Safe Expanding Forecast Horizons

An exact stationary lag criterion with expansion bounds

Albaraa Mesawi

Preprint
22 October 2026

Abstract

For a known countable family of stationary binary laws, I characterize when one fixed predictor's conditional forecasts admit uniformly safe expanding horizons. Each dependent candidate copies one remote bit with a prescribed bias. For every coherent predictor that converges at each fixed horizon under each candidate, safe expansion is possible exactly when the biases tend to zero. The positive rule controls the probability of any inaccurate approval and expands almost surely under every candidate. With the specified likelihood weights, biases $b_\ell = \ell^{-\gamma}/8$, and strong merging under independence, the ideal rule's horizon has almost-sure order $n^{2\gamma}$ for $0 < \gamma < 1/2$ and $n/\log n$ for $\gamma \geq 1/2$. For the chronological Brier aggregate, a current-weight certificate preserves these guarantees and rates without evaluating future continuations. Its implementation uses $O(n^2)$ arithmetic operations per date and has polynomial bit cost for rationally supplied amplitudes and parameters. An explicit finite-date bound controls approval under independence. Every fixed-date-safe selector obeys an asymptotic approval-probability cap on the same scales, with sharp error level. A same-predictor comparison separates conditional approval from stationary-marginal approval.

1 The question and its scope

A convergent predictor need not tell a user how far to trust its current forecast. Consider a forecaster and a certifier who observe the same binary record and know the possible laws. The forecaster supplies a coherent probability law for the next block. The certifier chooses a horizon or abstains. An accurate approval approximates every event probability in that block within a fixed tolerance. Simultaneous safety bounds the probability of ever approving an inaccurate block under each possible truth. The horizon should also grow without bound.

This note gives an exact criterion in a countable stationary family. Each dependent law copies a bit at one prescribed lag with a known bias. A sufficiently long lag is invisible in the observed record but can affect the conditional future. Safe expansion exists exactly when those lag effects tend to zero across the catalog. The theorem applies to every fixed coherent predictor with fixed-horizon convergence under each candidate. Its obstruction needs only fixed-date safety and expansion under independence. Under the specified weights, for biases $b_\ell = \ell^{-\gamma}/8$ and a fixed predictor strongly merging under independence, the positive rule's horizon has almost-sure order $n^{2\gamma}$ for $0 < \gamma < 1/2$ and $n/\log n$ for $\gamma \geq 1/2$. Every fixed-date-safe selector obeys an asymptotic approval-probability cap on the same scale. A same-predictor comparison with stationary marginals shows that the obstruction is specific to the conditional target.

Santhanam et al. (2022) already distinguish pointwise convergence from data-derived accuracy guarantees. Their criterion concerns lossless compression of iid sources on a countable alphabet. Wu and Santhanam (2021) ask when a stopping rule can certify, at any prescribed confidence, that predictions scored by a binary loss have made their last error; they characterize this property for iid classes and leave the non-iid case open. The present target is the total-variation accuracy of one fixed predictor's conditional future block at the observed history. The selector may change the approved horizon at every date, and the horizon must grow without bound. Section 6 relates this target to prefix redundancy, structural inference, and other dependent-prediction tasks. Appendix A states reusable sufficient conditions, applies them to a different stationary Markov family, and gives a class where rare discrepancies permit exact approval outside those conditions. Section 6 ends with open questions about history-dependent classes and exact computation for other predictors.

The tolerance criterion and the expansion objective are stipulations. A fixed horizon, one terminal event, or average decision loss requires separate analysis. The examples establish no empirical forecasting claim. Appendix B supplies an explicit strongly merging chronological Brier predictor and an all-history bound from its current weights. The conservative selector in Appendix C combines that bound with likelihood confidence sets. It preserves the ideal rule's expansion rates using $O(n^2)$ arithmetic operations per date, and has polynomial bit cost for rational inputs or a polynomial-time amplitude approximation oracle. Uniformly computable amplitudes alone give finite computation without a uniform time bound.

2 The model family and the convergence premise

Let $X_t \in \{0,1\}$ and $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. All logarithms are natural. For probability laws on a finite space, use

$$d_{\mathrm{TV}}(P,Q) = \sup_A |P(A) - Q(A)| = \frac{1}{2}\sum_x |P(x) - Q(x)|.$$

For positive finite laws, relative entropy is

$$D_{\mathrm{KL}}(P \,\|\, Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}.$$

For Bernoulli laws, write

$$\mathrm{kl}(p,q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}.$$

Fix a specified sequence $0 < b_\ell \le 1/8$ and the candidate class

$$\mathcal{P}_b = \{P_0, P_1, P_2, \dots\}. \tag{1}$$

Under P_0, the observations are independent fair bits. Under P_ℓ, initialize $X_{1:\ell}$ uniformly on $\{0,1\}^\ell$ and set

$$P_\ell(X_t = 1 \mid X_{1:t-1}) = \begin{cases} 1/2, & t \le \ell, \\ 1/2 + b_\ell(2X_{t-\ell} - 1), & t > \ell. \end{cases} \tag{2}$$

The match-or-flip innovations after initialization are independent. Each bit copies the bit ℓ places earlier with probability $1/2 + b_\ell$. The parameter stays fixed within each candidate.

For a concrete instance, compare P_0 with P_6 when $b_6 = 1/8$. The record $h = (1,0,1,1,0,0)$ has probability $1/64$ under either law. After that record, the next bit has conditional probability $1/2$ of

being one under P_0 and 5/8 under P_6, because it can copy the first recorded bit. The certifier sees the same record in both cases. Knowing both model formulas does not reveal which conditional probability governs the next observation.

Lemma 1. *Every P_ℓ is stationary and ergodic. Every consecutive block of length ℓ is uniform, and*

$$P_\ell(X_{1:n} = h) = P_0(X_{1:n} = h) = 2^{-n} \qquad (n \le \ell). \tag{3}$$

The laws in $\mathcal{P}_b$ are pairwise distinct and mutually singular on infinite sequences.

Proof. Use the latest ℓ bits as the state of a finite Markov chain. A transition shifts the state and appends a noisy copy of its oldest bit. Each destination has two predecessor states, distinguished by that oldest bit. Their probabilities of appending the required bit sum to one. The transition matrix therefore preserves the uniform state distribution.

Every appended bit has positive conditional probability. Any state can reach any other state in ℓ steps. The all-zero state also has a positive self-loop. The finite state chain is irreducible and aperiodic, and its uniform initialization gives a stationary ergodic output law. Marginalizing a uniform ℓ-bit block proves (3).

For $1 \le j < k$, the stationary probability of $X_{t+j} = X_t$ is $1/2 + b_j$ under P_j. It is $1/2$ under P_k, since the relevant block has length $j + 1 \le k$. The probability is also $1/2$ under P_0. Ergodic averages of these match indicators distinguish P_j from P_k almost surely. The lag-k match frequency distinguishes P_k from P_0. The corresponding disjoint almost-sure frequency events prove mutual singularity for every pair. A positive dependence on the oldest state bit also prevents representing P_ℓ by a smaller memory order. □

B. Ryabko (2015) gives an earlier construction of dependent order-k Markov laws with uniform k-bit blocks. The single-copy lag here permits the exact conditional-distance calculation below.

For a history h of length n, write $P_{i,h,H}$ for the conditional law of the next H bits under P_i. Fix a coherent predictor R, supplied by one probability kernel on every finite history, and write $R_{h,H}$ for its conditional future law. This specification also covers histories to which the induced law R assigns probability zero. Set $d_i(h, H) = d_{\mathrm{TV}}(P_{i,h,H}, R_{h,H})$ and $d_i(h, 0) = 0$. The main equivalence assumes

$$d_i(X_{1:n}, H) \longrightarrow 0 \quad P_i\text{-almost surely} \qquad \text{for every fixed } i \text{ and finite } H. \tag{4}$$

Condition (4) concerns each fixed candidate and each fixed horizon. It supplies no date, uniform over the class, at which a forecast is accurate. Strong merging in the sense of Blackwell and Dubins (1962) additionally makes this convergence uniform over all finite future horizons, as expressed in (41). Fixed-horizon convergence is weak merging in the sense of Kalai and Lehrer (1994). The stronger premise will yield the extra conclusion $H_n = n$ eventually under each dependent candidate. Appendix B proves both properties for one explicit chronological Brier aggregate. The criterion itself applies to any fixed predictor satisfying (4).

3 The exact conditional-horizon boundary

Definition 2 (Full-access selector)**.** A selector is a sequence $(H_n)_{n\ge1}$ of random variables with values in $\{0, 1, 2, \ldots\}$ such that H_n is $\sigma(X_{1:n}, U)$-measurable, where U is a random seed independent of the observations with the same law under every candidate. Zero denotes abstention. The selector may depend on $\mathcal{P}_b$, on R, on τ, and on α.

The selector receives no indicator of the true index. Model simulations add no observations of the target process; their randomness belongs to U.

Fix $0 < \tau < 1$ and $0 < \alpha < 1$. The tolerance τ bounds probability approximation error. The error level α bounds erroneous approval. Define simultaneous safety by

$$P_i\{d_i(X_{1:n}, H_n) \le \tau \text{ for every } n \ge 1\} \ge 1 - \alpha \qquad \text{for every } i. \tag{5}$$

The weaker fixed-date requirement is

$$P_i\{d_i(X_{1:n}, H_n) > \tau\} \le \alpha \qquad \text{for every } i, n. \tag{6}$$

The safety probability averages over histories and the independent seed under each fixed true law. It controls approximation to the conditional law at the observed history, without requiring a separate error probability bound for each history. The minimum expansion objective is

$$H_n \longrightarrow \infty \quad \text{in } P_0\text{-probability}. \tag{7}$$

The expansion objective requires no rate or monotonicity and permits temporary abstention.

Total variation gives a decision interpretation of the stipulated target. On a finite future space,

$$d_{\mathrm{TV}}(P, Q) = \sup_{0 \le g \le 1} |\mathbb{E}_P g - \mathbb{E}_Q g|.$$

To verify this identity, indicators attain the event supremum, and $g(y) = \int_0^1 \mathbf{1}\{g(y) > u\}\, du$ bounds every other such function. A safe horizon therefore controls the error in every bounded future-loss expectation by τ, including a loss chosen using the observed history. For a finite action set with losses in $[0, 1]$, an action minimizing forecast expected loss has true conditional expected loss within 2τ of the minimum. Applying the expectation bound once to the selected action and once to a true minimizing action proves this last claim. These are consequences of the probability-approximation criterion.

Theorem 3 (Recognition boundary)**.** *Fix $\mathcal{P}_b$, $0 < \tau < 1$, and $0 < \alpha < 1$. Let R be one known coherent predictor with the all-history kernels specified above and satisfying* (4)*. The following statements are equivalent, and the selector in (iii) can be taken deterministic:*

(i) $b_\ell \to 0$ *as* $\ell \to \infty$.

(ii) Some full-access selector satisfies fixed-date safety (6) *and expansion* (7)*.*

(iii) Some full-access selector satisfies simultaneous safety (5) *and* $H_n \to \infty$ *almost surely under every* $P_i \in \mathcal{P}_b$*.*

If b_ℓ does not tend to zero, every selector satisfying (7) *has this property: for each $\rho > 0$ and each N_0, some $n \ge N_0$ satisfies*

$$P_n\{d_n(X_{1:n}, H_n) > \tau\} \ge 1 - \rho. \tag{8}$$

If $b_\ell \to 0$ and R also satisfies strong merging (41)*, the selector constructed below has $H_n = n$ eventually almost surely under every fixed P_j with $j \ge 1$.*

The theorem applies to the Brier aggregate in Appendix B. Decay concerns effects across the catalog; each fixed true process retains its own bias. The positive rule retains plausible models and approves only when their conditional futures are close to the fixed forecast.

