Rate-Optimal Polynomial-Time Prediction of Stationary Renewal Processes

Abstract

Han, Jiang, and Wu [11] established the minimax $\Theta(n^{-1/2})$ one-step Kullback–Leibler (KL) prediction rate for stationary binary renewal processes and left polynomial-time attainment open. We resolve this computational question under only a finite-mean gap law, without a known uniform mean bound, higher-moment condition, or mixing assumption. Our exact multiset dynamic program evaluates a queried prefix continuation mass of a renewal likelihood envelope in $O(N^4)$ integer operations. With $N = n+1$, the full batch required by a compression-to-prediction reduction takes $O(N^5)$ integer operations and polynomial bit time, yielding the optimal $O(n^{-1/2})$ risk averaged over stationary histories. A computable mixture over occupied gap types retains this worst-case rate and adapts to simpler gap laws: it achieves $O(s \log n/n)$ for s-point support, $O(\log^2 n/n)$ for exponential tails, and $O(n^{-(1-1/\alpha)} \log n)$ when $\mu(d) \leq Cd^{-\alpha}$ with $\alpha > 2$. The block assignment also gives $O(\sqrt{N})$ pathwise log-loss regret at a known horizon; dyadic restarting yields a horizon-free polynomial-time predictor with minimax-order $O(\sqrt{N})$ cumulative KL risk. A matching fixed-time lower bound holds even for aperiodic finite-support laws satisfying $T \leq 4\mathbb{E}T$. Finally, we distinguish the genuine stationary conditional on an all-zero history from the ordinary post-renewal hazard, whose substitution changes the integrated minimax risk to the nonvanishing value $\log 2$.

Note. This paper was generated entirely by AI, including MiMo, using an automated research pipeline developed by Chenghua Liu and Hanyu Li.

1 Introduction

Renewal processes are simple enough to expose the computational content of prediction, yet their unbounded age and countably supported gap law make them genuine infinite-memory models. We ask whether their optimal fixed-time prediction rate can be attained by one exact polynomial-time algorithm under only a pointwise finite-mean assumption.

We study excess one-step logarithmic loss. For the true stationary conditional $p_\mu(X^n)$ and a prediction $q(X^n)$, this excess is the stationary-history average of $D_{\mathrm{KL}}(\mathrm{Bern}(p_\mu(X^n)) \| \mathrm{Bern}(q(X^n)))$. The link between log-loss prediction and universal compression is classical [6, 13, 22]. For stationary renewal processes, Han et al. [11] proved the minimax rate $\Theta(n^{-1/2})$ and left polynomial-time attainment open. Classical renewal coding already gives $\Theta(\sqrt{N})$ block redundancy [5, 7]; the missing step is to evaluate a queried prefix continuation mass without enumerating exponentially many future renewal patterns.

Contributions. This paper gives a constructive, rate-optimal answer.

- An exact multiset recurrence evaluates one arbitrary queried prefix continuation mass in $O(N^4)$ integer operations. A specified batch of $O(N)$ queries, including the suffix batch used for fixed-time prediction, takes $O(N^5)$ operations on polynomial-bit integers. The resulting predictor answers the polynomial-time question of Han et al. [11].

- We refine the assignment by mixing over the number of distinct completed gaps. The resulting data-dependent pointwise regret is the minimum of the worst-case $O(\sqrt{N})$ term and an $O((V_N+1)\log N)$ term, where V_N is the observed gap-type count. This gives distribution-adaptive fixed-time rates without knowing a support size or tail exponent.

- The block assignment also gives pathwise cumulative log-loss regret. Restarting it on dyadic epochs yields a single horizon-free predictor with $O(\sqrt{N}+\log^2 N)=O(\sqrt{N})$ cumulative KL risk and a rate-matching lower bound.

- We strengthen the fixed-time lower bound while retaining both stationary censoring factors. The hard laws are uniform on $\Theta(\sqrt{n})$ points, aperiodic, and satisfy $T \le 4\mathbb{E}T$ almost surely. Thus the worst-case rate is not caused by heavy tails or extreme standardized moments.

We also separate a boundary issue from this computational question. The formal risk of Han et al. [11] uses the genuine stationary conditional, as do our main results. The informal all-zero use of their Eq. (5), p. 4, instead substitutes the ordinary post-renewal hazard for the equilibrium residual hazard. This observation does not concern their formal minimax risk or Theorem 1.4. If the substituted hazard is made the target, its integrated minimax KL risk is exactly $\log 2$.

Section 2 defines the stationary model and criteria. Sections 3 and 4 give the construction, algorithm, and statistical guarantees, followed by related work and limitations in Section 5. The technical appendices after the references give the boundary propositions, block assignment, exact query recurrence, compression bridge, adaptive analysis, lower bounds, anytime construction, and a bounded-support comparison.

2 Model, boundary condition, and main result

This section defines the stationary (equilibrium) renewal model, distinguishes the genuine all-zero conditional from an auxiliary hazard substitution, and states the fixed-time minimax result. It also fixes the probability and computational quantifiers used below.

Let T be positive-integer-valued with law μ and finite mean

$$m_\mu := \mathbb{E}_\mu T < \infty.$$

Write

$$\overline{F}_\mu(a) := \mathbb{P}_\mu(T > a), \qquad a \in \{0, 1, \ldots\}.$$

A stationary binary renewal process is obtained by taking an iid sequence $T_1, T_2, \ldots \sim \mu$ and an initial wait T_0 independent of that sequence, and setting $X_t = 1$ exactly at the times $T_0, T_0+T_1, T_0+T_1+T_2, \ldots$. Stationarity is equivalent to the equilibrium-delay law

$$\mathbb{P}_\mu(T_0 = t) = \frac{\mathbb{P}_\mu(T \ge t)}{m_\mu} = \frac{\overline{F}_\mu(t-1)}{m_\mu}, \qquad t \ge 1. \tag{2.1}$$

Let P_μ denote the induced process law, P_μ^N its marginal on X_1^N, and $P_\mu(x^N) := P_\mu^N(\{x^N\})$ its block pmf. For $1 \le i \le j$, write $X_i^j = (X_i, \ldots, X_j)$, and similarly for x_i^j; write $X^j = X_1^j$, and interpret a slice with $i > j$ as the empty string. All logarithms are natural. Products indexed by gap lengths are over positive counts unless stated otherwise. We use $0 \log(0/q) = 0$ and $p \log(p/0) = +\infty$ for $p > 0$. For $p, q \in [0,1]$, abbreviate

$$d(p\|q) := D_{\mathrm{KL}}(\mathrm{Bern}(p), \mathrm{Bern}(q)).$$

If a renewal has been observed and the current age is a, then every positive-probability such history has $\overline{F}_\mu(a) > 0$, and its true next-step probability is the ordinary hazard

$$h_\mu(a) := \frac{\mu(a+1)}{\overline{F}_\mu(a)}. \tag{2.2}$$

The all-zero history has a different boundary law. Put

$$R_n := \sum_{a \geq n} \overline{F}_\mu(a) = \mathbb{E}_\mu[(T-n)_+].$$

Whenever $\mathbb{P}_\mu(X^n = 0^n) > 0$, equivalently $R_n > 0$, (2.1) gives

$$\mathbb{P}_\mu(X_{n+1} = 1 \mid X^n = 0^n) = \mathbb{P}_\mu(T_0 = n+1 \mid T_0 > n) = \frac{\overline{F}_\mu(n)}{R_n}, \tag{2.3}$$

which is generally not $h_\mu(n) = \mu(n+1)/\overline{F}_\mu(n)$ when the latter ratio is defined. Conditional probabilities are defined on P_μ-positive histories; their values on null histories are arbitrary and do not affect the integrated risks below. Thus, on every positive-probability history,

$$p_\mu(x^n) := P_\mu(X_{n+1} = 1 \mid X^n = x^n) = \begin{cases} h_\mu(n - \max\{1 \leq t \leq n : x_t = 1\}), & x^n \neq 0^n, \\ \overline{F}_\mu(n)/R_n, & x^n = 0^n. \end{cases} \tag{2.4}$$

A possibly randomized statistical predictor is a probability kernel from $x^n \in \{0,1\}^n$ to an output in $[0,1]$. Its stationary conditional KL risk is

$$\mathcal{R}_n(\widehat{q}_n, \mu) := \mathbb{E}_\mu D_{\mathrm{KL}}(P_\mu(X_{n+1} \in \cdot \mid X^n), \mathrm{Bern}(\widehat{q}_n(X^n))), \tag{2.5}$$

where the expectation includes private randomness. Let $\mathfrak{R}_n$ be the infimum of $\sup_{\mu : m_\mu < \infty} \mathcal{R}_n$ over these kernels. A deterministic family $\widehat{q} = (\widehat{q}_n)_{n \geq 1}$ is *uniform polynomial-bit-time* if one Turing machine, given (n, x^n), outputs an exact rational encoded as a binary integer pair (a, b), with $0 \leq a \leq b$ and $b > 0$, using $\mathrm{poly}(n)$ bit operations. The pair need not be reduced. Requiring one machine for all horizons rules out fixed-horizon hard-coding.

For the auxiliary formulation, let

$$\tau_n := n - \max\{1 \leq t \leq n : X_t = 1\}$$

when a renewal is visible, and set $\tau_n = n$ on 0^n. The *ordinary-hazard substitution* uses $\widetilde{p}_\mu(x^n) := h_\mu(\tau_n)$ on every history, including 0^n, whenever the denominator in (2.2) is positive; on a null history where it vanishes, define $\widetilde{p}_\mu$ arbitrarily. Denote the corresponding minimax integrated risk by

$$\mathfrak{R}_n^{\mathrm{haz}} := \inf_{\widehat{q}_n} \sup_{\mu : m_\mu < \infty} \mathbb{E}_\mu d(\widetilde{p}_\mu(X^n) \| \widehat{q}_n(X^n)). \tag{2.6}$$

This auxiliary target formalizes the informal all-zero substitution; it is not the formal conditional risk of Han et al. [11].

Question 2.1 (Following Han–Jiang–Wu (paraphrased))**.** Can a polynomial-time predictor attain the information-theoretically optimal $\Theta(n^{-1/2})$ fixed-time KL risk for every stationary renewal process under only the pointwise condition $m_\mu < \infty$?

The efficient context-tree route discussed in this literature incurs an additional logarithmic factor; see Garivier [8] for the corresponding pointwise redundancy bound.

For later use, set

$$c_{\mathrm{p}} := \pi\sqrt{\frac{2}{3}}, \qquad B_N := c_{\mathrm{p}}\sqrt{N} + \log(N+1) + \log\left(1 + \frac{2\sqrt{N}}{c_{\mathrm{p}}}\right). \tag{2.7}$$

The boundary substitution and the genuine conditional loss lead to different answers.

Theorem 2.2 (Fixed-time minimax rate and polynomial-time attainment)**.** *For every integer $n \geq 1$, even with randomized and computationally unrestricted predictors,*

$$\mathfrak{R}_n^{\mathrm{haz}} = \log 2. \tag{2.8}$$

For the stationary conditional KL risk, a single deterministic uniform polynomial-bit-time family $\widehat{q}$ satisfies, for every $n \geq 1$,

$$\mathfrak{R}_n \leq \sup_{\mu : m_\mu < \infty} \mathcal{R}_n(\widehat{q}_n, \mu) \leq \frac{B_{n+1} + \log(n+2)}{n} \leq \left(\frac{2\pi}{\sqrt{3}} + 3\sqrt{2}\right)\frac{1}{\sqrt{n}}. \tag{2.9}$$

Conversely, for every $n \geq 600$,

$$\mathfrak{R}_n \geq \frac{11(\log 2 - 1/2)}{40600\sqrt{n}}. \tag{2.10}$$

Thus the fixed-time minimax rate is $\Theta(n^{-1/2})$, and one uniform polynomial-bit-time algorithm attains it.

Proof architecture. A completed-gap likelihood envelope has normalization $\exp(O(\sqrt{N}))$. A first-renewal mixture converts it into a stationary block assignment, and a multiset recurrence computes each queried prefix mass exactly. The compression-to-prediction bridge then yields the upper bound. For the lower bound, a random-support prior leads from a visible good event to posterior membership uncertainty and then to a Bernoulli KL radius. Separate two-point laws establish the exact value for the auxiliary hazard substitution.

3 Construction and exact query algorithm

We first build a block assignment with $O(\sqrt{N})$ pointwise redundancy, then describe the multiset dynamic program that evaluates its queried prefix masses. The construction separates the two stationary boundaries from the ordinary renewal segment between them.

