+1}} L^{-\frac{2}{2\alpha+1}} \left(\frac{(\kappa_I^\circ)^2\, m \log(mT/\eta)}{\delta}\right)^{\frac{1}{2\alpha+1}}$$

and rate

$$\mathrm{Err} \asymp L^{\frac{1}{2\alpha+1}} (\kappa_I^\circ)^{\frac{2\alpha}{2\alpha+1}} \left(\frac{m \log(mT/\eta)}{\delta T} \right)^{\frac{\alpha}{2\alpha+1}},$$

recovering the Lipschitz case of Theorem 1 at $\alpha = 1$. We state the full dependence on L, δ, m, and κ_I° explicitly here, rather than assert, as an earlier draft did, that the Lipschitz case extends "verbatim" with these factors left out.

Dependence beyond conditional independence. Assumption 2 treats the action at a recognized visit as conditionally independent of the past given the visitation event. Relaxing this to allow genuine temporal dependence in the action process — for instance, a mixing condition on the round-indexed chain — is a natural further direction, but it requires either a uniform contraction condition shared across the entire time-varying family of kernels or an explicit mixing assumption stated directly on the non-stationary process actually in force, together with a Bernstein-type inequality suited to that setting, such as that of [3]. We have not derived such an extension here and do not claim the rate above continues to hold under mixing alone without further argument.

Unresolved side questions. A hierarchical Dirichlet-process prior on the drift, and a parallel change-point monitor for abrupt (non-Lipschitz) jumps charged against a separate variation budget, are plausible directions for relaxing Assumption 3, but neither is derived here, and we no longer assert specific log-factor costs for them.

Related work on drifting distributions. Independent of the game-theoretic motivation, [2] establish matching minimax rates for online estimation of a drifting discrete distribution under a bounded per-step drift assumption, a setting closely related to the emission-distribution tracking problem of Section 2 once the game-theoretic structure (visitation, information sets, deconvolution) is stripped away; the $1/3$ exponent recurring there is the same one this section derives from a Lipschitz-in-continuous-time drift assumption rather than a bounded-per-step one.

A more intrinsic channel complexity. The worst-case quantity κ_I° is clean and interpretable, but it ignores the actual covariance structure of the deconvolved estimator, whose local covariance is approximately $N^{-1} B_I^+ (\mathrm{diag}(q) - qq^\top)(B_I^+)^\top$ rather than a quantity controlled uniformly by κ_I° alone. This suggests a statistical channel radius $\chi_I := \sup_{q \in B_I \Delta_k} \| B_I^+ (\mathrm{diag}(q) - qq^\top)^{1/2} \|_{2\to 2}$, which a refined theorem might use in place of the cruder combination $\kappa_I^\circ \sqrt{m}$, with a corresponding lower bound phrased via Fisher information along the least-informative tangent direction rather than the binary construction of Proposition 2. We do not develop χ_I further here; unifying the upper-bound variance, the binary attenuation ρ, and general-channel lower bounds through this single quantity is a natural further research direction rather than a claim made by the present theorem.

Fixed-horizon versus anytime. Theorem 1 is a fixed-horizon result: the window $w^\star$ is chosen once, using knowledge of T, and the guarantee holds uniformly over $t \in [w^\star, T]$ for that fixed choice. It is not an anytime guarantee usable without advance knowledge of the horizon. A genuine anytime corollary — recalibrating the window at the start of each of a sequence of dyadic epochs $[1, 2], [3, 4], [5, 8], \ldots$ and applying Theorem 1 within each — is the natural route to such a result, at the cost of constant factors from the epoch doubling, but we do not carry out that construction here.