3.1 Necessity: identical records and separated conditional laws

Lemma 4. *For $0 < b \le 1/8$, define*

$$J_m(b) = \frac{1}{2} \sum_{k=0}^{m} \binom{m}{k} \left| (1/2+b)^k (1/2-b)^{m-k} - 2^{-m} \right|. \tag{9}$$

For every $n \ge m$ and every history $h \in \{0,1\}^n$,

$$d_{\mathrm{TV}}(P_{n,h,m}, P_{0,h,m}) = J_m(b_n).$$

If $b_n \ge \beta > 0$, define the future event

$$E_{n,m}(h) = \left\{ \sum_{s=1}^{m} \mathbf{1}\{X_{n+s} = h_s\} > m(1/2 + \beta/2) \right\}.$$

Then

$$P_{n,h,m}(E_{n,m}(h)) - P_{0,h,m}(E_{n,m}(h)) \ge \Delta_m(\beta) := 1 - 2e^{-m\beta^2/2}. \tag{10}$$

Proof. The map from a future block to its match indicators against $h_1, \dots, h_m$ is a bijection. Under P_n, those indicators are independent Bernoulli variables of parameter $1/2 + b_n$; under P_0, they are independent fair bits. Under P_n, every indicator vector with k matches has probability $(1/2+b_n)^k(1/2-b_n)^{m-k}$; under P_0, every such vector has probability 2^{-m}. Grouping the total-variation sum by k proves the stated identity with (9), which is also the distance between the corresponding binomial count laws.

Under P_0, the m match indicators are independent Bernoulli variables of parameter $1/2$. Under P_n, they reference the already observed bits $h_1, \dots, h_m$ and are independent Bernoulli variables of parameter $1/2 + b_n$. The bounded-variable tail inequality gives

$$P_{0,h,m}(E_{n,m}(h)) \le e^{-m\beta^2/2}, \qquad P_{n,h,m}(E_{n,m}(h)^c) \le e^{-m\beta^2/2}.$$

The strict event causes no rounding problem: its fair-law upper tail is contained in the event of a deviation at least $m\beta/2$, and its complement under P_n has a downward deviation at least that size. Subtracting gives (10). □

Proof of necessity in Theorem 3. If b_ℓ does not tend to zero, some $\beta > 0$ has $b_n \ge \beta$ for infinitely many n. Choose a fixed m with $\Delta_m(\beta) > \tau$ and a number $0 < \delta < \Delta_m(\beta) - \tau$. For each such $n \ge m$, consider

$$A_n = \{H_n \ge m,\ d_0(X_{1:n}, m) \le \delta\}.$$

Equation (3) gives the same distribution of $X_{1:n}$ under P_0 and P_n. The class, predictor, computations, and independent seed also stay fixed. It follows that $P_n(A_n) = P_0(A_n)$.

On A_n, the triangle inequality and (10) give

$$d_n(X_{1:n}, m) \ge \Delta_m(\beta) - d_0(X_{1:n}, m) > \tau.$$

Total variation between block laws cannot decrease when the block is enlarged: an event in its first m coordinates is still available in the larger block. Every approval on A_n therefore fails under P_n, and

$$P_n\{d_n(X_{1:n}, H_n) > \tau\} \ge P_0(H_n \ge m) - P_0\{d_0(X_{1:n}, m) > \delta\}. \tag{11}$$

Expansion makes the first term tend to one. Fixed-horizon convergence under P_0 makes the second term tend to zero. Along the infinite set of indices with $b_n \ge \beta$, the right side tends to one. This proves (8) and rules out (6). □

In (8), the fixed pool contains the adverse laws, but the true index witnessing a late failure varies with n. The result does not assert infinitely many failures under one fixed truth. Under the additional strong-merging condition (41), all selected finite horizons meet the tolerance eventually almost surely under any fixed P_i. What fails is one finite-date error bound valid across the whole class.

For constant amplitude $b_\ell = \beta$, the same transfer argument with $0 < \delta < J_m(\beta) - \tau$ gives, for every fixed-date-safe selector,

$$\limsup_{n\to\infty} P_0(H_n \geq m) \leq \alpha \qquad \text{whenever } J_m(\beta) > \tau.$$

At $\beta = 1/8$ and $\tau = 1/4$, the smallest such horizon is six: $J_5(1/8) = 3683/16384 < 1/4$ and $J_6(1/8) = 33069/131072 > 1/4$. The event that at least four of the next six bits match the first six recorded bits has probabilities $11/32$ and $78125/131072$ under the two conditional laws. Their difference is $J_6(1/8)$, although the one-step gap is only $1/8$.

3.2 Sufficiency: a valid confidence set and a decay bound

The rule approves a horizon when every retained model is within tolerance of the fixed forecast. Coverage of the true model gives safety; expansion requires separate tail control.

Fix positive weights a_i with $\sum_i a_i = 1$; the choice in (35) is one example. Use a likelihood mixture as a testing reference:

$$M = \sum_{i\geq 0} a_i P_i, \qquad C_n = \{i : P_i(X_{1:n}) \geq \alpha M(X_{1:n})\}. \tag{12}$$

The mixture M is a testing reference. The forecasts being certified remain those of R, including when R is the different chronological Brier aggregate of Appendix B.

Under each true P_i, the ratio $M(X_{1:n})/P_i(X_{1:n})$ is a nonnegative martingale starting at one. Conditional summation over the next bit verifies the martingale identity. Ville's inequality (Ville, 1939) gives

$$P_i\{i \in C_n \text{ for every } n\} \geq 1 - \alpha. \tag{13}$$

Apply the stopped martingale identity at the first crossing of $1/\alpha$ up to a bounded date. Its expectation remains one, so the probability of crossing by that date is at most α. Increasing the date gives the asserted inequality. No union over the true index is required. The set C_n is nonempty because its defining threshold is smaller than the weighted average $M(X_{1:n})$.

Define

$$U_n(H) = \sup_{i\in C_n} d_i(X_{1:n}, H), \quad U_n(0) = 0, \qquad H_n^* = \max\{0 \leq H \leq n : U_n(H) \leq \tau\}. \tag{14}$$

Every distance is a function on finitely many histories at each fixed date and horizon. The countable supremum and finite maximum therefore define a measurable selector. On the coverage event, every chosen horizon is accurate under the true candidate. This proves simultaneous safety. Expansion requires a separate argument. The rule uses no seed.

Lemma 5. *Under each fixed P_i, the likelihood weights*

$$v_j(n) = \frac{a_j P_j(X_{1:n})}{M(X_{1:n})}$$

satisfy $v_i(n) \to 1$ and $v_j(n) \to 0$ for every fixed $j \neq i$. The true index eventually belongs to C_n, and each fixed false index is eventually excluded.

Proof. On a joint space, draw J with probabilities (a_i) and then draw a sequence with law P_J. Countably many pairwise singular laws admit disjoint measurable typical sets of full measure under their respective laws. One construction intersects, for each index, its probability-one separating events against every other index, and then removes overlaps in index order. The infinite observation sequence therefore determines J almost surely.

The weights satisfy $v_j(n) = \mathbb{E}[\mathbf{1}\{J = j\} \mid \mathcal{F}_n]$. Lévy's upward theorem gives $v_j(n) \to \mathbf{1}\{J = j\}$ almost surely; this is Doob's consistency argument for a countable class of pairwise singular laws. Conditioning on the event $J = i$, which has positive probability, yields the claims under P_i. In particular,

$$\frac{M(X_{1:n})}{P_i(X_{1:n})} = \frac{a_i}{v_i(n)} \to a_i < 1/\alpha, \qquad \frac{P_j(X_{1:n})}{M(X_{1:n})} = \frac{v_j(n)}{a_j} \to 0.$$

These inequalities establish the membership conclusions. □

Proof of sufficiency in Theorem 3. Assume $b_\ell \to 0$. Every transition of P_ℓ lies within b_ℓ of $1/2$. The bound $\mathrm{kl}(q, 1/2) \le 4(q - 1/2)^2$, the relative-entropy chain rule, and Pinsker's inequality give

$$d_{\mathrm{TV}}(P_{\ell,h,H}, P_{0,h,H}) \le \sqrt{2H}\, b_\ell \tag{15}$$

for every history and finite horizon. For each fixed $j \ge 1$, the independent match innovations give

$$\frac{1}{n} \log \frac{P_j(X_{1:n})}{P_0(X_{1:n})} \longrightarrow D_j := \mathrm{kl}(1/2 + b_j, 1/2) > 0 \quad P_j\text{-almost surely.} \tag{16}$$

For every record,

$$\log \frac{P_\ell(X_{1:n})}{P_0(X_{1:n})} \le n \log(1 + 2b_\ell) \le 2nb_\ell. \tag{17}$$

Together with Lemma 1, these bounds verify the sufficient condition of Theorem 16 in Appendix A, with $c_H = \sqrt{2H}$, $\varepsilon_\ell = b_\ell$, and $\lambda_\ell = 2b_\ell$. That theorem gives simultaneous safety, almost-sure expansion under every candidate, and the stated strong-merging conclusion. Statement (iii) implies (ii); necessity gives (ii) implies (i). □

Under P_0, the set behaves differently. Lemma 5 eventually includes index zero. Equation (3) then includes every $\ell \ge n$ as well. The confidence set remains infinite and retains unbounded possible memory orders. The decay assumption makes those alternatives close on each fixed horizon.

4 Expansion rates under independence

The rate result concerns the ideal rule (14), capped at n, and assumes strong merging under P_0. Both predictors in the appendices satisfy that premise.

4.1 Eventual lower bounds and a polynomial cap

Proposition 6 (Polynomially decaying amplitudes)**.** *Let $b_\ell = \ell^{-\gamma}/8$ with $\gamma > 0$, and use the weights* (35)*. Suppose R satisfies* (4) *and strongly merges under P_0. Set $K = 1024$ and, for $n \ge 3$,*

$$L_n = \left\lfloor \min \left\{ \frac{n}{2}, \left(\frac{n}{K \log n} \right)^{1/(2\gamma)} \right\} \right\rfloor.$$

Almost surely under P_0, for all sufficiently large n,

$$H_n^* \geq \left\lfloor \min\{n,\, 8\tau^2(L_n+1)^{2\gamma}\} \right\rfloor. \tag{18}$$

In particular, the eventual lower bounds are

$$H_n^* \geq \begin{cases} \lfloor 2^{3-2\gamma}\tau^2 n^{2\gamma} \rfloor, & 0 < \gamma < 1/2, \\ \lfloor \tau^2 n/(128 \log n) \rfloor, & \gamma \geq 1/2. \end{cases} \tag{19}$$

For $0 < \gamma < 1/2$, put

$$C_\tau = 128 \log \frac{4}{1-\tau}, \qquad m_n = \lceil C_\tau n^{2\gamma} \rceil.$$

Every selector satisfying fixed-date safety obeys

$$\limsup_{n\to\infty} P_0(H_n \geq m_n) \leq \alpha. \tag{20}$$

The probability cap permits arbitrary finite horizons, including horizons greater than n.

Proof. For a fixed lag $\ell < n$, the map from $X_{1:n}$ to its first ℓ bits and the indicators $\mathbf{1}\{X_t = X_{t-\ell}\}$, $t = \ell+1, \ldots, n$, is a bijection. Under P_0 all these coordinates are independent fair bits. If $Z_{\ell,n} = P_\ell(X_{1:n})/P_0(X_{1:n})$, then

$$\mathbb{E}_0 \sqrt{Z_{\ell,n}} = A(b_\ell)^{n-\ell}, \qquad A(b) = \frac{\sqrt{1+2b} + \sqrt{1-2b}}{2}.$$

Since $A(b)^2 = (1 + \sqrt{1-4b^2})/2 \leq 1 - b^2$, we have $A(b) \leq e^{-b^2/2}$. Membership in C_n requires $Z_{\ell,n} \geq \alpha a_0 = \alpha/2$, because $M \geq a_0 P_0$. Markov's inequality gives

$$P_0(\ell \in C_n) \leq \sqrt{2/\alpha}\, \exp\{-(n-\ell)b_\ell^2/2\}. \tag{21}$$

For $1 \leq \ell \leq L_n$ and all sufficiently large n,

$$(n-\ell)b_\ell^2 \geq \frac{n}{128} L_n^{-2\gamma} \geq \frac{K}{128} \log n.$$

A union bound over at most n lags makes the probability that any such lag is retained at most $\sqrt{2/\alpha}\, n^{1-K/256} = \sqrt{2/\alpha}\, n^{-3}$. The sum over n is finite. The Borel–Cantelli lemma proves that eventually all indices $1, \ldots, L_n$ are excluded, almost surely.