Suppose a renewal is known at time zero and $y \in \{0,1\}^L$ is observed. Let c_d count its completed gaps of length d, let $K = \sum_d c_d$, and let the final zero run be censored. Define

$$M_L(y) = \begin{cases} 1, & K = 0, \\ \displaystyle\prod_{d : c_d > 0} (c_d/K)^{c_d}, & K > 0, \end{cases} \qquad Z_L = \sum_y M_L(y), \qquad q_L = M_L/Z_L,$$

with $M_0(\varnothing) = Z_0 = q_0(\varnothing) = 1$. The ordinary-renewal likelihood is at most M_L, while an integer-partition argument gives

$$Z_L \leq \left(1 + \frac{2\sqrt{L}}{c_{\mathrm{p}}}\right) e^{c_{\mathrm{p}}\sqrt{L}}, \qquad c_{\mathrm{p}} = \pi\sqrt{2/3}.$$

Table 1: Multiset-DP states for one queried continuation mass.

State	Invariant
(k, u)	Number and total length of future completed gaps using lengths at most d.
$G_d(k, u)$	Weighted mass of all such future multisets, including their orderings.
$S_d(k, u)$	The same mass, marked by the number of gap copies whose length exceeds the current censored age A.

To handle the left boundary, Q_N mixes uniformly over the first visible renewal $J = 1, \dots, N$ and a no-renewal component. Given J, it emits $0^{J-1}1$ followed by q_{N-J}; the last component emits 0^N. Thus $J = N$ uses $q_0(\varnothing) = 1$, and Q_N has full support. For every finite-mean μ and every P_μ-positive word,

$$\log \frac{P_\mu(x^N)}{Q_N(x^N)} \le B_N := c_{\mathrm{p}}\sqrt{N} + \log(N+1) + \log\left(1 + \frac{2\sqrt{N}}{c_{\mathrm{p}}}\right). \tag{3.1}$$

It remains to answer prefix queries without enumerating extensions. For a prefix $z \in \{0,1\}^\ell$ of the ordinary block, let a_d count its completed gaps, set $r = \sum_d a_d$, $s_\star = \sum_d d a_d$, $A = \ell - s_\star$, and $R = L - \ell$. The continuation mass is $W_L(z) = \sum_v M_L(zv)$. The two DP aggregates have the following invariants; $a = (a_1, \dots, a_L)$ and A are fixed during a run.

For $0 \le d \le L$ and $0 \le k \le u \le A + R$, put $G_0(0,0) = 1$, all other $G_0 = 0$, and all $S_0 = 0$; states outside this domain are zero. With $B_d(b;k) = \binom{k}{b}(a_d + b)^{a_d + b}$, the transitions are

$$\begin{aligned} G_d(k,u) &= \sum_{b=0}^{\min\{k, \lfloor u/d \rfloor\}} B_d(b;k) G_{d-1}(k-b, u-db), \\ S_d(k,u) &= \sum_{b=0}^{\min\{k, \lfloor u/d \rfloor\}} B_d(b;k) S_{d-1}(k-b, u-db) \\ &\quad + \sum_{b=0}^{\min\{k, \lfloor u/d \rfloor\}} B_d(b;k) \mathbf{1}\{d > A\} b G_{d-1}(k-b, u-db). \end{aligned}$$

Writing M_{nf} for the unique extension with no future renewal, the multiset-ordering identity in Appendix C gives

$$W_L(z) = M_{\mathrm{nf}} + \sum_{k=1}^{A+R} \sum_{u=k}^{A+R} \frac{S_L(k,u)}{k(r+k)^{r+k}}.$$

For $k \ge 1$, $S_L(k,u)/k$ is an integer: it counts ordered future gap sequences after marking the first gap copy that can exceed A. If the first one of a queried Q_N-prefix y^t is at J, set $L = N - J$ and $z = y_{J+1}^t$ ($z = \varnothing$ when $J = t$); then

$$Q_N(X_{t+1} = 1 \mid X^t = y^t) = \frac{W_L(z1)}{W_L(z0) + W_L(z1)}.$$

An all-zero prefix has conditional $1/(N - t + 1)$.

The interface is deliberately query based. One arbitrary queried conditional takes $O(N^4)$ integer operations; any specified $O(N)$-query batch takes $O(N^5)$. This covers both the suffix-prefix batch used by the fixed-time predictor and the forward on-path batch used to evaluate regret on a supplied block; it does not enumerate an exponential prefix tree. Pascal recurrences, exact integer exponentiation, and exact division suffice. Operands have polynomial length, rolling workspace is $O(N^3 \log N)$ bits, and a safe schoolbook bound for the compression-averaged predictor is $O(N^{11} \log^2 N)$ bit operations. Complete invariants, divisibility, and operand-growth details appear in Appendix C.

4 Statistical guarantees

We next connect block regret to one-step prediction, refine the assignment to the observed occupied-gap complexity, and state the lower and online results. The key bridge is conditional information carried by the unobserved part of a stationary past.

For $N = n + 1$, define

$$\widehat{q}_n(x^n) = \frac{1}{n} \sum_{t=1}^{n} Q_N\bigl(X_{t+1} = 1 \mid X_1^t = x_{n-t+1}^n\bigr). \tag{4.1}$$

For any stationary binary process law P, conditional KL decomposition and convexity yield

$$\mathcal{R}_n(\widehat{q}_n, P) \le \frac{1}{n} D_{\mathrm{KL}}(P^N, Q_N) + \frac{1}{n} \sum_{t=1}^{n} I_P(X_{n+1}; X_1^{n-t} \mid X_{n-t+1}^n).$$

For a renewal process, let C_i be the inclusive countdown from position i to the next renewal. If the first renewal lies inside the suffix, regeneration removes the earlier past; if it lies beyond the suffix, the suffix is all zero and the next bit is determined by the countdown. Hence $X_{n+1} \perp X_1^{n-t} \mid (X_{n-t+1}^n, C_{n-t+1})$. Conditional data processing, stationarity, and the mutual-information chain rule give

$$\sum_{t=1}^{n} I_P(X_{n+1}; X_1^{n-t} \mid X_{n-t+1}^n) \le I_P(C_1; X_1^{n+1}) \le \log(n+2).$$

Together with (3.1), this proves the upper half of Theorem 2.2; Appendix D gives the complete conditional-law argument.

For adaptation, let $V_N(x^N)$ be the number of distinct completed gap lengths after the first visible renewal, and define the population proxy

$$v_N(\mu) = \sum_{d=1}^{N-1} \min\biggl\{1, \frac{(N-d)\mu(d)}{m_\mu}\biggr\}. \tag{4.2}$$

This quantity controls expected occupied gap types in a stationary block; it is neither support cardinality nor parameter dimension. Restricting the envelope to at most s occupied lengths and mixing over $s = 1, \ldots, N$ gives an exactly computable full-support assignment Q_N^{ad}. A single j-refined table gives every restriction level s by prefix sums in j; the DP is not rerun for each s. A specified $O(N)$-query batch takes $O(N^6)$ integer operations, polynomial bit time, and $O(N^4 \log N)$ rolling bit space. Its fixed-time predictor satisfies

$$\mathcal{R}_n(\widehat{q}_n^{\mathrm{ad}}, \mu) \le \frac{1}{n}\Bigl[\log 2 + \log(n+2) + \min\{B_{n+1}, (2v_{n+1}(\mu) + 6)\log(n+2)\}\Bigr].$$

Consequently, while never exceeding the worst-case order, the risk is

- $O(s \log n/n)$ for s-point laws;
- $O(\log^2 n/n)$ if $\mu(d) \le Ce^{-cd}$; and

- $O(n^{-(1-1/\alpha)} \log n)$ if $\mu(d) \le Cd^{-\alpha}$ with $\alpha > 2$.

Appendix E proves the pointwise oracle, refined-DP projection identities, and Jensen step. The logarithmic finite-support endpoint is necessary even for an unknown deterministic period; see Proposition E.6.

The fixed-time lower bound uses a random support of size $a = 24\lceil\sqrt{n}\rceil$ in $[2a]$. Every hard law is aperiodic and satisfies $T \le 4m_\mu$. The proof follows five explicit steps:

$$\begin{aligned}\text{random support} &\longrightarrow \text{visible completed gaps} \longrightarrow \mathcal{E}\\ &\longrightarrow \text{posterior membership uncertainty} \longrightarrow \text{Bernoulli KL radius.}\end{aligned}$$

Both stationary boundary factors are retained in the exact likelihood. On a good event of probability at least 1/8, membership and nonmembership of the candidate next gap each have posterior mass bounded below by a universal constant, which gives the lower half of Theorem 2.2. The posterior stability lemma and its constants are proved in Appendix F.

Finally, let $\pi = (\pi_t)_{t\ge0}$ be a sequential predictor, with $\pi_0(\varnothing) = Q_1(X_1 = 1) = 1/2$, and define

$$\mathcal{C}_N(\pi,\mu) = \sum_{t=0}^{N-1} \mathbb{E}_\mu d\big(P_\mu(X_{t+1} = 1 \mid X^t) \| \pi_t(X^t)\big)\,.$$

At a known horizon, the on-path conditionals of Q_N have pathwise log-loss regret at most B_N. Restarting on dyadic epochs gives one horizon-free uniform polynomial-bit-time predictor π^{any} with

$$\sup_{\mu: m_\mu<\infty} \mathcal{C}_N(\pi^{\mathrm{any}},\mu) \le c_{\mathrm{p}}(2+\sqrt{2})\sqrt{N} + O(\log^2 N).$$

A separate occupancy prior gives the matching-order $\Omega(\sqrt{N})$ lower bound, including randomized horizon-aware predictors. Thus the anytime minimax cumulative KL risk is $\Theta(\sqrt{N})$ as $N \to \infty$. Appendix G supplies the dependence and randomization arguments.

5 Related work, scope, and limitations

This section positions the contribution against renewal coding and efficient prediction, then records what the results do and do not claim. The distinction between existence, exact query computation, and practical scalability is important here.

5.1 Coding, prediction, and renewal models

The coding–prediction connection goes back at least to the universal coding program of Rissanen [22] and to universal sequential prediction for individual sequences [6, 13]. Normalized maximum likelihood and context-tree weighting are canonical probability assignments for pointwise log loss [23, 25]. More recently, Han et al. [10] developed fixed-time KL prediction for finite-order Markov chains, and Han et al. [11] treated infinite-memory hidden Markov and renewal models. The latter work proves the $\Theta(n^{-1/2})$ renewal rate and explicitly identifies efficient attainment as open.

The underlying $\Theta(\sqrt{N})$ redundancy of renewal processes is classical [5]. The completed-gap multinomial envelope and its reduction to integer partitions are explicit in the analytic coding work of Flajolet and Szpankowski [7]; those ingredients are not new. Garivier's context-tree weighting construction is efficient but has $\Theta(\sqrt{N}\log N)$ pointwise redundancy on renewal sources [8]. Its logarithm comes from the tradeoff between a finite-memory approximation and context estimation, rather than from a known computational barrier. Enumerative coding [4] and exact multinomial stochastic-complexity recurrences [15] provide nearby algorithmic precedents. The step missing for the rate-optimal renewal assignment is an exact computation of arbitrary prefix continuation masses, including the stationary left boundary; our multiset recurrence supplies it. To our knowledge, the cited methods do not provide the same combination of a uniform finite-time stationary conditional-KL guarantee, the minimax $n^{-1/2}$ rate, and exact polynomial-bit-time queried conditionals.

The adaptive result is related in spirit to universal coding over unknown or countably infinite alphabets [2, 3, 21]. Distinct-symbol counts and expected occupancy are classical complexity measures in that literature [1, 14]. Here the observed objects are not an iid sample of gaps: the initial and terminal gaps are censored, and which internal gaps are completed depends on their lengths. Our occupied-gap complexity is therefore a renewal-weighted occupancy count, and the new claim is a stationary fixed-time KL oracle inequality whose restricted assignments retain exact polynomial-time prefix marginals.

There is also a substantial literature on estimating interarrival or residual waiting-time laws under length-biased and censored renewal sampling [9, 24]. Morvai and Weiss [16, 17, 18, 19] establish several asymptotic, intermittent, or residual-time guarantees for binary renewal and more general stationary processes. The random-context method of Oliveira [20] is adaptive and efficient, and for renewal processes gives rates under a uniformly bounded $(2+\gamma)$-moment condition. These results target different losses or asymptotic modes of convergence. To our knowledge, they do not provide the present uniform finite-time KL guarantee over the entire finite-mean class.

5.2 Scope and limitations

The upper bounds assume only pointwise finite mean; they do not require a known uniform mean bound, a higher moment, tail decay, mixing time, support bound, or lower-hazard condition. The guarantee is averaged over stationary histories. A uniform guarantee over every positive-probability history is impossible, and the auxiliary ordinary-hazard substitution on an all-zero block has minimax value $\log 2$. The polynomial exponents are conservative Turing-model existence bounds, not claims of practical scalability or computational lower bounds. Faster exact convolution and reuse across overlapping suffixes remain open algorithmic directions. For comparison, Appendix H gives a direct add-one hazard estimator under a known support bound $B \le n/2$, with risk at most $2B^2/(m_\mu n)$.