7 Significance for Education

The setting of Sections 2–6 is directly motivated by adaptive tutoring, and by the companion analysis of consistent student modeling, which treats a student's knowledge state as static. A learning student's mastery instead drifts, gradually, over practice opportunities, and an instructor observes only a noisy signal of it: the same algebraic form as the guess/slip emission channel B_I of Section 2, with mastery playing the role of the hidden action A_s. The corrected comparison, stated plainly: relative to a fully observed, static-mastery benchmark — which a simple average would estimate at the parametric rate $T^{-1/2}$ — allowing the student to keep learning changes the achievable rate to $T^{-1/3}$, the cost of tracking a moving target rather than an artifact of partial observation. Partial observation through a known, identifiable guess/slip channel costs a further, explicit penalty of order $(\kappa_I^\circ)^{2/3}$ (not a full extra factor of κ_I°, since the channel conditioning enters only the estimation-noise term of Theorem 1, not the drift bias), worsening with how close the channel is to non-identifiable, but not changing the polynomial exponent, and Proposition 2 shows this $(\kappa_I^\circ)^{2/3}$ dependence is itself unavoidable for a binary channel, not an artifact of this estimator. The practical prescription is the window size $w^\star \asymp ((\kappa_I^\circ)^2 m T^2 \log(mT/\eta)/(\delta L^2))^{1/3}$: an adaptive tutoring system should size its recent-performance window using the visitation rate δ of the specific knowledge component, the conditioning κ_I° of its guess/slip channel, and the estimated learning rate L, rather than a fixed or linearly growing window. This is an *oracle* prescription, and should be understood as one: it presumes L, δ, and B_I are known exactly, whereas in practice L is generally unknown, δ may be unknown or itself policy-dependent, and B_I is typically estimated rather than given. A natural, undeveloped extension would replace the oracle window with a Lepski-type adaptive procedure — computing estimates at a geometric sequence of window sizes $w \in \{1, 2, 4, 8, \dots\}$ and selecting the largest whose estimates remain mutually compatible — to adapt to unknown L, paired with a conservative empirical lower confidence bound in place of a known δ; we do not carry out that construction here, and the window formula above is offered as a scaling principle rather than an implementable, assumption-free algorithm. We stress that w throughout is a window of *global* practice rounds across all skills, not a count of attempts at the specific skill in question; a window of w rounds contains, in expectation, only δw recognized visits to a skill with visitation rate δ, and it is this smaller effective count that governs estimation precision.

7.1 A Worked Illustration

The scaling is easiest to feel with numbers, so we work through

$$w^\star = \left(\frac{(\kappa_I^\circ)^2 \, m \, T^2 \log(mT/\eta)}{\delta L^2} \right)^{1/3}$$

for two contrasting knowledge components after $T = 1000$ global practice interactions, with a binary correct/incorrect signal ($m = 2$) and 95% simultaneous confidence ($\eta = 0.05$), taking $L = 1$ in both cases (a skill whose true mastery probability could plausibly swing across its full range over the whole course). We stress that the constants suppressed by $\asymp$ in Theorem 1 are not pinned down here; the following numbers illustrate how $w^\star$ *scales* with δ, κ_I°, and T, not a literal, calibration-free guarantee.

	Skill A (frequent, clean)	**Skill B** (rare, noisy)
Visitation rate δ	0.2 (practiced on 1 in 5 global interactions)	0.05 (practiced on 1 in 20 global interactions)
Channel conditioning κ_I°	2 (well-separated guess/slip rates)	4 (guess/slip rates closer together)
Formula's $w^\star$ (global rounds)	≈ 751	≈ 1893
Truncation ($w^\star \le T$?)	yes — interior regime	no — exceeds $T = 1000$
Expected recognized visits within window	$\delta w^\star \approx 150$	$\delta T = 50$ (using $w^\star = T$)
Practical instruction	use the ∼750 most recent *global* interactions, expected to contain ∼150 recognized observations of Skill A	use all 1000 global interactions collected so far, expected to contain only ∼50 recognized observations of Skill B; even that is not enough for a tight estimate

Skill B illustrates the truncation clause of Theorem 1 concretely, not just as a technicality: when a skill is rarely practiced and its assessment is noisy, the formula's unconstrained optimum can legitimately exceed the amount of data collected so far, and the correct response is $w^\star = T$ — use everything available — while accepting a looser confidence bound than Skill A's, rather than inventing additional data or shrinking the window below what the truncation analysis of Theorem 1 permits. Because the constant C_0 in Theorem 1's feasibility condition $\delta T \ge C_0 \log(mT/\eta)$ is not pinned down, we cannot certify from these illustrative numbers alone that Skill B's parameters actually lie inside the theorem's admissible regime at the stated 95% confidence; Skill B's $w^\star = T$ recommendation should accordingly be read as a scaling illustration of what the truncation clause implies, not a certified guarantee for these specific numbers. The distinction between the window w (global rounds) and the effective sample size δw (recognized visits to the specific skill) is central to what δ means here, and collapsing the two — treating $w^\star$ itself as a count of skill-specific attempts — would overstate how much information the window actually contains.