Strong merging gives $\sup_H d_0(X_{1:n}, H) \leq \tau/2$ eventually. Every retained positive index then has $b_\ell \leq (L_n+1)^{-\gamma}/8$, so (15) and the triangle inequality give

$$U_n(H) \leq \frac{\tau}{2} + \frac{\sqrt{2H}}{8(L_n+1)^\gamma}.$$

Every integer $H \leq \min\{n, 8\tau^2(L_n+1)^{2\gamma}\}$ qualifies. This proves (18). If $\gamma < 1/2$, eventually $L_n = \lfloor n/2 \rfloor$, and the stated polynomial term is less than n. If $\gamma \geq 1/2$, eventually the other term defines L_n, and $(L_n+1)^{2\gamma} \geq n/(K \log n)$. These observations prove (19).

For the converse, the affinity between the fair and biased product laws in Lemma 4 equals $A(b)^m$. For finite laws P, Q, $\sum_x \min\{P(x), Q(x)\} \leq \sum_x \sqrt{P(x)Q(x)}$, so

$$J_m(b) \geq 1 - A(b)^m \geq 1 - e^{-mb^2/2}.$$

Since $\gamma < 1/2$, eventually $m_n \le n$. The choice of C_τ gives $m_n b_n^2/2 \ge \log[4/(1-\tau)]$, and therefore $J_{m_n}(b_n) \ge (3+\tau)/4$. Put $\delta = (1-\tau)/2$. Then $J_{m_n}(b_n) > \tau + \delta$ with the fixed margin $(1-\tau)/4$. The finite-date transfer argument (11), now used at m_n, gives

$$P_0(H_n \ge m_n) \le \alpha + P_0\{d_0(X_{1:n}, m_n) > \delta\}.$$

Strong merging makes the last term tend to zero. This proves (20). □

The converse caps approval probability at the prescribed error level α on the scale $n^{2\gamma}$. It gives no almost-sure upper bound, expectation bound, or conclusion $H_n = O_{P_0}(n^{2\gamma})$ for an arbitrary fixed-date-safe selector, and Remark 10 shows that no such bound holds for every such selector. Fixed-horizon convergence alone would not remove its moving-horizon error term. For $\gamma \ge 1/2$, the lower bound $n/\log n$ and the trivial cap n leave a logarithmic factor. Section 4.2 removes that factor in the same approval-probability sense and proves matching almost-sure upper bounds for the ideal rule in both regimes. The constants are conservative, and Remark 11 shows that the eventual date cannot be uniform over predictors.

4.2 Visible lags and matching orders

For $\gamma \ge 1/2$, the transfer argument with lags at least n gives no cap below a constant multiple of n. By (15), such a lag separates conditional futures by more than τ only at horizons above $32\tau^2 n^{2\gamma}$. A sharper converse must use lags shorter than the record. The record then carries evidence about those lags, and exact prefix equality fails. An average over many visible lags restores approximate equality with independence. The first step extends the exact distance of Lemma 4 to every visible lag and every horizon.

Lemma 7. *For every lag $1 \le \ell \le n$, every history $h \in \{0,1\}^n$, and every horizon $H \ge 1$,*

$$d_{\mathrm{TV}}(P_{\ell,h,H}, P_{0,h,H}) = J_H(b_\ell).$$

Proof. Write $z = (h, y)$ for the history followed by a future block $y \in \{0,1\}^H$, and set $\xi_s = \mathbf{1}\{z_{n+s} = z_{n+s-\ell}\}$ for $1 \le s \le H$. Since $\ell \le n$, the coordinate $z_{n+s-\ell}$ lies in h or in $y_1, \dots, y_{s-1}$. Given h, the map $y \mapsto \xi$ is therefore a bijection of $\{0,1\}^H$: y_s is recovered from ξ_s and earlier coordinates. Under P_ℓ, (2) gives conditional match probability $1/2 + b_\ell$ at every date after ℓ, so $\xi_1, \dots, \xi_H$ are independent Bernoulli variables of parameter $1/2 + b_\ell$ given h. Under P_0, y is uniform, and so is ξ. The bijection preserves total variation, and the grouping argument of Lemma 4 gives $J_H(b_\ell)$. Lemma 4 is the case $\ell = n$ and $H \le n$. □

The next proposition bounds the rule (14) by the smallest visible lag that it retains.

Proposition 8 (Retained lags bound the rule)**.** *Let $b_\ell = \ell^{-\gamma}/8$ with $\gamma > 0$, put $\kappa_\tau = \log[2/(1-\tau)]$, and write $\bar{d}_n = \sup_{H \ge 1} d_0(X_{1:n}, H)$.*

(a) If $\ell \in C_n$, $1 \le \ell \le n$, and $\bar{d}_n \le (1-\tau)/2$, then

$$H_n^* \le \lfloor 128 \kappa_\tau \ell^{2\gamma} \rfloor. \tag{22}$$

(b) Suppose R strongly merges under P_0. Almost surely under P_0, for all sufficiently large n, the index n belongs to C_n, $H_n^* \le \lfloor 128\kappa_\tau n^{2\gamma} \rfloor$, and the smallest retained positive lag $\ell_{\min}(n) = \min\{\ell \ge 1 : \ell \in C_n\}$ satisfies

$$\min\{n, \lfloor 8\tau^2 \ell_{\min}(n)^{2\gamma} \rfloor\} \le H_n^* \le \lfloor 128 \kappa_\tau \ell_{\min}(n)^{2\gamma} \rfloor. \tag{23}$$

Proof. (a) Lemma 7, the triangle inequality, and the bound $J_H(b) \geq 1 - e^{-Hb^2/2}$ from the proof of Proposition 6 give, for every $H \geq 1$,

$$U_n(H) \geq d_\ell(X_{1:n}, H) \geq J_H(b_\ell) - d_0(X_{1:n}, H) \geq 1 - e^{-Hb_\ell^2/2} - \frac{1-\tau}{2}.$$

If $H > 2\kappa_\tau / b_\ell^2 = 128\kappa_\tau \ell^{2\gamma}$, then $e^{-Hb_\ell^2/2} < (1-\tau)/2$ and $U_n(H) > \tau$. No such H qualifies in (14).

(b) Under P_0, Lemma 5 gives $M(X_{1:n})/P_0(X_{1:n}) \to a_0 < 1/\alpha$, and (3) gives $P_n(X_{1:n}) = P_0(X_{1:n})$. These limits put n in C_n eventually. Strong merging gives $\bar{d}_n \leq \min\{\tau/2, (1-\tau)/2\}$ eventually. At such n, part (a) with $\ell = n$ gives the bound by $n^{2\gamma}$, and part (a) with $\ell = \ell_{\min}(n) \leq n$ gives the upper half of (23). Every retained positive index satisfies $\ell \geq \ell_{\min}(n)$, so (15) and the triangle inequality give

$$U_n(H) \leq \frac{\tau}{2} + \sqrt{2H}\, b_{\ell_{\min}(n)},$$

as in the proof of Proposition 6. Every integer $H \leq \min\{n, 8\tau^2 \ell_{\min}(n)^{2\gamma}\}$ therefore qualifies, which proves the lower half. □

Under the hypotheses of Proposition 6 with $0 < \gamma < 1/2$, (19) and part (b) give

$$\lfloor 2^{3-2\gamma} \tau^2 n^{2\gamma} \rfloor \leq H_n^* \leq \lfloor 128\kappa_\tau n^{2\gamma} \rfloor$$

eventually almost surely under P_0. The rule's horizon therefore has almost-sure order $n^{2\gamma}$. Since $128\kappa_\tau < C_\tau$, the rule also satisfies $H_n^* < m_n$ eventually, with m_n from (20). For $\gamma \geq 1/2$, part (b) bounds H_n^* below n only when $128\kappa_\tau \ell_{\min}(n)^{2\gamma} < n$. The index $\ell_{\min}(n)$ depends on the confidence set alone; the predictor enters (23) only through the eventual condition on $\bar{d}_n$.

Proposition 9 (Logarithmic scale)**.** *Fix $\gamma \geq 1/2$ and $0 < \alpha, \tau < 1$. Let $b_\ell = \ell^{-\gamma}/8$, and suppose one fixed predictor R strongly merges under P_0. With $C_\tau = 128 \log[4/(1-\tau)]$ as in Proposition 6, put*

$$m_n = \left\lceil \frac{9^\gamma \gamma C_\tau n}{4 \log n} \right\rceil \qquad (n \geq 3).$$

(a) Every selector satisfying fixed-date safety obeys

$$\limsup_{n \to \infty} P_0(H_n \geq m_n) \leq \alpha. \tag{24}$$

(b) Under the weights (35)*, the ideal rule* (14) *satisfies $H_n^* < m_n$ at every sufficiently large integer date, almost surely under P_0.*

Proof. Put $x_n = (\gamma n/(4 \log n))^{1/(2\gamma)}$, $L_n = \lceil x_n \rceil$, and $S_n = \{L_n, L_n + 1, \ldots, 2L_n\}$. Since $1/(2\gamma) \leq 1$, $x_n/n \leq (\gamma/(4 \log n))^{1/(2\gamma)} \to 0$, while $x_n \to \infty$. Consider n so large that $x_n \geq 2$ and $2L_n \leq n$. Then $2L_n \leq 2x_n + 2 \leq 3x_n$. Every $\ell \in S_n$ therefore satisfies $1 \leq \ell \leq n$, $nb_\ell^2 \leq n x_n^{-2\gamma}/64 = \log n/(16\gamma)$, and

$$\frac{m_n b_\ell^2}{2} \geq \frac{m_n}{128(3x_n)^{2\gamma}} \geq \frac{9^\gamma \gamma C_\tau n}{4 \log n} \cdot \frac{4 \log n}{128 \cdot 9^\gamma \gamma n} = \log \frac{4}{1-\tau}.$$

Separation. For every $\ell \in S_n$ and every history $h \in \{0,1\}^n$, Lemma 7 and $J_m(b) \geq 1 - e^{-mb^2/2}$ give $d_{\mathrm{TV}}(P_{\ell,h,m_n}, P_{0,h,m_n}) = J_{m_n}(b_\ell) \geq (3+\tau)/4$. Put $\delta = (1-\tau)/2$ and

$$A_n = \{H_n \geq m_n,\ d_0(X_{1:n}, m_n) \leq \delta\}.$$

On A_n, block monotonicity and the triangle inequality give, for every $\ell \in S_n$,

$$d_\ell(X_{1:n}, H_n) \geq d_\ell(X_{1:n}, m_n) \geq J_{m_n}(b_\ell) - \delta \geq \tau + \frac{1-\tau}{4}.$$

Fixed-date safety therefore gives $P_\ell(A_n) \leq \alpha$ for every $\ell \in S_n$.

Transfer. Write $Z_\ell = P_\ell(X_{1:n})/P_0(X_{1:n})$, $\bar{Z}_n = |S_n|^{-1} \sum_{\ell \in S_n} Z_\ell$, and $Q_n = |S_n|^{-1} \sum_{\ell \in S_n} P_\ell$. Then $Q_n(A_n) \leq \alpha$ and $\mathbb{E}_0 \bar{Z}_n = 1$. The seed is independent of the record and has a common law. For each seed value, the section of A_n is an event in $\mathcal{F}_n$, so its probabilities under Q_n and P_0 differ by at most $d_{\mathrm{TV}}(Q_n|_{\mathcal{F}_n}, P_0|_{\mathcal{F}_n})$. Integration over the seed and the Cauchy–Schwarz inequality give

$$P_0(A_n) \leq Q_n(A_n) + d_{\mathrm{TV}}(Q_n|_{\mathcal{F}_n}, P_0|_{\mathcal{F}_n}) = Q_n(A_n) + \tfrac{1}{2}\mathbb{E}_0|\bar{Z}_n - 1| \leq \alpha + \tfrac{1}{2}\sqrt{\mathbb{E}_0 \bar{Z}_n^2 - 1}.$$

Second moments. Put $s_t = 2X_t - 1$. By (2), the conditional probability under P_ℓ that X_t takes its observed value is $(1 + 2b_\ell s_t s_{t-\ell})/2$ for $t > \ell$, and the first ℓ bits are fair, so

$$Z_\ell = \prod_{t=\ell+1}^{n} (1 + 2b_\ell s_t s_{t-\ell}).$$

Under P_0, the signs $s_1, \ldots, s_n$ are independent and uniform on $\{-1, 1\}$. The bijection in the proof of Proposition 6 makes the products $s_t s_{t-\ell}$, $\ell < t \leq n$, independent uniform signs, so

$$\mathbb{E}_0 Z_\ell^2 = (1 + 4b_\ell^2)^{n-\ell} \leq e^{4nb_\ell^2} \leq n^{1/(4\gamma)}.$$

For distinct $\ell, \ell' \in S_n$, consider the graph on $\{1, \ldots, n\}$ with the disjoint edge sets $E_\ell = \{\{t-\ell, t\} : \ell < t \leq n\}$ and $E_{\ell'}$. Give each edge in E_ℓ the weight $2b_\ell$ and each edge in $E_{\ell'}$ the weight $2b_{\ell'}$. Expanding both products writes $Z_\ell Z_{\ell'}$ as a sum, over edge sets $F \subseteq E_\ell \cup E_{\ell'}$, of the weight product $w(F)$ times $\prod_{\{u,v\} \in F} s_u s_v$. That sign product has P_0-expectation one when every vertex has even degree in F, and zero otherwise. Therefore $\mathbb{E}_0[Z_\ell Z_{\ell'}]$ equals the sum of $w(F)$ over the even edge sets F. A nonempty even edge set contains a cycle, and removing that cycle leaves an even set. Every even set is therefore a union of edge-disjoint cycles, and its weight is the product of their weights. Distinct even sets give distinct collections of cycles, so

$$\mathbb{E}_0[Z_\ell Z_{\ell'}] \leq \prod_C \{1 + w(C)\} \leq \exp\Big\{\sum_C w(C)\Big\},$$

where C ranges over the cycles of the graph. Every vertex has degree at most four, so at most $4n3^{k-2}$ cycles have length $k \geq 3$: choose a starting vertex, a first neighbor, and each later vertex among at most three neighbors. Every edge weight is at most $2b_{L_n} \leq 1/4$. Therefore

$$\sum_C w(C) \leq \sum_{k \geq 3} 4n3^{k-2}(2b_{L_n})^k = \frac{4n}{9} \cdot \frac{(6b_{L_n})^3}{1 - 6b_{L_n}} \leq 384\, n b_{L_n}^3.$$

Limit. Since $|S_n| = L_n + 1 > x_n$,

$$\mathbb{E}_0 \bar{Z}_n^2 - 1 \leq \frac{\max_\ell \mathbb{E}_0 Z_\ell^2}{|S_n|} + \max_{\ell \neq \ell'} \mathbb{E}_0[Z_\ell Z_{\ell'}] - 1 \leq \Big(\frac{4 \log n}{\gamma}\Big)^{1/(2\gamma)} n^{-1/(4\gamma)} + e^{384 n b_{L_n}^3} - 1.$$

The first term tends to zero. The bound $nb_{L_n}^2 \leq \log n/(16\gamma)$ and $b_{L_n} \leq x_n^{-\gamma}/8$ give

$$n b_{L_n}^3 \leq \frac{\log n}{128\gamma}\Big(\frac{4 \log n}{\gamma n}\Big)^{1/2} \longrightarrow 0.$$

Therefore $\mathbb{E}_0 \bar{Z}_n^2 \to 1$.

Conclusion. For (a),

$$P_0(H_n \geq m_n) \leq P_0(A_n) + P_0\{d_0(X_{1:n}, m_n) > \delta\} \leq \alpha + \tfrac{1}{2}\sqrt{\mathbb{E}_0 \bar{Z}_n^2 - 1} + P_0\{d_0(X_{1:n}, m_n) > \delta\}.$$

Strong merging makes the last term tend to zero, which proves (24).

The rule at every date. For (b), put $n_k = \lceil (3/2)^k \rceil$ and $N_k = n_{k+1}$. For all sufficiently large k, $N_k \leq 7n_k/4$, $x_{n_k} \geq 2$, and $2L_{n_k} \leq n_k$. Keep S_{n_k} fixed throughout the block of dates $n_k \leq t \leq N_k$, and define

$$Z_{k,t} = \frac{1}{|S_{n_k}|} \sum_{\ell \in S_{n_k}} \frac{P_\ell(X_{1:t})}{P_0(X_{1:t})} \qquad (0 \leq t \leq N_k).$$

Each summand is a mean-one square-integrable martingale under P_0, so the same holds for $Z_{k,t}$, with $Z_{k,0} = 1$. The preceding diagonal calculation, now at date N_k, gives

$$\mathbb{E}_0 \left[\left\{ \frac{P_\ell(X_{1:N_k})}{P_0(X_{1:N_k})} \right\}^2 \right] \leq e^{4N_k b_{L_{n_k}}^2} \leq n_k^{7/(16\gamma)}.$$

The cycle calculation bounds each cross moment for distinct lags in S_{n_k} by $e^{384 N_k b_{L_{n_k}}^3} \leq e^{672 n_k b_{L_{n_k}}^3}$. Since $|S_{n_k}| > x_{n_k}$, averaging gives

$$\mathbb{E}_0 (Z_{k,N_k} - 1)^2 \leq \left(\frac{4 \log n_k}{\gamma} \right)^{1/(2\gamma)} n_k^{-1/(16\gamma)} + e^{672 n_k b_{L_{n_k}}^3} - 1.$$

Also,

$$672 n_k b_{L_{n_k}}^3 \leq \frac{21}{2} \gamma^{-3/2} (\log n_k)^{3/2} n_k^{-1/2}.$$

Both variance terms are summable in k: n_k grows geometrically, and $e^y - 1 \leq 2y$ for $0 \leq y \leq 1$.

Write $W_{k,t} = Z_{k,t} - 1$. Let T_k be the first $t \leq N_k$ with $|W_{k,t}| \geq 1/2$, taking $T_k = N_k$ if no crossing occurs. Conditioning on each event $\{T_k = t\}$ gives the bounded stopping identity $W_{k,T_k} = \mathbb{E}_0[W_{k,N_k} \mid \mathcal{F}_{T_k}]$. Conditional Jensen and $|W_{k,T_k}| \geq 1/2$ on the crossing event give

$$P_0 \left\{ \max_{0 \leq t \leq N_k} |Z_{k,t} - 1| \geq \frac{1}{2} \right\} \leq 4 \mathbb{E}_0 W_{k,T_k}^2 \leq 4 \mathbb{E}_0 W_{k,N_k}^2.$$

The Borel–Cantelli lemma gives $Z_{k,t} > 1/2$ throughout every sufficiently late block $n_k \leq t \leq N_k$, almost surely under P_0.

Under the weights (35), Lemma 5 gives $M(X_{1:t})/P_0(X_{1:t}) \to 1/2 < 1/(2\alpha)$, and strong merging gives $\bar{d}_t \leq (1-\tau)/2$ eventually, almost surely under P_0. At every date t in every sufficiently late block, some $\ell \in S_{n_k}$ therefore satisfies $P_\ell(X_{1:t}) \geq P_0(X_{1:t})/2 > \alpha M(X_{1:t})$. This lag belongs to C_t and is visible because $\ell \leq 2L_{n_k} \leq n_k \leq t$. Proposition 8(a) gives

$$\begin{aligned} H_t^* &\leq 128 \kappa_\tau (2L_{n_k})^{2\gamma} \leq \frac{9^\gamma \gamma \cdot 128 \kappa_\tau}{4} \frac{n_k}{\log n_k} \\ &< \frac{9^\gamma \gamma C_\tau}{4} \frac{t}{\log t} \leq m_t. \end{aligned}$$

Here $2L_{n_k} \leq 3x_{n_k}$, $128\kappa_\tau < C_\tau$, and $t/\log t$ increases for $t \geq 3$. The blocks cover all sufficiently large integer dates. The argument uses no monotonicity of C_t or H_t^*. □

Under the hypotheses of Proposition 6 with $\gamma \geq 1/2$, (19) and part (b) give $\lfloor \tau^2 n/(128 \log n) \rfloor \leq H_n^* < m_n$ eventually almost surely under P_0. The rule's horizon therefore has almost-sure order $n/\log n$ at all sufficiently large integer dates. Part (a) caps every fixed-date-safe selector on the same scale. These assertions fix γ, α, τ, and R; the exceptional null set and onset date may depend on them. Since $m_n/n \to 0$, $H_n^*/n \to 0$ almost surely.

The cap in part (a) is asymptotic. In its proof, the bound on $P_0(H_n \geq m_n)$ exceeds α by $\frac{1}{2}(\mathbb{E}_0 \bar{Z}_n^2 - 1)^{1/2}$ plus the merging term $P_0\{d_0(X_{1:n}, m_n) > \delta\}$, whose rate depends on R. The bound on $\mathbb{E}_0 \bar{Z}_n^2 - 1$ is of order $(\log n)^{1/(2\gamma)} n^{-1/(4\gamma)} + (\log n)^{3/2} n^{-1/2}$. At $\gamma = 1/2$ and $n = 10^6$ it exceeds one, so the explicit bound on $P_0(H_n \geq m_n)$ at that date exceeds $\alpha + 0.6$. The proposition gives no informative finite-sample bound at moderate n.

Remark 10 (The level α is sharp). The level α in (20) and (24) cannot be lowered, even for selectors with simultaneous safety and almost-sure expansion. Assume the hypotheses of Proposition 6, and let (m_n) be the sequence in (20) when $\gamma < 1/2$ and in (24) when $\gamma \geq 1/2$. Fix $0 < q < \alpha$, integers $h_n \geq m_n$, and $\alpha' = (\alpha - q)/(1 - q) \in (0, \alpha)$. Let H_n' denote the rule (14) at error level α'. Take the seed U uniform on $[0, 1]$, and set $H_n = h_n$ if $U \leq q$ and $H_n = H_n'$ otherwise. Independence of U and simultaneous safety of the rule at level α' give, under each P_i,

$$P_i\{d_i(X_{1:n}, H_n) > \tau \text{ for some } n\} \leq q + (1 - q)\alpha' = \alpha.$$

The sufficiency proof of Theorem 3 at level α' gives $H_n' \to \infty$ almost surely under every candidate, and $h_n \geq m_n \to \infty$. Yet $P_0(H_n \geq m_n) \geq q$ for every $n \geq 3$, and $q < \alpha$ is arbitrary. Since (h_n) is also arbitrary, no deterministic sequence r_n satisfies $H_n = O_{P_0}(r_n)$ for every selector with simultaneous safety and almost-sure expansion. In contrast, the rule (14) itself has $P_0(H_n^* \geq m_n) \to 0$ in both regimes, by Proposition 8 and Proposition 9(b).

Remark 11 (No predictor-uniform onset date). The date after which (18) holds depends on R, and no date serves every strongly merging predictor. Fix $N \geq 1$. Let $R^{(N)}$ predict a one with probability one at every history of length less than N, and use the one-step kernels of the mixture M at every other history. For $n \geq N$, the distance $d_i(X_{1:n}, H)$ uses only kernels at histories of length at least n. The predictor $R^{(N)}$ therefore inherits (4) and strong merging (41) under every candidate from M (Appendix D). For $n < N$ and $1 \leq H \leq N - n$, $R^{(N)}_{X_{1:n}, H}$ is the point mass at 1^H. Every candidate gives 1^H conditional probability at most $(5/8)^H$, because its one-step probabilities lie in $[3/8, 5/8]$. Therefore $d_i(X_{1:n}, H) \geq 1 - (5/8)^H$ for every i. Let H_τ be the least integer with $(5/8)^{H_\tau} < 1 - \tau$. Since C_n is nonempty, $U_n(H_\tau) > \tau$ for every $n \leq N - H_\tau$, and block monotonicity extends this to every $H \geq H_\tau$. The rule therefore satisfies $H_n^* < H_\tau$ for every $n \leq N - H_\tau$, on every record. For $\tau = 1/4$, $H_\tau = 1$, so the rule abstains at every date before N. The right side of (18) does not depend on R and is at least H_τ for all $n \geq n_1$, for some n_1. Given any date N_0, the choice $N = H_\tau + \max\{N_0, n_1\}$ makes (18) fail for $R^{(N)}$ at the date $N - H_\tau \geq N_0$.

5 Uniform learning of stationary block probabilities

Stationary block probabilities remain uniformly learnable throughout the same family. The proof bounds the total influence of a bit along its residue chain, uniformly over the lag. Corollary 13 then holds the predictor fixed while changing its target from the conditional future to a stationary marginal.

For $1 \leq H \leq n$, let $\mu_{i,H}$ be the stationary distribution of H consecutive bits under P_i. At horizon zero, $\mu_{i,0}$ and every conditional block law are the point mass on the empty word. Define

the empirical sliding-block distribution by

$$\widehat{\mu}_{n,H}(A) = \frac{1}{n-H+1} \sum_{s=1}^{n-H+1} \mathbf{1}\{X_{s:s+H-1} \in A\}, \qquad A \subseteq \{0,1\}^H. \tag{25}$$

Its expectation is $\mu_{i,H}(A)$ by stationarity. The target $\mu_{i,H}$ is the unconditional block marginal. It differs from the next-block law conditional on the observed history. The observations and overlapping blocks need not be independent.

Theorem 12 (Uniform block-marginal learning)**.** *For every sequence $0 < b_\ell \le 1/8$, define*

$$A_{n,H}(\alpha) = (2^H + 1)\log 2 + \log \frac{n(n+1)H(H+1)}{\alpha}, \tag{26}$$

$$B_{n,H}(\alpha) = \min\left\{1, \frac{4H}{3(n-H+1)}\sqrt{\frac{n}{2}A_{n,H}(\alpha)}\right\}. \tag{27}$$

For each $P_i \in \mathcal{P}_b$,

$$P_i\left\{d_{\mathrm{TV}}(\widehat{\mu}_{n,H}, \mu_{i,H}) \le B_{n,H}(\alpha) \text{ for all } n \ge 1,\ 1 \le H \le n\right\} \ge 1 - \alpha. \tag{28}$$

The radius tends to zero for each fixed H. It also tends to zero for $H = \lfloor c \log_2 n \rfloor$ with any fixed $0 < c < 1$. The deterministic rule

$$G_n = \max(\{0\} \cup \{1 \le H \le n : B_{n,H}(\alpha) \le \tau\})$$

expands and selects a simultaneously valid block-marginal estimate under every possible truth, with distance zero at horizon zero.

Proof. First let $F : \{0,1\}^n \to \mathbb{R}$ change by at most L when one coordinate changes. Under P_ℓ, the process comprises ℓ independent residue chains, each with copy probability $1/2 + b_\ell$. Fix a prefix through date $s-1$ and compare the two possible values of X_s. Couple the future identically outside the residue chain containing s. In that chain, when the copies disagree, their next Bernoulli parameters differ by $2b_\ell$. A maximal coupling leaves a disagreement with probability $2b_\ell$. Once the copies agree, couple them identically thereafter. The disagreement probability at $s + k\ell$ is $(2b_\ell)^k$, including $k = 0$. This construction also applies to an initially fair residue-chain root.

The expected number of differing coordinates is at most

$$\sum_{k \ge 0} (2b_\ell)^k = \frac{1}{1 - 2b_\ell} \le \frac{4}{3}.$$

Changing the coordinates one by one bounds the difference between the two conditional expectations of F by $4L/3$. Each difference of the Doob martingale $\mathbb{E}_i[F \mid \mathcal{F}_s]$ has conditional range length at most $4L/3$. This is a bound on range length, sufficient for the conditional exponential-moment estimate.

The conditional bounded-variable lemma states that a centered variable of range length c has exponential moment at most $\exp(\lambda^2 c^2/8)$. Iteration over the martingale differences, exponential Markov inequality, and minimization over λ give

$$P_i\{|F - \mathbb{E}_i F| \ge u\} \le 2\exp\left(-\frac{2u^2}{n(4L/3)^2}\right). \tag{29}$$

For P_0, independent coupling gives the smaller range bound L, so (29) still holds. This is a specialization of the coupling and martingale concentration method (Chazottes et al., 2007; Kontorovich and Ramanan, 2008); the lag-specific calculation supplies the uniform factor 4/3.

For $F = \widehat{\mu}_{n,H}(A)$, changing one bit changes at most H of the $N = n - H + 1$ summands. Set $L = H/N$ in (29) and use stationarity to obtain

$$P_i\{|\widehat{\mu}_{n,H}(A) - \mu_{i,H}(A)| \geq u\} \leq 2\exp\left(-\frac{9(n-H+1)^2u^2}{8nH^2}\right). \tag{30}$$

There are 2^{2^H} subsets of $\{0,1\}^H$. Allocate

$$\alpha_{n,H} = \frac{\alpha}{n(n+1)H(H+1)}$$

to their union at date n and horizon H. Substitution of the untruncated radius in (27) makes the exponent in (30) equal $A_{n,H}(\alpha)$. The union bound is $2^{2^H+1}e^{-A_{n,H}(\alpha)} = \alpha_{n,H}$. If that radius exceeds one, the total-variation bound one is deterministic. Since

$$\sum_{n\geq 1}\sum_{H=1}^{n}\frac{1}{n(n+1)H(H+1)} \leq \left(\sum_{n\geq 1}\frac{1}{n(n+1)}\right)^2 = 1,$$

a second union bound proves (28).

For fixed H, the untruncated radius is $O(\sqrt{\log n/n})$. For $H = \lfloor c\log_2 n\rfloor$, it is bounded by a constant times

$$(\log n)\sqrt{\frac{n^c + \log n}{n}},$$

which tends to zero when $c < 1$. Every fixed horizon therefore qualifies eventually for G_n, and $G_n \to \infty$. All probability guarantees hold separately under every true P_i. □

The bound is conservative. With $\alpha = 0.05$ and $\tau = 1/4$, $B_{n,1}(\alpha) \leq \tau$ first holds at $n = 238$, whereas $B_{n,6}(\alpha) \leq \tau$ first holds at $n = 37{,}305$. The theorem claims neither an optimal rate nor useful accuracy at small samples. The empirical distributions at different horizons need not form a coherent chronological forecast law. The following corollary applies to any fixed supplied coherent predictor R; the Brier aggregate in Appendix B provides one explicit example.

Corollary 13 (A fixed predictor and a marginal target)**.** *For any sequence $0 < b_\ell \leq 1/8$, let R be a fixed supplied coherent predictor with specified kernels on every history. Define*

$$H_n^{\mathrm{mar}} = \max\Big(\{0\} \cup \{1 \leq H \leq n : d_{\mathrm{TV}}(\widehat{\mu}_{n,H}, R_{X_{1:n},H}) + B_{n,H}(\alpha) \leq \tau\}\Big). \tag{31}$$

Then, for every i,

$$P_i\{d_{\mathrm{TV}}(\mu_{i,H_n^{\mathrm{mar}}}, R_{X_{1:n},H_n^{\mathrm{mar}}}) \leq \tau \text{ for every } n\} \geq 1 - \alpha.$$

If R satisfies fixed-horizon convergence under P_0, then $H_n^{\mathrm{mar}} \to \infty$ almost surely under P_0.

Proof. The simultaneous empirical-radius event and triangle inequality give safety for every selected horizon, without a convergence premise on R. Under P_0, each fixed-H empirical law converges almost surely to $\mu_{0,H}$ by ergodicity. The true conditional law equals $\mu_{0,H}$, so fixed-horizon convergence makes $R_{X_{1:n},H}$ approach the same law. Since the radius vanishes, every fixed horizon qualifies eventually. Intersect over integer horizons. □

For $b_\ell \equiv 1/8$, every fixed R satisfying the main convergence premise permits safe expanding marginal approval under P_0, whereas Theorem 3 rules out conditional approval with the same class, predictor, tolerance, error level, and expansion objective. The Brier law in Appendix B is an explicit strongly merging example. The corollary asserts no expansion under dependent truths. The empirical rule G_n does expand under every truth while directly estimating its stationary marginal.

6 Related tasks and limits

The criterion concerns approximate conditional use rather than a finite declaration of the true structure. Gallo and Leonardi (2015) rule out nontrivial upper confidence bounds for context trees over the stationary ergodic class, including its finite-context subclass, and construct lower confidence bounds with finite-sample coverage and consistency under summable tree-metric weights. Within the narrower lag family, the following argument separates approval from a declaration of independence.

Proposition 14. *For any allowed positive amplitude sequence, suppose a procedure declares "the true law is P_0" at a possibly randomized stopping time T and $P_\ell(T < \infty) \le \alpha$ for every $\ell \ge 1$. Then $P_0(T < \infty) \le \alpha$.*

Proof. For every n, choose $\ell \ge n$. The common record and seed laws give $P_0(T \le n) = P_\ell(T \le n) \le \alpha$. Increase n to infinity. □

When amplitudes decay, safe approximation can therefore expand while uniformly reliable finite declaration remains unavailable. This does not rule out identification in the limit.

Santhanam et al. (2022) certify when normalized unconditional prefix relative entropy stays below a tolerance. For iid sources, they characterize this property through local redundancy and relate it to certified single-letter marginal estimation. The corresponding prefix quantity is uniformly controlled throughout the present dependent family:

Proposition 15. *Under weights* (35)*, every allowed amplitude sequence satisfies*

$$\sup_{i \ge 0} \frac{1}{n} D_{\mathrm{KL}}(P_i|_{\mathcal{F}_n} \,\|\, M|_{\mathcal{F}_n}) \le \frac{\log(2n(n+1))}{n} \longrightarrow 0.$$

Proof. For $1 \le i < n$, domination gives $D_{\mathrm{KL}}(P_i|_{\mathcal{F}_n} \,\|\, M|_{\mathcal{F}_n}) \le -\log a_i \le \log(2n(n+1))$. For $i \ge n$, the prefix law is $P_0|_{\mathcal{F}_n}$ and domination by $a_0 P_0$ gives the bound $\log 2$, also valid for $i = 0$. □

A deterministic waiting date therefore certifies this loss with zero premature approval, even for constant amplitudes where conditional-horizon approval fails. This uses the uniform-bound implication already noted by Santhanam et al. (2022).

Morvai and Weiss (2005) consider one-step prediction at stopping times that eventually become consecutive under every finite-order stationary ergodic binary Markov law. They prove positive-probability persistent error under one fixed stationary ergodic adverse process in the larger class. The present result assumes a convergent predictor on a known countable class and concerns uniform finite-date conditional-block approval; its adverse truth can vary with the date. Shalizi and Kontorovich (2013) instead learn expectations conditional on the invariant sigma-field and preserve a common sample-complexity bound under mixtures. Under ergodicity that target is the stationary expectation used in the marginal comparison, rather than the current conditional future.

Class-membership testing also succeeds on both sides of the criterion. Enumerate finite binary words B_k and use $d_{\mathrm{dist}}(P,Q) = \sum_k c_k |P(B_k) - Q(B_k)|$ with positive summable weights. The prefix identity gives $P_\ell \to P_0$ in this metric, so $\mathcal{P}_b$ is closed and every member is ergodic. Its closure's ergodic decompositions are concentrated on members of $\mathcal{P}_b$. The criterion of D. Ryabko (2014) therefore gives tests with Type I error at most any prescribed $\gamma \in (0,1)$ at every date under $\mathcal{P}_b$ and eventual rejection almost surely under each fixed stationary ergodic alternative outside it. The same holds for the singleton $\{P_0\}$. These are rejection guarantees, without a finite-date error bound for acceptance uniform over alternatives.

The criterion also requires less than one shared summable continuity envelope of the form used by Asadi et al. (2014). Their class satisfies

$$\left| \frac{p(a \mid bu)}{p(a \mid b'u)} - 1 \right| \le d(|u|), \qquad \sum_{k \ge 1} d(k) < \infty.$$

For the lag-ℓ kernel, contexts $1u$ and $0u$ with $|u| = \ell - 1$ force

$$d(\ell - 1) \ge \frac{4b_\ell}{1 - 2b_\ell} \qquad (\ell \ge 2).$$

A common envelope therefore requires $\sum_\ell b_\ell < \infty$. The positive case $b_\ell = 1/(8\sqrt{\ell})$ violates that requirement. Each fixed law still admits its own finite-memory envelope. Asadi and colleagues estimate aggregated transition parameters on sufficiently frequent contexts; this comparison of assumptions establishes no impossibility for their different task.

The exact criterion and rates concern a designed family with known amplitudes and kernels. The marginal comparison keeps the predictor fixed; its probability target changes. Appendix A gives sufficient conditions with a different stationary application and a rare-trigger limitation, rather than a characterization of history-dependent classes. The logarithmic-scale result combines a second-moment comparison for visible-lag mixtures with a martingale bound over blocks of dates. It gives the ideal rule's matching almost-sure rate under the specified weights and a fixed strongly merging predictor. The computed certificate in Theorem 19 preserves those rates for the chronological Brier predictor; Corollary 20 gives a finite-date approval bound for that predictor. The arbitrary-selector result remains an asymptotic approval-probability cap. Remark 11 shows that no single date starts the rate bound for every strongly merging predictor.

Open problems. Two questions remain outside the proved claims. First, characterize the countable classes that give every finite word positive probability and predictors satisfying (4) for which some full-access selector satisfies fixed-date safety (6) and expansion (7). Theorem 16 and the path-dependent abstention of the rare-trigger class in Appendix A are two sufficient mechanisms, and Theorem 17 is one obstruction. Appendix A shows that the sufficient conditions and the obstruction match exactly in the lag family; no such match is proved for other history-dependent classes. Wu and Santhanam (2021) leave the analogous non-iid characterization open for certified final errors under a binary loss. Second, characterize supplied predictors for which the exact ideal selector (14) admits a polynomial-time implementation. The Brier certificate computes a conservative horizon with the same asymptotic orders; it need not compute the ideal maximum. The present results give no polynomial algorithm for that exact maximum or for arbitrary supplied coherent predictors.

A Reusable conditions and their limits

The confidence-set argument extends to countable classes through finite false-model exclusion and two uniform tail bounds. The following sufficient conditions require no stationarity.

Let $\mathcal{A}$ be a finite alphabet and let $P_0, P_1, \ldots$ be laws on $\mathcal{A}^{\mathbb{N}}$ that give positive probability to every finite word. Define d_i, safety, and selectors as before, with $\mathcal{A}$ replacing $\{0,1\}$. A fixed coherent R satisfies (4) under each P_i. For normalized weights $a_i > 0$, $\sum_i a_i = 1$, define M, C_n, U_n, and H_n^* by (12) and (14).

Theorem 16 (A countable-class sufficient condition)**.** *Suppose the laws are pairwise mutually singular. Suppose nonnegative sequences $\varepsilon_\ell \to 0$ and $\lambda_\ell \to 0$ satisfy, for every n, every history $h \in \mathcal{A}^n$, every $\ell \geq 1$, and every fixed finite H,*

$$d_{\mathrm{TV}}(P_{\ell,h,H}, P_{0,h,H}) \leq c_H \varepsilon_\ell, \qquad c_H < \infty, \tag{32}$$

$$\log \frac{P_\ell(h)}{P_0(h)} \leq n\lambda_\ell. \tag{33}$$

For each $j \geq 1$, assume a constant $D_j > 0$ satisfies

$$\frac{1}{n} \log \frac{P_j(X_{1:n})}{P_0(X_{1:n})} \longrightarrow D_j \qquad P_j\text{-almost surely.} \tag{34}$$

Then H_n^* satisfies simultaneous safety and tends to infinity almost surely under every P_i. Under each P_j with $j \geq 1$, $C_n = \{j\}$ eventually almost surely. If R strongly merges under P_j, then $H_n^* = n$ eventually under that law.

Proof. Positivity makes $M(X_{1:n})/P_i(X_{1:n})$ a mean-one nonnegative martingale under each P_i. The proof of (13) gives simultaneous coverage, and approval using the supremum over C_n gives safety. The proof of Lemma 5 uses only countability, positive weights, and pairwise singularity. It therefore gives eventual exclusion of each fixed false index and eventual inclusion of the true index here.

Under P_0, fix H and L. After indices $1, \ldots, L$ leave the confidence set, the triangle inequality gives

$$U_n(H) \leq d_0(X_{1:n}, H) + c_H \sup_{\ell > L} \varepsilon_\ell.$$

The bound also covers the case where index zero is absent. First increase n, using fixed-horizon convergence, and then increase L. This proves $U_n(H) \to 0$ almost surely. Countably many fixed horizons imply $H_n^* \to \infty$.

Under P_j with $j \geq 1$, choose $L \geq j$ such that $\sup_{\ell > L} \lambda_\ell \leq D_j/4$. Equation (34) gives $P_j/P_0 \geq e^{nD_j/2}$ eventually. For every $\ell > L$, (33) then gives $P_\ell/P_j \leq e^{-nD_j/4}$. Since $M \geq a_j P_j$, this whole tail leaves C_n eventually. Fixed-index exclusion and eventual inclusion of j now give $C_n = \{j\}$. Fixed-horizon convergence proves expansion. Strong merging, when assumed, makes $d_j(X_{1:n}, n) \leq \tau$ eventually and proves the last assertion. □

The obstruction needs fewer assumptions. It uses neither singularity nor likelihood separation. The next statement keeps the all-history positive-cylinder setting, but requires convergence only under P_0 at one fixed horizon.

Theorem 17 (Indistinguishable records and separated futures)**.** *Let $v_k \to \infty$ be integer dates and let $i_k \geq 1$ be candidate indices. Suppose $P_{i_k}|_{\mathcal{F}_{v_k}} = P_0|_{\mathcal{F}_{v_k}}$. Suppose a fixed horizon m, a constant $\delta > 0$, and sets $B_k \subseteq \mathcal{A}^{v_k}$ satisfy $P_0(X_{1:v_k} \in B_k) \to 0$ and*

$$d_{\mathrm{TV}}(P_{i_k,h,m}, P_{0,h,m}) \geq \tau + \delta \qquad (h \notin B_k).$$

If $d_0(X_{1:n}, m) \to 0$ in P_0-probability, every selector with $H_n \to \infty$ in P_0-probability satisfies

$$P_{i_k}\{d_{i_k}(X_{1:v_k}, H_{v_k}) > \tau\} \longrightarrow 1.$$

No such selector has fixed-date safety for an error level $\alpha < 1$.

Proof. Use the event

$$A_k = \{H_{v_k} \geq m,\ d_0(X_{1:v_k}, m) < \delta,\ X_{1:v_k} \notin B_k\}.$$

Its probability under P_0 tends to one. Prefix agreement and the common independent seed give $P_{i_k}(A_k) = P_0(A_k)$. On this event the triangle inequality gives $d_{i_k}(X_{1:v_k}, m) > \tau$. Coherence makes the distance nondecreasing in the forecast horizon, so every approval on A_k fails. This proves the assertion. □

For the lag family, (15), (16), and (17) verify the sufficient hypotheses when $b_\ell \to 0$. Failure of decay gives a subsequence $b_{i_k} \geq \beta > 0$; set $v_k = i_k$ and choose a fixed m with $\Delta_m(\beta) > \tau$. The prefix identity and Lemma 4 then verify the obstruction with no exceptional histories. This explains the exact match in the lag criterion. The next two examples give a different application and a limit of these sufficient assumptions.

A history-sensitive stationary Markov family. Let Q_0 be fair iid bits. For distinct $\beta_\ell \in (0, 1/8]$ tending to zero, let Q_ℓ be the stationary first-order Markov law with

$$Q_\ell(1 \mid 0) = \tfrac{1}{2} + \beta_\ell, \qquad Q_\ell(1 \mid 1) = \tfrac{1}{2} + 2\beta_\ell.$$

Its stationary masses are $\pi_{\ell,1} = (1/2 + \beta_\ell)/(1 - \beta_\ell)$ and $\pi_{\ell,0} = (1/2 - 2\beta_\ell)/(1 - \beta_\ell)$. Positive transitions give ergodicity and positive cylinders. Ergodic transition frequencies distinguish all parameters, so the laws are pairwise singular.

Coupling each next bit with a fair bit by a common uniform variable gives disagreement probability at most $2\beta_\ell$ at each step. A union bound gives

$$d_{\mathrm{TV}}(Q_{\ell,h,H}, Q_{0,h,H}) \leq 2H\beta_\ell$$

for every nonempty history. The bound also covers the empty history, since $|\pi_{\ell,1} - 1/2| = 3\beta_\ell/[2(1 - \beta_\ell)] \leq 2\beta_\ell$. Every transition density ratio to a fair bit is at most $1 + 4\beta_\ell$. The largest initial ratio is $(1 + 2\beta_\ell)/(1 - \beta_\ell) \leq 1 + 4\beta_\ell$. The record ratio satisfies $Q_\ell(h)/Q_0(h) \leq (1 + 4\beta_\ell)^n$, so its log is at most $4n\beta_\ell$.

Under each fixed Q_j, the ergodic theorem applied to log transition ratios gives

$$\frac{1}{n} \log \frac{Q_j(X_{1:n})}{Q_0(X_{1:n})} \longrightarrow \pi_{j,0}\, \mathrm{kl}(1/2 + \beta_j, 1/2) + \pi_{j,1}\, \mathrm{kl}(1/2 + 2\beta_j, 1/2) > 0.$$

The finite initial log ratio vanishes after division by n. Theorem 16 applies with $c_H = 2H$, $\varepsilon_\ell = \beta_\ell$, and $\lambda_\ell = 4\beta_\ell$ to every supplied predictor satisfying its convergence premise. For example, a positive normalized mixture of these laws strongly merges: its posterior true-index weight tends to one by Lemma 5's argument, and every conditional-block distance to the truth is bounded by one minus that weight. This application has a last-state-dependent discrepancy and no exact prefix invisibility. It is a worked application, not a general Markov-estimation result.

Rare discrepancies outside the sufficient conditions. The following nonstationary example shows that the sufficient assumption package as a whole is not necessary for safe expansion; let S_0 be fair iid bits. For each $\ell \geq 1$, let S_ℓ have fair conditional bits except at time $\ell + 1$, when the next-bit probability is $3/4$ if $X_{1:\ell} = 1^\ell$ and $1/2$ otherwise. Every later bit is conditionally fair. These positive-cylinder laws are mutually absolutely continuous with S_0. At date ℓ and history 1^ℓ, their one-step distance from S_0 is $1/4$, so no vanishing all-history closeness bound holds. Singularity and a positive likelihood-rate advantage also fail.

Take $R = S_0$ and approve $H_n = n$ after the first observed zero, with $H_n = 0$ beforehand. After a zero, each candidate's possible modification is already past or can no longer be triggered. Every approval is exact. A zero occurs in finite time almost surely under every candidate, so the rule expands almost surely with zero error. The exceptional all-ones histories at the visibility dates have vanishing probability, which also prevents the obstruction's separation premise. Path-dependent abstention can therefore neutralize rare discrepancies without identifying the law. Wu and Santhanam (2021) use a similar waiting rule in a two-state Markov example to show that their sufficient condition for certified final errors is not necessary outside iid classes. The sufficient theorem here does not cover this mechanism, and the first open problem in Section 6 includes it.

B An explicit predictor with regret and strong merging

The main theorem assumes convergence and asks whether current data can certify a useful horizon. This appendix supplies one predictor satisfying that assumption for every permitted amplitude sequence. Its proof therefore ensures that failure of conditional approval is compatible with strong predictive convergence. The same predictor is held fixed in Corollary 13.

Candidate P_i provides $f_{i,t} = P_i(X_t = 1 \mid \mathcal{F}_{t-1})$ on every finite history. All its conditional probabilities belong to $[3/8, 5/8]$, and every finite history has positive probability. Fix positive weights a_i with $\sum_i a_i = 1$ and a rate $0 < \eta \leq 1/2$. One concrete choice is

$$a_0 = \frac{1}{2}, \qquad a_\ell = \frac{1}{2\ell(\ell+1)} \quad (\ell \geq 1). \tag{35}$$

Define scalar binary Brier loss and the exponential-weight aggregate by

$$\ell(q, x) = (q - x)^2, \qquad L_i(n) = \sum_{t=1}^{n} \ell(f_{i,t}, X_t), \tag{36}$$

$$w_i(n) = \frac{a_i e^{-\eta L_i(n)}}{\sum_j a_j e^{-\eta L_j(n)}}, \qquad r_{n+1} = \sum_i w_i(n) f_{i,n+1}. \tag{37}$$

The forecast r_t uses weights computed after observing $X_{1:t-1}$. The outcome X_t then updates losses and weights. The coherent law R multiplies these forecasts along each possible continuation, updating after each hypothetical outcome. Its future block probabilities differ in general from $\sum_i w_i(n) P_i(\cdot \mid \mathcal{F}_n)$, which freezes today's weights throughout a block. All results below concern chronological R.

Proposition 18 (Regret, current-weight bounds, and strong merging)**.** *The aggregate satisfies, on every observation sequence,*

$$L_R(N) - L_i(N) \leq C_i := \frac{-\log a_i}{\eta} \qquad \textit{for every fixed } i, N, \tag{38}$$

where $L_R(N) = \sum_{t=1}^{N} \ell(r_t, X_t)$. *For history* $h = X_{1:n}$, *let* $P_{i,h,H}$ *and* $R_{h,H}$ *be the conditional laws of the next* H *outcomes, and put*

$$d_i(h, H) = d_{\mathrm{TV}}(P_{i,h,H}, R_{h,H}), \qquad d_i(h, 0) = 0. \tag{39}$$

Put $\epsilon = 3/8$, $c = 2\eta\epsilon(1-\epsilon)$, *and*

$$g_i(h) = \min\left\{1, \sqrt{\frac{-\log w_i(n)}{c}}\right\}. \tag{40}$$

For every finite history h *and every finite* H, $d_i(h, H) \le g_i(h)$. *Under each* P_i, $w_i(n) \to 1$ *almost surely, so*

$$\sup_{H \ge 1} d_i(X_{1:n}, H) \longrightarrow 0 \qquad P_i\text{-almost surely.} \tag{41}$$

Proof. For $x \in \{0, 1\}$ and $q \in [0, 1]$,

$$\frac{d^2}{dq^2} e^{-\eta(q-x)^2} = 2\eta e^{-\eta(q-x)^2}\{2\eta(q-x)^2 - 1\} \le 0.$$

Set $W_n = \sum_i a_i e^{-\eta L_i(n)}$, with $W_0 = 1$. The square loss on $[0, 1]$ is η-exp-concave for $\eta \le 1/2$ (Cesa-Bianchi and Lugosi, 2006), and the aggregate is an instance of Vovk's aggregating algorithm (Vovk, 1990). Jensen's inequality gives

$$e^{-\eta\ell(r_t, X_t)} \ge \sum_i w_i(t-1) e^{-\eta\ell(f_{i,t}, X_t)} = \frac{W_t}{W_{t-1}}.$$

The same inequality holds for countable mixtures by bounded convergence. Multiplication and $W_N \ge a_i e^{-\eta L_i(N)}$ give (38). This calculation uses scalar loss (36); a two-coordinate vector Brier loss is twice this value. The result asserts the displayed bound for this arithmetic aggregate, without asserting that it is the optimal aggregation rule.

Restart the same multiplication at a history h of length n, with initial weights $w_i(n)$. On every continuation through date $N > n$,

$$\sum_{t=n+1}^{N} \{\ell(r_t, X_t) - \ell(f_{i,t}, X_t)\} \le \frac{-\log w_i(n)}{\eta}.$$

Under the conditional law $P_i(\cdot \mid h)$, the proper-score identity is

$$\mathbb{E}_i[\ell(r_t, X_t) - \ell(f_{i,t}, X_t) \mid \mathcal{F}_{t-1}] = (r_t - f_{i,t})^2.$$

Taking conditional expectations and increasing N gives the all-history bound

$$V_i(h) := \mathbb{E}_i\left[\sum_{t>n} (r_t - f_{i,t})^2 \,\middle|\, h\right] \le \frac{-\log w_i(n)}{\eta}.$$

For Bernoulli parameters, $\log u \le u - 1$ gives

$$\mathrm{kl}(p, r) \le \frac{(p-r)^2}{r(1-r)}.$$

Every candidate and R have one-step probabilities in $[\epsilon, 1-\epsilon]$. The conditional relative-entropy chain rule and Pinsker's inequality yield

$$d_i(h, H) \le \min\left\{1, \sqrt{\frac{V_i(h)}{2\epsilon(1-\epsilon)}}\right\} \le g_i(h).$$

It remains to prove concentration of the Brier weights. Fix the truth P_i. For every k, set $T_{ki}(n) = \exp\{-\eta[L_k(n) - L_i(n)]\}$. With $D_t = \ell(f_{k,t}, X_t) - \ell(f_{i,t}, X_t)$ and $d_t = f_{k,t} - f_{i,t}$, the conditional mean of D_t is d_t^2 and its conditional range has length $2|d_t|$. The conditional bounded-variable lemma gives

$$\mathbb{E}_i[e^{-\eta D_t} \mid \mathcal{F}_{t-1}] \le \exp\{-(\eta - \eta^2/2)d_t^2\} \le 1.$$

Each T_{ki} is therefore a nonnegative supermartingale starting at one.

For $k \neq i$, put $m = \max\{i, k\} \ge 1$. After date m, the latest m bits form a finite Markov chain under P_i. Every prescribed sequence of m appended bits has probability at least $(3/8)^m$, so its m-step transition matrix is strictly positive. Its invariant distribution gives positive mass to every state. The two forecast kernels differ on some state: if one index is zero their squared difference is the other index's b^2; otherwise take opposite bits at the two distinct lag positions. The ergodic theorem gives $n^{-1}\sum_{t\le n} d_t^2 \to \delta_{ki} > 0$. The differences $D_t - d_t^2$ are bounded martingale differences. Their averages tend to zero almost surely, by martingale convergence of $\sum_t (D_t - d_t^2)/t$ and Kronecker's lemma. It follows that

$$\frac{L_k(n) - L_i(n)}{n} \longrightarrow \delta_{ki} > 0, \qquad T_{ki}(n) \longrightarrow 0 \quad P_i\text{-almost surely.}$$

For $m \ge i$, conditional monotone convergence makes $S_m(n) = \sum_{k>m} a_k T_{ki}(n)$ a nonnegative supermartingale with initial value $\sum_{k>m} a_k$. Ville's inequality gives

$$P_i\{\sup_n S_m(n) > u\} \le \frac{\sum_{k>m} a_k}{u} \qquad (u > 0).$$

The suprema decrease with m; the displayed bound implies $\sup_n S_m(n) \to 0$ almost surely. Split $\sum_{k\neq i} a_k T_{ki}(n)$ into a finite sum and this uniformly small tail. The finite sum tends to zero, so

$$w_i(n) = \frac{a_i}{a_i + \sum_{k\neq i} a_k T_{ki}(n)} \longrightarrow 1 \quad P_i\text{-almost surely.}$$

The bound (40), uniform in H, now proves (41). □

The regret constant depends on the comparison index. Under (35), it equals $\log(2i(i+1))/\eta$ for $i \ge 1$. A countable pool has no common finite constant of this form. The bound (40) uses established exp-concavity, the proper-score identity, and conditional relative entropy. The next appendix combines that bound with the lag family's confidence and decay calculations to compute safe horizons.

C A computed Brier certificate

Fix the chronological Brier predictor (37) and weights (35). At a history h of length n, abbreviate (40) by $g_i = g_i(h)$. For $i \geq 1$, put

$$k_i(n,H) = \min\{H, (n+H-i)_+\}, \qquad (x)_+ = \max\{x, 0\},$$

$$\Delta_i(n,H) = \min\{1, \sqrt{2k_i(n,H)}\, b_i\}, \qquad \Delta_0(n,H) = 0, \tag{42}$$

$$T_i(n,H) = \min\{g_i, g_0 + \Delta_i(n,H), 1\}. \tag{43}$$

Only $k_i(n,H)$ of the future transitions can be biased. The proof of (15) with that number of transitions gives $d_{\mathrm{TV}}(P_{i,h,H}, P_{0,h,H}) \leq \Delta_i(n,H)$. Proposition 18 and the triangle inequality give $d_i(h,H) \leq T_i(n,H)$ on every history. For $i \geq n+H$, the conditional future law is exactly fair and $d_i(h,H) = d_0(h,H) \leq g_0$.

Let $\varepsilon_n = \min\{\alpha/4, 1/(n+1)\}$. Approximate each distinct likelihood ratio $z_i = P_i(h)/M(h)$ by a rational $\widehat{z}_i$ with error at most ε_n, using one common approximation for all $i \geq n$. Define

$$\widetilde{C}_n = \{i : \widehat{z}_i + \varepsilon_n \geq \alpha\}. \tag{44}$$

Writing $C_n(\beta) = \{i : z_i \geq \beta\}$, these choices ensure

$$C_n(\alpha) \subseteq \widetilde{C}_n \subseteq C_n(\alpha/2). \tag{45}$$

Choose a rational numerical allowance $0 < \xi \leq \tau/4$, with $\xi = \tau/4$ as the default. For each $1 \leq H \leq n$, compute rational upper bounds for $T_i(n,H)$ when $i \in \widetilde{C}_n$ and $i < n+H$, and for g_0 if some $i \geq n+H$ is retained. Each upper bound must lie between its analytic bound and that bound plus ξ. Let $\widetilde{U}_n(H)$ be the maximum of these finitely many upper bounds, set $\widetilde{U}_n(0) = 0$, and define

$$\widehat{H}_n = \max\{0 \leq H \leq n : \widetilde{U}_n(H) \leq \tau\}. \tag{46}$$

An unretained tail contributes no bound. Although all $i \geq n$ have identical observed likelihoods, indices $n \leq i < n+H$ have different future laws and retain their individual bounds.

Theorem 19 (Effective approval and matching rates)**.** *Suppose (b_i) is uniformly computable and η, α, τ, ξ are rational in the ranges above. The rule* (46) *has a finite implementation at every date, satisfies simultaneous safety, and obeys $\widehat{H}_n \leq H_n^*$. If $b_i \to 0$, then $\widehat{H}_n \to \infty$ almost surely under every candidate, and $\widehat{H}_n = n$ eventually almost surely under each fixed P_j with $j \geq 1$. No convergence modulus for (b_i) is needed.*

For $b_i = i^{-\gamma}/8$ with fixed computable $\gamma > 0$, the rule has the same eventual lower bounds (18)–(19) *as the ideal rule, and has almost-sure order $n^{2\gamma}$ for $0 < \gamma < 1/2$ and $n/\log n$ for $\gamma \geq 1/2$, under P_0.*

At date n, evaluating the certificate formulas uses $O(n^2)$ arithmetic operations and $O(n)$ elementary-function evaluations, without evaluating future continuations. This is an algebraic evaluation count; precision control and elementary approximation contribute separate bit costs. If $b_1, \ldots, b_{2n-1}$ are supplied as binary rationals, the implementation's bit cost is polynomial in n and the total input length, including the parameters. Uniform computability alone supplies no uniform time bound. A polynomial-time approximation oracle for the amplitude catalog gives the corresponding polynomial-time implementation for nonrational amplitudes.

Proof. The outer set contains C_n, and every analytic certificate bounds its true conditional distance. The coverage event (13) proves simultaneous safety. Each approved horizon also satisfies $U_n(H) \leq \tau$, which proves $\widehat{H}_n \leq H_n^*$.

Under P_0, each fixed positive index leaves $\widetilde{C}_n$ eventually, by Lemma 5 and $\varepsilon_n \to 0$. Proposition 18 gives $g_0 \to 0$. For each fixed H, the remaining positive indices have $\Delta_i(n,H) \le \sqrt{2H}b_i$, whose supremum tends to zero after exclusion of a growing finite initial set. The extra allowance $\xi \le \tau/4$ leaves every fixed horizon eligible eventually. Under P_j, $j \ge 1$, the proof of Theorem 16, applied at confidence level $\alpha/2$, gives $C_n(\alpha/2) = \{j\}$ eventually. Equation (45) and eventual inclusion of j give $\widetilde{C}_n = \{j\}$ eventually. Since $g_j \to 0$, every horizon through n is then eligible.

For the rate lower bound, retention of ℓ implies $P_\ell(h)/P_0(h) \ge \alpha/4$. The affinity calculation (21) now gives

$$P_0(\ell \in \widetilde{C}_n) \le \frac{2}{\sqrt{\alpha}} e^{-(n-\ell)b_\ell^2/2} \qquad (1 \le \ell < n). \tag{47}$$

For L_n in Proposition 6, the same summable union bound excludes $1, \dots, L_n$ eventually. Also $g_0 \le \tau/4$ eventually. Every retained positive index then has $\Delta_i(n,H) \le \tau/2$ whenever $H \le \min\{n, 8\tau^2(L_n+1)^{2\gamma}\}$. The analytic certificate is at most $3\tau/4$, so its numerical upper bound is at most τ. This proves the unchanged lower bounds. The inequality $\widehat{H}_n \le H_n^*$ and the ideal rule's proved upper bounds give both matching orders.

For computation, write $s_t = 2h_t - 1$. The current scores and likelihoods are

$$L_i(n) = \frac{n}{4} + (n-i)b_i^2 - b_i \sum_{t=i+1}^{n} s_t s_{t-i} \qquad (1 \le i < n),$$

$$P_i(h) = 2^{-n} \prod_{t=i+1}^{n} (1 + 2b_i s_t s_{t-i}) \qquad (1 \le i < n).$$

For $i = 0$ and $i \ge n$, they equal $n/4$ and 2^{-n}. The exact fair-tail mass is $\sum_{i \ge n} a_i = 1/(2n)$. These formulas compute the Brier normalizer and $M(h)$ with $O(n^2)$ arithmetic operations. Current weights for all $i < 2n$ then require $O(n)$ further operations and exponentials or logarithms. Cache $\sqrt{2k}$ for $0 \le k \le n$. Evaluating (42)–(46) for all $H \le n$ uses $O(n^2)$ further arithmetic operations. This calculation includes amplitudes only through b_{2n-1}.

Here are explicit precision bounds for conservative evaluation. First restrict amplitude intervals to $[0, 1/8]$. At date n,

$$W_n \ge a_0 e^{-\eta n}, \qquad M(h) \ge a_0 2^{-n}, \qquad w_i(n) \ge a_i e^{-\eta n} \ge a_i 2^{-n} \quad (i < 2n).$$

Also $a_i \ge 1/(8n^2)$ for these indices, and $W_n \le 1$. Refine rational interval enclosures until each z_i enclosure has width at most $2\varepsilon_n$, and take its midpoint for $\widehat{z}_i$. For weights, restrict the enclosures to the known interval $[2^{-n}/(8n^2), 1]$ before taking $-\log$. Restrict the resulting log enclosures to $[0, \infty)$. On this weight interval, $-\log$ has derivative in absolute value at most $8n^2 2^n$; division by c and $|\sqrt{x} - \sqrt{y}| \le \sqrt{|x-y|}$ show that log errors at most $c(\xi/8)^2$ suffice for square-root errors at most $\xi/8$. Evaluate the amplitude products and the other square roots to the same accuracy. Minima and maxima do not increase the largest absolute error. Refining to width at most ξ for each analytic certificate then gives the required rational upper endpoint. Positive lower bounds for every divisor make this a finite procedure. Membership and approval compare rational numbers at fixed accuracy; neither comparison waits to determine the sign of an exact real equality.

For rational inputs of total bit length B, the likelihood products have polynomial bit length. Each has at most n factors bounded by $5/4$, so perturbing each amplitude by at most δ changes a scaled likelihood by at most $2n(5/4)^{n-1}\delta$. The score formula changes by at most $5n\delta/4$. Together with the displayed divisor, log, and square-root bounds, these estimates require only polynomially many working bits in $n + B$: losses and exponential arguments are $O(n)$, and $\log(1/c)$, $\log(1/\varepsilon_n)$,

and $\log(1/\xi)$ are $O(B+\log n)$. They also bound rounding errors through the polynomial number of operations. One can choose this polynomial precision in advance and evaluate the formulas once. Range reduction and convergent power series for exponentials and logarithms, and rational bisection for square roots, compute their interval enclosures in polynomial time at this precision. The resulting bit cost is polynomial, without asserting $O(n^2)$ bit operations. For a uniformly computable catalog the same refinement terminates, but obtaining its amplitude enclosures can take arbitrarily long. A polynomial-time approximation oracle removes that source of unbounded running time. □

Corollary 20 (A finite-date approval bound)**.** *Suppose b_i decreases with i. Set*

$$q_\ell = e^{-\eta b_\ell^2}\cosh(\eta b_\ell), \qquad F_n = \sum_{\ell=1}^{n-1} a_\ell q_\ell^{n-\ell} + \frac{1}{2n}.$$

Fix $n \geq 2$, $1 \leq L < n$, and $u, v > 0$ with $u + v + \xi \leq \tau$. For every integer $0 \leq h \leq \min\{n, \lfloor v^2/(2b_{L+1}^2)\rfloor\}$, the rule above satisfies

$$P_0(\widehat{H}_n \geq h) \geq 1 - \frac{2}{\sqrt{\alpha}}\sum_{\ell=1}^{L} e^{-(n-\ell)b_\ell^2/2} - \frac{F_n}{a_0(e^{cu^2}-1)}.$$

Proof. For each fixed lag $\ell < n$, the fair match indicators in Proposition 6 are independent. The score formula therefore gives $\mathbb{E}_0 e^{-\eta(L_\ell(n)-L_0(n))} = q_\ell^{n-\ell}$. The terms with $\ell \geq n$ equal one. Tonelli's theorem gives

$$\mathbb{E}_0 S_n = F_n, \qquad S_n = \sum_{\ell \geq 1} a_\ell e^{-\eta(L_\ell(n)-L_0(n))}.$$

Since $w_0 = a_0/(a_0+S_n)$ and $u < 1$, the event $g_0 > u$ equals $\{S_n > a_0(e^{cu^2}-1)\}$. Markov's inequality bounds its probability by the last term. Equation (47) and a union bound control retention of any $1 \leq \ell \leq L$ by the other error term. Outside those two events, every retained positive index has $\Delta_i(n,h) \leq v$, and every relevant analytic certificate is at most $u+v$. Its numerical upper bound is at most $u+v+\xi \leq \tau$, so $\widehat{H}_n \geq h$. The case $h = 0$ follows directly. □

D A second merging predictor

The recognition boundary is separate from a particular Brier update. The ordinary likelihood-mixture law $M = \sum_i a_i P_i$ also satisfies strong merging in this countable truth-containing class; this is the theorem of Blackwell and Dubins (1962), since $P_i \ll M$ with density at most $1/a_i$. For every N, domination gives

$$D_{\mathrm{KL}}(P_i|_{\mathcal{F}_N} \,\|\, M|_{\mathcal{F}_N}) \leq -\log a_i.$$

The chain rule bounds the expected sum of its one-step conditional relative entropies by $-\log a_i$. Let Z be this nonnegative integrable sum. Martingale convergence gives $\mathbb{E}_i[Z \mid \mathcal{F}_n] \to Z$ almost surely; subtracting its observed partial sum shows that its conditional expected tail tends to zero. The future relative-entropy chain rule and Pinsker's inequality then establish (41) with $R = M$. In (12), M remains a testing reference; the computed certificate above concerns chronological Brier forecasts.

Acknowledgments

AI-assisted writing. I designed and directed this paper, made the editorial decisions, and edited and verified the manuscript. OpenAI GPT-6 Astra and GPT-6.1 Sol assisted with drafting and manuscript review. I take full responsibility for the paper.

References

Sandro Gallo and Florencia Leonardi. "Nonparametric Statistical Inference for the Context Tree of a Stationary Ergodic Process." *Electronic Journal of Statistics* 9(2) (2015): 2076–2098. https://doi.org/10.1214/15-EJS1065.

Jean-René Chazottes, Pierre Collet, Christof Külske, and Frank Redig. "Concentration Inequalities for Random Fields via Coupling." *Probability Theory and Related Fields* 137 (2007): 201–225. https://doi.org/10.1007/s00440-006-0026-1.

Narayana Santhanam, Venkatachalam Anantharam, and Wojciech Szpankowski. "Data-Derived Weak Universal Consistency." *Journal of Machine Learning Research* 23(27) (2022): 1–55. https://jmlr.org/papers/v23/20-644.html.

Changlong Wu and Narayana Santhanam. "Prediction with Finitely Many Errors Almost Surely." *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, PMLR 130 (2021): 3223–3231. https://proceedings.mlr.press/v130/wu21c.html.

Daniil Ryabko. "A Criterion for Hypothesis Testing for Stationary Processes." Expanded author report, arXiv:0905.4937v4 (27 December 2014). https://arxiv.org/abs/0905.4937v4.

Meysam Asadi, Ramezan Paravi Torghabeh, and Narayana P. Santhanam. "Stationary and Transition Probabilities in Slow Mixing, Long Memory Markov Processes." *IEEE Transactions on Information Theory* 60(9) (2014): 5682–5701. https://doi.org/10.1109/TIT.2014.2334594.

Boris Ryabko. "Two-Faced Processes and Random Number Generators." Author manuscript, arXiv:1512.06961v1 (22 December 2015). https://arxiv.org/abs/1512.06961v1.

Gusztáv Morvai and Benjamin Weiss. "Limitations on Intermittent Forecasting." *Statistics & Probability Letters* 72(4) (2005): 285–290. https://doi.org/10.1016/j.spl.2004.12.016.

Cosma Rohilla Shalizi and Aryeh Kontorovich. "Predictive PAC Learning and Process Decompositions." *Advances in Neural Information Processing Systems* 26 (2013): 1619–1627. https://papers.nips.cc/paper_files/paper/2013/hash/28267ab848bcf807b2ed53c3a8f8fc8a-Abstract.html.

Jean Ville. *Étude critique de la notion de collectif*. Gauthier-Villars, Paris, 1939.

David Blackwell and Lester Dubins. "Merging of Opinions with Increasing Information." *Annals of Mathematical Statistics* 33(3) (1962): 882–886.

Ehud Kalai and Ehud Lehrer. "Weak and Strong Merging of Opinions." *Journal of Mathematical Economics* 23(1) (1994): 73–86.

Leonid (Aryeh) Kontorovich and Kavita Ramanan. "Concentration Inequalities for Dependent Random Variables via the Martingale Method." *Annals of Probability* 36(6) (2008): 2126–2158.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Vladimir Vovk. "Aggregating Strategies." *Proceedings of the Third Annual Workshop on Computational Learning Theory* (1990): 371–383.