The construction therefore answers a computational question rather than an information-theoretic one: rate-optimal renewal assignments were known, but their queried conditionals were not known to be polynomial-time computable in this setting. Dyadic restarting preserves the $\sqrt{N}$ order with a constant-factor increase in the leading redundancy term and an additional lower-order dependence term. Extensions to projectively consistent adaptive assignments and Markov-renewal models are natural next steps.

References

[1] A. Ben-Hamou, S. Boucheron, and M. I. Ohannessian. Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications. *Bernoulli*, 23(1):249–287, 2017. doi:10.3150/15-BEJ743.

[2] D. Bontemps, S. Boucheron, and E. Gassiat. About adaptive coding on countable alphabets. *IEEE Transactions on Information Theory*, 60(2):808–821, 2014. doi:10.1109/TIT.2013.2288914.

[3] S. Boucheron, A. Garivier, and E. Gassiat. Coding on countably infinite alphabets. *IEEE Transactions on Information Theory*, 55(1):358–373, 2009. doi:10.1109/TIT.2008.2008150.

[4] T. M. Cover. Enumerative source encoding. *IEEE Transactions on Information Theory*, 19(1):73–77, 1973. doi:10.1109/TIT.1973.1054929.

[5] I. Csiszár and P. C. Shields. Redundancy rates for renewal and other processes. *IEEE Transactions on Information Theory*, 42(6):2065–2072, 1996. doi:10.1109/18.556596.

[6] M. Feder, N. Merhav, and M. Gutman. Universal prediction of individual sequences. *IEEE Transactions on Information Theory*, 38(4):1258–1270, 1992. doi:10.1109/18.144706.

[7] P. Flajolet and W. Szpankowski. Analytic variations on redundancy rates of renewal processes. *IEEE Transactions on Information Theory*, 48(11):2911–2921, 2002. doi:10.1109/TIT.2002.804115.

[8] A. Garivier. Redundancy of the context-tree weighting method on renewal and Markov renewal processes. *IEEE Transactions on Information Theory*, 52(12):5579–5586, 2006. doi:10.1109/TIT.2006.885484.

[9] R. D. Gill and N. Keiding. Product-limit estimators of the gap time distribution of a renewal process under different sampling patterns. *Lifetime Data Analysis*, 16(4):571–579, 2010. doi:10.1007/s10985-010-9156-y.

[10] Y. Han, S. Jana, and Y. Wu. Optimal prediction of Markov chains with and without spectral gap. *IEEE Transactions on Information Theory*, 69(6):3920–3959, 2023. doi:10.1109/TIT.2023.3239508.

[11] Y. Han, T. Jiang, and Y. Wu. Prediction from compression for models with infinite memory, with applications to hidden Markov and renewal processes. In *Proceedings of the 37th Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 2270–2307, 2024. PMLR:han24a.

[12] G. H. Hardy and S. Ramanujan. Asymptotic formulae in combinatory analysis. *Proceedings of the London Mathematical Society*, s2-17(1):75–115, 1918. doi:10.1112/plms/s2-17.1.75.

[13] D. Haussler, J. Kivinen, and M. K. Warmuth. Sequential prediction of individual sequences under general loss functions. *IEEE Transactions on Information Theory*, 44(5):1906–1925, 1998. doi:10.1109/18.705569.

[14] S. Karlin. Central limit theorems for certain infinite urn schemes. *Indiana University Mathematics Journal*, 17(4):373–401, 1967. doi:10.1512/iumj.1968.17.17020.

[15] P. Kontkanen and P. Myllymäki. A linear-time algorithm for computing the multinomial stochastic complexity. *Information Processing Letters*, 103(6):227–233, 2007. doi:10.1016/j.ipl.2007.04.003.

[16] G. Morvai and B. Weiss. On universal estimates for binary renewal processes. *The Annals of Applied Probability*, 18(5):1970–1992, 2008. doi:10.1214/07-AAP512.

[17] G. Morvai and B. Weiss. Inferring the residual waiting time for binary stationary time series. *Kybernetika*, 50(6):869–882, 2014. doi:10.14736/kyb-2014-6-0869.

[18] G. Morvai and B. Weiss. A versatile scheme for predicting renewal times. *Kybernetika*, 52(3):348–358, 2016. doi:10.14736/kyb-2016-3-0348.

[19] G. Morvai and B. Weiss. On universal algorithms for classifying and predicting stationary processes. *Probability Surveys*, 18:77–131, 2021. doi:10.1214/20-PS345.

[20] R. I. Oliveira. Stochastic processes with random contexts: A characterization and adaptive estimators for the transition probabilities. *IEEE Transactions on Information Theory*, 61(12):6910–6925, 2015. doi:10.1109/TIT.2015.2496200.

[21] A. Orlitsky, N. P. Santhanam, and J. Zhang. Universal compression of memoryless sources over unknown alphabets. *IEEE Transactions on Information Theory*, 50(7):1469–1481, 2004. doi:10.1109/TIT.2004.830761.

[22] J. Rissanen. Universal coding, information, prediction, and estimation. *IEEE Transactions on Information Theory*, 30(4):629–636, 1984. doi:10.1109/TIT.1984.1056936.

[23] Y. M. Shtarkov. Universal sequential coding of single messages. *Problems of Information Transmission*, 23(3):175–186, 1987.

[24] Y. Vardi. Nonparametric estimation in the presence of length bias. *The Annals of Statistics*, 10(2):616–620, 1982. doi:10.1214/aos/1176345802.

[25] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens. The context-tree weighting method: Basic properties. *IEEE Transactions on Information Theory*, 41(3):653–664, 1995. doi:10.1109/18.382012.

A The auxiliary ordinary-hazard substitution at the boundary

This appendix computes the minimax risk of the auxiliary all-zero hazard substitution and then contrasts it with the genuine history-wise criterion. Neither proposition changes the formal stationary integrated risk used in Theorem 2.2.

Proposition A.1 (Auxiliary integrated boundary value)**.** *For every $n \geq 1$, the ordinary-hazard substitution satisfies $\mathfrak{R}_n^{\mathrm{haz}} = \log 2$.*

Proof. The constant predictor $\widehat{q}_n \equiv 1/2$ gives the upper bound because $d(p\|1/2) = \log 2 - H_2(p) \leq \log 2$, where $H_2(p) = -p \log p - (1-p)\log(1-p)$.

For the converse, fix $\delta \in (0,1)$ and $M \in \mathbb{Z}_{>n+1}$. Let δ_M denote the point mass at M, and consider

$$\mu_0 = \delta_M, \qquad \mu_1 = (1-\delta)\delta_{n+1} + \delta\delta_M.$$

On the common history $E = \{X^n = 0^n\}$, the substituted targets are

$$h_{\mu_0}(n) = 0, \qquad h_{\mu_1}(n) = 1 - \delta =: p.$$

The equilibrium-delay formula gives

$$\mathbb{P}_{\mu_0}(E) = \frac{M-n}{M}, \qquad \mathbb{P}_{\mu_1}(E) = \frac{(1-\delta) + \delta(M-n)}{(1-\delta)(n+1) + \delta M}. \tag{A.1}$$

For fixed δ, both probabilities tend to one as $M \to \infty$.

Let Q be the predictor's possibly random output on the deterministic input 0^n; its law is the same under the two models. If ρ_M is the smaller probability in (A.1), then the larger of the two risks is at least

$$\rho_M \max\{\mathbb{E}d(0\|Q), \mathbb{E}d(p\|Q)\} \geq \rho_M \mathbb{E}\frac{d(0\|Q) + d(p\|Q)}{2}$$
$$\geq \rho_M \inf_{q\in[0,1]} \frac{d(0\|q) + d(p\|q)}{2}.$$

The part of the last objective that depends on q is

$$-\frac{p}{2}\log q - \frac{2-p}{2}\log(1-q).$$

Its derivative vanishes only at $q = p/2$, and strict convexity on $(0,1)$ makes this the unique minimizer. The resulting value is the binary Jensen–Shannon divergence

$$H_2(p/2) - \frac{1}{2}H_2(p). \tag{A.2}$$

First let $M \to \infty$, and then let $\delta \downarrow 0$. The expression in (A.2) tends to $H_2(1/2) = \log 2$. Since the supremum over finite-mean gap laws dominates every member of this family, the lower bound follows. The argument is pointwise in the realized output, so private randomization cannot improve the value. □

The proposition concerns only the substituted target. For example, the genuine conditionals on E above are $1/(M-n)$ and $\overline{F}_{\mu_1}(n)/R_n$, respectively, and stationary probability downweights difficult all-zero blocks. The following complementary fact shows why that integration over histories cannot be removed.

Proposition A.2 (Genuine conditional under a history-wise criterion)**.** *Let*

$$\mathfrak{R}_n^{\mathrm{pw}} := \inf_{\widehat{q}_n} \sup_{\mu: m_\mu<\infty} \sup_{x^n: P_\mu(x^n)>0} \mathbb{E}d(p_\mu(x^n)\|\widehat{q}_n(x^n)).$$

Then $\mathfrak{R}_n^{\mathrm{pw}} = \log 2$ *for every* $n \geq 1$.

Proof. Again $1/2$ gives the upper bound. On 0^n, the genuine conditional for $\mu = \delta_{n+1}$ is one, whereas for $\mu = \delta_M$ it is $1/(M-n)$. Both histories have positive probability. Letting $M \to \infty$ reduces the lower bound to the minimax KL radius of the two Bernoulli endpoints, which is $\log 2$, also for randomized outputs by the same averaging argument as in Proposition A.1. □

B A rate-optimal assignment for stationary renewal blocks

This appendix proves the envelope normalization and stationary first-renewal mixture claims summarized in Section 3. We first treat an ordinary renewal block and then add the left boundary.

B.1 Completed-gap envelope

Suppose a renewal is known at time 0, and observe $y \in \{0,1\}^L$. Let

$$0 = s_0 < s_1 < \cdots < s_K \leq L$$

be its renewal times, let $d_i = s_i - s_{i-1}$ be the completed gaps, and put $c_d = |\{i : d_i = d\}|$. The terminal censored age is $A = L - s_K$, with $s_K = 0$ when $K = 0$. Let $P_\mu^{\mathrm{ord},L}$ denote the length-L law conditional on the renewal at time zero. Its exact likelihood is

$$P_\mu^{\mathrm{ord},L}(y) = \left(\prod_{i=1}^{K} \mu(d_i)\right) \overline{F}_\mu(A). \tag{B.1}$$

Define the completed-gap envelope

$$M_L(y) := \begin{cases} 1, & K = 0, \\ \displaystyle\prod_{d:c_d>0} \left(\frac{c_d}{K}\right)^{c_d}, & K > 0. \end{cases} \tag{B.2}$$

Since the survival factor in (B.1) is at most one and the multinomial likelihood is maximized at c_d/K,

$$P_\mu^{\mathrm{ord},L}(y) \le \prod_{i=1}^{K} \mu(d_i) \le M_L(y). \tag{B.3}$$

Thus M_L is an upper envelope based on completed gaps; it is not the maximum likelihood of the censored model in (B.1).

Set

$$Z_L := \sum_{y\in\{0,1\}^L} M_L(y), \qquad q_L(y) := \frac{M_L(y)}{Z_L}, \tag{B.4}$$

with $M_0(\varnothing) = Z_0 = q_0(\varnothing) = 1$.

The next lemma controls the normalization penalty that becomes the $O(\sqrt{L})$ block-redundancy term.

Lemma B.1 (Partition bound)**.** *For $L = 0$, $Z_0 = 1$. For every $L \ge 1$,*

$$Z_L \le \sum_{s=0}^{L} p(s) \le \left(1 + \frac{2\sqrt{L}}{c_{\mathrm{p}}}\right) \exp(c_{\mathrm{p}}\sqrt{L}), \tag{B.5}$$

where $p(s)$ is the integer-partition number and $p(0) = 1$. Moreover, Z_L is nondecreasing in L.

Proof. The case $s = 0$ consists only of the all-zero word and contributes one. Now fix a total completed-gap length $s = \sum_d dc_d \in \{1, \dots, L\}$. Each count vector (c_d) is a partition of s, with $K \ge 1$. Its distinct gap orderings contribute in total

$$\frac{K!}{\prod_{d:c_d>0} c_d!} \prod_{d:c_d>0} \left(\frac{c_d}{K}\right)^{c_d} \le 1,$$

because the expression is a single multinomial point probability. Once the ordered gaps are fixed, the remaining symbols form the unique terminal zero run. Summing first over count vectors and then over s gives the first inequality. For completeness, the classical partition bound [12]

$$p(s) \le \exp(c_{\mathrm{p}}\sqrt{s}), \qquad s \ge 1, \tag{B.6}$$

follows directly from Euler's product. Indeed, for every $\theta > 0$,

$$p(s)e^{-\theta s} \le \prod_{j\ge 1}(1-e^{-\theta j})^{-1},$$

$$\log \prod_{j\ge 1}(1-e^{-\theta j})^{-1} = \sum_{m\ge 1} \frac{1}{m(e^{\theta m}-1)} \le \frac{1}{\theta}\sum_{m\ge 1}\frac{1}{m^2} = \frac{\pi^2}{6\theta}.$$

Optimizing at $\theta = \pi/\sqrt{6s}$ proves (B.6). Monotonicity of the exponential now gives

$$\begin{aligned}\sum_{s=0}^{L} e^{c_{\mathrm{p}}\sqrt{s}} &\le e^{c_{\mathrm{p}}\sqrt{L}} + \int_0^L e^{c_{\mathrm{p}}\sqrt{x}}\,dx \\ &\le \left(1+\frac{2\sqrt{L}}{c_{\mathrm{p}}}\right) e^{c_{\mathrm{p}}\sqrt{L}},\end{aligned}$$

where the last line follows after setting $u = \sqrt{x}$ and bounding $u \le \sqrt{L}$. Finally, the injection $y \mapsto y0$ preserves all completed-gap counts and hence preserves M_L, proving $Z_{L+1} \ge Z_L$. □

B.2 A first-renewal mixture for stationary blocks

The stationary left boundary is handled by mixing over the location J of the first visible renewal. Conditional on J, the suffix after that renewal is an ordinary renewal block, whereas an additional no-renewal component emits the all-zero word. Equal weights on these $N+1$ cases uniformly dominate the two stationary boundary factors.

Fix a block length N. Define Q_N by choosing uniformly among the $N+1$ cases

$$J = 1,\dots,N, \qquad \text{or no renewal in the block.}$$

In the last case output 0^N. In case J, output $0^{J-1}1$, followed by a length-$(N-J)$ word drawn from q_{N-J}. In particular, $J = N$ uses $q_0(\varnothing) = 1$. Equivalently,

$$Q_N(0^N) = \frac{1}{N+1}, \qquad Q_N(x^N) = \frac{1}{N+1} q_{N-J}(x_{J+1}^N) \tag{B.7}$$

when the first one of x^N occurs at J. This is a normalized, full-support assignment. No consistency between different horizons N is required.

This pointwise guarantee is the input to both the fixed-time compression reduction and the known-horizon sequential result.

Proposition B.2 (Pointwise block-regret bound)**.** *For every $N \ge 1$,*

$$\sup_{\mu: m_\mu<\infty} \ \sup_{x^N: P_\mu(x^N)>0} \log \frac{P_\mu(x^N)}{Q_N(x^N)} \le B_N. \tag{B.8}$$

The same bound therefore holds for $\sup_\mu D_{\mathrm{KL}}(P_\mu^N, Q_N)$.

Proof. If $x^N \ne 0^N$ has first renewal J, internal completed gaps $d_1,\dots,d_K$, and terminal age A, then

$$P_\mu(x^N) = \frac{\overline{F}_\mu(J-1)}{m_\mu}\left(\prod_{i=1}^{K}\mu(d_i)\right)\overline{F}_\mu(A) \le M_{N-J}(x_{J+1}^N). \tag{B.9}$$

The two boundary factors are probabilities and hence at most one. From (B.7), Lemma B.1, and monotonicity of Z_L,

$$\frac{P_\mu(x^N)}{Q_N(x^N)} \le (N+1)Z_{N-J} \le (N+1)Z_N.$$

For $x^N = 0^N$, the ratio is at most $N+1$. Taking logarithms proves (B.8); expectation under P_μ^N gives the KL redundancy bound. □

C Exact prefix marginals in polynomial bit complexity

The regret proof alone does not make the predictor efficient: it remains to evaluate conditionals of Q_N without summing exponentially many extensions. For an ordinary-renewal prefix $z \in \{0,1\}^\ell$, $0 \le \ell \le L$, define its unnormalized continuation mass

$$W_L(z) := \sum_{v \in \{0,1\}^{L-\ell}} M_L(zv). \tag{C.1}$$

Parse the completed observed gaps in z. For $d = 1, \ldots, L$, let $a_d \in \mathbb{Z}_{\ge 0}$ be their counts, and set $a_d = 0$ outside this range. Put

$$r := \sum_d a_d, \qquad s_\star := \sum_d d a_d, \qquad A := \ell - s_\star, \qquad R := L - \ell. \tag{C.2}$$

Here A is the current censored age and R is the remaining horizon. The unique extension with no future renewal contributes

$$M_{\mathrm{nf}} := \begin{cases} 1, & r = 0, \\ r^{-r} \prod_d a_d^{a_d}, & r > 0. \end{cases} \tag{C.3}$$

The main simplification is to sum over the entire multiset of future completed gaps at once and mark how many gap copies can occur first. Throughout this DP, $a = (a_1, \ldots, a_L)$ and A are fixed. For $d = 0, 1, \ldots, L$, let

$$G_d(k,u) := \sum_{\substack{b_1,\ldots,b_d \ge 0 \\ \sum_{j\le d} b_j = k \\ \sum_{j \le d} j b_j = u}} \frac{k!}{\prod_{j \le d} b_j!} \prod_{j=1}^{d} (a_j + b_j)^{a_j + b_j}, \tag{C.4}$$

$$S_d(k,u) := \sum_{\substack{b_1,\ldots,b_d \ge 0 \\ \sum_{j\le d} b_j = k \\ \sum_{j \le d} j b_j = u}} \frac{k!}{\prod_{j \le d} b_j!} \prod_{j=1}^{d} (a_j + b_j)^{a_j + b_j} \left(\sum_{A < j \le d} b_j \right). \tag{C.5}$$

We use $0^0 = 1$. The valid state domain is

$$0 \le d \le L, \qquad 0 \le k \le u \le A + R = L - s_\star.$$

States outside this domain are zero. The base cases are $G_0(0,0) = 1$, $G_0(k,u) = 0$ for $(k,u) \ne (0,0)$, and $S_0(k,u) = 0$ for all (k,u).

Lemma C.1 (Multiset prefix recurrence)**.** *For $d = 1, \dots, L$,*

$$G_d(k,u) = \sum_{b=0}^{\min\{k,\lfloor u/d\rfloor\}} \binom{k}{b} (a_d+b)^{a_d+b} G_{d-1}(k-b, u-db), \tag{C.6}$$

$$\begin{aligned} S_d(k,u) = &\sum_{b=0}^{\min\{k,\lfloor u/d\rfloor\}} \binom{k}{b} (a_d+b)^{a_d+b} \\ &\times \Big[S_{d-1}(k-b,u-db) + \mathbf{1}\{d>A\} b G_{d-1}(k-b,u-db)\Big]. \end{aligned} \tag{C.7}$$

Moreover,

$$W_L(z) = M_{\mathrm{nf}} + \sum_{k=1}^{A+R} \sum_{u=k}^{A+R} \frac{S_L(k,u)}{k(r+k)^{r+k}}. \tag{C.8}$$

For every integer $k \geq 1$ and every u, $S_L(k,u)/k$ is a nonnegative integer.

Proof. Separating $b_d = b$ in (C.4) and using

$$\binom{k}{b} \frac{(k-b)!}{\prod_{j<d} b_j!} = \frac{k!}{b! \prod_{j<d} b_j!}$$

gives (C.6). In (C.5), the marked count $\sum_{j>A} b_j$ is the previous marked count plus $\mathbf{1}\{d>A\}b$; this gives (C.7).

It remains to justify (C.8). Fix a future completed-gap multiset $b = (b_1, \dots, b_L)$, and write

$$k = \sum_d b_d, \qquad u = \sum_d d b_d, \qquad s_A(b) = \sum_{d>A} b_d.$$

An ordering of this multiset can follow the currently censored age A exactly when its first future completed gap has length greater than A and $u - A \leq R$. The number of such orderings is

$$\sum_{e>A: b_e>0} \frac{(k-1)!}{(b_e-1)! \prod_{d\neq e} b_d!} = \frac{s_A(b)}{k} \frac{k!}{\prod_d b_d!}. \tag{C.9}$$

After an ordering is fixed, the unused positions form one terminal zero run. The resulting envelope value is

$$\frac{\prod_d (a_d+b_d)^{a_d+b_d}}{(r+k)^{r+k}}.$$

Conversely, every extension with a future renewal has one such multiset and ordering. Summing (C.9) over all $k \geq 1$ and $u \leq A+R$, and then adding M_{nf}, proves (C.8). The left side of (C.9) is an integer for each multiset. Hence the aggregated numerator $S_L(k,u)$ is divisible by k, as claimed. □

Proposition C.2 (Block conditionals)**.** *Each queried conditional of Q_N is exactly computable from two child continuation masses in* (C.8)*. More precisely, for $0 \leq t < N$,*

$$Q_N(X_{t+1} = 1 \mid X^t = 0^t) = \frac{1}{N-t+1}. \tag{C.10}$$

If the first one of $y \in \{0,1\}^t$ is at J, put $L = N - J$ and $z = y_{J+1}^t$, with $z = \varnothing$ when $J = t$. Then

$$Q_N(X_{t+1} = 1 \mid X^t = y) = \frac{W_L(z1)}{W_L(z0) + W_L(z1)}. \tag{C.11}$$

Proof. For an all-zero prefix, the compatible mixture components in (B.7) are "no renewal" and $J = t+1, \dots, N$, altogether $N-t+1$ equally weighted cases. Only $J = t+1$ has next symbol one, proving (C.10). Once the first renewal J is visible, its mixture component is known. Both child masses contain the same factors $(N+1)^{-1}Z_L^{-1}$, which cancel and give (C.11). The denominator is positive because M_L is positive on every word. □

C.1 The explicit predictor

For clarity, the fixed-time algorithm obtained from these identities is the following.

Algorithm PredictRenewal(n, x^n)**.** Set $N = n+1$. For each $t = 1, \dots, n$, regard the suffix $y = x^n_{n-t+1}$ as a length-t prefix of Q_N. If $y = 0^t$, set $q_t = 1/(N-t+1)$. Otherwise locate its first one J, put $L = N - J$ and $z = y^t_{J+1}$, and compute $W_L(z0)$ and $W_L(z1)$ by (C.6)–(C.8). Set

$$q_t = \frac{W_L(z1)}{W_L(z0) + W_L(z1)}$$

and return $n^{-1}\sum_{t=1}^n q_t$.

Proposition C.3 (Arithmetic and bit complexity)**.** *For a fixed prefix, $W_L(z)$ is exactly evaluable with $O(N^4)$ integer arithmetic operations and $O(N^2)$ rolling table entries. All conditionals in any specified $O(N)$-query batch require $O(N^5)$ arithmetic operations. This includes both the suffix batch used by* PredictRenewal *and the forward on-path batch for a supplied block; it does not enumerate all prefixes. Every intermediate integer has* $\mathrm{poly}(N)$ *bits, so the algorithm is deterministic uniform polynomial-bit-time, with $O(N^3 \log N)$ rolling bit space.*

Proof. Tabulate $\binom{k}{b}$, for $0 \le b \le k \le N$, by Pascal's recurrence, and tabulate c^c, for $0 \le c \le N$, by repeated squaring, with $0^0 = 1$. Fill (C.6) and (C.7) in increasing d. The four indices d, k, u, b are at most N, so one continuation mass uses $O(N^4)$ updates. Only layers $d-1$ and d are needed, giving $O(N^2)$ table space. There are two child masses for each of n suffixes.

For bit lengths, feasibility gives

$$r + k \le r + A + R = r + L - s_\star \le L \le N,$$

because every completed gap has length at least one and hence $r \le s_\star$. A fixed state aggregates at most 2^N ordered positive-gap sequences. If $c_j = a_j + b_j$ and $K_d = \sum_{j \le d} c_j \le N$, then

$$\prod_{j \le d} c_j^{c_j} \le K_d^{K_d} \le N^N.$$

Thus $G_d(k,u) \le 2^N N^N$, $S_d(k,u) \le N G_d(k,u)$, and both have $O(N \log N)$ bits. By Lemma C.1, divide $S_L(k,u)$ by k exactly before forming (C.8). Every remaining denominator is K^K for some $K \le N$, and hence divides

$$D_N := \prod_{K=1}^{N} K^K, \qquad \log_2 D_N = O(N^2 \log N). \tag{C.12}$$

Within one W_L evaluation, putting all terms over D_N gives a numerator and denominator of $O(N^2 \log N)$ bits. Averaging the n resulting rational conditionals uses at most $O(N^3 \log N)$-bit integers; component ratios and the final average are retained as unreduced integer pairs and need not have denominators dividing D_N. Pascal's recurrence, exact integer exponentiation, addition, multiplication, comparison, and the exact divisions certified by Lemma C.1 are the only arithmetic primitives. Standard exact integer algorithms therefore give polynomial Turing time; the deliberately crude schoolbook bound $O(N^{11}\log^2 N)$ bit operations already suffices. Releasing the previous d-layer and reusing the workspace across queries gives the stated $O(N^3 \log N)$ rolling bit-space bound. □

Corollary C.4 (Efficient rate-optimal sequential assignment)**.** *For every N, each requested value Z_L, block mass $Q_N(x^N)$, or queried sequential conditional of Q_N is exactly computable in polynomial bit complexity. A supplied path's N on-path conditionals form a specified $O(N)$-query batch covered by Proposition C.3. Moreover, for every μ and every P_μ-positive word x^N,*

$$\sum_{t=0}^{N-1} \log \frac{P_\mu(x_{t+1} \mid x^t)}{Q_N(x_{t+1} \mid x^t)} = \log \frac{P_\mu(x^N)}{Q_N(x^N)} \le B_N. \tag{C.13}$$

Proof. The identity $Z_L = W_L(\varnothing)$ computes every normalizer. Equations (B.7) and (C.11) then compute the mass and conditionals of Q_N. The chain rule gives the equality in (C.13), and Proposition B.2 gives the inequality. □

D Compression to prediction

This appendix makes the compression-to-prediction bridge self-contained. It first separates the loss due to a block assignment from the information in the unobserved past, then bounds the latter by a truncated renewal countdown.

Let P be the law of a stationary binary process, let P^N be its length-N marginal, where $N = n+1$, and let Q_N be any full-support assignment. Define

$$\widehat{q}_n(x^n) := \frac{1}{n}\sum_{t=1}^{n} Q_N\big(X_{t+1} = 1 \mid X_1^t = x_{n-t+1}^n\big). \tag{D.1}$$

For a stationary renewal process, define the inclusive countdown

$$C_i := \min\{j \ge 1 : X_{i+j-1} = 1\}.$$

Finite positive-integer gaps imply $C_i < \infty$ almost surely.

Lemma D.1 (Suffix regeneration)**.** *For $1 \le t \le n$, with $Y = X_{n+1}$, $O_t = X_1^{n-t}$, and $S_t = X_{n-t+1}^n$,*

$$Y \perp O_t \mid (S_t, C_{n-t+1}).$$

Proof. If $C_{n-t+1} \le t$, a renewal occurs inside the suffix; conditional on its location and the observed suffix after it, regeneration makes the later process independent of O_t. If $C_{n-t+1} > t$, then $S_t = 0^t$ and $Y = \mathbf{1}\{C_{n-t+1} = t+1\}$, so Y is determined once the countdown is given. These two cases prove the conditional independence. □

Lemma D.2 (Compression-to-prediction)**.** *For every stationary process law P, the predictor* (D.1) *satisfies*

$$\begin{aligned} &\mathbb{E}_P D_{\mathrm{KL}}(P(X_{n+1} \in \cdot \mid X^n), \mathrm{Bern}(\widehat{q}_n(X^n))) \\ &\quad\le \frac{1}{n} D_{\mathrm{KL}}(P^N, Q_N) + \frac{1}{n}\sum_{t=1}^{n} I_P(X_{n+1}; X_1^{n-t} \mid X_{n-t+1}^n). \end{aligned} \tag{D.2}$$

For the class of stationary renewal processes, the second term is at most $\log(n+2)/n$*, uniformly over* μ.

Proof. Convexity of KL in its second argument bounds the left side by the average loss against the n conditionals appearing in (D.1). Fix t, and write

$$Y := X_{n+1}, \qquad S_t := X_{n-t+1}^{n}, \qquad O_t := X_1^{n-t}.$$

Let $q_t(\cdot \mid S_t)$ be the conditional of Q_N that treats the realized length-t suffix as its prefix. Conditional KL has the exact Pythagorean decomposition

$$\begin{aligned} &\mathbb{E}_P D_{\mathrm{KL}}\big(P_{Y|O_t,S_t}, q_t(\cdot \mid S_t)\big) \\ &\quad = I_P(Y; O_t \mid S_t) + \mathbb{E}_P D_{\mathrm{KL}}\big(P_{Y|S_t}, q_t(\cdot \mid S_t)\big)\,. \end{aligned} \tag{D.3}$$

Indeed, on each positive-probability conditioning event, expand both KL terms on the common binary outcome space and cancel the logarithm of $P(Y \mid S_t)$. Null-event versions are arbitrary, and full support of Q_N makes the predictor terms finite. By stationarity, the second term in (D.3) has the same law as

$$D_{\mathrm{KL}}\big(P(X_{t+1} \in \cdot \mid X_1^t), Q_N(X_{t+1} \in \cdot \mid X_1^t)\big)\,.$$

Summing over t and using the KL chain rule gives

$$D_{\mathrm{KL}}(P^N, Q_N) - D_{\mathrm{KL}}(P^1, Q_N^1) \le D_{\mathrm{KL}}(P^N, Q_N),$$

where Q_N^1 is the first-coordinate marginal of Q_N. This proves (D.2).

For a renewal process, Lemma D.1 and conditional data processing give the first line below; stationarity gives the equality:

$$\begin{aligned} I_P(X_{n+1}; X_1^{n-t} \mid X_{n-t+1}^n) &\le I_P(X_{n+1}; C_{n-t+1} \mid X_{n-t+1}^n) \\ &= I_P(X_{t+1}; C_1 \mid X_1^t). \end{aligned}$$

Summing, then applying the mutual-information chain rule and nonnegativity, gives

$$\sum_{t=1}^{n} I_P(X_{n+1}; X_1^{n-t} \mid X_{n-t+1}^n) \le \sum_{t=1}^{n} I_P(C_1; X_{t+1} \mid X_1^t) \le I_P(C_1; X^N).$$

Set $\widetilde{C}_1 = \min\{C_1, n+2\}$. On $\{\widetilde{C}_1 \le n+1\}$, this truncation retains C_1 exactly; on its remaining value $n+2$, there is no renewal in the block X^N, where $N = n+1$, and hence $X^N = 0^N$. The conditional law of X^N given C_1 therefore depends only on $\widetilde{C}_1$, proving the Markov chain $C_1 \to \widetilde{C}_1 \to X^N$. Hence

$$I_P(C_1; X^N) \le H(\widetilde{C}_1) \le \log(n+2),$$

as claimed. $\square$

Combining pointwise block regret, exact queried conditionals, and the preceding information bound gives the promised fixed-time algorithm.

Theorem D.3 (Polynomial-time rate-optimal predictor)**.** *For every $n \ge 1$, the deterministic predictor* (D.1)*, with $N = n+1$ and Q_N defined by* (B.7)*, is computable in polynomial bit complexity and satisfies*

$$\sup_{\mu : m_\mu < \infty} \mathcal{R}_n(\widehat{q}_n, \mu) \le \frac{B_{n+1} + \log(n+2)}{n}. \tag{D.4}$$

Proof. Apply Lemma D.2 and Proposition B.2 with $N = n + 1$. Polynomial-bit computability follows from Propositions C.2 and C.3. For the explicit constant, $\sqrt{n+1}/n \le \sqrt{2}/\sqrt{n}$. Also, $1 + 2\sqrt{n+1}/c_{\mathrm{p}} \le n + 2$ and $\log(n + 2)/n \le \sqrt{2}/\sqrt{n}$ for every $n \ge 1$. Thus the square-root term contributes at most $2\pi/(\sqrt{3}\sqrt{n})$, and the two normalizer logarithms plus the memory logarithm contribute at most $3\sqrt{2}/\sqrt{n}$, proving the last inequality in (2.9). □

Corollary D.4 (The all-zero contribution is lower order)**.** *Let $H_m = \sum_{j=1}^{m} 1/j$. On the all-zero input, the predictor in Theorem D.3 outputs*

$$\widehat{q}_n(0^n) = \frac{H_{n+1} - 1}{n}. \tag{D.5}$$

Uniformly over μ, the contribution of $\{X^n = 0^n\}$ to $\mathcal{R}_n(\widehat{q}_n, \mu)$ is $O(\log n/n)$.

Proof. Equation (C.10) gives

$$\widehat{q}_n(0^n) = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{n-t+2} = \frac{H_{n+1} - 1}{n} =: q.$$

If $R_n = 0$, the event has probability zero and its contribution vanishes. Assume henceforth that $R_n > 0$, and write $w = P_\mu(X^n = 0^n) = R_n/m_\mu$ and $p = \overline{F}_\mu(n)/R_n$. The tail-sum identity and monotonicity give

$$m_\mu = \sum_{a\ge 0} \overline{F}_\mu(a) \ge (n + 1)\overline{F}_\mu(n).$$

Consequently,

$$wp = \frac{\overline{F}_\mu(n)}{m_\mu} \le \frac{1}{n+1}.$$

Using $p \le 1$, $w \le 1$, and $\log((1 - p)/(1 - q)) \le -\log(1 - q)$,

$$wd(p\|q) \le \frac{\log(1/q)}{n+1} - \log(1 - q) = O\left(\frac{\log n}{n}\right).$$

This is precisely the stationary weighting absent from the ordinary-hazard substitution in Proposition A.1. □

E Adaptation to stationary occupied-gap complexity

The worst-case envelope must allow arbitrarily many gap lengths. On an individual block, however, only a small number of distinct completed gaps may appear. We now exploit this fact without asking the predictor to know a support size or tail class.

For $y \in \{0, 1\}^L$, using the completed-gap parse from Appendix B, let

$$\nu(y) := |\{d : c_d(y) > 0\}|$$

be its number of distinct completed-gap lengths, with $\nu(0^L) = 0$ and $\nu(\varnothing) = 0$. For $s \ge 1$, define

$$M_{L,s}(y) := M_L(y)\mathbf{1}\{\nu(y) \le s\}, \qquad Z_{L,s} := \sum_y M_{L,s}(y), \qquad q_{L,s} := \frac{M_{L,s}}{Z_{L,s}}. \tag{E.1}$$

For $L = 0$, set $M_{0,s}(\varnothing) = Z_{0,s} = q_{0,s}(\varnothing) = 1$. Let $Q_{N,s}$ be the first-renewal mixture (B.7) with $q_{N-J,s}$ in place of q_{N-J}. Although $Q_{N,s}$ need not have full support when $s < N$, $Q_{N,N} = Q_N$.

Lemma E.1 (Restricted normalization)**.** *For $L \geq 0$ and $s \geq 1$, $Z_{0,s} = 1$, and for $L \geq 1$,*

$$Z_{L,s} \leq 1 + \sum_{j=1}^{\min\{s,L\}} \binom{L}{j} L^j \leq (L+1)^{2s+1}. \tag{E.2}$$

Proof. The case $L = 0$ is the stated convention. Assume $L \geq 1$. As in Lemma B.1, all orderings associated with one completed-gap count vector have total envelope mass at most one. A vector using exactly j gap lengths is specified by choosing those lengths from $[L]$ and then assigning each a positive count at most L, giving at most $\binom{L}{j}L^j$ possibilities. The all-zero vector contributes one. For the second inequality, each summand is at most $(L+1)^{2s}$, there are at most s of them, and $s+1 \leq L+1$ when $s \leq L$; if $s > L$, replace s by L in the first bound. $\square$

For $1 \leq s \leq N$, set

$$\omega_{s,N} := \begin{cases} \dfrac{1}{2s(s+1)}, & s < N, \\ \dfrac{1}{2} + \dfrac{1}{2N}, & s = N, \end{cases} \qquad Q_N^{\mathrm{ad}} := \sum_{s=1}^{N} \omega_{s,N} Q_{N,s}. \tag{E.3}$$

The weights sum to one by telescoping, and the $s = N$ component makes Q_N^{ad} full support. If $x^N \neq 0^N$ has first renewal J, define

$$V_N(x^N) := \nu(x_{J+1}^N),$$

and set $V_N(0^N) = 0$ and $\kappa_N(x^N) = \max\{1, V_N(x^N)\}$.

Theorem E.2 (Data-dependent pointwise oracle)**.** *For every $N \geq 1$, every finite-mean renewal law μ, and every x^N with $P_\mu(x^N) > 0$,*

$$\log \frac{P_\mu(x^N)}{Q_N^{\mathrm{ad}}(x^N)} \leq \log 2 + \min\Big\{B_N, (2\kappa_N(x^N) + 2)\log(N+1) + 2\log(\kappa_N(x^N) + 1)\Big\}. \tag{E.4}$$

Proof. The component $s = N$ has weight at least $1/2$, so Proposition B.2 gives the first branch. For the second, take $s = \kappa_N(x^N)$. For $N \geq 2$, this choice satisfies $s < N$: if $V_N = 0$, then $s = 1 < N$; otherwise the s distinct positive completed gap lengths have sum at least $s(s+1)/2$ and fit within a block of length N. The case $N = 1$ follows directly from the universal $s = N$ component. If the word is nonzero, the likelihood domination (B.9) and Lemma E.1 give

$$\frac{P_\mu(x^N)}{Q_{N,s}(x^N)} \leq (N+1) Z_{N-J,s} \leq (N+1)^{2s+2}.$$

The all-zero word satisfies the same bound because its ratio is at most $N+1$. For $s < N$, $-\log \omega_{s,N} \leq \log 2 + 2\log(s+1)$. Taking the better component proves (E.4); the universal branch covers the remaining $N = 1$ case. $\square$

E.1 Exact adaptive prefix masses

The restricted assignments remain polynomial-time computable. Refine (C.4)–(C.5) by adding a state j equal to

$$|\{i \le d : a_i + b_i > 0\}| .$$

Write the resulting tables as $G_d(k,u,j)$ and $S_d(k,u,j)$, and let $\epsilon_d(b) = \mathbf{1}\{a_d + b > 0\}$. Their recurrences are

$$G_d(k,u,j) = \sum_b \binom{k}{b} (a_d + b)^{a_d+b} G_{d-1}(k-b, u-db, j-\epsilon_d(b)), \tag{E.5}$$

$$\begin{aligned} S_d(k,u,j) = \sum_b \binom{k}{b} (a_d + b)^{a_d+b} \Big[& S_{d-1}(k-b, u-db, j-\epsilon_d(b)) \\ & + \mathbf{1}\{d > A\} b G_{d-1}(k-b, u-db, j-\epsilon_d(b)) \Big], \end{aligned} \tag{E.6}$$

where $b = 0, \dots, \min\{k, \lfloor u/d \rfloor\}$, out-of-range states are zero, $G_0(0,0,0) = 1$, all other G_0 entries are zero, and every S_0 entry is zero. If $\nu(a) = |\{d : a_d > 0\}|$, then

$$W_{L,s}(z) := \sum_v M_{L,s}(zv) = \mathbf{1}\{\nu(a) \le s\} M_{\mathrm{nf}} + \sum_{k=1}^{A+R} \sum_{u=k}^{A+R} \frac{\sum_{j=0}^{s} S_L(k,u,j)}{k(r+k)^{r+k}}. \tag{E.7}$$

This follows from the same multiset bijection as Lemma C.1, now sorting the final count vectors by their number of occupied coordinates. Summing over that coordinate projects the refined tables onto the ordinary ones:

$$\sum_{j=0}^{L} G_d(k,u,j) = G_d(k,u), \qquad \sum_{j=0}^{L} S_d(k,u,j) = S_d(k,u). \tag{E.8}$$

All entries are nonnegative, so each refined entry inherits the ordinary table's bit-length bound.

For a zero prefix of length t, every $Q_{N,s}$ has prefix mass $(N-t+1)/(N+1)$. Once its first one J is visible, with $L = N - J$ and suffix prefix z,

$$Q_{N,s}(X^t = y) = \frac{1}{N+1} \frac{W_{L,s}(z)}{Z_{L,s}}, \qquad Z_{L,s} = W_{L,s}(\varnothing). \tag{E.9}$$

The conditional of Q_N^{ad} is obtained by first mixing these *prefix masses* and then taking the ratio of its two child masses. It is not, in general, the prior-weighted average of the component conditionals. The universal $s = N$ component guarantees that the mixed denominator is positive.

The five recurrence indices d, k, u, j, b are bounded by N, so one refined table takes $O(N^5)$ integer operations. Prefix sums in j extract $W_{L,s}$ for every $s = 1, \dots, N$ from that single table; the DP is not rerun separately for each s. The fixed-time suffix batch uses at most $2n$ child-prefix tables and $O(n)$ normalizer tables, for $O(N^6)$ integer operations in total. By (E.8), table entries have the bit lengths from Proposition C.3. Unreduced component ratios and mixtures remain polynomial length. A safe schoolbook bound is $O(N^{14} \log^2 N)$ bit operations and $O(N^4 \log N)$ rolling bit space. We have proved:

Proposition E.3 (Computability of the adaptive assignment)**.** *Each queried prefix mass or conditional of Q_N^{ad}, any specified $O(N)$-query suffix or forward-path batch, and the compression-averaged predictor formed from the suffix batch are exactly computable by a deterministic uniform polynomial-bit-time algorithm.*

E.2 Distribution-dependent fixed-time rates

Define the stationary occupied-gap complexity at block length N by

$$v_N(\mu) := \sum_{d=1}^{N-1} \min\left\{1, \frac{(N-d)\mu(d)}{m_\mu}\right\}. \tag{E.10}$$

This quantity controls the expected number of distinct completed gap types in a stationary block: long gaps have fewer opportunities to be completed inside it. It is neither the support cardinality nor an estimated parameter dimension.

Theorem E.4 (Adaptive fixed-time prediction)**.** *Let $\widehat{q}_n^{\mathrm{ad}}$ be* (D.1) *with Q_{n+1}^{ad} in place of Q_{n+1}. For every $n \geq 1$ and every finite-mean μ,*

$$\mathcal{R}_n(\widehat{q}_n^{\mathrm{ad}}, \mu) \leq \frac{1}{n}\Big[\log 2 + \log(n+2) + \min\big\{B_{n+1}, (2v_{n+1}(\mu) + 6)\log(n+2)\big\}\Big]. \tag{E.11}$$

The predictor receives none of $v_{n+1}(\mu)$, a support size, or a tail parameter as input.

Proof. Set $N = n+1$. Let C_d count completed gaps of length d lying wholly in a stationary block of length N. A renewal occurs at each possible start with intensity $1/m_\mu$, and its following gap has law μ, so

$$\mathbb{E}_\mu C_d = \frac{(N-d)\mu(d)}{m_\mu}, \qquad 1 \leq d < N.$$

Therefore

$$\mathbb{E}_\mu V_N(X^N) = \sum_{d<N} \mathbb{P}_\mu(C_d \geq 1) \leq v_N(\mu). \tag{E.12}$$

Take expectations in (E.4). Define, for $0 \leq k \leq N$,

$$f_N(k) := (2k+2)\log(N+1) + 2\log(k+1), \qquad g_N(k) := \min\{B_N, f_N(k)\}.$$

The function f_N is increasing and concave. At its intersection with the constant B_N, the slope drops from a nonnegative value to zero, so g_N is also increasing and concave. Since $\kappa_N \leq V_N + 1$, (E.12) and Jensen's inequality give

$$\begin{aligned}
\mathbb{E}_\mu g_N(\kappa_N) &\leq g_N(\mathbb{E}_\mu \kappa_N) \leq g_N(v_N(\mu)+1) \\
&\leq \min\{B_N, (2v_N(\mu)+4)\log(N+1) + 2\log(v_N(\mu)+2)\} \\
&\leq \min\{B_N, (2v_N(\mu)+6)\log(N+1)\},
\end{aligned}$$

where the last step uses $v_N(\mu) \leq N-1$. Substitution in the pointwise oracle inequality yields

$$D_{\mathrm{KL}}(P_\mu^N, Q_N^{\mathrm{ad}}) \leq \log 2 + \min\{B_N, (2v_N(\mu)+6)\log(N+1)\}. \tag{E.13}$$

Apply Lemma D.2 with $N = n+1$ and add its memory term $\log(n+2)/n$. $\square$

Corollary E.5 (Automatic tail adaptation)**.** *The predictor in Theorem E.4 has the following rates, with constants allowed to depend on the displayed tail parameters:*

1. *if $|\mathrm{supp}(\mu)| \leq s$, then*

$$\mathcal{R}_n(\widehat{q}_n^{\mathrm{ad}}, \mu) = O\left(\min\left\{n^{-1/2}, \frac{(s+1)\log n}{n}\right\}\right);$$

2. *if* $\mu(d) \le Ce^{-cd}$*, then the risk is* $O_{C,c}(\log^2 n/n)$*;*

3. *if* $\mu(d) \le Cd^{-\alpha}$ *for* $\alpha > 2$*, then the risk is* $O_{C,\alpha}(n^{-(1-1/\alpha)}\log n)$*.*

Proof. The first statement uses $v_N(\mu) \le s$. Since $m_\mu \ge 1$, the other two follow from

$$v_N(\mu) \le \sum_{d\ge 1} \min\{1, N\mu(d)\}.$$

Splitting the exponential sum at $d \asymp \log N$ gives $O(\log N)$. Splitting the power-law sum at $d \asymp N^{1/\alpha}$ gives $O(N^{1/\alpha})$. Substitute these bounds into (E.11). ☐

The logarithmic finite-support endpoint is not merely an artifact of the upper bound, even for a deterministic gap law.

Proposition E.6 (Unknown deterministic period)**.** *Let* $\mathcal{D} = \{\delta_d : d \ge 1\}$*. For all sufficiently large* n*,*

$$\inf_{\widehat{q}_n} \sup_{\mu\in\mathcal{D}} \mathcal{R}_n(\widehat{q}_n, \mu) = \Theta\left(\frac{\log n}{n}\right), \tag{E.14}$$

where the infimum permits randomized predictors.

Proof. The upper bound is the $s = 1$ case of Corollary E.5. For the lower bound take $\mu_0 = \delta_{n+1}$, $\mu_1 = \delta_{2n}$, and $E = \{X^n = 0^n\}$. Explicitly,

$$P_{\mu_0}(E) = \frac{1}{n+1}, \quad p_{\mu_0}(0^n) = 1; \qquad P_{\mu_1}(E) = \frac{1}{2}, \quad p_{\mu_1}(0^n) = \frac{1}{n}.$$

For a deterministic output q, put $q_\star = 4\log(en)/n$, which is at most one for all sufficiently large n. If $q \le q_\star$, then

$$\frac{1}{n+1} d(1\|q) \ge \frac{1}{n+1} \log \frac{n}{4\log(en)} = \Omega\left(\frac{\log n}{n}\right).$$

If $q > q_\star$, use $d(p\|q) \ge (1-p)q - H_2(p)$ and $H_2(1/n) \le \log(en)/n$ to obtain the same order from $\frac{1}{2}d(1/n\|q)$. Thus the sum of the two weighted losses is pointwise $\Omega(\log n/n)$. The dichotomy holds for every realization of a privately randomized output; taking its expectation and using that the maximum risk dominates the two-point Bayes average proves the randomized lower bound. ☐

F A matching-order lower bound with both boundary factors

We give a self-contained Bayes lower bound that retains both stationary boundary factors. The proof passes from a random support to visible completed gaps, a positive-probability good event, posterior membership uncertainty, and finally a Bernoulli KL radius. The construction uses only finite-support laws, so it applies within the class of the theorem.

Theorem F.1 (Finite-support minimax lower bound retaining both boundaries)**.** *For every* $n \ge 600$*, every possibly randomized predictor obeys*

$$\sup_{\mu: m_\mu < \infty} \mathcal{R}_n(\widehat{q}_n, \mu) \ge \frac{11(\log 2 - 1/2)}{40600\sqrt{n}}. \tag{F.1}$$

The displayed lower bound already holds when the supremum is restricted to aperiodic laws that are uniform on $a = 24\lceil\sqrt{n}\rceil$ *points in* $[2a]$ *and satisfy* $T \le 4m_\mu$ *almost surely.*

Let $b = \lceil \sqrt{n} \rceil$ and $a = 24b$, and write $[a] = \{1, \dots, a\}$. For $n \geq 600$, $a \leq n$ and $a \leq 25\sqrt{n}$. Independently choose

$$\mathsf{L} \sim \mathrm{Unif}\binom{[a]}{a/2}, \qquad \mathsf{H} = \{2a-1, 2a\} \cup \mathsf{H}', \quad \mathsf{H}' \sim \mathrm{Unif}\binom{\{a+1, \dots, 2a-2\}}{a/2-2}, \qquad \mathsf{S} = \mathsf{L} \cup \mathsf{H}, \tag{F.2}$$

and let μ_{S} be uniform on S. Let Π denote the resulting joint Bayes law of S and the observations. Write

$$m(\mathsf{S}) = \frac{1}{a} \sum_{d \in \mathsf{S}} d, \qquad U_t(\mathsf{S}) := |\{d \in \mathsf{S} : d \geq t\}|.$$

Here $a = 24b$ is even, $|\mathsf{S}| = a$, and $\mathsf{S} \subseteq [2a]$. Moreover, $\{2a-1, 2a\} \subseteq \mathsf{S}$, so every law is aperiodic, while $m(\mathsf{S}) \geq a/2$ implies $T \leq 2a \leq 4m(\mathsf{S})$ almost surely. We also have $m(\mathsf{S}) \leq 2a$, and the stationary initial wait satisfies

$$\mathbb{P}_{\mathsf{S}}(T_0 = t) = \frac{U_t(\mathsf{S})}{am(\mathsf{S})}, \qquad t \geq 1.$$

Lemma F.2 (Stationary cycle representation)**.** *In a two-sided stationary renewal process, if D_0 is the length of the cycle containing a fixed time and $A_0 \in \{0, \dots, D_0 - 1\}$ is that time's age, then*

$$\mathbb{P}(D_0 = d, A_0 = j) = \frac{\mu(d)}{m_\mu}, \qquad 0 \leq j < d. \tag{F.3}$$

Thus D_0 is size biased, its conditional position is uniform, and the completed gaps strictly to its left are iid μ and independent of (D_0, A_0).

Proof. Construct the process by choosing (D_0, A_0) according to (F.3), attaching independent iid μ-gaps on both sides, and placing the origin at position A_0. Every cycle-position pair has mass $\mu(d)/m_\mu$, which is invariant under a unit shift, while regeneration at the adjacent endpoints gives the stated independence. Restricting this two-sided construction to times $t \geq 1$ has exactly the one-sided stationary law defined in Section 2. □

For the hard prior, let g be the current age at time n, plus one. Its stationary law is

$$\mathbb{P}_{\mathsf{S}}(g = t) = \frac{U_t(\mathsf{S})}{am(\mathsf{S})}. \tag{F.4}$$

When $g \leq a$, the last renewal $K = n + 1 - g$ is visible. Since $a \leq n$, it lies in $\{1, \dots, n\}$, and

$$g = n + 1 - \max\{t \leq n : X_t = 1\}.$$

Write the visible renewals as

$$1 \leq t_0 = J < t_1 < \cdots < t_r = K \leq n, \qquad d_i = t_i - t_{i-1},$$

and let $\mathsf{A} = \{d_1, \dots, d_r\}$ be the set of distinct internal completed gaps. Every quantity in the following event is therefore determined by X^n:

$$\mathcal{E} := \{g \leq a,\ r \leq b,\ g \notin \mathsf{A}\}. \tag{F.5}$$

Lemma F.3 (Visible good event)**.** *For every support S in* (F.2), $\mathbb{P}_{\mathsf{S}}(\mathcal{E}) \geq 1/8$.

Proof. Half of the support lies above a, so

$$\mathbb{P}_{\mathsf{S}}(g \le a) = \frac{\mathbb{E}_{\mathsf{S}} \min\{T, a\}}{m(\mathsf{S})} \ge \frac{1}{4}. \tag{F.6}$$

Stationary intensity gives

$$\mathbb{E}_{\mathsf{S}} \sum_{t=1}^{n} X_t = \frac{n}{m(\mathsf{S})} \le \frac{2n}{a}, \qquad \mathbb{P}_{\mathsf{S}}(g \le a,\ r > b) \le \mathbb{P}_{\mathsf{S}}\left(\sum_{t=1}^{n} X_t > b\right) \le \frac{2n}{ab} \le \frac{1}{12}.$$

By Lemma F.2, the completed gaps to the left of the interval containing time n are iid μ_{S} and independent of the current age. Denote them, backward from the current interval, by $D_{-1}, D_{-2}, \ldots$. Pathwise on $\{g \le a, r \le b\}$, $\mathsf{A} \subseteq \{D_{-1}, \ldots, D_{-b}\}$. Thus, for every $1 \le g \le a$, conditional on g and S,

$$\mathbb{P}_{\mathsf{S}}(g \in \mathsf{A},\ r \le b \mid g) \le \mathbb{P}_{\mathsf{S}}(D_{-j} = g \text{ for some } j \le b \mid g) \le \frac{b}{a} = \frac{1}{24}.$$

Consequently,

$$\mathbb{P}_{\mathsf{S}}(g \le a,\ g \in \mathsf{A},\ r \le b) \le \frac{b}{a} \mathbb{P}_{\mathsf{S}}(g \le a) \le \frac{1}{24}.$$

Combining this with (F.6) gives, for every S,

$$\mathbb{P}_{\mathsf{S}}(\mathcal{E}) \ge \frac{1}{4} - \frac{1}{12} - \frac{1}{24} = \frac{1}{8}. \tag{F.7}$$

□

Lemma F.4 (Exact two-boundary likelihood)**.** *For every realized block $x \in \mathcal{E}$ with positive prior-predictive probability, its exact likelihood under S is*

$$P_{\mathsf{S}}(X^n = x) = \mathbf{1}\{\mathsf{A} \subseteq \mathsf{S}\} a^{-(r+2)} \frac{U_J(\mathsf{S}) U_g(\mathsf{S})}{m(\mathsf{S})}. \tag{F.8}$$

Proof. Indeed, the three factors are respectively the equilibrium initial wait $U_J/(am)$, the r internal gaps $a^{-r}\mathbf{1}\{\mathsf{A} \subseteq \mathsf{S}\}$, and the terminal survival U_g/a. In particular, the left-censoring factor is retained. □

Lemma F.5 (Posterior stability)**.** *For every $x \in \mathcal{E}$ with positive prior-predictive probability, the posterior membership and nonmembership probabilities are both at least $\eta := 11/203$.*

Proof. Let Π_{A} be the support prior conditioned on $\mathsf{A} \subseteq \mathsf{S}$. Under Π_{A}, the low and high supports remain independent. Put $\ell = |\mathsf{A} \cap [a]|$. Since $\ell \le b \le a/24$, and since $g \notin \mathsf{A}$,

$$\Pi_{\mathsf{A}}(g \in \mathsf{L}) = \frac{a/2 - \ell}{a - \ell} \in [11/23, 1/2]. \tag{F.9}$$

Relative to this conditional prior, (F.8) weights a compatible support by

$$w_x(\mathsf{S}) = \frac{U_J(\mathsf{S}) U_g(\mathsf{S})}{m(\mathsf{S})}. \tag{F.10}$$

If $J \le a$, then $U_J, U_g \in [a/2, a]$ and $m \in [a/2, 2a]$, so the ratio of the largest and smallest compatible weights is at most 16. If $J > a$, first fix a compatible high support H. Then U_J is constant as L varies, U_g varies by at most a factor two, and m varies by at most a factor two: indeed,

$$\sum_{h \in \mathsf{H}} h \ge a^2/2, \qquad \sum_{\ell' \in \mathsf{L}} \ell' \le a^2/2.$$

Thus the conditional weight ratio across low supports is at most four. More explicitly, for compatible low supports $\mathsf{L}_1, \mathsf{L}_2$ and every compatible H, either $U_J(\mathsf{H}) = 0$, so both weights vanish, or

$$w_x(\mathsf{L}_1, \mathsf{H}) \le 4 w_x(\mathsf{L}_2, \mathsf{H}).$$

Since the conditional law of H under Π_{A} is the same for every L, define the marginalized weight

$$\overline{w}_x(\mathsf{L}) := \mathbb{E}_{\Pi_{\mathsf{A}}}[w_x(\mathsf{L}, \mathsf{H}) \mid \mathsf{L}].$$

This quantity is positive for every compatible L. Indeed, because x has positive Bayes probability, some high support containing $\mathsf{A} \cap \{a+1, \dots, 2a\}$ has $U_J > 0$. That high support has positive conditional prior mass for every compatible L, while $U_g/m > 0$. Averaging the pointwise comparison gives $\overline{w}_x(\mathsf{L}_1) \le 4\overline{w}_x(\mathsf{L}_2)$ for every pair of compatible low supports. Thus, when $J > a$, the posterior marginal on L is its Π_{A}-marginal reweighted by quantities with ratio at most four. The membership event in (F.9) depends only on L.

In general, if an event has prior probability p and the largest weight is at most C times the smallest, its reweighted probability and complementary probability are at least

$$\frac{p}{p + C(1-p)}, \qquad \frac{1-p}{Cp + 1 - p},$$

respectively. For $J \le a$, use $C = 16$ and $p \in [11/23, 1/2]$ to obtain the bounds $11/203$ and $1/17$. For $J > a$, the marginalized comparison uses $C = 4$, giving the still stronger bounds $11/59$ and $1/5$. Hence the uniform constant $\eta = 11/203$ safely gives

$$\Pi(g \in \mathsf{S} \mid X^n = x) \ge \eta, \qquad \Pi(g \notin \mathsf{S} \mid X^n = x) \ge \eta. \tag{F.11}$$

□

Lemma F.6 (Bernoulli posterior radius)**.** *For $x \in \mathcal{E}$, the true next-step parameter is*

$$\theta_{\mathsf{S}}(x) = \frac{\mathbf{1}\{g \in \mathsf{S}\}}{U_g(\mathsf{S})}. \tag{F.12}$$

For every proposed output $q \in [0, 1]$,

$$\mathbb{E}_{\Pi(\cdot \mid X^n = x)}[d(\theta_{\mathsf{S}}(x) \| q)] \ge \frac{\eta(\log 2 - 1/2)}{a}. \tag{F.13}$$

Proof. Because $g \le a$, the entire high half of the support contributes to U_g. Hence $\theta_{\mathsf{S}} = 0$ when $g \notin \mathsf{S}$, while $\theta_{\mathsf{S}} \in [1/a, 2/a]$ when $g \in \mathsf{S}$.

If $q \ge 1/(2a)$, the second event in (F.11), on which the parameter is zero, gives conditional risk at least $\eta d(0\|q) = \eta[-\log(1-q)] \ge \eta q \ge \eta/(2a)$. This is stronger than the common lower bound $\eta(\log 2 - 1/2)/a$ used below. If $q < 1/(2a)$, then $q \le \theta/2$ on the positive-parameter event. Since $d(\theta\|q)$ decreases in $q < \theta$,

$$\begin{aligned} d(\theta\|q) &\ge d(\theta\|\theta/2) \\ &= \theta \log 2 + (1-\theta) \log \frac{1-\theta}{1-\theta/2} \\ &\ge (\log 2 - 1/2)\theta \ge \frac{\log 2 - 1/2}{a}. \end{aligned}$$

For the penultimate inequality, write the second logarithm as $\log(1 - \theta/(2-\theta))$ and use $\log(1-u) \ge -u/(1-u)$; its contribution is at least $-\theta/2$. Thus, for every q, the posterior conditional risk on $x \in \mathcal{E}$ is at least $\eta(\log 2 - 1/2)/a$. The assertion is pointwise in q, so private randomization cannot improve it. □

Proof of Theorem F.1. Average Lemma F.6 over x, use Lemma F.3, and then use $a \le 25\sqrt{n}$. The Bayes risk is at least

$$\frac{1}{8}\frac{\eta(\log 2 - 1/2)}{a} \ge \frac{11(\log 2 - 1/2)}{40600\sqrt{n}}.$$

The prior may depend on n, as is standard in a fixed-horizon minimax lower bound. The maximum risk dominates this Bayes average. Every prior model has the aperiodicity and standardized-support properties stated in the theorem, so the restricted-class assertion follows as well. ☐

Proof of Theorem 2.2. The hazard-substitution statement is Proposition A.1. The genuine upper bound is Theorem D.3, and Theorem F.1 supplies the matching lower bound from $n = 600$ onward. Since $\mathfrak{R}_n$ is the infimum over all predictors, it is at most the worst-case risk of the displayed uniform polynomial-bit-time family. Together these statements give (2.9). ☐

Remark F.7 (Why conditioning on the high support matters). When $J > a$, the factor $U_J(\mathsf{S})$ can vary substantially across different high supports. It cannot be discarded, nor can one compare all compatible complete supports by a single constant ratio. The proof above first fixes H, compares only L, and then marginalizes. This is the step that simultaneously preserves the stationary left boundary and keeps the event $\mathcal{E}$ observable.

G Anytime cumulative log-loss prediction

The prefix algorithm also solves a different, genuinely online problem. For a sequential predictor $\pi = (\pi_t)_{t\ge 0}$, define its cumulative conditional KL risk through time N by

$$\mathcal{C}_N(\pi,\mu) := \sum_{t=0}^{N-1} \mathbb{E}_\mu d\big(P_\mu(X_{t+1} = 1 \mid X^t)\|\pi_t(X^t)\big)\,. \tag{G.1}$$

For a randomized predictor, the expectation also averages over its private randomness. When a full binary conditional pmf is needed, write

$$\pi_t(b \mid x^t) := \pi_t(x^t)^b\big(1 - \pi_t(x^t)\big)^{1-b}, \qquad b \in \{0,1\}.$$

At a known horizon, the sequential conditionals of Q_N satisfy the pathwise bound (C.13) by Corollary C.4. We next remove knowledge of the terminal horizon.

Partition the prediction times into epochs of lengths $L_j = 2^j$, $j = 0, 1, \ldots$, where epoch j starts after $s_j = 2^j - 1$ observations. At its start, discard the earlier data for the purpose of choosing probabilities and use the sequential conditionals of Q_{L_j} on the within-epoch prefix. The schedule is fixed, so this defines one horizon-free predictor π^{any}, with $\pi_0^{\text{any}}(\varnothing) = Q_1(X_1 = 1) = 1/2$.

Theorem G.1 (Anytime cumulative KL risk)**.** *The horizon-free predictor π^{any} is uniform polynomial-bit-time and, for every $N \ge 1$,*

$$\begin{aligned}\sup_{\mu: m_\mu < \infty} \mathcal{C}_N(\pi^{\text{any}},\mu) &\le \sum_{j=0}^{\lfloor \log_2 N\rfloor} \big[B_{2^j} + \log(2^j + 1)\big] \\ &\le c_{\mathrm{p}}(2+\sqrt{2})\sqrt{N} + \frac{3\log 2}{2}(J_N+1)(J_N+2),\end{aligned} \tag{G.2}$$

where $J_N = \lfloor \log_2 N\rfloor$. In particular, the bound is $O(\sqrt{N} + \log^2 N) = O(\sqrt{N})$. Conversely,

$$\inf_{\pi} \sup_{\mu: m_\mu < \infty} \mathcal{C}_N(\pi,\mu) = \Theta(\sqrt{N}) \qquad (N \to \infty), \tag{G.3}$$

even when the infimum allows horizon-aware randomized predictors.

Proof. Consider an epoch of scheduled length L for which only its first $m \le L$ symbols have been observed by the terminal time. Write

$$A = X^s, \qquad Y = X_{s+1}^{s+m},$$

where s is the number of observations before the epoch, and let Q_L^m be the length-m prefix marginal of Q_L. The expected cumulative conditional loss in this part of the epoch is exactly

$$\mathbb{E} \log \frac{P_\mu(Y \mid A)}{Q_L^m(Y)} = I_\mu(A; Y) + D_{\mathrm{KL}}(P_\mu^m, Q_L^m), \tag{G.4}$$

where stationarity identifies the marginal law of Y with P_μ^m.

The full-block pointwise domination in Proposition B.2 can be summed over all length-$(L-m)$ extensions, giving

$$P_\mu^m(y^m) = \sum_{v \in \{0,1\}^{L-m}} P_\mu^L(y^m v) \le e^{B_L} \sum_{v \in \{0,1\}^{L-m}} Q_L(y^m v) = e^{B_L} Q_L^m(y^m).$$

Hence the second term of (G.4) is at most B_L. For the first term, let

$$Z := \min\{r \ge 1 : X_{s+r} = 1\}$$

be the inclusive countdown from the first position of the epoch to its next renewal; it is finite almost surely because the interarrival law is supported on the positive integers and has finite mean. Conditional on $Z \le m$, the zeros before the renewal and the renewal at position Z are fixed, and regeneration makes the remaining part of Y independent of A. Conditional on $Z > m$, $Y = 0^m$. Hence $A \perp Y \mid Z$, or equivalently $A \to Z \to Y$. Moreover, with $\widetilde{Z} = \min\{Z, m+1\}$, the conditional law of Y given Z depends only on $\widetilde{Z}$: the exact first-renewal position is retained when it lies in the epoch, and all values $Z > m$ give the same zero word. Therefore

$$I_\mu(A; Y) \le I_\mu(Z; Y) \le H(\widetilde{Z}) \le \log(m+1).$$

Every complete or final partial epoch consequently costs at most $B_L + \log(L+1)$.

For $L \ge 1$, the inequality $1 + 2\sqrt{L}/c_{\mathrm{p}} \le L + 1$ gives $B_L \le c_{\mathrm{p}}\sqrt{L} + 2\log(L+1)$. The geometric sum obeys

$$\sum_{j=0}^{J_N} 2^{j/2} \le (2 + \sqrt{2})\sqrt{N},$$

and $\log(2^j + 1) \le (j+1)\log 2$. Summing proves (G.2). By Proposition C.3, all conditionals in a length-L epoch use $O(L^5)$ arithmetic operations; summing the fifth powers of all complete epochs and the final partial epoch gives $O(N^5)$ operations through time N, because its scheduled length is at most N.

Finally, every deterministic sequential predictor induces the assignment $Q_\pi(x^N) = \prod_{t=0}^{N-1} \pi_t(x_{t+1} \mid x^t)$, and $\mathcal{C}_N(\pi, \mu) = D_{\mathrm{KL}}(P_\mu^N, Q_\pi)$. Conversely, every full-support assignment factors into sequential conditionals.

For this cumulative problem, use a separate occupancy prior from the one in the fixed-time lower bound. Let $a = \lceil \sqrt{N} \rceil$, choose an a-subset V uniformly from $[2a]$, and let μ_{V} be uniform on it. Put $k = \lfloor (N - 2a)/(2a) \rfloor$. The equilibrium delay and every subsequent gap are at most $2a$, so X^N reveals the first k complete iid gaps $D_1, \dots, D_k$ following its first visible renewal. Given a realization

$D^{i-1} = d^{i-1}$, let $r(d^{i-1})$ be its number of distinct values. The posterior on V is uniform over all compatible supports, and a direct entropy calculation gives

$$I(\mathsf{V}; D_i \mid D^{i-1} = d^{i-1}) = \left(1 - \frac{r(d^{i-1})}{a}\right) \log \frac{2a - r(d^{i-1})}{a - r(d^{i-1})}. \tag{G.5}$$

Indeed, a previously seen symbol has predictive mass $1/a$, whereas each of the $2a - r(d^{i-1})$ unseen candidates has mass $(a - r(d^{i-1}))/(a(2a - r(d^{i-1})))$. Since $r(d^{i-1}) \le i - 1 \le k \le a/2$, the right-hand side of (G.5) is at least $(\log 2)/2$ for every realized history. Averaging over D^{i-1} and applying the mutual-information chain rule, for $N \ge 100$, $k \ge \sqrt{N}/4$, whence

$$I(\mathsf{V}; X^N) \ge I(\mathsf{V}; D^k) \ge \frac{\log 2}{8}\sqrt{N}.$$

For every assignment Q, the Bayes identity

$$\mathbb{E}_{\mathsf{V}} D_{\mathrm{KL}}(P_{\mathsf{V}}^N, Q) = I(\mathsf{V}; X^N) + D_{\mathrm{KL}}(P_{X^N}, Q)$$

therefore gives the required $\Omega(\sqrt{N})$ Bayes lower bound for every deterministic assignment. If a horizon-aware randomized predictor has private seed R, set

$$Q_R(x^N) := \prod_{t=0}^{N-1} \pi_{t,R}(x_{t+1} \mid x^t), \qquad \overline{Q} := \mathbb{E}_R Q_R.$$

Convexity of KL in its second argument, followed by the preceding deterministic assignment bound, gives

$$\mathbb{E}_{\mathsf{V},R} D_{\mathrm{KL}}(P_{\mathsf{V}}^N \| Q_R) \ge \mathbb{E}_{\mathsf{V}} D_{\mathrm{KL}}(P_{\mathsf{V}}^N \| \overline{Q}) \ge \frac{\log 2}{8}\sqrt{N}.$$

The maximum risk dominates this Bayes average, so private randomization cannot improve the order. Together with the upper bound, this also recovers the classical renewal redundancy rate [5, 7]. □

H Why direct hazard estimation is delicate

The coding construction avoids a stopping bias that complicates the more obvious empirical estimator. For an ordinary renewal process, set $S_0 = 0$ and $S_j = \sum_{i=1}^{j} T_i$. The number of completed sampled gaps of length g by a time budget H is

$$C_g(H) := \sum_{j \ge 0} \mathbf{1}\{S_j \le H - g,\ T_{j+1} = g\}.$$

This definition excludes both censored boundary intervals. Let

$$U(t) := \sum_{j \ge 0} \mathbb{P}(S_j \le t)$$

be the renewal function, with $U(t) = 0$ for $t < 0$. For $1 \le g \le H$, direct summation over the index of the gap gives

$$\mathbb{E} C_g(H) = \sum_{j \ge 0} \mathbb{P}(S_j \le H - g)\mu(g) = \mu(g) U(H - g). \tag{H.1}$$

The multiplier depends on g, so ratios of completed-gap tail counts are biased. Incorporating the boundary-crossing gap as censored survival data corrects the corresponding likelihood contribution, but a sharp KL analysis under this outcome-dependent stopping rule remains nontrivial. A uniform pseudo-count over gap lengths also assigns artificial mass across long zero-hazard plateaus, so without a support or tail cutoff it does not directly give a uniform KL guarantee.

Proposition H.1 (Known bounded-support plug-in)**.** *Suppose* $\operatorname{supp}(\mu) \subseteq [B]$*, where* $B \le n/2$ *is known. There is a direct add-one hazard predictor* $\widehat{h}$ *satisfying*

$$\mathbb{E}d(h_\mu(A)\|\widehat{h}(A)) \le \frac{B-1}{m_\mu \lfloor n/B \rfloor} \le \frac{B(B-1)}{m_\mu(n-B)} \le \frac{2B^2}{m_\mu n} \le \frac{2B^2}{n}, \tag{H.2}$$

where A *is the age at time* n*.*

Proof. Put $k = \lfloor n/B \rfloor - 1$. The last renewal $K \le n$ satisfies $K > n - B$, because the current gap has length at most B and crosses time n. If $D_{-1}, \ldots, D_{-k}$ are its preceding k gaps, then

$$K - \sum_{i=1}^{k} D_{-i} \ge K - kB > n - B - kB \ge 0.$$

The left side is an integer renewal time and hence is at least one; all k gaps are therefore fully visible in X_1^n. By (F.3), these backward gaps are iid μ and independent of the current age $A := n - K$. Moreover, $T_0 \le B \le n/2$ almost surely, so the all-zero history has probability zero and the genuine next-step conditional is $h_\mu(A)$. Let $\widehat{\mu}$ be their add-one estimator and let $\widehat{h}$ be its hazard sequence. For any law ν with full support on $[B]$, the hazard chain rule gives

$$D_{\mathrm{KL}}(\mu, \nu) = \sum_{a=0}^{B-1} \overline{F}_\mu(a) d(h_\mu(a)\|h_\nu(a)), \tag{H.3}$$

with zero-survival terms interpreted as zero. Indeed, the factors $h(a)$ and $1 - h(a)$ are the transition probabilities in the sequential survival representation of a draw from the law, so (H.3) is the KL chain rule for that representation.

For completeness, if N_j is the count of value j among the k backward gaps, then $N_j \sim \mathrm{Bin}(k, \mu(j))$, $\widehat{\mu}(j) = (N_j + 1)/(k + B)$ and, for $\mu(j) > 0$,

$$\mathbb{E}\frac{1}{N_j + 1} = \frac{1 - (1 - \mu(j))^{k+1}}{(k+1)\mu(j)}. \tag{H.4}$$

This follows by writing $\binom{k}{r}/(r+1) = \binom{k+1}{r+1}/(k+1)$ and summing the binomial expansion. Jensen's inequality and (H.4) yield

$$\begin{aligned}
\mathbb{E}D_{\mathrm{KL}}(\mu, \widehat{\mu}) &= \sum_{j:\mu(j)>0} \mu(j)\mathbb{E}\log\frac{\mu(j)(k+B)}{N_j+1} \\
&\le \sum_{j:\mu(j)>0} \mu(j)\log\left(\mu(j)(k+B)\mathbb{E}\frac{1}{N_j+1}\right) \\
&= \sum_{j:\mu(j)>0} \mu(j)\log\left[\frac{k+B}{k+1}\left(1-(1-\mu(j))^{k+1}\right)\right] \\
&\le \log\frac{k+B}{k+1} \le \frac{B-1}{k+1}.
\end{aligned}$$

Since the stationary age law is $\mathbb{P}(A = a) = \overline{F}_\mu(a)/m_\mu$, (H.3) therefore gives

$$\begin{aligned}
\mathbb{E}d(h_\mu(A)\|\widehat{h}(A)) &= \frac{1}{m_\mu}\mathbb{E}D_{\mathrm{KL}}(\mu, \widehat{\mu}) \le \frac{B-1}{m_\mu(k+1)} \\
&= \frac{B-1}{m_\mu\lfloor n/B \rfloor} \le \frac{B(B-1)}{m_\mu(n-B)} \\
&\le \frac{2B^2}{m_\mu n} \le \frac{2B^2}{n}.
\end{aligned} \tag{H.5}$$

Thus the direct empirical method reaches the worst-case $O(n^{-1/2})$ benchmark whenever $B^2/m_\mu = O(\sqrt{n})$, and in particular whenever $B = O(n^{1/4})$ without using a lower bound on the mean. The assignment construction removes the known-support restriction entirely. $\square$