7.2 Design Implications

Three consequences follow directly from the theorem and are not simply restatements of common tutoring-system practice.

First, for a fixed planning horizon T, the optimal window scales *sublinearly* in T, as $T^{2/3}$ rather than linearly or at a fixed size: using all historical data with equal weight throughout the course treats a student's genuine recent progress as noise to be averaged away, while a small fixed window is dominated by guess/slip noise, and the theorem gives the correct exponent in between rather than leaving it to be tuned by trial and error, as recency-weighting parameters typically are in deployed spaced-repetition and mastery-learning systems. Theorem 1 itself is a fixed-horizon statement, choosing $w^\star$ once using T; an online system that does not know its effective horizon in advance would need the dyadic-epoch recalibration sketched in Section 7 to recover an analogous elapsed-time scaling, a corollary we have not proved here.

Second, different knowledge components warrant different window sizes, and the theorem says by how much: a skill assessed only occasionally, or through a question format prone to guessing (few multiple-choice options, true/false), needs a larger window — or, as Skill B shows, may need the system to candidly report lower confidence rather than a falsely tight estimate — while a frequently practiced, cleanly assessed skill can be tracked tightly with comparatively little recent data.

Third, the identifiability condition of Section 2 is a genuine constraint on assessment design, not a mathematical nicety: if guess and slip rates are close enough that B_I loses full column rank, mastery is not recoverable from the observed response pattern at *any* sample size, however large. Two-option true/false items are the setting most at risk of this, though the risk comes specifically from how distinguishable the resulting emission distributions are, not from the raw number of answer options; assessment formats that produce more *distinguishable* response distributions across the mastered/not-mastered states improve channel conditioning, and merely adding answer options does not guarantee this unless it actually achieves that separation.

7.3 A Cautionary Note on Claims of Instant Tracking

The lower bounds of Section 6 are the deliberately unglamorous half of this result, and are worth stating plainly for a non-technical audience: without additional structural assumptions linking the target skill to other skills, other students, shared item parameters, or a low-dimensional latent representation — assumptions that can genuinely help, and that this section has not assumed — no estimator can track a genuinely still-learning student, from noisy per-question signals on that skill alone, faster than the $T^{-1/3}$ rate, nor evade the $(\kappa_I^\circ)^{2/3}$ penalty of a noisy assessment channel. This is a reason for mild skepticism toward claims that an adaptive system tracks student mastery in near-real time with high confidence from a handful of multiple-choice responses *and no such cross-task structure*: some part of any such claim is either relying on assessment items with little guessing (small κ_I°), restricting to skills practiced often (large δ), leaning on structural assumptions across skills or students that go beyond what is analyzed here, reporting a rate that degrades exactly as this theorem predicts once examined closely, or simply not being able to distinguish, from its own data, a confidently wrong estimate from a genuinely accurate one.

AI-Assistance Disclosure

This section was developed with substantial assistance from Claude (Anthropic) across several rounds of detailed referee review, including corrections to a probability-theory error in the concentration proof (invalid conditioning on visit times, repaired with a supermartingale argument), an algebra error in the optimized rate, and a mislabeled guess/slip channel formula. The author reviewed the derivations, citations, and final presentation and takes full responsibility for the content. A detailed round-by-round revision record is maintained separately.

References

[1] Howard, S. R., Ramdas, A., McAuliffe, J., & Sekhon, J. (2020). Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17, 257–317.

[2] Mazzetto, A., & Upfal, E. (2023). Nonparametric density estimation under distribution drift. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, PMLR 202, 24251–24270.

[3] Merlevède, F., Peligrad, M., & Rio, E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High Dimensional Probability V: The Luminy Volume*, IMS Collections, Vol. 5, 273–292. DOI: 10.1214/09-IMSCOLL518.

[4] Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, 8(6), 1348–1360.

[5] Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer.