THE PRICE OF A FIXED SCALE: EXACT REGRET CONSTANTS AND ADAPTIVE SCALING FOR TSALLIS-INF

Anonymous authors
Paper under double-blind review

ABSTRACT

Best-of-both-worlds bandit algorithms are usually compared by regret rates, but their stochastic leading constants determine whether robustness preserves asymptotic efficiency. We resolve this question for reduced variance $1/2$-Tsallis-INF. For every fixed scale $c > 0$ and fixed K, its Bernoulli small gap coefficient is exactly $C(c) = 4/c^2 + 3/8 + c^2/128$ under an explicit admissibility condition. The optimum is $(3 + 2\sqrt{2})/8 \approx 0.7286$, strictly above the local Lai–Robbins constant $1/2$, while the standard choice $c = 4$ yields $3/4$. Thus every fixed scale under this limit misses local Lai–Robbins efficiency. The formula points to local variance as the missing quantity. Guided by this decomposition, we adapt the scale using sparse probes that require neither instance parameters nor arm order. For two stochastic arms, a fixed clip family approaches $2m(1-m)$ for every interior mean, and a moving clip learner attains $1/2$ on the central family under ordered limits. The proof identifies a harmonic occupation law and controls the singular x^{-2} regret cost using recovery estimates and inverse moments. The adaptive learners do not retain adversarial guarantees, and we do not provide a joint finite horizon and small gap rate.

1 INTRODUCTION

Best of both worlds bandit algorithms use one policy in adversarial and stochastic environments. Tsallis-INF is a canonical example: without knowing the regime, it achieves minimax adversarial regret and logarithmic stochastic regret (Zimmert & Seldin, 2021). The logarithmic rate does not say how often a suboptimal arm is pulled; that information lies in its leading constant.

For Bernoulli bandits, the Lai–Robbins lower bound gives the optimal constant for each instance (Lai & Robbins, 1985). Near mean $1/2$, a suboptimal arm with gap Δ_i contributes $1/(2\Delta_i)$. We ask whether reduced variance $1/2$-Tsallis-INF can attain this constant by choosing the scale c in $\eta_t = c/\sqrt{t}$. Existing analyses establish the logarithmic rate but do not answer this question: the original bound is within a factor two of the benchmark (Zimmert & Seldin, 2021), a later analysis gives leading coefficient 1 before $\sum_{i\neq 1} \Delta_i^{-1} \log T$ (Masoudian & Seldin, 2021), and adaptive learning rates have been analyzed at the level of regret rates rather than the exact constant of this recursion (Ito et al., 2024).

We answer with an equality rather than a bound. Fix K and $c > 0$, take the horizon limit, and then let the gaps vanish along any admissible Bernoulli sequence from Theorem 1. The normalized coefficient is exactly

$$C(c) = \frac{4}{c^2} + \frac{3}{8} + \frac{c^2}{128}.$$

The first and last terms favor opposite choices of c, and their balance at $c = 4\,2^{1/4}$ gives $C(c) = (3 + 2\sqrt{2})/8 \approx 0.7286$. Under the normalization of Zimmert & Seldin (2021) the standard reduced variance choice is $c = 4$, with coefficient $3/4$. The middle term carries no c at all, so tuning the scale cannot remove the excess over $1/2$, and the reported factor two is in fact 1.457. A one sided bound cannot separate these two statements, because the obstruction is a term of the exact coefficient rather than a slack in an estimate.

The decomposition also suggests how to change the learner. A fixed baseline leaves residual variance $1/4$, whereas the local information bound depends on $m(1-m)$. We let the scale depend on a

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

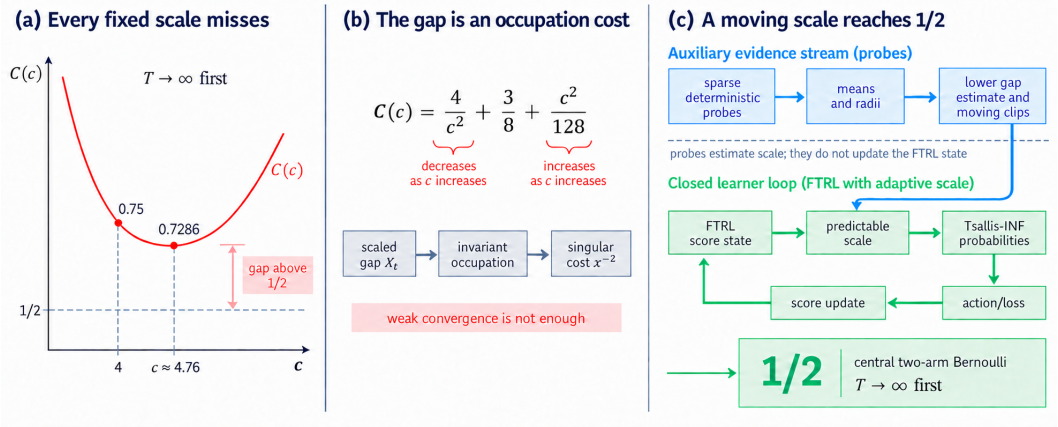

Figure 1: Fixed scales and the adaptive repair. (a) For every fixed scale, the exact small gap coefficient $C(c)$ stays above the local Lai–Robbins value $1/2$; the curve is analytic, not empirical. (b) The coefficient is a singular inverse square cost under the invariant occupation law of the scaled gap, so weak convergence alone does not identify regret. (c) Sparse deterministic probes estimate the gap and choose moving clips without updating the FTRL score state. The resulting predictable scale gives $\lim_{\Delta\downarrow 0}\lim_{T\to\infty}\Delta R_T/\log T = 1/2$ on the central two arm Bernoulli family.

predictable estimate of the gap to variance ratio. Sparse probes estimate this ratio without the true arm order, gap, mean, or variance. A fixed clip family approaches $2m(1-m)$ when the horizon, gap, and clip limits are taken in order. For means $1/2$ and $1/2+\Delta$, moving the clips gives one policy with $\lim_{\Delta\downarrow 0}\lim_{T\to\infty}\Delta R_T/\log T = 1/2$.

Figure 1 summarizes the fixed scale obstruction, its occupation mechanism, and the adaptive repair.

The main technical difficulty is the singular regret cost. On logarithmic time, the scaled loss gap converges to a piecewise deterministic Markov process, and the regret coefficient is the integral of x^{-2} against its invariant law. Since this integrand diverges near zero, weak convergence of occupation measures is insufficient. We first prove that the learner spends only finite expected time with the wrong leading arm, then control the singularity with recovery estimates and inverse moment potentials. A diffusion limit evaluates the invariant integral. The same argument also gives an exact coefficient for plain importance weighting.

2 Related Work and Scope

For stochastic bandits, the Lai–Robbins lower bound (Lai & Robbins, 1985) sets the instance-dependent benchmark, while UCB gives a finite-time logarithmic guarantee (Auer et al., 2002a). The benchmark is attained by KL-UCB (Cappé et al., 2013), Thompson sampling (Kaufmann et al., 2012; Agrawal & Goyal, 2013), and KL-Maillard Sampling (Qin et al., 2023). The behavior of Thompson sampling can depend on its prior outside one-parameter models (Honda & Takemura, 2014). Other work combines asymptotic and minimax guarantees (Jin et al., 2021), identifies optimal fixed-confidence allocations (Garivier & Kaufmann, 2016), studies asymptotically optimal allocation and information-directed sampling in linear bandits (Lattimore & Szepesvári, 2017; Kirschner et al., 2021), or develops a non-asymptotic language for instance optimality (Wagenmaker & Foster, 2023). These results specify what an efficient learner can achieve. We ask whether one fixed Tsallis recursion actually attains that benchmark.

The adversarial lineage starts with Exp3 and its importance-weighted loss estimator (Auer et al., 2002b). Implicit exploration sharpens high-probability control without uniform mixing (Neu, 2015a), while first-order analyses adapt to small cumulative loss (Neu, 2015b; Gaillard et al., 2014). Parameter-free expert methods provide a parallel form of stochastic adaptivity (Luo & Schapire, 2015), and Tsallis regularization also appears in delayed composite-feedback bandits (Cesa-Bianchi et al., 2021). These papers analyze regret bounds. Our occupation calculation instead resolves the long-run sampling cost of a specific importance-weighted recursion.

The best of both worlds line asks one learner to perform well under both stochastic and adversarial losses. Early constructions tested the regime and switched policies (Bubeck & Slivkins, 2012; Auer & Chiang, 2016); Exp3 variants then used arm-dependent exploration and gap estimates (Seldin & Slivkins, 2014; Seldin & Lugosi, 2017). Adaptive regularization and hybrid bounds removed further tuning requirements (Wei & Luo, 2018; Ito, 2021), and black-box reductions now convert FTRL or mirror-descent algorithms into best of both worlds learners (Dann et al., 2023).

This program extends beyond ordinary multi-armed bandits. Feedback-graph structure changes the attainable adversarial rates (Alon et al., 2015), and Kong, Zhou, and Li obtain simultaneous stochastic and adversarial guarantees for general graph feedback (Kong et al., 2022). Chen, Zhao, and Li study the same objective for position-based ranking feedback (Chen et al., 2022). For linear bandits, near instance-optimal and minimax guarantees can be combined (Lee et al., 2021); FTRL, variance adaptation, and exploration by optimization yield best-of-three-worlds results (Kong et al., 2023; Ito, 2023; Ito & Takemura, 2023). Related guarantees now cover linear contextual bandits (Kuroki et al., 2024). Adaptive learning-rate rules and real-time stability-penalty matching further connect the stability and penalty terms to best of both worlds and data-dependent bounds (Tsuchiya & Ito, 2024; Nguyen et al., 2025). Our result is narrower in model class but sharper in its target: it identifies the exact coefficient and shows that no fixed scale reaches the stochastic benchmark.

Algorithm-specific failure results are the closest methodological precedent. Kong et al. show that Thompson sampling with a greedy approximation oracle can incur a sharp additional cost in combinatorial semi-bandits (Kong et al., 2021). Kalvit and Zeevi characterize UCB arm-sampling rates and their small-gap diffusion behavior (Kalvit & Zeevi, 2021); recent work derives precise and sometimes non-deterministic sampling asymptotics for variance-aware UCB (Fan et al., 2025). These works motivate studying the algorithm itself, rather than stopping at a worst-case guarantee.

Zimmert & Seldin (2021) analyze the reduced variance estimator considered here and prove a stochastic leading constant within a factor two of the benchmark. Their regularizer is $\Psi(p) = -4\sum_i \sqrt{p_i}$ and their learning rate is $\eta_t = c/\sqrt{t}$; their standard reduced variance setting corresponds to $c = 4$ in our notation. Tsallis regularization has also been adapted to decoupled exploration and exploitation (Rouyer & Seldin, 2020). Subsequent analyses sharpen stochastic and corruption bounds (Masoudian & Seldin, 2021), obtain hybrid data-dependent guarantees (Ito, 2021), and derive variance-dependent bounds with optimistic FTRL (Ito et al., 2022). Stability and information-ratio analyses give another general view of mirror descent (Lattimore & György, 2021), while adaptive learning rates can be chosen by balancing upper-bound terms (Ito et al., 2024). None of these results computes the invariant law or exact coefficient of the fixed-scale recursion.

For plain importance weighting, Zhan et al. (2025) prove a last iterate lower bound that already implies a qualitative failure of the Lai–Robbins constant over part of their scale range. Their estimator and target quantity differ from ours, and their bound does not give a positive floor over all scales. A later paper uses a heuristic diffusion of the same gap process to design a Lyapunov function for best arm identification (Zhan et al., 2026). Here the process limit and its invariant regret cost are proved. Diffusion approximations for Thompson sampling consider a different regime in which the gap shrinks jointly with a finite horizon (Fan & Glynn, 2021); our horizon tends to infinity before the gap vanishes.

The occupation argument also uses four sources from stochastic process theory. Semimartingale characteristics and their predictable parametrization follow Jacod & Shiryaev (2003). The right process and extended generator facts for piecewise deterministic Markov processes come from Davis (1993). We verify the hypotheses of the martingale problem occupation theorem of Bhatt et al. (2006), and use the continuous time Foster–Lyapunov criteria of Meyn & Tweedie (1993) for positive Harris recurrence. Appendix A records the exact results used and checks their assumptions for our process.

Thus the paper's scope is the exact efficiency of Tsallis/FTRL sampling: a fixed-K obstruction for reduced variance and plain importance weighting, a singular occupation method that identifies the coefficient, and a two-arm variance-adaptive repair under ordered limits.

Question	Earlier results	Result proved here
Exact RV constant	Tsallis-INF upper bounds; plain IW lower bound	Fixed K, admissible Bernoulli sequences, every fixed scale
Information efficiency	KL-UCB, Thompson sampling, KL-MS	Tsallis/FTRL construction, two stochastic arms
Adaptive FTRL scale	Bounds from stability and penalty	Exact occupation coefficient and ordered local limit
Process approximation	Bandit diffusion limits; heuristic Tsallis diffusion	Invariant occupation law after fixing the gap

3 SETUP AND ASSUMPTIONS

We consider $K \geq 2$ arms with independent Bernoulli losses of means $\mu_1, \ldots, \mu_K$. Arm 1 is uniquely optimal, and we write $\Delta_i = \mu_i - \mu_1$ for $i \geq 2$. The exact learner, initialization, filtration, probability map, baseline, and estimator are defined next.

3.1 THE EXACT LEARNER AND ITS PROBABILITY MAP

Let $\Delta_K = \{p \in [0,1]^K : \sum_{a=1}^K p_a = 1\}$. The losses $\ell_{t,a}$ are independent Bernoulli random variables with means μ_a, independent across rounds and of the learner's randomization. Write $\mathcal{F}_{t-1}$ for the information available before round t. Initialize $\widehat{L}_0 = 0$. At round t, with $\eta_t = c/\sqrt{t}$, the learner chooses the unique minimizer

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\{ \langle p, \widehat{L}_{t-1} \rangle - \frac{4}{\eta_t} \sum_{a=1}^K \sqrt{p_a} \right\}. \tag{1}$$

Thus p_t is $\mathcal{F}_{t-1}$-measurable and $\mathbb{P}(I_t = a \mid \mathcal{F}_{t-1}) = p_{t,a}$. After observing ℓ_{t,I_t}, set

$$B_t(a) = \frac{1}{2}\mathbf{1}\{p_{t,a} \geq \eta_t^2\}, \qquad \widehat{\ell}_{t,a} = B_t(a) + \frac{\mathbf{1}\{I_t = a\}}{p_{t,a}}(\ell_{t,a} - B_t(a)), \qquad \widehat{L}_t = \widehat{L}_{t-1} + \widehat{\ell}_t. \tag{2}$$

All coordinates of p_t are strictly positive, so the estimator is defined; moreover $\mathbb{E}[\widehat{\ell}_{t,a} \mid \mathcal{F}_{t-1}] = \mu_a$. The objective in (1) is the symmetric order-$1/2$ Tsallis regularizer: on the simplex it differs only by a constant from $\eta_t^{-1} \sum_a (-4\sqrt{p_a} + 2p_a)$.

Exact KKT representation. Let $M_t = \min_a \widehat{L}_{t-1,a}$, $\overline{L}_{t-1,a} = \widehat{L}_{t-1,a} - M_t$, and choose any $m_t \in \operatorname{argmin}_a \widehat{L}_{t-1,a}$ for notation. There is a unique $w_t > 0$ satisfying

$$p_{t,a} = \frac{4}{(\eta_t \overline{L}_{t-1,a} + w_t)^2}, \qquad \sum_{a=1}^K \frac{4}{(\eta_t \overline{L}_{t-1,a} + w_t)^2} = 1. \tag{3}$$

Indeed, the derivative of the objective in (1) is $\widehat{L}_{t-1,a} - 2/(\eta_t\sqrt{p_a})$; the derivative tends to $-\infty$ at the boundary, so the minimizer is interior, and the Lagrange equations give (3). In particular,

$$w_t = \frac{2}{\sqrt{p_{t,m_t}}}, \qquad 2 \leq w_t \leq 2\sqrt{K}. \tag{4}$$

Every empirical minimizer has the same probability, so the representation is independent of the choice of m_t. This also fixes all ties, including $p_{1,a} = 1/K$ for every a.

For $i \geq 2$, put

$$D_{t-1,i} = \widehat{L}_{t-1,i} - \widehat{L}_{t-1,1}, \qquad X_{t-1,i} = \frac{D_{t-1,i}}{\Delta_i(t-1)} \quad (t \geq 2), \qquad \varepsilon_t = \sum_{i=2}^K p_{t,i}. \tag{5}$$

More generally, if $q_t = -\min\{0, \min_{i \geq 2} D_{t-1,i}\}$, then $\overline{L}_{t-1,1} = q_t$ and $\overline{L}_{t-1,i} = D_{t-1,i} + q_t$. Hence (3) determines p_t from (t, X_{t-1}) and defines the transition kernel used below, which varies with time. On

$$G_t = \{D_{t-1,i} \geq 0 \text{ for every } i \geq 2\}, \tag{6}$$

arm 1 is an empirical minimizer, $q_t = 0$, and

$$\frac{1}{\sqrt{p_{t,i}}} = \frac{\eta_t D_{t-1,i}}{2} + \frac{1}{\sqrt{p_{t,1}}}, \qquad p_{t,i} D_{t-1,i} = \frac{2}{\eta_t}\left(\sqrt{p_{t,i}} - \frac{p_{t,i}}{\sqrt{p_{t,1}}}\right), \qquad w_t = \frac{2}{\sqrt{1-\varepsilon_t}}. \tag{7}$$

If $0 \le \varepsilon_t \le 1/2$, then the last identity obeys the sharper bound

$$2 \le w_t \le 2 + 2\varepsilon_t. \tag{8}$$

Threshold at a finite time and compact asymptotics. On G_t, the condition for the active baseline is exactly

$$B_t(i) = \frac{1}{2} \quad \Longleftrightarrow \quad X_{t-1,i} \le x^{\mathrm{sw}}_{t,i}(w_t) := \frac{2t/c^2 - w_t\sqrt{t}/c}{\Delta_i(t-1)}; \tag{9}$$

equality is active. Thus $x^{\mathrm{sw}}_{t,i}(w_t) \to x^0_i := 2/(c^2\Delta_i)$ whenever $w_t = O(1)$.

For a compact set $Q \subset (0,\infty)^{K-1}$, write $a_i = \min_{x\in Q} x_i > 0$. If $X_{t-1} \in Q$ and G_t holds, then, for $t \ge 2$,

$$p_{t,i} \le \frac{16}{c^2\Delta_i^2 a_i^2 t}, \qquad \varepsilon_t \le \frac{C_Q}{t}, \qquad C_Q := \frac{16}{c^2}\sum_{i=2}^{K}\frac{1}{\Delta_i^2 a_i^2}. \tag{10}$$

Consequently, once $t \ge 2C_Q$, (8) applies and, uniformly for $x = X_{t-1} \in Q$,

$$t p_{t,i} = \frac{4}{c^2\Delta_i^2 x_i^2(1-1/t)^2 + 2w_t c\Delta_i x_i(1-1/t)/\sqrt{t} + w_t^2/t} = \frac{4}{c^2\Delta_i^2 x_i^2} + O_Q(t^{-1/2}). \tag{11}$$

In particular, on compact sets separated from the hyperplane $x_i = x_i^0$, (9) eventually selects the same baseline branch as $x_i \le x_i^0$.

Exact coordinate update. For later conditioning arguments let $Z_{t,i} = D_{t,i} - D_{t-1,i} = \widehat{\ell}_{t,i} - \widehat{\ell}_{t,1}$. Whenever $B_t(1) = 1/2$, its possible values are

played arm	$B_t(i) = 1/2$	$B_t(i) = 0$
$I_t = 1$	$-(\ell_{t,1} - 1/2)/p_{t,1}$	$-1/2 - (\ell_{t,1} - 1/2)/p_{t,1}$
$I_t = i$	$(\ell_{t,i} - 1/2)/p_{t,i}$	$\ell_{t,i}/p_{t,i} - 1/2$
$I_t = j \notin \{1, i\}$	0	$-1/2$

and the rescaled state satisfies the exact identity

$$X_{t,i} = \left(1 - \frac{1}{t}\right) X_{t-1,i} + \frac{Z_{t,i}}{\Delta_i t}. \tag{12}$$

Equations (3), (2), and (12) therefore specify the full transition kernel for one step; no additional tie rule is needed.

Its cumulative pseudo regret is $R_T = \sum_{i=2}^{K} \Delta_i \sum_{t=1}^{T} \mathbb{E} p_{t,i}$.

Put

$$r = \frac{c^2}{16}, \qquad C(c) = \frac{(r+1)(r+2)}{8r}.$$

For $0 < \Delta \le 1/2$, define the admissible scalar parameter set

$$\Theta_\Delta(c) = \left\{\theta : \ \frac{1}{2} - \theta\Delta \in [0, 1-\Delta], \quad \theta < \frac{2}{c^2\Delta}\right\}. \tag{13}$$

4 Fixed scale asymptotics

Theorem 1 (Exact coefficient for fixed K in the small gap limit)**.** *Fix $K \ge 2$ and $c > 0$. For each Bernoulli instance suppose that*

$$0 \le \mu_1 < \mu_i = \mu_1 + \Delta_i \le 1, \qquad 0 < \Delta_i \le \frac{1}{2}, \qquad c^2\Big(\frac{1}{2} - \mu_1\Big) < 2 \tag{14}$$

for every $i \geq 2$. Then, for that fixed instance,

$$\lim_{T\to\infty} \frac{R_T}{\log T} = \sum_{i=2}^{K} \frac{4}{c^2\Delta_i} \int_0^\infty x^{-2}\, \pi_{\Delta_i,\theta_i}(dx), \qquad \theta_i = \frac{1/2-\mu_1}{\Delta_i}, \tag{15}$$

where $\pi_{\Delta,\theta}$ is the unique invariant law of the scalar piecewise deterministic Markov process defined in (24) *below. Consequently, for every sequence of instances satisfying* (14)*, with the same K, c and $\max_{i\geq 2} \Delta_i \to 0$,*

$$\lim_{n\to\infty} \frac{\lim\limits_{T\to\infty} R_T^{(n)}/\log T}{\sum\limits_{i=2}^{K} (\Delta_i^{(n)})^{-1}} = C(c). \tag{16}$$

Moreover,

$$\inf_{c>0} C(c) = \frac{3+2\sqrt{2}}{8}, \qquad \textit{attained when } c^2/16 = \sqrt{2}.$$

The scientific point is architectural: no tuning of the single fixed scale can match the local information geometry. The coefficient in (16) can be written as

$$C(c) = \frac{4}{c^2} + \frac{3}{8} + \frac{c^2}{128}.$$

The order of limits is part of the statement: the horizon tends to infinity before the gaps vanish along an admissible sequence. The terms $4/c^2$ and $c^2/128$ move in opposite directions as the scale changes. The arithmetic–geometric mean inequality gives $4/c^2 + c^2/128 \geq 1/(2\sqrt{2})$, with equality only at $c = 4\,2^{1/4}$. Thus even the best fixed scale has coefficient strictly larger than $1/2$.

The sequence formulation in (16) is essential. The scope condition is equivalent to $\theta_i \in \Theta_{\Delta_i}(c)$. This permits $\theta_i < 0$ when $\mu_1 > 1/2$ and $\theta_i > 1$ when both means lie below $1/2$, provided the switching surface remains strictly transversal.

Corollary 1 (Finite expected time with an incorrect leading arm)**.** *For each fixed instance and scale in the theorem,*

$$\mathbb{E} \sum_{t\geq 1} \mathbf{1}\{\min_{i\geq 2} D_{t-1,i} \leq 0\} < \infty.$$

This includes an incorrect empirical leader and a tie involving arm 1. Ties only among suboptimal arms are not counted.

Appendix C derives this corollary from (53) after verifying the hypotheses of Theorem 4.

To see where the coefficient comes from, let D_t denote an estimated loss gap. The KKT equations give $tp_t \asymp t^2/D_t^2$, so regret is governed by a harmonic occupation average of an inverse square cost. On logarithmic time, the scaled gap converges to a piecewise deterministic Markov process. Weak convergence alone cannot pass this singular cost to the limit. A comparison potential first controls recovery of the full gap vector; strip estimates and inverse moment bounds then give the required uniform integrability. The invariant integral is finally evaluated through a diffusion limit that is uniform as the gap vanishes. Appendices A, B, and C contain these three steps.

The same occupation argument also isolates the effect of the loss estimator. Theorem 6 evaluates plain importance weighting for every valid mean parameter and bounds the gap error explicitly. For the common regime $\mu_1 \to 1/2$, $\max_i \Delta_i \to 0$ after the horizon limit, write $C_{\mathrm{RV}} = C$. Then

$$C_{\mathrm{IW}}(c, 1/2) = \frac{4}{c^2} + \frac{3}{4} + \frac{c^2}{32} = 2C_{\mathrm{RV}}(\sqrt{2}c). \tag{17}$$

Thus the best attainable fixed scale constants are $(3+2\sqrt{2})/4$ for IW and $(3+2\sqrt{2})/8$ for RV, at respective scales $2^{7/4}$ and $2^{9/4}$. Variance reduction halves that optimum, although the remaining RV constant still exceeds the Lai–Robbins benchmark. This comparison keeps the FTRL probability map fixed and optimizes each estimator's scale separately. Appendix G gives a finite accuracy bound for bounded RV tests; its constants are too large for a practical rate.

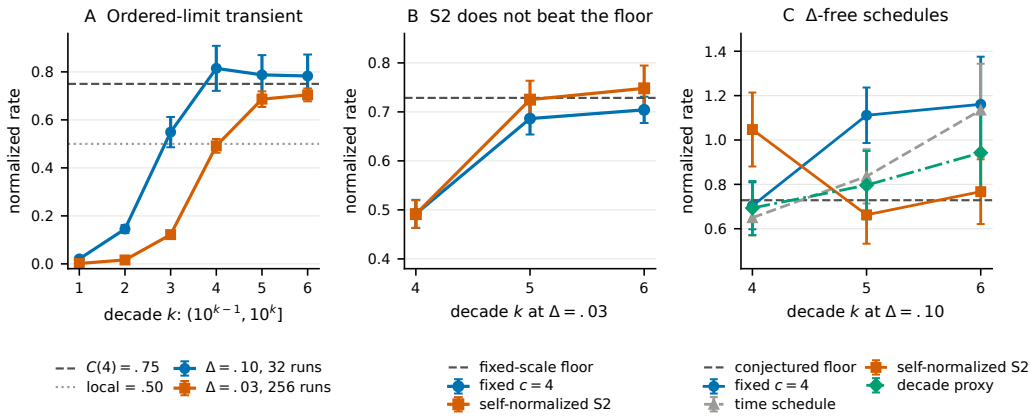

Figure 2: Three retained-data checks using the per-decade normalized rate $\Delta^2 \sum_{t\in(10^{k-1},10^k]} p_{t,2}/\log 10$; bars are standard errors. (A) The fixed minimizer approaches $C(4)$ differently at $\Delta = .10$ and .03. (B) At $\Delta = .03$, the self-normalized S2 schedule ($a_n = \mathrm{clip}(16/\hat{x}_n, 1, 256)$ after 10^3 burn in rounds, $\hat{x}_n = (D_n/n)/\hat{D}_n$, with $\hat{D}_n$ the harmonic average of D_m/m over $m \le n$) does not beat fixed $c = 4$ in the late decades. (C) At $\Delta = .10$, no tested Δ-free schedule resolves below the conjectured floor. These are finite-horizon consistency and falsification checks, not evidence for a theorem.

Figure 2 shows why a whole horizon average is a poor finite horizon diagnostic for a limit taken after the horizon. The theorem constrains the cumulative rate $\Delta R_T/\log T$, which at $T = 10^k$ is the average of the decade resolved rates; those rates are therefore a diagnostic, and if they converge their limit must equal the cumulative coefficient. In these runs the decade rate settles near $C(4)$ only once the scaled gap process has mixed.

5 Adaptive scaling for two arms

The two scale dependent terms in $C(c)$ suggest adapting the scale to the variance seen by the learner. Simply substituting an estimated optimal scale is insufficient: predictability, recovery, and the singular occupation cost must be controlled together. We first analyze such a learner for two stochastic arms. Its clip limits are taken after the horizon and gap limits.

Here s denotes elapsed time, $n(s)$ the number of ordinary FTRL updates completed by time s, and T the regret horizon on that clock; probe regret is included in R_T. If $K_P(T)$ denotes the number of probes by time T, then $K_P(T) = o(\log T)$ and $n(T) = T - K_P(T)$ is asymptotically interchangeable with T in harmonic normalizations.

At deterministic calendar times $\tau_k = \lceil k^k \rceil$, $k \ge 2$, the learner alternates forced probe pulls between the two arms. Probe outcomes update only counts N_i^P and sums S_i^P; they do not update the FTRL state. Before an ordinary learner round n, define

$$b_{n,i} = \begin{cases} 1/2, & N_i^P = 0, \\ S_i^P/N_i^P, & N_i^P > 0, \end{cases} \qquad \mathrm{rad}(N) = \sqrt{\frac{2\log(\max\{N,2\})}{N}},$$

with $\mathrm{rad}(0) = +\infty$, and put

$$\widehat{\Delta}_{L,n} = \left(|b_{n,2} - b_{n,1}| - \sum_{i=1}^{2} \mathrm{rad}(N_i^P) \right)_+ . \tag{18}$$

Let $\widehat{j}_n$ be an arm with larger probe mean, using a fixed tie rule, and set

$$\nu_n = \frac{1}{4\log(e + \min_i N_i^P)}, \qquad \widehat{v}_n = \max\{b_{n,\widehat{j}_n}(1 - b_{n,\widehat{j}_n}), \nu_n\}.$$

For fixed $0 < L < U < \infty$ and $0 < \kappa < 1/2$, write $D_{n-1} = \widehat{L}_{n-1,2} - \widehat{L}_{n-1,1}$ and define

$$r_n = \min\{\widehat{\Delta}_{L,n}/\widehat{v}_n, \kappa\}, \qquad a_n = \begin{cases} L, & D_{n-1} = 0 \text{ or } r_n = 0, \\ \mathrm{clip}\left(4nr_n/|D_{n-1}|, L, U\right), & \text{otherwise.} \end{cases} \tag{19}$$

On ordinary rounds the learner uses the exact two arm FTRL map (1) with $\eta_n = \sqrt{a_n/n}$, followed by the predictable centered baseline and unbiased estimate

$$B_{n,i} = b_{n,i}\mathbf{1}\{p_{n,i} \geq a_n/n\}, \qquad \widehat{\ell}_{n,i} = B_{n,i} + \frac{\mathbf{1}\{I_n = i\}}{p_{n,i}}(\ell_{n,i} - B_{n,i}). \tag{20}$$

Every input in (19)–(20) is known before the arm draw. The construction uses neither the true ordering, gap, mean, nor variance.

Theorem 2 (Adaptive occupation and information constant)**.** *Consider two Bernoulli arms with $0 < \mu_1 = m < \mu_2 = m + \Delta < 1$, and let $v_2 = \mu_2(1 - \mu_2)$. If*

$$\Delta/v_2 < \kappa, \qquad L\Delta < 2, \tag{21}$$

then the learner (18)–(20) *has a unique invariant occupation law $\pi^{\mathrm{VA}}_{\Delta,L,U}$ and*

$$\lim_{T\to\infty} \frac{R_T}{\log T} = \frac{1}{\Delta}\int_0^\infty \frac{4}{q_\Delta(x)x^2}\,\pi^{\mathrm{VA}}_{\Delta,L,U}(dx), \qquad q_\Delta(x) = \mathrm{clip}\left(\frac{4}{v_2 x}, L, U\right). \tag{22}$$

Moreover, for every fixed interior m,

$$\lim_{L\downarrow 0}\lim_{U\uparrow\infty}\lim_{\Delta\downarrow 0}\lim_{T\to\infty} \frac{\Delta R_T}{\log T} = 2m(1-m). \tag{23}$$

The four limits in (23) *are ordered as displayed.*

The right side of (23) is exactly the local Bernoulli Lai–Robbins coefficient because $\mathrm{kl}(m+\Delta, m) = \Delta^2/[2m(1-m)] + O(\Delta^3)$. The mechanism is visible in the diffusion obtained as the gap vanishes. Its generator is

$$\mathcal{L}f(x) = (1-x)f'(x) + \frac{m(1-m)q(x)x^2}{8}f''(x).$$

In the unclipped region $q(x) = 4/[m(1-m)x]$, this becomes $(1-x)f'(x) + (x/2)f''(x)$, whose invariant law is $\mathrm{Gamma}(2,2)$ and whose sampling cost is $m(1-m)\mathbb{E}[X^{-1}] = 2m(1-m)$.

Appendix H proves Theorem 2, including the passage from the finite gap process to its unbounded stationary cost.

The fixed clip theorem takes separate outer limits in U and L. A second learner, denoted MC, moves both clips using confidence bounds from the sparse probe stream while retaining the central reduced variance estimator. Its occupation law at each fixed gap and a diagonal invariant limit yield one policy for the central family.

Theorem 3 (Central efficiency of one adaptive learner)**.** *For the MC learner of Appendix I, let $0 < \alpha < 1$ and $M > 6\alpha + 1$. On $\mu_1 = 1/2$, $\mu_2 = 1/2 + \Delta$,*

$$\lim_{\Delta\downarrow 0}\lim_{T\to\infty} \frac{\Delta R_T}{\log T} = \frac{1}{2}.$$

Appendix I proves the occupation identity at each fixed gap and the diagonal limit for the singular cost. The result gives no joint (T, Δ) rate.

6 What the floor is made of

Theorem 3 shows that $1/2$ is reachable, so the floor $(3 + 2\sqrt{2})/8$ is not a property of the regularizer or of the estimator. The MC learner reaches it by calibrating its clips to the gap through independent probes, and the calibration is what the fixed scale family lacks. This raises the question the fixed scale theorem alone cannot answer: is gap knowledge necessary, or only sufficient?

The question has teeth because of a heuristic tension between precision and regret. A learner that keeps $\Delta R_T = O(\log T)$ plays the suboptimal arm of order $\log T/\Delta^2$ times in expectation. Its standard error is then $o(\Delta)$, but a confidence radius valid with probability $1 - T^{-c}$, the level needed to trust a gap estimate for the rest of the horizon, is at least of order Δ. The MC learner sidesteps this requirement with a sparse probe schedule that does not depend on the ordinary sampling budget; its probe regret is still counted in R_T.

Conjecture 1 (Scale free floor)**.** *Consider the two arm reduced variance Tsallis-INF learner that plays its FTRL map on every round, with no forced probe pulls, and uses $\eta_n = \sqrt{a_n/n}$, where a_n is predictable with respect to its own history and independent of Δ except through that history. Then for the central Bernoulli family,*

$$\liminf_{\Delta \downarrow 0} \liminf_{T \to \infty} \frac{\Delta R_T}{\log T} \;\geq\; \frac{3 + 2\sqrt{2}}{8}.$$

Two consequences make the conjecture worth settling. If it is true, the factor the literature reports for RV Tsallis-INF is 1.457 rather than 2, and it is a price for not knowing the gap rather than a price for robustness, which separates two explanations that are currently conflated. If it is false, the witness is an implementable learner, since the conjecture restricts only predictability and forced exploration.

7 Limitations

All exact constants are asymptotic, and Theorem 1 holds on the admissible window $c^2(1/2 - \mu_1) < 2$ with $\Delta_i \leq 1/2$. The horizon tends to infinity before the gap vanishes, and the adaptive result takes the clip limits afterward. We do not give a joint rate in these parameters. The explicit finite occupation bounds explain which estimates would be needed for such a rate, but their constants are too large to predict a practical crossover horizon.

The adaptive learners are analyzed only for two stochastic Bernoulli arms. The moving clip result covers the central family, and neither adaptive learner inherits the adversarial guarantee of Tsallis-INF. Extending one policy to general means, more than two arms, and adversarial losses remains open.

8 Conclusion

The logarithmic regret guarantee for Tsallis-INF hides a fixed scale loss in its leading constant. Even after optimization it is $(3 + 2\sqrt{2})/8$, rather than the local benchmark $1/2$. The decomposition of the coefficient points to the missing quantity: the scale must respond to local variance.

For two stochastic arms, sparse probes make this response possible without knowing the gap, mean, variance, or arm order. The resulting occupation law matches the local Bernoulli information constant under the stated limits. The same analysis explains both results: regret is a singular stationary cost of the scaled gap process, and variance determines the scale at which that cost is information efficient.

Reproducibility statement

The appendices give the full proofs, state the theorem scope, and identify the external stochastic process results used in the argument. Numerical scripts do not support any theorem.

AI use disclosure

Generative AI tools assisted with literature search and organization, manuscript drafting and editing. The authors checked the cited sources, reviewed all AI-assisted text. The authors take responsibility for the final content of this work, including text and artifacts produced with the aid of generative AI.

References

Shipra Agrawal and Navin Goyal. Further optimal regret bounds for thompson sampling. In *Artificial Intelligence and Statistics*, 2013.

Noga Alon, Nicolò Cesa-Bianchi, Ofer Dekel, and Tomer Koren. Online learning with feedback graphs: Beyond bandits. In *Conference on Learning Theory*, 2015.

Peter Auer and Chao-Kai Chiang. An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits. In *Conference on Learning Theory*, 2016.

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2002a.

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002b.

Abhay G. Bhatt, Rajeeva L. Karandikar, and B. V. Rao. On characterisation of markov processes via martingale problems, 2006. arXiv:math/0607613.

Sébastien Bubeck and Aleksandrs Slivkins. The best of both worlds: Stochastic and adversarial bandits, 2012. arXiv:1202.4473.

Olivier Cappé, Aurélien Garivier, Odalric-Ambrym Maillard, Rémi Munos, and Gilles Stoltz. Kullback–leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, 41(3):1516–1541, 2013. doi: 10.1214/13-AOS1119.

Nicolò Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Claudio Gentile, and Yishay Mansour. Nonstochastic bandits with composite anonymous feedback, 2021. arXiv:2112.02866.

Cheng Chen, Canzhe Zhao, and Shuai Li. Simultaneously learning stochastic and adversarial bandits under the position-based model. In *AAAI Conference on Artificial Intelligence*, 2022.

Christoph Dann, Chen-Yu Wei, and Julian Zimmert. A blackbox approach to best of both worlds in bandits and beyond. In *Conference on Learning Theory*, 2023.

M. H. A. Davis. *Markov Models and Optimization*. Chapman & Hall, 1993. doi: 10.1007/978-1-4899-4483-2.

Lin Fan and Peter W. Glynn. Diffusion approximations for thompson sampling in the small gap regime, 2021. arXiv:2105.09232.

Yingying Fan, Yuxuan Han, Jinchi Lv, Xiaocong Xu, and Zhengyuan Zhou. Precise asymptotics and refined regret of variance-aware UCB, 2025. arXiv:2412.08843.

Pierre Gaillard, Gilles Stoltz, and Tim van Erven. A second-order bound with excess losses. In *Conference on Learning Theory*, 2014.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, 2016.

Junya Honda and Akimichi Takemura. Optimality of thompson sampling for gaussian bandits depends on priors. In *Artificial Intelligence and Statistics*, 2014.

Shinji Ito. Parameter-free multi-armed bandit algorithms with hybrid data-dependent regret bounds. In *Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pp. 1–32, 2021.

Shinji Ito. Best-of-three-worlds linear bandit algorithm with variance-adaptive regret bounds. In *Conference on Learning Theory*, 2023.

Shinji Ito and Kei Takemura. An exploration-by-optimization approach to best of both worlds in linear bandits. In *Advances in Neural Information Processing Systems*, 2023.

Shinji Ito, Taira Tsuchiya, and Junya Honda. Adversarially robust multi-armed bandit algorithm with variance-dependent regret bounds. In *Conference on Learning Theory*, 2022.

Shinji Ito, Taira Tsuchiya, and Junya Honda. Adaptive learning rate for follow-the-regularized-leader: Competitive analysis and best-of-both-worlds, 2024. arXiv:2403.00715.

Jean Jacod and Albert N. Shiryaev. *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2 edition, 2003. doi: 10.1007/978-3-662-05265-5.

Tianyuan Jin, Pan Xu, Xiaokui Xiao, and Quanquan Gu. Double explore-then-commit: Asymptotic optimality and beyond. In *Conference on Learning Theory*, 2021.

Anand Kalvit and Assaf Zeevi. A closer look at the worst-case behavior of multi-armed bandit algorithms. In *Advances in Neural Information Processing Systems*, 2021.

Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In *Algorithmic Learning Theory*, pp. 199–213, 2012. doi: 10.1007/978-3-642-34106-9_18.

Johannes Kirschner, Tor Lattimore, Claire Vernade, and Csaba Szepesvári. Asymptotically optimal information-directed sampling. In *Conference on Learning Theory*, 2021.

Fang Kong, Yueran Yang, Wei Chen, and Shuai Li. The hardness analysis of thompson sampling for combinatorial semi-bandits with greedy oracle. In *Advances in Neural Information Processing Systems*, 2021.

Fang Kong, Yichi Zhou, and Shuai Li. Simultaneously learning stochastic and adversarial bandits with general graph feedback. In *International Conference on Machine Learning*, 2022.

Fang Kong, Canzhe Zhao, and Shuai Li. Best-of-three-worlds analysis for linear bandits with follow-the-regularized-leader algorithm. In *Conference on Learning Theory*, 2023.

Yuko Kuroki, Alberto Rumi, Taira Tsuchiya, Fabio Vitale, and Nicolò Cesa-Bianchi. Best-of-both-worlds algorithms for linear contextual bandits. In *Artificial Intelligence and Statistics*, 2024.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985. doi: 10.1016/0196-8858(85)90002-8.

Tor Lattimore and András György. Mirror descent and the information ratio. In *Conference on Learning Theory*, 2021.

Tor Lattimore and Csaba Szepesvári. The end of optimism? an asymptotic analysis of finite-armed linear bandits. In *Artificial Intelligence and Statistics*, 2017.

Chung-Wei Lee, Haipeng Luo, Chen-Yu Wei, Mengxiao Zhang, and Xiaojin Zhang. Achieving near instance-optimality and minimax-optimality in stochastic and adversarial linear bandits simultaneously. In *International Conference on Machine Learning*, 2021.

Haipeng Luo and Robert E. Schapire. Achieving all with no parameters: AdaNormalHedge. In *Conference on Learning Theory*, 2015.

Saeed Masoudian and Yevgeny Seldin. Improved analysis of the tsallis-inf algorithm in stochastically constrained adversarial bandits and stochastic bandits with adversarial corruptions. In *Conference on Learning Theory*, 2021.

Sean P. Meyn and Richard L. Tweedie. Stability of markovian processes III: Foster–lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25(3):518–548, 1993. doi: 10.2307/1427522.

Gergely Neu. Explore no more: Improved high-probability regret bounds for non-stochastic bandits. In *Advances in Neural Information Processing Systems*, 2015a.

Gergely Neu. First-order regret bounds for combinatorial semi-bandits. In *Conference on Learning Theory*, 2015b.

Quan Nguyen, Shinji Ito, Junpei Komiyama, and Nishant A. Mehta. Data-dependent bounds with T-optimal best-of-both-worlds guarantees in multi-armed bandits using stability-penalty matching. In *Conference on Learning Theory*, 2025.

Hao Qin, Kwang-Sung Jun, and Chicheng Zhang. Kullback–leibler maillard sampling for multi-armed bandits with bounded rewards. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2642.

Chloé Rouyer and Yevgeny Seldin. Tsallis-INF for decoupled exploration and exploitation in multi-armed bandits. In *Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 1–23, 2020.

Yevgeny Seldin and Gábor Lugosi. An improved parametrization and analysis of the EXP3++ algorithm for stochastic and adversarial bandits. In *Conference on Learning Theory*, 2017.

Yevgeny Seldin and Aleksandrs Slivkins. One practical algorithm for both stochastic and adversarial bandits. In *International Conference on Machine Learning*, 2014.

Taira Tsuchiya and Shinji Ito. A simple and adaptive learning rate for FTRL in online learning with minimax regret of $\Theta(T^{2/3})$ and its application to best-of-both-worlds. In *Advances in Neural Information Processing Systems*, 2024.

Andrew Wagenmaker and Dylan J. Foster. Instance-optimality in interactive decision making: Toward a non-asymptotic theory. In *Conference on Learning Theory*, 2023.

Chen-Yu Wei and Haipeng Luo. More adaptive algorithms for adversarial bandits. In *Conference on Learning Theory*, 2018.

Jingxin Zhan, Yuze Han, and Zhihua Zhang. Last-iterate analyses of ftrl with the 1/2-tsallis entropy in stochastic bandits, 2025. arXiv:2510.22819.

Jingxin Zhan, Yuze Han, and Zhihua Zhang. Does 1/2-Tsallis-INF also work well for best-arm identification?, 2026. arXiv:2608.15365.

Julian Zimmert and Yevgeny Seldin. Tsallis-inf: An optimal algorithm for stochastic and adversarial bandits. *Journal of Machine Learning Research*, 22(28):1–49, 2021.

GUIDE TO THE RESULTS AND PROOFS

The main results separate one negative theorem from two adaptive repairs. The tables below give the shortest route from each statement to its proof. The detailed assumptions remain in the statements themselves.

Main results.

Result	Proof location	Main ingredients
Theorem 1	Appendices A, B, and C	KKT map, invariant law, harmonic occupation, joint recovery
Corollary 1	Appendix C	Recovery entry sum from Theorem 4
Theorem 2	Appendix H	Probe consistency, controlled PDMP, clip removal
Theorem 3	Appendix I	Moving recovery, compensated occupation, diagonal limit

Supporting formal results.

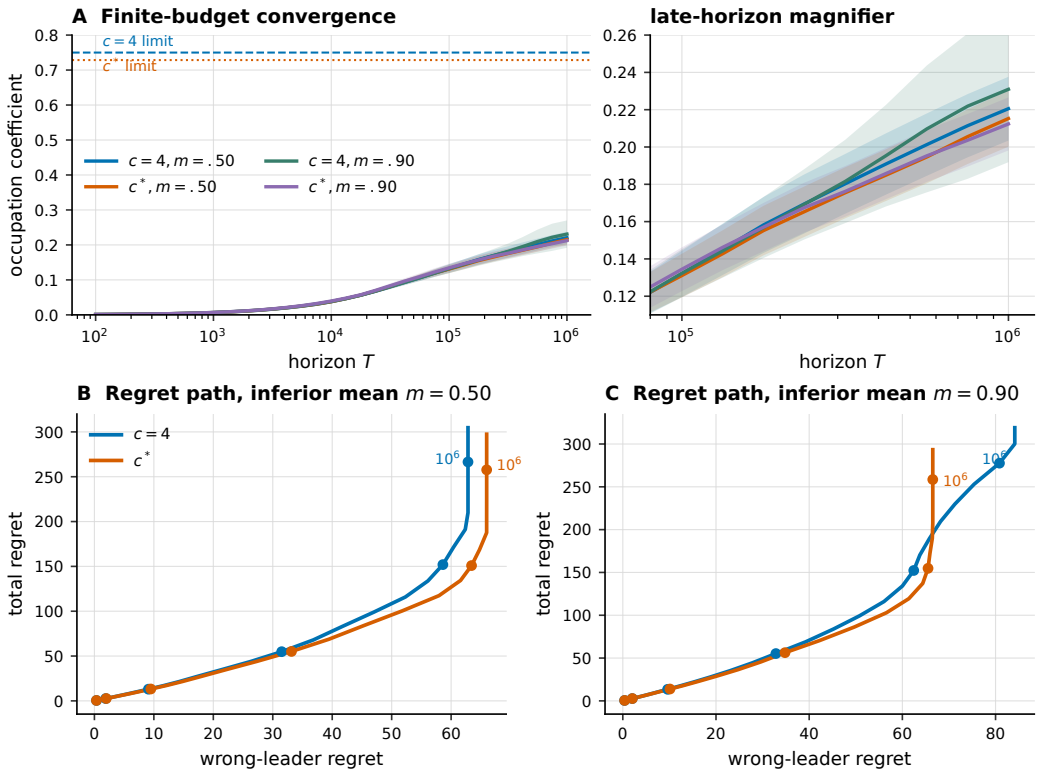

Figure 3: Finite-budget behavior of the two fixed-scale learners. (A) The occupation-coefficient trajectories, with a late-horizon magnifier, remain far below their scale-specific asymptotic limits through horizon 10^6; ribbons show descriptive Monte Carlo ± 1.96 standard errors. (B–C) Total regret versus wrong-leader regret along the same horizon grid for inferior means 0.50 and 0.90. The terminal points are marked 10^6. These panels visualize finite-budget approach and regret composition; they do not estimate a new asymptotic constant.

Result	Proof location	Role
Theorem 4	Appendix B	Converts recovery and moment bounds into equality of occupation costs
Corollary 2	Appendix D	Verifies recovery and uniform integrability for plain importance weighting
Theorem 5	Appendix D	Gives the exact occupation coefficient for plain importance weighting
Theorem 6	Appendix D	Evaluates its small gap limit uniformly in the mean
Theorem 7	Appendix F	Gives a finite occupation error bound for bounded tests
Corollary 3	Appendix F	Shows why bounded Borel tests cannot converge uniformly
Proposition 1	Appendix G	Instantiates one explicit Poisson approximation
Proposition 2	Appendix I	Supplies the fixed gap identity used by Theorem 3

A complete index with assumptions and dependencies is included with the submission source as `THEOREM PROOF INDEX.md`.

A THE FIXED SCALE PDMP AND ITS INVARIANT LAW

The proof of Theorem 1 needs three facts about the scalar process. Lemma 1 establishes its invariant law, Lemma 2 identifies harmonic occupation limits, and Lemma 4 evaluates the singular moment as the gap vanishes. Appendix C then proves equality and Corollary 1.

A.1 THE LIMITING PDMP

For fixed admissible (Δ, θ) define

$$\gamma_\Delta = \frac{c^2\Delta}{8}, \quad x_0 = \frac{2}{c^2\Delta}, \quad \mu_- = \frac{1}{2} - \theta\Delta, \quad \mu_+ = \frac{1}{2} + (1-\theta)\Delta,$$

$$a = \frac{\mu_-}{\Delta} = \frac{1}{2\Delta} - \theta, \qquad \lambda(x) = \frac{4}{c^2\Delta^2 x^2}, \qquad j(x) = \gamma_\Delta x^2.$$

With the active convention at $x = x_0$, its generator is

$$\begin{aligned}\mathcal{A}_{\Delta,\theta} f(x) &= (\theta - x) f'(x) \\ &\quad + \lambda(x)\mu_+[f(x + j(x)) - f(x)] \\ &\quad + \lambda(x)(1-\mu_+)[f(x - j(x)) - f(x)], \quad x \le x_0, \\ &= (-a - x) f'(x) \\ &\quad + \mu_+\lambda(x)[f(x + 2j(x)) - f(x)], \quad x > x_0.\end{aligned} \tag{24}$$

The downward landing is positive because $\gamma_\Delta x \le 1/4$ on the active branch. Condition (14) is precisely $c^2\Delta_i\theta_i < 2$, so $\theta_i < x_{0,i}$ and the two one sided flows cross the switching surface in only one direction.

Lemma 1 (Well posed invariant law)**.** *For every $0 < \Delta \le 1/2$ and $\theta \in \Theta_\Delta(c)$,* (24) *has a conservative, càdlàg, Borel right strong Markov realization. It is positive Harris recurrent and has a unique invariant probability $\pi_{\Delta,\theta}$. For $V_q(x) = x + x^{-q}$, $q > 2$, there are $\kappa, b > 0$ and a compact petite set C such that, on every compact stopped truncation,*

$$\mathcal{A}_{\Delta,\theta} V_q \le -\kappa V_q + b\mathbf{1}_C. \tag{25}$$

Proof. When $\theta \ge 0$, use the mode coordinate $y = \log x$, with state

$$(\{A\} \times (-\infty, y_0]) \sqcup (\{I\} \times (y_0, \infty)), \qquad y_0 = \log x_0.$$

The mode fields are $\theta e^{-y} - 1$ and $-1 - ae^{-y}$, respectively. They are locally Lipschitz; the outgoing inactive boundary (I, y_0) is sent deterministically to (A, y_0). The jump rates and log landing maps are Borel and locally integrable along flows. Compact stopping with $V_3 = x + x^{-3}$ gives an integrable martingale and

$$\mathbb{E}_x V_3(X_{t\wedge\tau_N}) \le e^{bt} V_3(x), \qquad \mathbb{P}_x(\tau_N \le t) \le e^{bt} V_3(x)/N.$$

Hence random events do not accumulate; their expected number is finite on bounded intervals. These facts verify Davis's four standard conditions, so Theorems 25.5 and 27.8 of Davis (1993) give a right continuous strong Markov and Borel right process on the lifted state. Its scalar projection is càdlàg by direct path analysis: it is continuous at the forced boundary transition and has finite left and right limits at random jumps. Strong Markovness and Borel rightness pass through the Borel state space isomorphism. When $\theta < 0$, instead use the x-coordinate lift with active component $(0, x_0]$ and inactive component (x_0, ∞). The active flow

$$\varphi_t(x) = \theta + (x - \theta)e^{-t}$$

has finite outgoing time $\tau_0 = \log((x-\theta)/(-\theta))$, but $\varphi_t(x) \asymp (-\theta)(\tau_0 - t)$ there. Consequently $\int_0^{\tau_0} \lambda(\varphi_s(x))\, ds = \infty$, so a random jump occurs before zero almost surely. This is precisely the finite boundary mechanism allowed by Davis's Example 24.7; no boundary kernel at zero is required. Moreover $R(x) \le C(1 + x^{-3})$. The stopped V_3 estimate therefore bounds the expected random event count, and each forced switch is bracketed by random events, ruling out both random and boundary event accumulation.

For each $q > 2$, flowwise absolute continuity, equality of the two boundary values, and jump integrability up to each compact stopping time put V_q in Davis's local extended generator domain by Theorem 26.14. More precisely, for every precompact exhaustion O_N the stopped residual is an integrable true martingale, so $V_q \in D(\mathcal{A}^{(N)}_{\Delta,\theta})$, where the superscript denotes stopping on exit from O_N.

Here is the drift calculation. On the active branch put $k = c^2/8$, $z = k\Delta x \in [0, 1/4]$, and

$$H_q(z) = \mu_+[(1+z)^{-q} - 1] + (1-\mu_+)[(1-z)^{-q} - 1].$$

Since $H_q'(0) = -2q(1-\theta)\Delta$, Taylor's theorem gives $H_q(z) \le -2q(1-\theta)\Delta z + C_q z^2$. Combining the jump term $\lambda(x)x^{-q}H_q(z)$ with $(\theta - x)(x^{-q})' = -q\theta x^{-q-1} + qx^{-q}$ yields

$$\mathcal{A}_{\Delta,\theta}x^{-q} \le -qx^{-q-1} + K_q x^{-q}, \qquad K_q = q + c^2 C_q/16. \tag{26}$$

On the inactive branch the upward jump decreases x^{-q}, while $a = \mu_-/\Delta \ge 0$ and $x > x_0$ imply $a/x < c^2\mu_-/2 \le c^2/2$, and hence $\mathcal{A}_{\Delta,\theta}x^{-q} \le q(1 + c^2/2)x^{-q}$. Direct substitution also gives $\mathcal{A}_{\Delta,\theta}x = 1 - x$ on both branches. Therefore (26) controls zero, $1 - x$ controls infinity, and continuity on the remaining compact gives (25).

We next verify the petite set and irreducibility. On the active branch the flow and signed landing maps are

$$\varphi_t(x) = \theta + (x - \theta)e^{-t}, \qquad T_\sigma(x) = x + \sigma\gamma_\Delta x^2, \qquad \sigma \in \{-1, +1\}.$$

Choose $x_\star \in (0, x_0)$ with $x_\star \ne 2\theta$: take $x_\star = \theta$ and the downward channel when $\theta > 0$, and take $x_\star \in (0, x_0/4)$ and the upward channel when $\theta \le 0$. There are a closed interval C_0 around $x_\star$, a jump time interval $I = [s_0, s_1]$, and an observation time T such that the entire one jump tube is active and avoids $\{0, 2\theta\}$. Conditional on exactly one jump of sign σ at time $s \in I$ and no later jump, the endpoint is

$$F_x(s) = \theta + [T_\sigma(\varphi_s(x)) - \theta]e^{-(T-s)}, \qquad F_x'(s) = \sigma\gamma_\Delta u(2\theta - u)e^{-(T-s)}, \qquad u = \varphi_s(x).$$

The derivative is bounded above and away from zero uniformly on $C_0 \times I$. The first jump density $\lambda(\varphi_s(x)) \exp\{-\int_0^s \lambda(\varphi_r(x))dr\}$, the chosen channel probability, and the no second jump probability are uniformly positive on the tube. Shrinking C_0, I makes all $F_x(I)$ contain one interval J, so change of variables gives

$$P_T(x, A) \ge \varepsilon \,\mathrm{Leb}(A \cap J), \qquad x \in C_0. \tag{27}$$

If $\theta > 0$, every no jump path reaches C_0 after an inactive path first crosses x_0. If $\theta \le 0$, points above C_0 reach it along a no jump path before the zero endpoint; from a point below it, finitely many scheduled upward jumps satisfy $z_{n+1} \ge z_n + (\gamma_\Delta/2)z_n^2$ and cross its lower endpoint. Each finite schedule has positive probability. Thus (27) gives irreducibility. From the compact drift set in (25), compactness bounds entrance times, path rates, and no jump probabilities uniformly; composing its entrance kernel with (27) makes that drift set petite. Meyn–Tweedie's continuous time Foster–Lyapunov theorem (Meyn & Tweedie, 1993), applied to the domains $D(\mathcal{A}^{(N)}_{\Delta,\theta})$, now gives positive Harris recurrence. Irreducibility gives uniqueness of the invariant probability. □

A.2 Occupation identification

Recall $D_{t,i}$ and $X_{t,i}$ from (5). Writing $X_t = (X_{t,2}, \ldots, X_{t,K})$, define its logarithmic occupation measure

$$\Gamma_T = \frac{1}{\log T}\sum_{t=2}^{T}\frac{1}{t}\mathcal{L}(X_{t-1}).$$

This is a finite measure of mass $\left(\sum_{t=2}^{T} t^{-1}\right)/\log T \to 1$; every weak limit below is therefore a probability measure.

Lemma 2 (Harmonic occupation limits)**.** *Suppose that along $T_n \to \infty$ every*

$$A^{(i)}_{T_n} = \frac{\Delta_i^2}{\log T_n}\sum_{t=2}^{T_n}\mathbb{E}p_{t,i}$$

is bounded. Then (Γ_{T_n}) *is tight. Every weak limit* Γ *is supported on* $(0,\infty)^{K-1}$*, gives zero mass to the threshold hyperplanes, is invariant for the* $(K-1)$*-dimensional PDMP with generator* $\mathcal{A}=\sum_{i=2}^{K}\mathcal{A}_i$*, and has marginals*

$$\Gamma_i = \pi_{\Delta_i,\theta_i}. \tag{28}$$

Proof. Write $\epsilon_t = \sum_{j=2}^{K} p_{t,j}$ and, at the beginning of round t,

$$x_i = \frac{D_{t-1,i}}{\Delta_i(t-1)}, \qquad z_i = \widehat{\ell}_{t,i} - \widehat{\ell}_{t,1}.$$

Thus the exact update is

$$X_{t,i} = \left(1-\frac{1}{t}\right)x_i + \frac{z_i}{\Delta_i t}. \tag{29}$$

Whenever $B_t(1)=1/2$, the baseline $B_t(i)=\frac{1}{2}\mathbf{1}\{p_{t,i}\ge c^2/t\}$ gives the following exhaustive table. This condition holds on all sufficiently large compact uncharged rounds used below.

played arm	$B_t(i)=1/2$	$B_t(i)=0$
1	$-(\ell_{t,1}-1/2)/p_{t,1}$	$-1/2-(\ell_{t,1}-1/2)/p_{t,1}$
i	$(\ell_{t,i}-1/2)/p_{t,i}$	$\ell_{t,i}/p_{t,i}-1/2$
$j\notin\{1,i\}$	0	$-1/2$

Here every displayed entry is z_i. Directly from (2), without assuming $B_t(1)=1/2$,

$$\mathbb{E}(|\widehat{\ell}_{t,j}| \mid \mathcal{F}_{t-1}) \le \frac{3}{2}, \qquad \mathbb{E}(|z_i| \mid \mathcal{F}_{t-1}) \le 3. \tag{30}$$

Together with (29), this yields

$$\mathbb{E}[|X_{t,i}| - |X_{t-1,i}| \mid \mathcal{F}_{t-1}] \le t^{-1}(3/\Delta_i - |X_{t-1,i}|),$$

and hence tightness of all coordinate first moments under the harmonic occupation measures.

We next make the switching and boundary exclusions quantitative. Put

$$x_i^0 = \frac{2}{c^2\Delta_i}, \quad \theta_i = \frac{1/2-\mu_1}{\Delta_i}, \quad d_i = x_i^0 - \theta_i > 0, \quad a_i = \frac{1}{2\Delta_i} - \theta_i,$$

and choose

$$\rho_i^0 = \min\{x_i^0/8, d_i/8\},$$
$$\delta_i = \min\left\{\frac{1}{4}, \frac{d_i}{4(x_i^0+2\rho_i^0)}, \frac{x_i^0+a_i-2\rho_i^0}{2(x_i^0+2\rho_i^0+1/(2\Delta_i))}\right\}. \tag{31}$$

Call $U_{t,i}=\{\epsilon_t>\delta_i\}$ an i-charged round. If $\Delta_{\min}=\min_{j\ge2}\Delta_j$ and $M=\sum_{j\ge2}\sup_n A_{T_n}^{(j)}$, then

$$\sum_{t=2}^{T_n}\mathbb{P}(U_{t,i}) \le \frac{M}{\Delta_{\min}^2\delta_i}\log T_n. \tag{32}$$

For $q_t\in[0,1]$ with $\sum_{t\le T}q_t\le B\log T$, rearrangement against $1/t$ gives $\sum_{t=2}^{T}q_t/t\le 1+\log(1+B\log T)$. Thus charged rounds have $o(\log T_n)$ harmonic mass.

On $U_{t,i}^c$, $p_{t,1}>1/2$, arm 1 is an empirical minimizer, and the Tsallis KKT equations give the exact switch at a finite time

$$x_{i,t}^0 = \frac{2t/c^2 - w_t\sqrt{t}/c}{\Delta_i(t-1)}, \qquad w_t = \frac{2}{\sqrt{1-\epsilon_t}} \le W_i := \frac{2}{\sqrt{1-\delta_i}}. \tag{33}$$

Consequently, for $t\ge2$,

$$|x_{i,t}^0 - x_i^0| \le \frac{2W_i}{c\Delta_i\sqrt{t}} + \frac{4}{c^2\Delta_i t}.$$

Fix $0 < \rho \le \rho_i^0$ and let ϕ_ρ be the decreasing $C^{1,1}$ ramp with $0 \le \phi_\rho \le 3\rho$, derivative -1 on $[x_i^0 - \rho, x_i^0 + \rho]$, derivative zero outside $[x_i^0 - 2\rho, x_i^0 + 2\rho]$, and $\mathrm{Lip}(\phi_\rho') = 1/\rho$. From the arm-1 cells of the table,

$$\begin{aligned} t\mathbb{E}[(X_{t,i} - x_i)\mathbf{1}\{I_t = 1\} \mid \mathcal{F}_{t-1}] &= \theta_i - p_{t,1}x_i, && B_t(i) = 1/2, \\ &= \theta_i - p_{t,1}(x_i + 1/(2\Delta_i)), && B_t(i) = 0. \end{aligned}$$

The choice (31) makes both expressions at most $-\kappa_i$ throughout the outer ramp, where

$$\kappa_i = \min\{d_i/2, (x_i^0 + a_i - 2\rho_i^0)/2\} > 0.$$

The own arm downward outcomes and all other arm outcomes cannot decrease the decreasing ramp. The own arm upward landing or overshoot has probability weighted cost at most $3\rho L_i/t$; the arm-1 Taylor remainder is at most $J_i^2/(2\rho t^2)$, where the finite constants

$$H_i = \frac{1}{2} + \frac{1}{2(1-\delta_i)}, \quad J_i = 2(x_i^0 + 2\rho_i^0) + \frac{2H_i}{\Delta_i},$$

$$L_i = \max\left\{ \frac{64}{c^2\Delta_i^2(x_i^0 - 2\rho_i^0)^2}, \frac{2}{\Delta_i(x_i^0 - 2\rho_i^0)} \right\}$$

follow directly from (29) and the KKT formula. Hence, for all sufficiently large t,

$$\begin{aligned} \mathbb{E}[\phi_\rho(X_{t,i}) - \phi_\rho(X_{t-1,i}) \mid \mathcal{F}_{t-1}] \ge \frac{\kappa_i}{t}\mathbf{1}\{|X_{t-1,i} - x_i^0| \le \rho, U_{t,i}^c\} \\ - \frac{3\rho L_i}{t} - \frac{J_i^2}{2\rho t^2} - 3\rho\mathbf{1}_{U_{t,i}}. \end{aligned} \tag{34}$$

Telescoping, using (32), gives a finite C_i with

$$\limsup_n \frac{1}{\log T_n} \sum_{t=2}^{T_n} \frac{1}{t}\mathbb{P}(|X_{t-1,i} - x_i^0| \le \rho) \le C_i\rho. \tag{35}$$

For $0 < e \le x_i^0/2$, the same KKT equation gives, eventually on uncharged rounds with $0 < X_{t-1,i} \le e$, $\Delta_i^2 t p_{t,i} \ge 1/(c^2e^2)$. Their normalized harmonic occupation is therefore at most $c^2e^2A_{T_n}^{(i)} + o(1)$. If $X_{t-1,i} = 0$, then $p_{t,i} = p_{t,1}$ and the round is charged; if $X_{t-1,i} < 0$, then $p_{t,i} > p_{t,1}$ and again it is charged. Sending first $n \to \infty$ and then $\rho, e \downarrow 0$ proves that every limit is supported on $E = (0,\infty)^{K-1}$ and gives no mass to $H = \bigcup_i \{x_i = x_i^0\}$.

Why weak convergence is insufficient. The preceding estimate excludes zero from weak occupation limits but is exactly critical for the unbounded cost x_i^{-2}, so it does not prove tail uniform integrability. The active scalar generator points to the sharper local potential $W_\beta(x) = e^{\beta/x}$: direct substitution gives

$$\lim_{x\downarrow 0} x^2 \frac{\mathcal{A}_{\Delta_i,\theta_i} W_\beta(x)}{W_\beta(x)} = -\beta + \frac{c^2}{32}\beta^2 + O(\beta^3). \tag{36}$$

Hence small β produces strict limiting drift near zero. We do not use this calculation to strengthen the theorem. A stopped version at a finite time must pay for returns through $D_{t-1,i} = 0$ and wrong leader states. When the optimal arm has probability $p_{t,1}$ and inactive baseline, its loss one channel has probability $\mu_1 p_{t,1}$ and gap jump of magnitude at least $1/p_{t,1} - 1/2$; consequently every fixed exponential moment contains the term

$$\mu_1 p_{t,1} \exp\{\beta(1/p_{t,1} - 1/2)\},$$

which is unbounded as $p_{t,1} \downarrow 0$. The charged round count above does not control the resulting entrance depths or total sampling cost. This is why this weak occupation bridge alone supplies only a lower bound. Appendices B and C supply the additional joint recovery and cost uniform integrability for every fixed K. The global FTRL comparison potential handles wrong leader states, and positive orthant coordinate strips control returns before a polynomial potential transfers the unbounded cost.

We derive the limiting transition mechanism from the one step kernel. On a compact $Q \Subset E \setminus H$, positivity of all $D_{t-1,i}$ for large t makes arm 1 the empirical minimizer. Uniformly on Q,

$$tp_{t,i} = \lambda_i(x_i) + O_Q(t^{-1/2}), \qquad \lambda_i(x) = \frac{4}{c^2\Delta_i^2 x^2}, \qquad \epsilon_t = O_Q(t^{-1}). \tag{37}$$

For a state $x \in \mathbb{R}^{K-1}$, reconstruct $D_i = \Delta_i(t-1)x_i$, set $q = -\min\{0, \min_{i\geq 2} D_i\}$, and define

$$\bar{L}_1 = q, \qquad \bar{L}_i = D_i + q, \qquad p_{t,a}(x) = \frac{4}{(\eta_t \bar{L}_a + w_t)^2}, \qquad \sum_{a=1}^{K} p_{t,a}(x) = 1.$$

For $a \in [K]$, $\ell \in \{0,1\}$, and every coordinate j, define the global baseline and estimator outcome

$$B_{t,j}(x) = \frac{1}{2}\mathbf{1}\{p_{t,j}(x) \geq c^2/t\}, \qquad \widehat{\ell}^{a,\ell}_{t,j}(x) = B_{t,j}(x) + \frac{\mathbf{1}\{a=j\}}{p_{t,j}(x)}(\ell - B_{t,j}(x)),$$

and let $z^{(i)}_{a,\ell}(x) = \widehat{\ell}^{a,\ell}_{t,i}(x) - \widehat{\ell}^{a,\ell}_{t,1}(x)$. Put

$$\Psi^{(i)}_{t,a,\ell}(x) = \left(1 - \frac{1}{t}\right) x_i + \frac{z^{(i)}_{a,\ell}(x)}{\Delta_i t}.$$

The exact transition kernel for one step is therefore

$$\mathsf{P}_t(x, dy) = \sum_{a=1}^{K} p_{t,a}(x) \left[\mu_a \delta_{\Psi_{t,a,1}(x)}(dy) + (1-\mu_a)\delta_{\Psi_{t,a,0}(x)}(dy)\right]. \tag{38}$$

Here Ψ denotes the full vector of its coordinate maps. For $h_t = \log(t/(t-1))$ and $f \in C^2_c(E)$, the table and (37) give uniformly on Q

$$h_t^{-1}(\mathsf{P}_t f(x) - f(x)) = \mathcal{A}f(x) + O_{Q,f}(t^{-1/2}), \tag{39}$$

where $\mathcal{A} = \sum_{i=2}^{K} \mathcal{A}_i$ and

$$\begin{aligned}
\mathcal{A}_i f(x) &= (\theta_i - x_i)\partial_i f(x) + \lambda_i(x_i)\{\mu_i[f(x + \gamma_i x_i^2 e_i) - f(x)] \\
&\qquad + (1-\mu_i)[f(x - \gamma_i x_i^2 e_i) - f(x)]\}, \qquad x_i \leq x_i^0, \\
\mathcal{A}_i f(x) &= (-\mu_1/\Delta_i - x_i)\partial_i f(x) + \mu_i \lambda_i(x_i)[f(x + 2\gamma_i x_i^2 e_i) - f(x)], \qquad x_i > x_i^0,
\end{aligned}$$

with $\gamma_i = c^2\Delta_i/8$. Thus the limiting kernel has channels

$$\begin{array}{lll}
q_{i,+} = \mu_i \lambda_i(x_i), & J_{i,+}(x) = x + \gamma_i x_i^2 e_i, & x_i \leq x_i^0, \\
q_{i,-} = (1-\mu_i)\lambda_i(x_i), & J_{i,-}(x) = x - \gamma_i x_i^2 e_i, & x_i \leq x_i^0, \\
q_{i,I} = \mu_i \lambda_i(x_i), & J_{i,I}(x) = x + 2\gamma_i x_i^2 e_i, & x_i > x_i^0.
\end{array} \tag{40}$$

Equivalently, with b equal to the displayed branch drift,

$$K(x, dz) = \sum_h q_h(x)\delta_{J_h(x)-x}(dz), \qquad \mathcal{A}f = b \cdot \nabla f + \int [f(x+z) - f(x)]K(x, dz).$$

To pass from local convergence to the occupation equation, first note the global source estimate. If $\operatorname{supp} f \subset \prod_i [a_i, b_i] \Subset E$, then conditioning on the played arm in the exact table and using the mean value theorem gives

$$\sup_{t\geq 2, x} |t(\mathsf{P}_t f(x) - f(x))| \leq \sum_{i=2}^{K} \|\partial_i f\|_\infty \left(2Kb_i + \frac{3(K+4)}{2\Delta_i}\right). \tag{41}$$

Moreover, every x for which the left side is nonzero lies in one compact landing source rectangle depending only on f. The exact martingale

$$M_m^f = f(X_m) - f(X_1) - \sum_{t=2}^{m} \{\mathsf{P}_t f(X_{t-1}) - f(X_{t-1})\}$$

has mean zero. Divide its expectation at $m = T_n$ by $\log T_n$. On compact subsets of $E \setminus H$, (39) replaces $t(\mathsf{P}_t - I)f$ by $\mathcal{A}f$; the total $O(t^{-1/2})$ error after harmonic weighting is $O(1)$. The compact source property, (41), and the boundary estimates above remove its complement. Since the endpoint term is $O(1/\log T_n)$,

$$\int_E \mathcal{A}f \, d\Gamma = 0, \qquad f \in C^2_c(E). \tag{42}$$

C^1 approximation on a common compact support extends (42) to $\mathcal{D} = \mathbb{R}1 + C_c^1(E)$: both the drift term and every finite landing difference converge uniformly on the same compact source hull. Thus

$$\int_E \mathcal{A}f \, d\Gamma = 0, \qquad f \in \mathcal{D}. \tag{43}$$

We record all remaining operator hypotheses needed for the invariant measure criterion. Let $R(x) = \sum_h q_h(x)$ and $\Phi(x) = 1 + |b(x)| + R(x)$. The first moment bound above and the compact KKT lower bound also give the missing inverse square moment: choose continuous $0 \le H_m \uparrow 4/(c^2 x_i^2)$ with compact support in $(0, \infty)$. The local rate limit, the vanishing charged mass, and the boundary controls give

$$\int H_m(x_i)\, \Gamma(dx) \le \sup_n A^{(i)}_{T_n}.$$

Monotone convergence, followed by the coordinate first moment bound, yields

$$\int_E \Phi \, d\Gamma < \infty, \qquad \Phi(x) \le C \left(1 + \sum_i x_i + \sum_i x_i^{-2}\right). \tag{44}$$

For $f = c + g \in \mathcal{D}$, the drift term is supported on $\operatorname{supp} \nabla g$, and the source of every nonzero landing difference lies in the compact source hull used in (41). Hence $|\mathcal{A}f| \le C_f \Phi$, which is BKR A1.

For A2, fix rational C^1 cutoffs χ_n that equal one on $[1/n, n]^{K-1}$ and vanish outside $[1/(2n), 2n]^{K-1}$. Rational constants plus finite rational combinations of $\chi_n p$, for rational coefficient polynomials p, form one countable set $\mathcal{D}_0$. Tensor polynomial approximation on a common compact gives, for every $f \in \mathcal{D}$, a sequence $f_m \in \mathcal{D}_0$ with

$$\|f_m - f\|_\infty + \|\nabla f_m - \nabla f\|_\infty \to 0, \qquad \|\Phi^{-1}(\mathcal{A}f_m - \mathcal{A}f)\|_\infty \to 0.$$

Thus $\mathcal{D}_0$ has the graph closure under bounded pointwise convergence required by A2. The domain $\mathcal{D}$ is an algebra containing constants and separating points, proving A8. At a nonnegative global maximum the gradient vanishes and every jump difference is nonpositive, so $\mathcal{A}f \le 0$; this is A9.

For A10–A11 use the Polish mark space $U = \mathbb{R}^{K-1} \times [0, \infty) \times E$ and

$$\widehat{A}f(x; \beta, r, y) = \beta \cdot \nabla f(x) + r[f(y) - f(x)], \qquad \widehat{\Phi}(x; \beta, r, y) = 1 + |\beta| + r.$$

When $R(x) > 0$, set

$$\eta_x = \sum_h \frac{q_h(x)}{R(x)} \delta_{(b(x), R(x), J_h(x))};$$

when $R(x) = 0$, set $\eta_x = \delta_{(b(x),0,x)}$. The branch indicators, rates, drift, and landing maps are Borel, so $x \mapsto \eta_x$ is a Borel kernel, while $\widehat{A}f$ and $\widehat{\Phi}$ are continuous on $E \times U$. Their η_x-averages are respectively $\mathcal{A}f(x)$ and $\Phi(x)$, which proves A10 and A11 and replaces the continuity assumption A7.

To turn (43) into invariance, let Y be an arbitrary continuous in probability solution of the $(\mathcal{A}, \delta_x)$ martingale problem satisfying the BKR integrability condition for $\Phi = 1 + |b| + K(\cdot, \mathbb{R}^{K-1})$. A countable separating $C_c^\infty(E)$ family $\{g_m\}$, uniformly dense on compacts in $C_0(E)$, separates the one point compactification E^Δ. For every m,

$$g_m(Y_t) = M_t^{g_m} + \int_0^t \mathcal{A}g_m(Y_s) ds.$$

The integral is continuous by the Φ bound and M^{g_m} has a càdlàg version after usual augmentation. Choose these versions simultaneously and embed E^Δ in the Hilbert cube using $g_m/(1 + \|g_m\|_\infty)$. Coordinatewise rational right limits give an E^Δ-valued càdlàg process. Continuity in probability of the original solution and separation by $\{g_m\}$ show that it is a modification at every deterministic time. With $V = 1 + \sum_i (x_i + x_i^{-2})$, set $G_m = \{V < m\}$ and $\tau_m = \inf\{s : Y_s \notin G_m\}$. Plateau cutoffs contain all states before exit and every positive rate landing. If $\chi_m \in C_c^\infty(E)$ equals one on that plateau, then $\mathcal{A}\chi_m = 0$ on G_m. Optional sampling gives $\mathbb{E}\chi_m(Y_{t \wedge \tau_m}) = 1$, so an exit cannot land outside the plateau or at the compactification point. Direct substitution in the channel generator gives

$\mathcal{A}x_i = 1 - x_i$ and $\mathcal{A}(x_i^{-2}) \le C_i(1 + x_i + x_i^{-2})$ after compact truncation. Summing, applying the stopped martingale identity to the truncations of V, and using Gronwall gives

$$\mathbb{P}(\tau_m \le T) \le e^{CT} V(x)/m. \tag{45}$$

Hence the modification is E-valued. Localized coordinate tests first make Y^{τ_m} a special semimartingale. If $\kappa(z) = z$ near zero, localized sine and cosine tests, whose plateau also contains every exit landing, identify by Jacod–Shiryaev Theorem II.2.42 the complete stopped characteristic triplet

$$\begin{aligned} C = 0, \qquad \nu(ds, dz) &= \mathbf{1}_{\{s \le \tau_m\}} K(Y_{s-}, dz)\, ds, \\ dB_s &= \mathbf{1}_{\{s \le \tau_m\}} \left[b(Y_{s-}) + \int \kappa(z) K(Y_{s-}, dz) \right] ds. \end{aligned} \tag{46}$$

This use of Fourier tests identifies the entire jump kernel, not merely its first two moments. The finite marked compensator is therefore

$$\nu^{\#}(ds, dh) = \sum_h q_h(Y_{s-}) \delta_h(dh) ds. \tag{47}$$

Before the first jump the branch ODE has a unique transversal flow $\varphi_s(x)$. Writing $R = \sum_h q_h$, compensation in (47) yields

$$\mathbb{P}_x(T_1 > s) = \exp\left\{ - \int_0^s R(\varphi_u(x)) du \right\}, \qquad \mathbb{P}_x(T_1 \in ds, H_1 = h) = \mathbb{P}_x(T_1 > s) q_h(\varphi_s(x)) ds. \tag{48}$$

No Markov property is assumed in this calculation.

If some $\theta_i < 0$, its active coordinate flow has the finite zero endpoint $\tau_{0,i} = \log((x_i - \theta_i)/(-\theta_i))$. As that endpoint approaches, $\varphi_s^{(i)}(x_i) \asymp (-\theta_i)(\tau_{0,i} - s)$, and hence

$$\int_0^{\tau_{0,i}} R(\varphi_s(x))\, ds \ge \int_0^{\tau_{0,i}} \lambda_i(\varphi_s^{(i)}(x_i))\, ds = \infty.$$

Thus (48) assigns zero probability to reaching that endpoint before a random jump. The sum of the V_3 estimates stopped on compact sets bounds the expected number of random jumps, while each deterministic threshold crossing is bracketed by random jumps. Therefore neither a zero endpoint nor a Zeno accumulation interrupts the finite marked jump recursion below.

It remains to justify restarting this recursion. Transfer the law to the Polish canonical space $D([0, \infty), E)$. For every bounded canonical stopping time σ, regular conditional laws of the shifted path exist. Apply optional sampling only to the true integrable martingales indexed by BKR's one countable graph core, rational times, and a countable bounded cylinder algebra. The conditional Φ integral is finite on every bounded interval. One common null set therefore suffices; monotone class closure and right continuity extend the identities to all times and all $f \in \mathcal{D}$. The conditional shifted law is again a continuous in probability solution started at Y_σ, since (46) has no atom at any deterministic time. Applying (48) after successive bounded jump times fixes every finite marked jump sequence. Compact stops have only finitely many jumps, and (45) removes the stops. Thus every solution from each initial point equals the canonical clock construction, proving BKR A3.

For a solution with arbitrary initial law λ, disintegrate on the same canonical space at Y_0. The same countable family gives, for λ-almost every x, a solution from each initial point with the required conditional Φ integrability. A3 identifies it with the measurable canonical law P_x, so $P(\cdot) = \int P_x(\cdot) \lambda(dx)$. This proves BKR A6 with the singleton initial law test $h_1 \equiv 0$. Together with (44), all hypotheses checked above are now available. Bhatt–Karandikar–Rao Theorem 4.2 and Remark 4.3 now promote (43) to invariance of Γ.

Finally, for a test depending only on coordinate i, $\mathcal{A}$ reduces to $\mathcal{A}_{\Delta_i, \theta_i}$. Thus the ith marginal of Γ is invariant for (24); Lemma 1 gives $\Gamma_i = \pi_{\Delta_i, \theta_i}$. No independence or product form assertion is used. □

Lemma 3 (Summed occupation lower bound)**.** *For every fixed instance satisfying* (14)*,*

$$\liminf_{T \to \infty} \frac{R_T}{\log T} \ge \sum_{i=2}^{K} \frac{4}{c^2 \Delta_i} \int_0^\infty x^{-2} \pi_{\Delta_i, \theta_i}(dx).$$

Proof. Choose a subsequence realizing the left hand liminf. If it is infinite there is nothing to prove. Otherwise nonnegativity and $R_T/\log T = \sum_i \Delta_i^{-1} A_T^{(i)} + o(1)$ imply that every coordinate coefficient is bounded on this subsequence. The compact KKT expansion gives $tp_{t,i} \to 4/(c^2\Delta_i^2 x_i^2)$ locally away from the boundaries. For compactly supported monotone truncations $H_m \uparrow 4/(c^2 x_i^2)$, its uniform domination and Lemma 2 give

$$\liminf_T \frac{\Delta_i^2}{\log T} \sum_{t=2}^{T} \mathbb{E} p_{t,i} \geq \int H_m(x_i)\, \Gamma(dx).$$

Let $m \to \infty$, use monotone convergence and (28), multiply by Δ_i^{-1}, and sum over the fixed finite set of arms. □

A.3 Uniform limit as the gap vanishes and proof of the theorem

Lemma 4 (Uniform invariant limit)**.** *For fixed $c > 0$, as $\Delta \downarrow 0$,*

$$\sup_{\theta \in \Theta_\Delta(c)} \left| \int x^{-2} \pi_{\Delta,\theta}(dx) - \frac{(r+1)(r+2)}{2} \right| \longrightarrow 0. \tag{49}$$

Proof. For $\Delta \leq \min\{1/2, 1/c^2\}$, the calculation for $V_3 = x + x^{-3}$ is uniform over $\theta \in \Theta_\Delta(c)$ and gives

$$\mathcal{A}_{\Delta,\theta} V_3 \leq -\frac{1}{4} V_3 + b\mathbf{1}_{[\varepsilon,R]}, \qquad \sup_{\substack{0<\Delta\leq \min\{1/2,1/c^2\} \\ \theta \in \Theta_\Delta(c)}} \pi_{\Delta,\theta}(V_3) < \infty. \tag{50}$$

This yields tightness at both zero and infinity. For $f \in C_c^\infty(0,\infty)$, Taylor expansion of (24) has a common compact source and, uniformly in $\theta \in \Theta_\Delta(c)$,

$$\mathcal{A}_{\Delta,\theta} f \longrightarrow \mathcal{L} f = (1-x) f'(x) + \frac{c^2 x^2}{32} f''(x). \tag{51}$$

Every subsequential invariant limit therefore annihilates $\mathcal{L}$. The diffusion $dX_t = (1 - X_t)dt + \sqrt{r} X_t\, dW_t$ has the unique invariant inverse Gamma density

$$p_0(x) = \frac{\beta^\gamma}{\Gamma(\gamma)} x^{-\gamma-1} e^{-\beta/x}, \qquad \gamma = 1 + \frac{2}{r}, \quad \beta = \frac{2}{r}.$$

Hence the convergence is uniform over θ by the sequential criterion. The x^{-3} bound in (50) makes x^{-2} uniformly integrable, and

$$\int x^{-2} p_0(x)\, dx = \frac{\gamma(\gamma+1)}{\beta^2} = \frac{(r+1)(r+2)}{2},$$

which proves (49). □

The preceding lemmas give the invariant limits and a liminf transfer. The equality step and completion of Theorem 1 are supplied in Appendix C, after the general occupation equality theorem.

B A general occupation equality theorem

Weak occupation limits do not by themselves control an unbounded sampling cost. The theorem below gives one step conditions under which recovery and uniform integrability supply the missing equality. Appendices C, D, and H verify these conditions for the three learners considered in the paper.

Let $D_n = (D_{n,1}, \ldots, D_{n,m})$ be adapted to $\mathcal{F}_n$, $t = n+1$, $Z_{n,i} = D_{n+1,i} - D_{n,i}$, and $G_n = \{D_{n,i} \geq 0\ \forall i\}$. Let $p_{n+1,i} \in [0,1]$ be $\mathcal{F}_n$-measurable. Constants below are deterministic and uniform over all sufficiently late times and realized histories.

One step assumptions. Fix $a > 0$, $h_n = a\sqrt{n+1}$, and assume the following.

1. The gap increments have uniform conditional first moments. On G_n, $\mathbb{E}[Z_{n,i} \mid \mathcal{F}_n] \geq \delta_i > 0$.
2. For fixed $d_0, C < \infty$ and $0 < \gamma \leq 1$, on $G_n \cap \{D_{n,i} = d \geq d_0\}$,
$$D_{n+1,i} \geq \gamma d, \qquad \mathbb{E}[Z_{n,i}^2 \mid \mathcal{F}_n] \leq C(1 + d^2/t), \qquad p_{n+1,i} \leq Ct/d^2.$$
For each fixed $b > 0$, $|Z_{n,i}|$ is uniformly bounded on $G_n \cap \{0 \leq D_{n,i} \leq b\sqrt{t}\}$, after a deterministic burn in.
3. There is a nonnegative adapted process W_n with uniform conditional first moments of its increments and finite expected value at the initial deterministic time, such that
$$\mathbb{E}[W_{n+1} - W_n \mid \mathcal{F}_n] \leq -\kappa \quad \text{if } \min_i D_{n,i} < 2h_n, \qquad W_n \leq C\sqrt{t} \quad \text{on } G_n, \tag{52}$$
for some $\kappa > 0$.

All gaps have finite first moments at the initial deterministic time. Assumption (52) is a global one step drift condition, not an assumed hitting time bound.

Theorem 4 (Occupation equality from joint recovery)**.** *Under these assumptions, define high episodes as passages from $\{D_i \geq 2h\ \forall i\}$ to $\{\min_i D_i \leq h\}$, with the intervening recoveries low. Include the initial recovery, or start high immediately if its upper target already holds. Write I_n for the predictable high indicator and τ_j for the times at which the process enters the high region. Then*

$$\mathbb{E}\sum_n (1 - I_n) < \infty, \qquad \mathbb{E}\sum_{j:\tau_j<\infty} (\tau_j + 1) < \infty. \tag{53}$$

For every coordinate, with $Y_{n,i} = (n+1)p_{n+1,i}$, the $3/2$ moment bound holds on high rounds:

$$\sum_{n<N} \frac{\mathbb{E}[I_n Y_{n,i}^{3/2}]}{n+1} \leq C(1 + \log(N+1)), \qquad \sum_{n<N} \mathbb{E}p_{n+1,i} = O(\log(N+1)). \tag{54}$$

Its full family of harmonic intensity laws is uniformly integrable.

Suppose separately that the high round normalized states X_n have weak harmonic limit that is a probability measure π on a positive open state space E, and that $Y_{n,i}$ converges to a continuous finite $g_i(X_n)$ uniformly on each compact subset of E. Then

$$\lim_{N\to\infty} \frac{\sum_{n<N} \mathbb{E}p_{n+1,i}}{\log N} = \int_E g_i \, d\pi < \infty. \tag{55}$$

The same statement holds along any subsequence with that weak limit. Weak convergence or uniqueness of the occupation law is not asserted by the one step assumptions.

Proof. Increase a deterministic n_0 so all needed local bounds hold and $h_n > d_0$. Finite stopping of (52) at
$$\tau = \inf\{n \geq \sigma : D_{n,i} \geq 2h_n\ \forall i\}$$
for any finite starting stopping time $\sigma \geq n_0$ gives
$$\mathbb{E}[\tau - \sigma \mid \mathcal{F}_\sigma] \leq W_\sigma/\kappa. \tag{56}$$
Use at most k relative increments, drop the nonnegative terminal potential, and let $k \to \infty$ by monotone convergence. Uniform conditional first moment control justifies each finite random start sum. The initial recovery therefore has finite mean.

All downward crossings of h and $2h$ during a high episode have bounded undershoot. Indeed, the fraction of steps without a crossing first localizes a departure coordinate that lands below bh_{n+1} to a fixed multiple of $\sqrt{n+1}$. Its local bounded increment condition then bounds the undershoot, including the one step motion of the curve. At a joint lower entrance ρ, all coordinates are consequently at least $h_\rho J > 0$. Assumption (52) and (56) give
$$\mathbb{E}[\tau_{\text{next}} - \rho \mid \mathcal{F}_\rho] \leq C\sqrt{\rho+1}. \tag{57}$$

For each coordinate, start an attempt when it enters below $2h$ during a high episode; every high entry equality at $2h$ also starts an attempt. The starting value at time s is at least $2h_s - J > h_s$. End the attempt when this coordinate exits $(h, 2h]$, or when any other coordinate ends the joint high episode. Inside the stopped strip, $D_{n,i} - h_n$ has bounded increments and conditional drift at least $\delta_i/2$. For small fixed $\lambda_i > 0$, its stopped negative exponential is a supermartingale. The elementary estimate

$$e^{-\lambda x} \le 1 - \lambda x + \lambda^2 B^2 e^{\lambda B}/2 \quad (|x| \le B)$$

proves this directly. Finite stopping and Fatou show

$$\mathbb{P}(\text{own lower exit} \mid \mathcal{F}_s) \le e^{-\lambda_i(h_s - J)}.$$

A simultaneous lower exit remains an own lower exit; stopping due to another coordinate can only discard nonnegative terms in this bound.

The exit time ζ has controlled moments. The martingale

$$\sum_{k=s}^{n-1} \mathbf{1}_{\{\zeta > k\}} \big(Z_{k,i} - \mathbb{E}[Z_{k,i} \mid \mathcal{F}_k]\big)$$

has bounded differences. Remaining in the strip for k steps forces a deviation at most $-\delta_i k/2$ once $k \ge C(\sqrt{s+1} + 1)$. The same bounded increment exponential estimate therefore gives $\mathbb{P}(\zeta - s > k \mid \mathcal{F}_s) \le e^{-c_i k}$ for those k. For each $v > 0$, this bounds $\mathbb{E}[(\zeta + 1)^{2v} \mid \mathcal{F}_s]$ by $C_v(s+1)^{2v}$. Cauchy–Schwarz with the exit probability yields

$$\mathbb{E}[(\zeta + 1)^v \mathbf{1}_{\{\text{own lower exit}\}} \mid \mathcal{F}_s] \le C_v(s+1)^v e^{-c_v\sqrt{s+1}}. \tag{58}$$

The indicator of an actual attempt at deterministic integer s is $\mathcal{F}_s$-measurable. Multiply (58) by it and sum over integers and the finitely many coordinates. Every joint lower entrance has at least one successful coordinate attempt. Simultaneous exits may be assigned to the smallest indexed exiting coordinate. The deterministic series is summable, so

$$\mathbb{E} \sum_{\text{joint lower entries } \rho} (\rho + 1)^v < \infty \quad (v > 0).$$

Together with (57), this proves finite expected total low duration. Also $\mathbb{E}[\tau_{\text{next}} + 1 \mid \mathcal{F}_\rho] \le \rho + 1 + C\sqrt{\rho + 1}$, which proves the first moment at entry to the high region sum in (53). No higher moment of an individual recovery duration has been assumed.

On high rounds use $V_{n,i}(d) = t^2/d^2$. The entire interpolation interval between d and $D_{n+1,i}$ lies in $[\gamma d, \infty)$, including the ending transition. Taylor's theorem and the variance bound give

$$\begin{aligned}\mathbb{E}[V_{n+1,i}(D_{n+1,i}) - V_{n,i}(d) \mid \mathcal{F}_n] &\le Ctd^{-2} - 2\delta_i t^2 d^{-3} + Ct^2 d^{-4} \\ &\le C'/t - \kappa_i t^2 d^{-3}.\end{aligned}$$

Absorb the last positive term since $d > h_n \to \infty$, then use boundedness above of $Cu^2 - \delta_i u^3/2$, $u \ge 0$.

Telescope with the predictable high indicator over truncated high episodes. Drop the nonnegative terminal potentials. Entry potentials are at most $(\tau_j + 1)/(4a^2)$, whose expected sum is finite by (53). At every fixed horizon, all active and terminal potentials are deterministically bounded, so the finite horizon expectations are justified. The drift budget is at most the full harmonic sum. Consequently

$$\sum_{n<N} \mathbb{E}[I_n t^2/D_{n,i}^3] \le C(1 + \log(N+1)).$$

The rare action assumption gives $Y_{n,i}^{3/2}/t \le Ct^2/D_{n,i}^3$ on high rounds, which proves the first bound in (54); the indicator I_n excludes low rounds there. For the second bound, split each $\mathbb{E}p_{n+1,i}$ at I_n. On high rounds $p_{n+1,i} \le (1 + Y_{n,i}^{3/2})/(n+1)$, and low rounds contribute at most their expected total count, which is finite by (53). No $3/2$ moment bound is claimed on low rounds; the uniform integrability argument below uses only (53) there.

For the tail, the high contribution is at most $CM^{-1/2}(1+\log(N+1))$. For the low contribution use $Y_{n,i} \le n+1$ and (53). Writing $\mathcal{H}_N = \sum_{n<N} 1/(n+1)$, its normalized tail is at most $\sum_{n+1>M} \mathbb{P}(I_n = 0)$, which tends to zero. The ratio $(1+\log(N+1))/\mathcal{H}_N$ is uniformly bounded. This proves uniform integrability for the full family, and also

$$\limsup_{N\to\infty} \frac{1}{\log N} \sum_{n<N} \mathbb{E}[p_{n+1,i}\mathbf{1}_{\{Y_{n,i}>M\}}] \le CM^{-1/2}.$$

Finally, a probability weak limit on the positive open state space provides tightness on its compact subsets. On each such compact the assumed local intensity approximation transfers bounded continuous tests of Y to tests of $g_i(X)$. The early harmonic terms and low mass vanish after normalization. Thus the intensity laws converge weakly to $(g_i)_{\#}\pi$. Uniform integrability transfers their first moments and proves (55), including finiteness of its right side. A vague or subprobability state limit would not suffice for this step. □

Scalar recovery alternative. For one coordinate, the auxiliary condition can instead be derived from global positive drift and a bounded negative to positive crossing overshoot, together with the local strip bounds. Finite stopping of $D_n - 2h_n$ gives mean recovery time at most $C(2h_\sigma - D_\sigma + J)$; at later lower entrances this is $O(\sqrt{\sigma+1})$. The same proof then applies. For the application to multiple arms below we verify the auxiliary condition directly using the FTRL objective.

C JOINT RECOVERY FOR THE EXACT TSALLIS LEARNER

We verify Theorem 4 for the reduced variance learner and all fixed finite K. Together with Lemmas 2 and 4, this proves Theorem 1 and Corollary 1. Here its coordinate i ranges over the suboptimal arms $2,\dots,K$, with gap $D_{n,i} = \widehat{L}_{n,i} - \widehat{L}_{n,1}$. The main new ingredient is global recovery of the whole gap vector.

The nonnegative comparison potential. For the exact objective at round t, write

$$F_t(L) = \min_{p\in\Delta_K} \{\langle p, L\rangle + \Psi_t(p)\}, \quad \Psi_t(p) = -\frac{4}{\eta_t}\sum_a \sqrt{p_a}, \quad \Phi_t(L) = L_1 - F_t(L) - \frac{4}{\eta_t}.$$

Taking $p = e_1$ and bounding the objective by its smallest coordinate give

$$0 \le \Phi_t(L) \le L_1 - \min_a L_a + \frac{4}{\eta_t}(\sqrt{K}-1). \tag{59}$$

Set $W_n = \Phi_{n+1}(\widehat{L}_n)$.

Lemma 5 (Global conditional drift)**.** *For the exact reduced variance learner, at every round,*

$$\mathbb{E}[W_{n+1} - W_n \mid \mathcal{F}_n] \le -\sum_{i\ge 2} \Delta_i p_{n+1,i} + \frac{C_{K,c}}{\sqrt{n+1}}, \qquad C_{K,c} = \frac{2c\sqrt{K}}{3} + \frac{2(\sqrt{K}-1)}{c}. \tag{60}$$

Proof. At fixed t, let p and q minimize the objectives at L and $L+v$. KKT and the simplex constraint give the exact remainder

$$\Phi_t(L+v) - \Phi_t(L) - \langle e_1 - p, v\rangle = \langle p - q, v\rangle - \sum_a \frac{2}{\eta_t\sqrt{p_a}}(\sqrt{q_a} - \sqrt{p_a})^2.$$

Relaxing the simplex coordinatewise, put $y = \sqrt{q_a/p_a}$, $A = 2\sqrt{p_a}/\eta_t$ and $B = p_a v_a$. Complete the square:

$$B(1-y^2) - A(y-1)^2 = \frac{B^2}{A+B} - (A+B)\left(y - \frac{A}{A+B}\right)^2.$$

The only negative estimator outcome is $v_a = 1/2 - 1/(2p_a)$ on a chosen active baseline zero loss. The threshold implies $\eta_t \le \sqrt{p_a}$, so $\eta_t\sqrt{p_a}v_a \ge -1/2$. Thus $A + B > 0$ and the remainder is at most

$$\sum_a \frac{\eta_t p_a^{3/2} v_a^2}{2 + \eta_t\sqrt{p_a}v_a}. \tag{61}$$

Direct enumeration of the estimator gives

$$\mathbb{E}[v_a^2 \mid \mathcal{F}_n] = \begin{cases} \mu_a - \frac{1}{4} + \frac{1}{4p_a}, & B_a = 1/2, \\ \mu_a/p_a, & B_a = 0, \end{cases} \qquad \mathbb{E}[v_a^2 \mid \mathcal{F}_n] \le 1/p_a.$$

Consequently the expected remainder is at most $(2/3)\eta_t \sum_a \sqrt{p_a} \le (2/3)\eta_t\sqrt{K}$. Unbiasedness makes the expected linear term $-\sum_{i\ge 2} \Delta_i p_i$.

At fixed updated L, the change from Φ_t to Φ_{t+1} is at most $4(\sqrt{t+1} - \sqrt{t})(\sqrt{K} - 1)/c \le 2(\sqrt{K} - 1)/(c\sqrt{t})$. This proves (60). □

Fix $a > 0$ and $h_n = a\sqrt{n+1}$. Outside the joint target $\{D_{n,i} \ge 2h_n \ \forall i\}$,

$$\sum_{i\ge 2} \Delta_i p_{n+1,i} \ge \delta := \frac{\Delta_{\min}}{(ca + \sqrt{2})^2}. \tag{62}$$

If $p_1 \le 1/2$, the left side is at least $\Delta_{\min}/2$. Otherwise arm 1 is the empirical minimizer and a coordinate $D_{n,i} < 2h_n$ has $p_i^{-1/2} = \eta_t D_{n,i}/2 + p_1^{-1/2} < ca + \sqrt{2}$. For large n, (60) is therefore at most $-\delta/2$ outside the target. On the positive orthant, (59) gives $W_n \le 4(\sqrt{K} - 1)\sqrt{n+1}/c$. The first moment requirements also hold: at fixed t, Φ_t is Lipschitz in L with respect to the coordinate sup norm with constant at most 2, and every estimator coordinate has uniformly bounded conditional absolute first moment. Its time increment is bounded. Thus all global recovery assumptions in (52) are verified, including deterministic time integrability.

Positive orthant coordinate bounds. Inside the positive orthant, arm 1 has maximal probability, $p_1 \ge 1/K$, and its baseline is active once $t \ge Kc^2$. For $d = D_{n,i} \ge 0$, KKT gives

$$p_i^{-1/2} = \frac{cd}{2\sqrt{t}} + p_1^{-1/2} \le \frac{cd}{2\sqrt{t}} + \sqrt{K}. \tag{63}$$

If $d \le b\sqrt{t}$, this coordinate and arm 1 have probabilities bounded below. Their baselines are eventually active and every possible increment of this coordinate is bounded, including third arm outcomes. Other coordinates may be arbitrarily large.

For a small active own probability $p_i \le 1/(4K)$, $\sqrt{p_i/p_1} \le 1/2$, and its own downward jump has ratio

$$\frac{1/(2p_i)}{d} = \frac{\eta_t}{4\sqrt{p_i}(1 - \sqrt{p_i/p_1})} \le 1/2.$$

For $p_i > 1/(4K)$ that jump is at most $2K$. The optimal arm centered update is at most $K/2$. With inactive own baseline, the own zero loss change is $-1/2$, the loss one change is positive, an optimal arm downward change has magnitude at most $(K+1)/2$, and a third arm change is $-1/2$. Hence, with $J_K = 2K + 1$,

$$D_{n+1,i} \ge \min\{d/2, d - J_K\}.$$

For $d \ge 2J_K$ this supplies the required no crossing fraction $1/2$. Unbiasedness gives mean Δ_i, and the second moment and rare action bounds follow from (63):

$$\mathbb{E}[Z_{n,i}^2 \mid \mathcal{F}_n] \le 2(1/p_i + 1/p_1) \le C(1 + d^2/t), \qquad p_i \le 4t/(c^2 d^2).$$

All assumptions of Theorem 4 are now verified.

Completion of Theorem 1. The general theorem proves logarithmic expected sampling and harmonic cost UI for every coordinate. Starting with any diverging horizon sequence, Lemma 2 therefore admits a further weak occupation subsequence. Every such limit has scalar marginals π_{Δ_i,θ_i}. Restricting to high rounds preserves this limit, since low rounds have finite expected total count and thus vanishing normalized harmonic mass. On compact positive state sets, (11) gives the continuous intensity map $g_i(x) = 4/(c^2\Delta_i^2 x_i^2)$. The equality transfer in Theorem 4 identifies each subsequential sampling constant with $4\int x^{-2}\pi_{\Delta_i,\theta_i}(dx)/(c^2\Delta_i^2)$. This value depends only on the already unique scalar marginal; uniqueness of the joint invariant law is unnecessary. All horizon sequences consequently

have the same constant. Multiplication by Δ_i and summation prove (15). Lemma 4 proves (16) by its uniform coordinate error, and the displayed minimization of $C(c)$ is algebraic.

A wrong empirical leader or a tie involving arm 1 implies $\min_{i\geq 2} D_{n,i} \leq 0$. Such a round is low for sufficiently large n, so (53) proves finite expected total count. Ties solely among suboptimal arms are not included in this assertion. □

D A SECOND APPLICATION: PLAIN IMPORTANCE WEIGHTING

Plain importance weighting separates the effect of the estimator from that of the FTRL probability map. Corollary 2 proves recovery, Subsection D.1 obtains the fixed gap coefficient, and Subsection D.2 evaluates the limit used in (17).

Keep the probability map (1), initialize $\widehat{L}_0 = 0$, and replace only the estimator by

$$\widehat{\ell}^{\mathrm{IW}}_{t,a} = \frac{\mathbf{1}_{\{I_t=a\}}\ell_{t,a}}{p_{t,a}}.$$

Here $N_i(T) = \sum_{t=1}^{T} \mathbf{1}_{\{I_t=i\}}$ counts pulls. Thus $\widehat{\ell}^{\mathrm{IW}}_{t,a} \geq 0$, with conditional first and second moments μ_a and $\mu_a/p_{t,a}$.

Corollary 2 (Plain IW recovery and harmonic uniform integrability)**.** *For every fixed finite K, $c > 0$, and Bernoulli instance with a unique optimal arm, the plain IW learner satisfies the one step assumptions of Theorem 4. Hence its joint low episodes have finite expected total duration, its entry into the high region times have a summable first moment, and its harmonic sampling intensity laws are uniformly integrable. In particular, for each suboptimal arm, $\mathbb{E}N_i(T) = O(\log(T+1))$, and*

$$\mathbb{E}\sum_{t\geq 1} \mathbf{1}_{\{\min_{i\geq 2} D_{t-1,i}\leq 0\}} < \infty.$$

No transversality at the switching boundary condition or restriction $\Delta_i \leq 1/2$ is required for these conclusions.

Proof. The FTRL comparison potential Φ_t from Appendix C applies without change because its exact Bregman remainder depends only on the objective. With plain IW, every estimator coordinate v_a is nonnegative, so the denominator $2 + \eta_t\sqrt{p_a}v_a$ is at least 2. The conditional expected stability cost is at most

$$\frac{\eta_t}{2}\sum_a p_{t,a}^{3/2}\mathbb{E}[v_a^2 \mid \mathcal{F}_{t-1}] = \frac{\eta_t}{2}\sum_a \mu_a\sqrt{p_{t,a}} \leq \frac{\eta_t\sqrt{K}}{2}.$$

For $W_n = \Phi_{n+1}(\widehat{L}_n)$, the same regularizer time increment therefore gives

$$\mathbb{E}[W_{n+1} - W_n \mid \mathcal{F}_n] \leq -\sum_{i\geq 2}\Delta_i p_{n+1,i} + \frac{C^{\mathrm{IW}}_{K,c}}{\sqrt{n+1}}, \qquad C^{\mathrm{IW}}_{K,c} = \frac{c\sqrt{K}}{2} + \frac{2(\sqrt{K}-1)}{c}.$$

Outside the joint upper target with $h_n = a\sqrt{n+1}$, exact KKT gives the same regret lower bound $\delta = \Delta_{\min}/(ca + \sqrt{2})^2$. Consequently this is a strictly negative eventual drift. The potential is nonnegative and at most $4(\sqrt{K}-1)\sqrt{n+1}/c$ on the positive orthant. Its conditional absolute increments are bounded in expectation, since $\mathbb{E}\sum_a v_a = \sum_a \mu_a \leq K$ and the potential is 2-Lipschitz for the coordinate sup norm.

Write $t = n+1$ and $d = D_{n,i}$. Inside the positive orthant, $p_1 \geq 1/K$ and the only gap updates are $-\ell_1/p_1$ on an optimal arm pull, ℓ_i/p_i on an own arm pull, and zero otherwise. In particular,

$$D_{n+1,i} \geq d - K \geq d/2 \quad (d \geq 2K).$$

The exact coordinate moments and KKT imply

$$\mathbb{E}[Z_{n,i} \mid \mathcal{F}_n] = \Delta_i, \qquad \mathbb{E}[|Z_{n,i}| \mid \mathcal{F}_n] = \mu_i + \mu_1 \leq 2,$$
$$\mathbb{E}[Z_{n,i}^2 \mid \mathcal{F}_n] = \mu_i/p_i + \mu_1/p_1 \leq c^2d^2/(2t) + 3K, \qquad p_i \leq 4t/(c^2d^2) \quad (d > 0).$$

On $0 \leq d \leq b\sqrt{t}$, every possible own coordinate increment has magnitude at most $(cb/2 + \sqrt{K})^2$. These verify every assumption of Theorem 4, with $d_0 = 2K$ and $\gamma = 1/2$. Its conclusions prove the corollary; the displayed wrong leader event is low after the deterministic burn in. □

D.1 THE IW OCCUPATION LAW AND EXACT COEFFICIENT

Theorem 5 (Plain IW occupation equality)**.** *For every instance in Corollary 2, fix $i \geq 2$ and put*

$$\Delta = \Delta_i, \quad u = \mu_1/\Delta, \quad q = c^2\Delta/4, \quad v = 4\mu_i/(c^2\Delta^2).$$

The scalar PDMP on $E = (0, \infty)$ with generator

$$A^{\mathrm{IW}} f(x) = (-x - u)f'(x) + \frac{v}{x^2}[f(x + qx^2) - f(x)] \tag{64}$$

is conservative and has a unique invariant probability π_i^{IW}. With $H = \sum_{t=2}^{T} t^{-1}$,

$$\Gamma_{T,i}^{\mathrm{IW}} = H^{-1} \sum_{t=2}^{T} t^{-1} \mathcal{L}\left(\frac{D_{t-1,i}}{\Delta_i(t-1)}\right) \Rightarrow \pi_i^{\mathrm{IW}}.$$

Moreover,

$$\lim_{T\to\infty} \frac{R_T}{\log T} = \sum_{i=2}^{K} \frac{4}{c^2\Delta_i} \int_E x^{-2} \pi_i^{\mathrm{IW}}(dx). \tag{65}$$

This statement requires neither RV transversality nor $\Delta_i \leq 1/2$.

Proof. We abbreviate $A = A^{\mathrm{IW}}$ and work with a single coordinate. Write $G_{t-1} = \{\min_j D_{t-1,j} \geq 0\}$. Corollary 2 gives $B = \sum_{t\geq 2} \mathbb{P}(G_{t-1}^c) < \infty$, $\mathbb{E}N_i(T) = O(\log T)$ and harmonic uniform integrability of $tp_{t,i}$.

Tightness and the exact scalar kernel. The conditional absolute gap increment has mean $\mu_i + \mu_1 \leq 2$, so $\sup_n \mathbb{E}|X_{n,i}| \leq 2/\Delta$. On G_{t-1}, $p_{t,1} \geq 1/K$ and

$$tp_{t,i} = \left[\frac{c\Delta}{2}(1 - 1/t)x + \frac{1}{\sqrt{tp_{t,1}}}\right]^{-2}. \tag{66}$$

For $0 \leq x \leq \epsilon$ and sufficiently large t, the right side is at least $(c\Delta\epsilon)^{-2}$. The logarithmic pull bound thus implies $\limsup_T \Gamma_{T,i}^{\mathrm{IW}}([0, \epsilon]) = O(\epsilon^2)$. Negative states cost at most B/H. Every weak limit ρ is a probability on E and, by compact truncation,

$$\rho x \leq 2/\Delta, \qquad \rho x^{-2} \leq \frac{c^2\Delta^2}{4} \limsup_T \frac{\mathbb{E}N_i(T)}{H} < \infty.$$

Replacing bad round states by 1 gives the same weak limits on E.

For $f \in C_c^2(E)$ extended by zero and $r = (1 - 1/t)x$, the exact full state conditional kernel is

$$\begin{aligned} P_t f = f(r) &+ \mu_i p_i[f(r + 1/(\Delta t p_i)) - f(r)] \\ &+ \mu_1 p_1[f(r - 1/(\Delta t p_1)) - f(r)]. \end{aligned} \tag{67}$$

Equation (66) gives, uniformly on scalar compact sets and G_{t-1}, $tp_i = 4/(c^2\Delta^2x^2) + O(t^{-1/2})$ and $1/(\Delta tp_i) = qx^2 + O(t^{-1/2})$. Taylor expand only the normalization and optimal arm terms in (67); the own arm macroscopic jump is retained exactly. Thus $t(P_t - I)f = Af + O_f(t^{-1/2})$.

This comparison is global on G_{t-1}. Every next state is at least $(1 - 1/t)x - K/(\Delta t)$, while an own upward landing is at most $x + qx^2 + c\sqrt{K}x/\sqrt{t} + K/(\Delta t)$. If $\operatorname{supp} f \subset [a, d]$, choose $l > 0$ with $l + ql^2 < a/2$ and $R > 2d + K/\Delta$. For large t, neither $x < l$ nor $x > R$ can contribute. The limiting source hull is also compact. No other normalized coordinate was restricted to a compact set. Telescoping the full chain, subtracting bad round increments, and summing $\sum_t t^{-3/2}$ yields

$$|\Gamma_{T,i}^{\mathrm{IW}} Af| \leq C_f/H. \tag{68}$$

Since Af is continuous and compactly supported, every limit annihilates A on $\mathcal{D} = \mathbb{R}1 + C_c^1(E)$, by common compact C^1 approximation.

Conservativeness and well posedness. The flow is $\phi_s(x) = (x+u)e^{-s} - u$ and the jump is $J(x) = x + qx^2$ at rate $\lambda(x) = v/x^2$. If $u > 0$, its integrated hazard diverges before the zero endpoint $\log((x+u)/u)$; if $u = 0$, the flow stays positive for every finite time. For $r \geq 2$, the Taylor inequality $(1+z)^{-r} \leq 1 - rz + r(r+1)z^2/2$ for $z \geq 0$ gives

$$Ax = 1 - x, \qquad A(x^{-r}) \leq -rx^{-r-1} + r\left(1 + \frac{(r+1)c^2\mu_i}{8}\right)x^{-r}. \tag{69}$$

For $V_r = 1 + x + x^{-r}$ this implies $AV_r \leq C_r V_r$. Cutoffs stopped on compact sets include the entire landing hull. Dynkin's formula and Gronwall therefore give

$$\mathbb{P}_x(\tau_m \leq T) \leq e^{C_r T} V_r(x)/m, \qquad \mathbb{E}_x V_r(X_T) \leq e^{C_r T} V_r(x), \quad \tau_m = \inf\{t : V_r(X_t) \geq m\}.$$

The clock construction is conservative; its expected finite horizon jump count is finite because $\lambda \leq CV_2$. Its bounded test identities are true martingales with envelope $\Phi = 1 + |b| + \lambda$.

We verify the arbitrary solution part of the martingale problem, rather than equating clock existence with well posedness. The compactification argument in the proof of Lemma 2 applies to a countable separating $C_c^\infty(E)$ family. Its martingale versions give a càdlàg modification on the one point compactification. A plateau $\chi_m = 1$ on $\{V_2 < m\}$ and all its J-landings has $A\chi_m = 0$ before exit. Optional sampling excludes a cemetery exit; (69) then removes all compact stops. Cutoff coordinate tests first make the stopped solution a special semimartingale. Cutoff Fourier tests and Jacod–Shiryaev II.2.42(b)⇒(a) identify

$$C = 0, \qquad \nu(ds, dz) = \mathbf{1}_{\{s \leq \tau_m\}} \lambda(Y_{s-}) \delta_{qY_{s-}^2}(dz)\, ds,$$

$$dB_s = \mathbf{1}_{\{s \leq \tau_m\}} [b(Y_{s-}) + \lambda(Y_{s-})\kappa(qY_{s-}^2)]\, ds.$$

The first jump survival is consequently $\exp[-\int_0^t \lambda(\phi_s(x))\, ds]$. On canonical Skorokhod space, optional sampling for the countable graph core, rational times and cylinder algebra gives conditional shifted martingale identities outside one common null set. Conditional Φ integrability, right continuity and graph closure extend them to the full domain. The atomless time compensator rules out fixed time jumps under the conditional laws. The conditional first jump recursion identifies all finite jump sequences; the compact exit bound removes the stops. This proves BKR A3.

For a general initial law, first condition on the initial filtration events $\{V_2(Y_0) \leq k\}$. Their stopped bounds are $e^{CT}k/m$; their union has probability one, so the regularized path is E-valued without an initial moment assumption. Now disintegrate on canonical path space at Y_0. The same countable argument and conditional integrability identify almost every conditional law with the point initial clock law. This proves A6 with $h_1 = 0$.

The remaining BKR conditions are direct: Φ and Af are continuous (A7); $|Af| \leq C_f\Phi$ (A1); rational common compact C^1 approximants give a countable weighted graph core (A2); and $\mathcal{D}$ is a constant containing separating algebra (A8). The moment bounds above give $\rho\Phi < \infty$. BKR Theorem 4.2 therefore makes ρ invariant (Bhatt et al., 2006). The discontinuous generator extension is unnecessary here.

Uniqueness of the invariant law. The Markov resolvent $\mathcal{R}(x, B) = \int_0^\infty e^{-t} P_t(x, B)\, dt$ is equivalent to Lebesgue measure for every $x > 0$. For absolute continuity, each finite jump path has strictly decreasing smooth flow segments with velocity $-(x+u)$, so its occupation on each segment has a density. For strict positivity, any positive measure B has a positive measure intersection with some $[a, d] \Subset E$. Finitely many sufficiently rapid upward jumps reach a state above d with positive probability, since $J(z) - z = qz^2 > 0$. The subsequent no jump flow across $[a, d]$ has positive survival and positive occupation density. This proves $\mathcal{R}(x, B) > 0$.

If π_1, π_2 were invariant, set $\bar{\pi} = (\pi_1 + \pi_2)/2$ and $h = d\pi_1/d\bar{\pi} \in [0, 2]$. Under $Q(dx, dy) = \bar{\pi}(dx)\mathcal{R}(x, dy)$, invariance gives $h(Y) = \mathbb{E}[h(X) \mid Y]$. Equal marginal laws force equality in conditional Jensen for the square, so $h(X) = h(Y)$ almost surely. Since Q is equivalent to $\bar{\pi} \otimes \bar{\pi}$, h is constant. Its mean is one, proving uniqueness.

All subsequential occupation limits consequently coincide. On positive compact sets (66) identifies the cost. Bad rounds have finite total expected probability cost, large x is controlled by $tp_i \leq 16/(c^2\Delta^2 x^2)$, and small x is controlled by the established harmonic UI together with the $O(\epsilon^2)$ occupation bound. Truncation gives (65). No joint product law assertion is needed. □

D.2 UNIFORM IW SMALL GAP LIMIT WITH AN EXPLICIT GAP ERROR

Write $\pi^{\mathrm{IW}}_{\Delta,\mu}$ for the invariant law in Theorem 5, with optimal mean μ and suboptimal mean $\mu+\Delta$, and define

$$k_c(\Delta,\mu)=\frac{4}{c^2}\int x^{-2}\pi^{\mathrm{IW}}_{\Delta,\mu}(dx),\qquad C_{\mathrm{IW}}(c,m)=\frac{4}{c^2}+\frac{3m}{2}+\frac{c^2m^2}{8}.$$

Theorem 6 (Uniform IW limit and finite gap enclosure)**.** *Fix $c>0$. For every $0<\Delta\le 1$ and $0\le\mu\le 1-\Delta$, with $a=c^2(\mu+\Delta)/4$,*

$$0\le C_{\mathrm{IW}}(c,\mu+\Delta)-k_c(\Delta,\mu)\le\frac{\Delta a(a+4)}{3}. \tag{70}$$

In particular, putting $L_c=3/2+7c^2/12+c^4/48$,

$$|k_c(\Delta,\mu)-C_{\mathrm{IW}}(c,\mu)|\le L_c\Delta. \tag{71}$$

The laws $\pi^{\mathrm{IW}}_{\Delta,\mu}$ converge, uniformly over the admissible μ in bounded Lipschitz distance, to ν_μ as $\Delta\downarrow 0$. Here $\nu_0=\delta_1$, and for $m>0$,

$$\nu_m(dx)=\frac{\beta^{\beta+1}}{\Gamma(\beta+1)}x^{-\beta-2}e^{-\beta/x}\,dx,\qquad \beta=\frac{8}{c^2m}. \tag{72}$$

For each fixed finite K instance, $S=\sum_{i\ge 2}\Delta_i^{-1}$ satisfies

$$\left|\frac{\lim_{T\to\infty}R_T/\log T}{S}-C_{\mathrm{IW}}(c,\mu_1)\right|\le L_c\max_{i\ge 2}\Delta_i. \tag{73}$$

The error in (73) *concerns the already taken horizon limit; it is not a finite horizon error estimate.*

Proof. Abbreviate $\pi=\pi^{\mathrm{IW}}_{\Delta,\mu}$ and $A=A^{\mathrm{IW}}$, and put $u=\mu/\Delta$, $q=c^2\Delta/4$ and $v=4(\mu+\Delta)/(c^2\Delta^2)$. Then $vq=u+1$ and $vq^2=a$.

Uniform moments without unbounded domain assumptions. For $V=x+x^{-3}$, (69) gives

$$AV\le 1-x-3x^{-4}+kx^{-3},\qquad k=3+\tfrac{3}{2}c^2.$$

The identity

$$\frac{z^4}{256}-zy^3+3y^4=\frac{(z-4y)^2((z+4y)^2+32y^2)}{256}\ge 0$$

implies

$$AV\le B_c-\tfrac{1}{2}V,\qquad B_c=1+(k+1/2)^4/256.$$

Choose smooth nondecreasing concave χ with $\chi(s)=s$ for $s\le 1$, $0\le\chi'\le 1$, and χ constant for $s\ge 2$. Then $f_n=n\chi(V/n)$ is a constant plus a compactly supported C^2 function on $(0,\infty)$. The flow chain rule and jump concavity give $Af_n\le\chi'(V/n)AV$. Invariance for this bounded legal test implies

$$\pi[V\chi'(V/n)]\le 2B_c.$$

Fatou, without a prior assumption that πV is finite, yields

$$\sup_{\Delta,\mu}\pi^{\mathrm{IW}}_{\Delta,\mu}(x+x^{-3})\le 2B_c. \tag{74}$$

Exact stationary inverse moments. The tests $x,\log x,x^{-1}$ are integrable under (74). Their absolute drift terms are respectively $x+u$, $1+u/x$, and $x^{-1}+ux^{-2}$. Their absolute jump compensator terms are bounded respectively by $u+1$, $(u+1)/x$, and $(u+1)x^{-2}$. All are integrable for each fixed gap. The finite jump path identity, compensation and Fubini therefore justify $\pi Ag=0$ for these three unbounded tests under stationary initialization. In particular $\pi x=1$.

For $z\ge 0$, let $R(z)=\log(1+z)-z+z^2/2$. Since $R'(z)=z^2/(1+z)$, $0\le R(z)\le z^3/3$ globally. Direct substitution gives

$$A\log x=x^{-1}-1-a/2+e(x),\qquad e(x)=vx^{-2}R(qx),\qquad 0\le e(x)\le aqx/3.$$

Thus $\pi x^{-1} = 1 + a/2 - \pi e$ and $0 \le \pi e \le aq/3$. The second exact identity is

$$A(x^{-1}) = (1+a)x^{-1} - x^{-2} - \frac{aq}{1+qx}.$$

Combining their stationary expectations yields

$$\begin{aligned}(1+a)(1+a/2) - \pi x^{-2} \\ = (1+a)\pi e + aq\,\pi[(1+qx)^{-1}] \;\in\; [0, aq(a+4)/3].\end{aligned} \tag{75}$$

Multiplication by $4/c^2$ proves (70). Moreover,

$$C_{\mathrm{IW}}(c, \mu+\Delta) - C_{\mathrm{IW}}(c,\mu) = \Delta[3/2 + c^2(2\mu+\Delta)/8] \le \Delta(3/2 + c^2/4),$$

and $a(a+4)/3 \le c^2/3 + c^4/48$. These bounds prove (71); weighting it by Δ_i^{-1}/S and using (65) proves (73).

Recognition of every weak limit. The limiting generator is

$$\mathcal{L}_m f = (1-x)f' + s_m x^2 f'', \qquad s_m = c^2 m/8.$$

For $f \in C_c^3(E)$ supported in $[l, R]$, the source hull of Af is contained uniformly in $[J_{\max}^{-1}(l), R]$, where $J_{\max}(x) = x + c^2x^2/4$. Upward jumps cannot enter the support from above it. Taylor expansion on this common compact, with exact cancellation of the first order terms, gives

$$\|A_{\Delta,\mu} f - \mathcal{L}_\mu f\|_\infty \le \Delta\left(\frac{c^2R^2}{8}\|f''\|_\infty + \frac{c^4R^4}{96}\|f'''\|_\infty\right).$$

For any sequence $\Delta_n \downarrow 0$, $\mu_n \to m$, (74) gives a subsequential probability limit ρ on E, and stationarity gives $\rho\mathcal{L}_m f = 0$.

We identify this measure directly. When $m > 0$, write $s = s_m$ and $w = sx^2\rho$ as a distribution. The stationary equation implies $[w' - (1-x)\rho]' = 0$, hence $w' - (1-x)w/(sx^2) = J\,dx$ for a constant J. With integrating factor $H_s = x^{1/s}e^{1/(sx)}$, the equation becomes $(H_s w)' = JH_s\,dx$. Thus ρ has density

$$\frac{C + J\int_1^x H_s(y)\,dy}{sx^2 H_s(x)}.$$

The primitive tends to $-\infty$ at zero and $+\infty$ at infinity. Nonnegativity forces $J = 0$, and normalization gives (72). When $m = 0$, the equation instead gives $(1-x)\rho = C\,dx$. Its sign on the two sides of 1 forces $C = 0$, so $\rho = \delta_1$.

For $m > 0$, x^{-1} under (72) is Gamma with shape $\beta + 1$ and rate β. Its first two moments are $1 + 1/\beta$ and $(1+1/\beta)(1+2/\beta)$. They tend to one as $m \downarrow 0$, giving continuity at ν_0. Continuity elsewhere follows from the Gamma family. Compactness of the mean parameter interval now turns the sequential identification into uniform bounded Lipschitz convergence. □

At $m > 0$, balancing $4/c^2$ and $c^2m^2/8$ gives

$$\inf_{c>0} C_{\mathrm{IW}}(c, m) = (3/2 + \sqrt{2})m, \qquad c^2 = 4\sqrt{2}/m.$$

At $m = 0$ the infimum is zero, unattained at finite c. In particular,

$$C_{\mathrm{IW}}(c, 1/2) = 2C_{\mathrm{RV}}(\sqrt{2}c), \qquad \min_c C_{\mathrm{IW}}(c, 1/2) = 2\min_c C_{\mathrm{RV}}(c). \tag{76}$$

The two minima use separate optimal scales and the same iterated horizon then small gap regime. They do not imply uniformity when the scale, gaps or number of arms change with the learner horizon.

E Explicit recovery and tail budgets

We now make the fixed instance recovery, entry, and tail constants explicit. Appendices F, G, and I use these estimates; the asymptotic constant in Theorem 1 does not. The bounds are conservative; they give no rate of convergence of the bounded test occupation measures. Assume the original initialization $\widehat{L}_0 = 0$, fixed K, c, and $\Delta_* = \min_{i \ge 2}\Delta_i > 0$. The estimates in this appendix allow every Bernoulli gap $0 < \Delta_i \le 1$. Fix $a > 0$; $a = 1/c$ is one convenient choice.

Burn in and geometry. Define

$$P = \frac{4(\sqrt{K}-1)}{c}, \quad C_0 = \frac{2c\sqrt{K}}{3} + \frac{2(\sqrt{K}-1)}{c}, \quad \delta = \frac{\Delta_*}{(ca+\sqrt{2})^2}, \quad J_0 = 2K+1,$$

$$Q = (2\sqrt{2}ca + \sqrt{K})^2, \qquad L = Q/2, \qquad B = L+1.$$

Take $n_0 = u_0 - 1$, where

$$u_0 = 1 + \left\lceil \max\left\{1, Kc^2, c^2Q, (2C_0/\delta)^2, (a/\Delta_*)^2, (4J_0/a)^2, (2B/a)^2, \left(\frac{2304K}{a\Delta_*}\right)^2\right\}\right\rceil. \quad (77)$$

The extra 1 makes the barrier positivity inequalities strict. The global recovery drift is then at most $-\delta/2$. On the positive orthant and $d \le 4\sqrt{2}a\sqrt{t}$, both relevant baselines are active, $p_i \ge 1/Q$, and $|Z_{n,i}| \le L$. The no skipping fraction localizes every downward crossing of h or $2h$ to this enlarged strip. The upper curve moves by at most 1 per step, so every crossing undershoots by at most B. In particular all coordinates at a joint lower entrance remain positive.

Explicit weighted sums at entry times. For each coordinate set

$$\lambda_i = \frac{\Delta_i}{2B^2}, \quad \kappa_i = \frac{\Delta_i^2}{8B^2}, \quad \beta_i = \frac{\Delta_i a}{8B^2}, \quad q_i = e^{-\kappa_i}, \quad d_i = \frac{8a}{\Delta_i} + \left(\frac{8a}{\Delta_i}\right)^2 + 1,$$

$$E_{1,i} = 1 + d_i + \frac{1}{1-q_i}, \qquad E_{2,i} = 2 + 2d_i^2 + 2\left\{\frac{2q_i}{(1-q_i)^2} + \frac{1}{1-q_i}\right\}.$$

For $j = 1, 2, 3$, put

$$F_j(\beta, z) = e^{-\beta z}\sum_{\ell=0}^{j} \frac{j!}{\ell!}\frac{z^\ell}{\beta^{j-\ell+1}}, \qquad z_0 = \sqrt{u_0 - 1},$$

$$S_{1/2}(\beta) = 2\{F_2(\beta, z_0) + F_1(\beta, z_0)\}, \qquad S_1(\beta) = 2\{F_3(\beta, z_0) + F_1(\beta, z_0)\}.$$

These are finite explicit expressions, not truncated numerical series. Define

$$\mathcal{A}_{1/2} = \sum_{i\ge 2}\sqrt{E_{1,i}}S_{1/2}(\beta_i), \qquad \mathcal{A}_1 = \sum_{i\ge 2}\sqrt{E_{2,i}}S_1(\beta_i).$$

The resulting budgets for low durations and entries into the high region are

$$R = \frac{2P}{\delta}, \quad R_0 = \frac{2(P + 2C_0\sqrt{n_0})}{\delta}, \quad \mathcal{L} = R_0 + R\mathcal{A}_{1/2}, \quad \mathcal{E} = u_0 + R_0 + \mathcal{A}_1 + R\mathcal{A}_{1/2}. \quad (78)$$

They satisfy

$$\mathbb{E}\sum_{n\ge n_0}(1 - I_n) \le \mathcal{L}, \qquad \mathbb{E}\sum_{\text{high entries } \tau}(\tau + 1) \le \mathcal{E}. \quad (79)$$

Here are the details of the series bound. For a coordinate attempt starting at s, let $u = s + 1$, let ζ be its stop, and let A_i denote an own lower exit. Inside the stopped strip, $D_i - h$ has increments bounded by B and mean at least $\Delta_i/2$. The exponential inequality for bounded variables with λ_i gives

$$\mathbb{P}(A_i \mid \mathcal{F}_s) \le e^{-\lambda_i(h_s - B)} \le e^{-\lambda_i a\sqrt{u}/2}.$$

The centered D_i increments also have magnitude at most B. Surviving k steps forces a negative deviation of at least $\Delta_i k/2$ once

$$k \ge \left\lceil \max\left\{8a\sqrt{u}/\Delta_i,\ (8a/\Delta_i)^2\right\}\right\rceil.$$

The bounded difference exponential estimate then gives $\mathbb{P}(\zeta - s > k \mid \mathcal{F}_s) \le e^{-\kappa_i k}$. Discrete first- and second moment tail sums imply

$$\mathbb{E}[\zeta + 1 \mid \mathcal{F}_s] \le E_{1,i}u, \qquad \mathbb{E}[(\zeta+1)^2 \mid \mathcal{F}_s] \le E_{2,i}u^2.$$

Cauchy–Schwarz consequently bounds the own lower weighted moments of orders $1/2$ and 1 by $\sqrt{E_{1,i}}u^{1/2}e^{-\beta_i\sqrt{u}}$ and $\sqrt{E_{2,i}}ue^{-\beta_i\sqrt{u}}$, respectively. For $v = 1/2, 1$, integrate $(x+1)^v e^{-\beta\sqrt{x}}$ over

each $[u-1,u]$ to dominate its discrete summand. Substitution $y=\sqrt{x}$ and $\sqrt{y^2+1}\le y+1$ give precisely S_v above. Thus summing actual predictable attempt starts gives the two lower entry moment budgets $\mathcal{A}_{1/2}$ and $\mathcal{A}_1$.

Finite stopping of the global comparison drift gives mean recovery at most $2W_\sigma/\delta$. Initially, $\mathbb{E}W_{n_0}\le P+2C_0\sqrt{n_0}$ by summing the global drift and dropping regret. At a later lower entrance ρ, $W_\rho\le P\sqrt{\rho+1}$. Summing these estimates proves (79), using only the first moment of each recovery duration.

A finite horizon tail estimate. Let $\mathcal{H}_T=\sum_{t=1}^{T}1/t$ and $\mathcal{H}_{u_0:T}=\sum_{t=u_0}^{T}1/t$, with the latter zero if $T<u_0$. Set

$$A=3+192c^2,\qquad C_i=\frac{16A^3}{27\Delta_i^2}.$$

On the positive orthant, $\mathbb{E}[Z_{n,i}^2\mid\mathcal{F}_n]\le c^2d^2/t+6K$. Taylor expansion of d^{-2} on $[d/2,\infty)$ has remainder at most $48Z^2/d^4$. For $V_n(d)=t^2/d^2$, this yields on high rounds

$$\mathbb{E}[\Delta V_n\mid\mathcal{F}_n]\le Atd^{-2}-2\Delta_i t^2d^{-3}+1152Kt^2d^{-4}.$$

Burn in (77) makes the last term at most $(\Delta_i/2)t^2d^{-3}$. The exact cubic maximum

$$\sup_{v\ge0}\{Av^2-(\Delta_i/2)v^3\}=\frac{16A^3}{27\Delta_i^2}$$

therefore gives $\mathbb{E}[\Delta V_n\mid\mathcal{F}_n]\le C_i/t-\Delta_i t^2d^{-3}$. Telescope over high episodes, charging their entries to $\mathcal{E}/(4a^2)$. KKT implies, for $Y_{n,i}=(n+1)p_{n+1,i}$,

$$\sum_{n=n_0}^{T-1}\frac{\mathbb{E}[I_nY_{n,i}^{3/2}]}{n+1}\le\frac{8}{c^3\Delta_i}\left\{\frac{\mathcal{E}}{4a^2}+C_i\mathcal{H}_{u_0:T}\right\}.$$

Early and low rounds then give, for every $T\ge1$ and $M>0$,

$$\sum_{t=1}^{T}\mathbb{E}[p_{t,i}\mathbf{1}_{\{tp_{t,i}>M\}}]\le n_0+\mathcal{L}+\frac{8}{c^3\Delta_i\sqrt{M}}\left\{\frac{\mathcal{E}}{4a^2}+C_i\mathcal{H}_{u_0:T}\right\}. \tag{80}$$

In particular,

$$\limsup_{T\to\infty}\frac{1}{\log T}\sum_{t=1}^{T}\mathbb{E}[p_{t,i}\mathbf{1}_{\{tp_{t,i}>M\}}]\le\frac{128(3+192c^2)^3}{27c^3\Delta_i^3\sqrt{M}}. \tag{81}$$

There is no additional K factor in this single coordinate asymptotic bound. Its finite transient budgets still depend on K.

For the full family, a nonempty intensity tail with $M\ge1$ requires $T>M$, hence $\mathcal{H}_T\ge\log(M+1)$. Dividing (80) by $\mathcal{H}_T$ gives the explicit modulus

$$\sup_{T\ge1}\frac{1}{\mathcal{H}_T}\sum_{t=1}^{T}\mathbb{E}[p_{t,i}\mathbf{1}_{\{tp_{t,i}>M\}}]\le\frac{n_0+\mathcal{L}}{\log(M+1)}+\frac{8}{c^3\Delta_i\sqrt{M}}\left\{\frac{\mathcal{E}}{4a^2\log(M+1)}+C_i\right\}. \tag{82}$$

Hölder's inequality also supplies

$$\sum_{t=1}^{T}\mathbb{E}p_{t,i}\le n_0+\mathcal{L}+\mathcal{H}_{u_0:T}^{1/3}\left[\frac{8}{c^3\Delta_i}\left\{\frac{\mathcal{E}}{4a^2}+C_i\mathcal{H}_{u_0:T}\right\}\right]^{2/3}.$$

These estimates quantify truncation and recovery costs. They do not bound the error for bounded test occupation convergence, so they do not give a time $T(\epsilon)$ for convergence to the exact coefficient. They also do not justify an exchange of the horizon, gap, and K limits.

F FINITE OCCUPATION ERROR BOUNDS FOR THE RV LEARNER

Harmonic occupation and the estimates in Appendix E give finite error bounds for branch matched generator tests and bounded targets. This result is separate from Theorem 1; Appendix G gives one explicit instance.

Fix an instance in the scope of Theorem 1 and one coordinate $i \geq 2$. Abbreviate

$$\Delta = \Delta_i, \quad \theta = (1/2 - \mu_1)/\Delta, \quad \gamma_\Delta = c^2\Delta/8, \quad x_0 = 2/(c^2\Delta), \quad a_0 = \mu_1/\Delta.$$

Let $\mathcal{A} = \mathcal{A}_{\Delta,\theta}$ be (24), with invariant law $\pi = \pi_{\Delta,\theta}$. Write $\mathcal{A}_A$ and $\mathcal{A}_I$ for its active and inactive branch formulas. We give an explicit $O(1/\log T)$ estimate for a class of generator tests, and a residual bound converting it into an occupation error bound for any specified bounded target. No quantitative mixing assumption is used.

Normalization and recovery budgets. Use $u_0 = n_0 + 1$, a, $\mathcal{L}$ and $\mathcal{E}$ from (77)–(79), and put

$$N = n_0 + \mathcal{L}, \qquad C_i = \frac{16(3 + 192c^2)^3}{27\Delta^2}, \qquad H = \sum_{t=2}^{T} t^{-1} \quad (T \geq 2).$$

Let $\iota_t = I_{t-1}$ on rounds $t \geq u_0$ and $\iota_t = 0$ earlier. Thus $\sum_{t\geq 2} \mathbb{P}(\iota_t = 0) \leq N$. Define probability measures

$$\widehat{\Gamma}_{T,i} = \frac{1}{H}\sum_{t=2}^{T} \frac{\mathcal{L}(X_{t-1,i})}{t}, \qquad \nu_T = \frac{1}{H}\sum_{t=2}^{T} \frac{1}{t}\left\{\mathbb{E}[\iota_t \delta_{X_{t-1,i}}] + \mathbb{E}[1-\iota_t]\delta_1\right\}. \tag{83}$$

For a bounded function h on $(0,\infty)$, define $h(x) = h(1)$ for $x \leq 0$. This preserves its supremum norm and gives

$$|\widehat{\Gamma}_{T,i}h - \nu_T h| \leq 2\|h\|_\infty N/H. \tag{84}$$

The measure $\widehat{\Gamma}_{T,i}$ is the ith marginal of Γ_T in Lemma 2, multiplied by $\log(T)/H$.

Constants for one test. Take $f \in C_c^2((0,\infty))$, extended by zero to $\mathbb{R}$, with support in $[u,v]$, $0 < u < v$, and $F_j = \|f^{(j)}\|_\infty$ for $j = 0,1,2$. Require the single linear matching condition

$$\mathcal{A}_A f(x_0) = \mathcal{A}_I f(x_0). \tag{85}$$

This condition is nonvacuous: every such f with $v < 3x_0/4$ satisfies it. Define

$$l = \min\{u/4, \sqrt{u/(32\gamma_\Delta)}\}, \quad R = 4v, \quad S = \frac{4\sqrt{K}}{c\Delta} + \frac{4}{c^2\Delta},$$
$$m = \max\left\{u_0, 2, 1 + \left\lceil \frac{8K}{\Delta u}\right\rceil, 1 + \lceil (2S/l)^2\rceil\right\}. \tag{86}$$

Set

$$z_- = c\Delta l, \quad z_+ = c\Delta R, \quad \Lambda = 16/z_-^2, \quad C_\lambda = 64(z_+ + 2\sqrt{K})/z_-^3,$$
$$C_J = \frac{(z_+ + 2\sqrt{K})(2z_+ + 2\sqrt{K})}{8\Delta},$$
$$C_{\mathrm{loc}} = 2F_0 C_\lambda + \Lambda F_1\left[2C_J + 2\left(R + \frac{1}{2\Delta}\right)\right] + \frac{F_2}{2}\left[R + \frac{K+1}{2\Delta}\right]^2. \tag{87}$$

For the threshold comparison put $l' = l/2$, $R' = 2R$, $\lambda' = 4/(c^2\Delta^2(l')^2)$, and

$$L_f = F_1 + (R' + |\theta| + a_0)F_2 + \frac{4\lambda'}{l'}F_0 + \lambda' F_1(2 + 4\gamma_\Delta R'), \qquad C_f = C_{\mathrm{loc}} + 2L_f S.$$

Finally, with $\mathcal{H}_m = \sum_{t=1}^{m} 1/t$, define

$$B_f = (R + |\theta| + a_0)F_1 + \frac{8F_0}{c^2\Delta^2 l^2},$$
$$D_f = 2F_0(1 + N + m) + B_f(\mathcal{H}_m + N) + \frac{2C_f}{\sqrt{m-1}}, \tag{88}$$
$$M_T = 2(1 + 3/\Delta + \Delta^2 C_i) + \frac{\Delta^2(\mathcal{E}/(4a^2) + C_i) + 2N}{H}.$$

All these constants are finite and specified by the instance, the chosen support interval and the derivative norms of f.

Theorem 7 (Finite RV occupation bound)**.** *For every f satisfying* (85) *and every $T \geq 2$,*

$$|\nu_T \mathcal{A} f| \leq D_f / H. \tag{89}$$

For any bounded Borel target $h : (0, \infty) \to \mathbb{R}$ and $b \in \mathbb{R}$, put

$$\eta(h, b, f) = \sup_{x>0} \frac{|h(x) - b - \mathcal{A}f(x)|}{1 + x + x^{-3}}. \tag{90}$$

Then

$$\left|\widehat{\Gamma}_{T,i} h - \pi h\right| \leq \frac{2\|h\|_\infty N + D_f}{H} + \eta(h, b, f) M_T. \tag{91}$$

In particular, $h = \mathcal{A}f$ has an explicit $O(1/\log T)$ bound on occupation error. For a general h, the second term in (91) *is an approximation residual, not an asserted vanishing quantity.*

Proof. We first prove a global scalar consistency estimate on high rounds. Write $x = X_{t-1,i}$ and let $\mathsf{P}_t f$ mean the exact full state kernel applied to the scalar function $f(x_i)$. On a high round, $p_{t,1} \geq 1/K$, the optimal baseline is active, and KKT gives

$$tp_{t,i} = \frac{4}{[c\Delta x(1 - 1/t) + w_t/\sqrt{t}]^2}, \qquad 2 \leq w_t \leq 2\sqrt{K}. \tag{92}$$

No bounds on the other normalized coordinates are needed. Indeed, writing I_t for the played arm, the exact outcome table gives

$$t\mathbb{E}[(X_{t,i} - x)\mathbf{1}_{\{I_t \neq i\}} \mid \mathcal{F}_{t-1}] = \begin{cases} \theta - (1 - p_{t,i})x, & B_t(i) = 1/2, \\ \theta - (1 - p_{t,i})(x + 1/(2\Delta)), & B_t(i) = 0. \end{cases} \tag{93}$$

The probabilities of the other suboptimal arms cancel in this sum. Its difference from the corresponding limiting drift is only $p_{t,i}x$ or $p_{t,i}(x + 1/(2\Delta))$.

For $t \geq m$, both $t(\mathsf{P}_t - I)f$ and $\mathcal{A}f$ vanish outside $[l, R]$ on high rounds. To see the lower exclusion, an own arm positive landing is at most $x + 4\gamma_\Delta x^2 + 2K/(\Delta t)$, which is less than u when $x < l$; the optimal arm's upward increment is at most $K/(2\Delta t)$. The other arms cannot increase x. For the upper exclusion, the high round no skipping bound gives $X_{t,i} \geq (1 - 1/t)x/2 \geq x/4$, including the last high transition. The limiting downward landing is at least $3x/4$ on its active branch. Thus no source outside the displayed interval can affect f.

On this interval let $z = c\Delta x$ and $v_t = z(1 - 1/t) + w_t/\sqrt{t}$. Then

$$v_t \geq z/2, \qquad |v_t - z| \leq (z_+ + 2\sqrt{K})/\sqrt{t}, \qquad tp_{t,i} \leq \Lambda.$$

Differentiating $4/v^2$ bounds the error in the rate by $C_\lambda/\sqrt{t}$. The active own arm jump magnitude is $v_t^2/(8\Delta)$, whose error relative to $\gamma_\Delta x^2$ is at most $C_J/\sqrt{t}$; the inactive magnitude and its error are twice as large. Compare these own arm landings with the Lipschitz bound F_1, retaining their exact contraction and baseline shifts. Expand by Taylor's theorem only the other channels. Their increments have magnitude at most $[R + (K + 1)/(2\Delta)]/t$, and their drift correction is given by (93). Consequently, whenever the finite and limiting branches agree,

$$|t(\mathsf{P}_t - I)f - \mathcal{A}f(x)| \leq C_{\mathrm{loc}}/\sqrt{t}.$$

The same calculation compares the exact kernel with its selected branch formula even when the limiting branch differs.

The exact switch in (9) satisfies $|x_{t,i}^{\mathrm{sw}} - x_0| \leq S/\sqrt{t}$. A mismatch in $[l, R]$ therefore has $|x - x_0| \leq S/\sqrt{t}$ and, by (86), $x_0 \in [l', R']$. Differentiating the two branch formulas, with the zero extension of f, bounds each derivative on this enlarged interval by L_f. The matching condition then bounds the difference of branch formulas by $2L_f S/\sqrt{t}$. We have proved

$$|t(\mathsf{P}_t - I)f - \mathcal{A}f(X_{t-1,i})| \leq C_f/\sqrt{t} \quad \text{on } \{\iota_t = 1\}, \quad t \geq m. \tag{94}$$

Also $\|\mathcal{A}f\|_\infty \leq B_f$ by the source interval above.

Telescope the full chain first and subtract its increments on low rounds. Each such increment is bounded by $2F_0$, so

$$\left|\sum_{t=2}^{T}\mathbb{E}[\iota_t(\mathsf{P}_t - I)f]\right| \le 2F_0(1+N).$$

Before m the comparison with the generator costs at most $2F_0m + B_f\mathcal{H}_m$. From m onward, (94) costs at most $C_f\sum_{t=m}^{\infty} t^{-3/2} \le 2C_f/\sqrt{m-1}$. Replacing low states by 1 in $\nu_T\mathcal{A}f$ costs at most B_fN. Division by H proves (89). Compact support, the compact source interval and bounded generator make the invariant identity $\pi\mathcal{A}f = 0$ legitimate.

It remains to control the weight in (90). The exact affine update and $\mathbb{E}[|Z_{t,i}| \mid \mathcal{F}_{t-1}] \le 3$ give $\sup_{n\ge1}\mathbb{E}|X_{n,i}| \le 3/\Delta$, including initialization. The potential telescoping in Appendix E gives

$$\sum_{t=u_0}^{T}\mathbb{E}\left[\iota_t\frac{t^2}{D_{t-1,i}^3}\right] \le \frac{1}{\Delta}\left\{\frac{\mathcal{E}}{4a^2} + C_i\mathcal{H}_{u_0:T}\right\}.$$

Since $x^{-3}/t = \Delta^3(t-1)^3/(tD_{t-1,i}^3) \le \Delta^3t^2/D_{t-1,i}^3$ and $\mathcal{H}_T = H+1$, it follows that

$$\nu_T x \le 3/\Delta + N/H, \qquad \nu_T x^{-3} \le \Delta^2 C_i + \frac{\Delta^2(\mathcal{E}/(4a^2) + C_i) + N}{H}. \tag{95}$$

Lemma 2, the finite sampling bound in Appendix E, and the vanishing replacement error identify every scalar weak limit as π. Applying Fatou through continuous compact truncations in (95) gives $\pi x \le 3/\Delta$ and $\pi x^{-3} \le \Delta^2 C_i$. Thus, for $W(x) = 1 + x + x^{-3}$, $(\nu_T + \pi)W \le M_T$. Writing $r = h - b - \mathcal{A}f$ now yields

$$|\nu_T h - \pi h| \le |\nu_T\mathcal{A}f| + \eta(h,b,f)(\nu_T + \pi)W.$$

The constant b cancels because both measures have mass one. Apply (89) and (84) to obtain (91). □

Scope of the residual bound. The residual (90) is a deterministic one dimensional supremum; it does not involve the unknown number πh. A certified upper bound for it suffices, and one may optimize (91) over proposed pairs (b, f). For $h_M(x) = \min\{4/(c^2\Delta^2x^2), M\}$, however, a vanishing rate requires a family with $\eta(h_M, b, f) \to 0$ and controlled D_f. We do not prove such a family, a quantitative Poisson regularity theorem, or a mixing bound. Positive Harris recurrence alone does not supply those constants. For unmatched f, the consistency argument retains a term proportional to

$$\frac{1}{H}\sum_{t=m}^{T}\frac{1}{t}\mathbb{P}\big(\iota_t = 1, |X_{t-1,i} - x_0| \le S/\sqrt{t}\big).$$

The qualitative argument on a strip around the switching boundary establishes its vanishing, but no numerical modulus is being assigned to it here.

Corollary 3 (A fixed Borel obstruction)**.** *There exists a bounded Borel $h : (0,\infty) \to [0,1]$, with the extension convention above, such that $\widehat{\Gamma}_{T,i}h = 1$ and $\pi h = 0$ for every $T \ge 2$. Thus occupation convergence cannot hold for all bounded Borel tests, or in total variation.*

Proof. At each round the Bernoulli learner has finitely many possible observed histories; its state is a deterministic function of that history. Hence $S = \bigcup_{n\ge1}\operatorname{supp}\mathcal{L}(X_{n,i})$ is countable.

The invariant PDMP law is atomless. For any fixed $y > 0$ other than a possible equilibrium $\theta > 0$, each path spends zero Lebesgue time at y: between its locally finitely many jumps, the flow crosses y with nonzero velocity. This includes x_0 by strict transversality. Stationarity and Fubini give $\pi\{y\} = 0$. If $0 < \theta < x_0$, the sources of jumps landing at θ form a finite set of non-θ points, obtained from the quadratic landing maps. They have zero stationary mass and finite jump rates, so compensation gives zero expected arrivals to θ. A flow started elsewhere cannot reach θ in finite time. A path started there leaves at rate $\lambda(\theta) > 0$, and all channels with positive rates move it. Thus stationarity implies $\pi\{\theta\} = e^{-t\lambda(\theta)}\pi\{\theta\}$, so this last possible atom is also absent.

Set $S_+ = (S\cap(0,\infty))\cup\{1\}$ and $h = \mathbf{1}_{S_+}$. Every positive learner state belongs to S_+; at nonpositive states the extension equals $h(1) = 1$. Atomlessness gives $\pi h = 0$, proving the claim. This obstruction does not apply to a specified continuous target such as h_M or to a class with a controlled continuity modulus. □

G AN EXPLICIT RV POISSON APPROXIMATION

For one specified instance, an explicit truncated intensity verifies the residual hypothesis of Theorem 7. This yields the finite accuracy guarantee stated in the main text, including the cutoff and threshold corrections. Its large constants are not intended as a practical horizon, and its positive residual floor does not give an arbitrary accuracy convergence rate.

Proposition 1 (An explicit finite accuracy bound)**.** *Consider the RV learner with*

$$K = 2, \qquad c = 1/100, \qquad \mu_1 = 1/2, \qquad \mu_2 = 1, \qquad \Delta = 1/2.$$

Set the recovery barrier parameter in Appendix E to $a = 100$. Let

$$h(x) = \min\{x^{-2}, 10^8\}, \quad s = 1/320000, \quad \kappa = 1 + 2s, \quad b = (1+s)\kappa.$$

There is an explicitly specified branch matched $f \in C_c^2((0,\infty))$, supported in $[10^{-8}, 2\cdot 10^{20}]$, satisfying

$$\eta(h, b, f) \le 5 \cdot 10^{-5}, \qquad F_0 \le 3 \cdot 10^8, \quad F_1 \le 7 \cdot 10^{16}, \quad F_2 \le 3 \cdot 10^{25}. \tag{96}$$

Evaluate D_f in (88) *using these derivative upper bounds, the displayed support, and this instance. With $N, C_2, \mathcal{E}$ as in Appendix F, put*

$$M_1 = \Delta^2 \left\{ \frac{\mathcal{E}}{4a^2} + C_2 \right\} + 2N, \qquad Q = 2 \cdot 10^8 N + D_f + 5 \cdot 10^{-5} M_1. \tag{97}$$

Then, with $H = \sum_{t=2}^{T} t^{-1}$,

$$|\widehat{\Gamma}_{T,2} h - \pi h| \le 0.00235 + Q/H. \tag{98}$$

In particular $H \ge 200Q$ implies an error at most $0.00735 < 0.01$; a sufficient ordinary horizon is $T \ge \lceil \exp(1 + 200Q) \rceil$.

Proof. For this instance,

$$\gamma_0 = 1/160000, \quad x_0 = 40000 = 1/(4\gamma_0), \quad \Lambda_0 = 160000 = 1/\gamma_0, \quad s = \gamma_0/2,$$

and the exact generator (24) becomes

$$\begin{aligned} \mathcal{A}_A g(x) &= -xg'(x) + \frac{\Lambda_0}{x^2}[g(x + \gamma_0 x^2) - g(x)], \qquad x \le x_0, \\ \mathcal{A}_I g(x) &= -(1+x)g'(x) + \frac{\Lambda_0}{x^2}[g(x + 2\gamma_0 x^2) - g(x)], \qquad x > x_0. \end{aligned} \tag{99}$$

All random jumps are upward, a property used in the cutoff bounds below. Write $W(x) = 1 + x + x^{-3}$.

Analytic starting function. Let $f_0(x) = -x^{-1} + \kappa \log x$. Its first three derivatives are

$$f_0' = x^{-2} + \kappa x^{-1}, \quad f_0'' = -2x^{-3} - \kappa x^{-2}, \quad f_0''' = 6x^{-4} + 2\kappa x^{-3}.$$

The last derivative is positive and decreasing. Direct substitution gives

$$(1-x) f_0' + s x^2 f_0'' = x^{-2} - b. \tag{100}$$

Compensating the linear jump term in (99) gives drift $1 - x$ on both branches. The quadratic Taylor coefficients are sx^2 and $4sx^2$, respectively. Taylor's third order bound, using the third derivative at the source to dominate the whole upward jump interval, therefore gives

$$\begin{aligned} 0 \le \mathcal{A}_A f_0 - (x^{-2} - b) &\le \gamma_0^2 (1 + \kappa x/3), \\ |\mathcal{A}_I f_0 - (x^{-2} - b)| &\le 3s(2/x + \kappa) + 8\gamma_0^2 (1 + \kappa x/3). \end{aligned}$$

These bounds hold even for large jump intervals. Since $W \ge 1 + x$, $\kappa < 3$, and the inactive branch has $x \ge 40000$,

$$\sup_{x>0} \frac{|\mathcal{A} f_0 - (x^{-2} - b)|}{W(x)} < 6 \cdot 10^{-10}. \tag{101}$$

In particular, the inactive upper expression after division is at most $3s(2/40000 + \kappa)/40000 + 8\gamma_0^2 < 6 \cdot 10^{-10}$.

Compact cutoffs. Define the globally C^2 step

$$S(z) = \begin{cases} 0, & z \le 0, \\ 10z^3 - 15z^4 + 6z^5, & 0 < z < 1, \\ 1, & z \ge 1. \end{cases}$$

It satisfies $0 \le S \le 1$, $\|S'\|_\infty \le 15/8$, and $\|S''\|_\infty \le 6$; indeed the last norm is $10\sqrt{3}/3$. Set

$$\varepsilon = 10^{-8}, \quad V = 10^{20}, \quad \chi_\ell(x) = S((x-\varepsilon)/\varepsilon), \quad \chi_u(x) = 1 - S((x-V)/V), \quad f_c = \chi_\ell \chi_u f_0.$$

Thus f_c is supported in $[\varepsilon, 2V]$ and equals f_0 on $[2\varepsilon, V]$.

For the lower cutoff write $g_\ell = (\chi_\ell - 1)f_0$. On $[\varepsilon, 2\varepsilon]$, $|f_0| \le 2/x$, $|f_0'| \le 2/x^2$, and $|f_0''| \le 3/x^3$. Here the logarithm term is controlled by $|\log \varepsilon| < 24$ and $\kappa x |\log x| < 1$. The product rule, also using $g_\ell = -f_0$ below ε, gives on $(0, 2\varepsilon]$

$$|g_\ell'| \le 10x^{-2}, \qquad |g_\ell''| \le 66x^{-3}.$$

For example the transition coefficients are bounded by $2 + (15/4)(x/\varepsilon) \le 10$ and $3 + (15/2)(x/\varepsilon) + 12(x/\varepsilon)^2 \le 66$. Above 2ε both derivatives vanish. Upward jumps imply $\mathcal{A}g_\ell = 0$ there. For smaller sources, the compensated active formula and its second order integral remainder give

$$|\mathcal{A}g_\ell(x)| \le 10(1+x)x^{-2} + 66s/x.$$

This also bounds jump intervals crossing the cutoff, since each point of the interval is at least the source x. Hence

$$\sup_{x>0} |\mathcal{A}g_\ell(x)|/W(x) \le 20\varepsilon(1+2\varepsilon) + 264s\varepsilon^2 < 2.1 \cdot 10^{-7}. \tag{102}$$

For the upper cutoff put $L = 2 \cdot 10^{12}$. The largest active landing is $5x_0/4$, and $L + 2\gamma_0 L^2 = 2 \cdot 10^{12} + 5 \cdot 10^{19} < V$. Thus its generator effect is exactly zero for sources below L. For $x \ge L$, the lower cutoff is one at the source and all landings. On $[L, 2V]$ we have

$$|f_c| \le 50, \qquad |f_c'| \le x^{-2} + \kappa/x + \frac{15 \cdot 50}{8V};$$

beyond $2V$ the function and derivative vanish. The bound 50 follows from $\log 10 < 12/5$, $\log 2 < 1$, and $\kappa < 1.00001$. The drift contribution is therefore less than 200 in absolute value where nonzero, and the jump contribution is at most $100\Lambda_0/x^2$. As $b < 2$,

$$\sup_{x \ge L} \frac{|x^{-2} - b - \mathcal{A}f_c(x)|}{W(x)} \le \frac{202}{L} + \frac{100\Lambda_0 + 1}{L^3} < 2 \cdot 10^{-10}.$$

Combining this with the exact source exclusion below L, (101), and (102) yields

$$\sup_{x>0} \frac{|x^{-2} - b - \mathcal{A}f_c(x)|}{W(x)} \le 6 \cdot 10^{-10} + 2.1 \cdot 10^{-7}. \tag{103}$$

Exact threshold correction. Take $d = 4000$ and the globally C^2 bump

$$\psi(x) = \begin{cases} (x - x_0)\big(1 - ((x-x_0)/d)^2\big)^3, & |x - x_0| < d, \\ 0, & |x - x_0| \ge d. \end{cases}$$

It satisfies $\psi(x_0) = 0$, $\psi'(x_0) = 1$, $\|\psi\|_\infty \le d$, $\|\psi'\|_\infty \le 1$, and $\|\psi''\|_\infty \le 120/d$. Indeed, for $z = (x - x_0)/d$ its derivative is $(1-z^2)^2(1 - 7z^2) \in [-32/49, 1]$, and $d\psi'' = -18z + 60z^3 - 42z^5$. Both threshold landings, 50000 and 60000, lie outside its support $[36000, 44000]$. Therefore $(\mathcal{A}_A - \mathcal{A}_I)\psi(x_0) = 1$. The cutoffs equal one at the switch and both landings. With $\rho = 1/40000$, define the exact coefficient

$$q = \frac{7}{15}\rho^2 + \kappa\rho\{1 + 4\log(5/6)\}, \qquad f = f_c - q\psi. \tag{104}$$

Substituting f_0 shows $q = (\mathcal{A}_A - \mathcal{A}_I)f_c(x_0)$, proving (85) exactly. The integral bounds $-1/5 \le \log(5/6) \le -1/6$ imply $0 < q < 9 \cdot 10^{-6}$.

The bump generator vanishes below 18000: its active upward landing is at most $18000 + \gamma_0 18000^2 = 20025 < 36000$, and the inactive branch starts at 40000. Elsewhere, its drift divided by W is at most $1 + 1/36000$ when nonzero; its jump term divided by W is at most $2d\Lambda_0/18000^3 < 0.00022$. Consequently

$$\sup_{x>0} |\mathcal{A}\psi(x)|/W(x) < 1.001, \qquad \sup_{x>0} |q\mathcal{A}\psi(x)|/W(x) < 9.009 \cdot 10^{-6}. \tag{105}$$

Residual and derivative bounds. For any $m > 0$,

$$\frac{(x^{-2} - m)_+}{W(x)} \le (x - mx^3)_+ \le \frac{2}{3\sqrt{3m}}.$$

At $m = 10^8$ this is less than $4 \cdot 10^{-5}$. Equations (103) and (105) now give

$$\eta(h, b, f) < 4 \cdot 10^{-5} + 2.1 \cdot 10^{-7} + 6 \cdot 10^{-10} + 9.009 \cdot 10^{-6} < 5 \cdot 10^{-5}.$$

For the derivative bounds, the lower transition satisfies

$$|f_c| \le 2/\varepsilon, \quad |f_c'| \le (2 + 15/4)\varepsilon^{-2}, \quad |f_c''| \le (3 + 15/2 + 12)\varepsilon^{-3}.$$

Raw derivatives away from this transition and the upper cutoff contributions are smaller; the two cutoff transitions are disjoint. Adding $q\|\psi^{(j)}\|_\infty$ proves the rounded constants in (96). Thus every hypothesis on f in Theorem 7 is satisfied.

Transfer to the learner. For this instance the constant part of the weight bound is

$$M_0 = 2(1 + 3/\Delta + \Delta^2 C_2) = 2\left\{7 + \frac{16}{27}(3 + 192/10000)^3\right\} < 47.$$

Equation (88) is $M_T = M_0 + M_1/H$. All its terms and those defining D_f are monotone in the derivative upper bounds used above. Theorem 7, with $\|h\|_\infty = 10^8$, therefore proves (98). The stated ordinary horizon follows from $H \ge \log(T + 1) - 1$. □

The original intensity in this instance is $h_M(x) = \min\{160000x^{-2}, 1.6 \cdot 10^{13}\} = 160000h(x)$. Multiplying (98) by 160000 gives the corresponding bound; the stated harmonic horizon guarantees an absolute error less than 1600, with the extension convention of Appendix F. No invariant mean is substituted to interpret this as relative error. This construction exhibits a fully specified finite accuracy horizon. The cutoff, derivative, and recovery constants make that horizon extremely conservative. The fixed positive floor in (98) remains: the construction does not establish an arbitrarily accurate family of RV Poisson approximations or a general fixed instance convergence rate.

H Proof of the adaptive variance theorem

We prove Theorem 2 in three steps. Probe consistency and recovery allow us to apply Theorem 4; the controlled PDMP identifies the fixed gap law; and the final subsection takes the ordered gap and clip limits. Throughout, time n counts ordinary learner rounds; the deterministic probe rounds do not update D_n.

H.1 Probe consistency and discrete recovery

Let $K_P(T) = |\{k \ge 2 : \lceil k^k \rceil \le T\}|$. Inverting $k \log k \le \log T$ gives

$$K_P(T) = O\left(\frac{\log T}{\log\log T}\right) = o(\log T). \tag{106}$$

Both arms receive infinitely many probe observations, and the probe regret is at most $\Delta K_P(T)$. Moreover, for $N \ge 2$, Hoeffding's inequality gives

$$\Pr(|b_{n,i} - \mu_i| > \mathrm{rad}(N)) \le 2e^{-2N\,\mathrm{rad}(N)^2} = 2N^{-4}.$$

Thus Borel–Cantelli and the strong law imply, almost surely,

$$b_{n,i} \to \mu_i, \qquad \widehat{\Delta}_{L,n} \to \Delta, \qquad \widehat{v}_n \to v_2, \qquad r_n \to \Delta/v_2. \tag{107}$$

The empirical worse arm label is eventually correct because the gap is fixed and positive. Assumption (21) makes the cap asymptotically inactive.

For later use, the exact gap recursion is

$$D_n = D_{n-1} + B_{n,2} - B_{n,1} + \frac{\mathbf{1}\{I_n = 2\}}{p_{n,2}}(\ell_{n,2} - B_{n,2}) - \frac{\mathbf{1}\{I_n = 1\}}{p_{n,1}}(\ell_{n,1} - B_{n,1}). \tag{108}$$

Every baseline is predictable, so $\mathbb{E}[D_n - D_{n-1} \mid \mathcal{F}_{n-1}] = \Delta$. Also, uniformly for $B \in [0,1]$,

$$\mathbb{E}\left|B + \frac{\mathbf{1}\{I=i\}}{p_i}(\ell_i - B)\right| \le B + \mathbb{E}|\ell_i - B| \le 2.$$

Hence the gap increment has conditional first absolute moment at most 4.

We next verify the crossing geometry required by Theorem 4. Suppose $d = D_{n-1} \ge 0$, and write $p = p_{n,2} \le 1/2$. If the rare baseline is active, then $p \ge \eta_n^2$. With $y = p^{-1/2}$ and $z = (1-p)^{-1/2} \le \sqrt{2}$, exact KKT gives

$$d = \frac{2(y-z)}{\eta_n} \ge 2y^2 - 2\sqrt{2}y,$$

and therefore

$$\frac{1}{p} \le \frac{d}{2} + \sqrt{d+1} + 1. \tag{109}$$

The most negative active rare outcome is a zero loss, for which

$$D_n \ge d - 1 - p^{-1} \ge d/2 - \sqrt{d+1} - 2.$$

For $d \ge 36$, this is at least $d/4$. All negative outcomes on the frequently sampled arm and all inactive rare negative outcomes reduce d by at most 2. Thus

$$d \ge 36 \quad \Longrightarrow \quad D_n \ge d/4. \tag{110}$$

Reflecting the argument, a transition from $D_{n-1} = -z \le 0$ to $D_n > 0$ satisfies

$$D_n \le -z + 1 + p^{-1} \le -z/2 + \sqrt{z+1} + 2 \le 3.$$

The opposite crossing obeys the same bound.

The remaining strip and moment hypotheses are direct KKT consequences. On $0 < d \le h\sqrt{n}$,

$$p^{-1/2} = \eta_n d/2 + (1-p)^{-1/2} \le \sqrt{U}h/2 + \sqrt{2}, \tag{111}$$

so both action probabilities are bounded below, both baselines are eventually active, and every gap increment is bounded. For arbitrary centers in $[0,1]$,

$$\mathbb{E}\widehat{\ell}_{n,i}^2 \le 2 + 2/p_{n,i},$$

while

$$p^{-1} \le Ud^2/(2n) + 4, \qquad p \le \frac{4n}{Ld^2}. \tag{112}$$

Equations (108)–(112) verify A1–A5 of Theorem 4, with crossing bound 3, high state threshold 36, and fractional retention $1/4$. The theorem consequently gives finite expected low state duration, logarithmic sampling, and harmonic uniform integrability of the intensity of the rarely sampled action for each fixed L, U.

H.2 The controlled finite gap PDMP

Put

$$q(x) = \mathrm{clip}\left(\frac{4}{v_2 x}, L, U\right), \quad \lambda(x) = \frac{4}{q(x)\Delta^2 x^2}, \quad c(x) = \frac{q(x)\Delta x^2}{4}, \quad x_0 = \frac{2}{L\Delta}. \tag{113}$$

Assumption (21) implies that the lower and middle clip regions are active and that the only switch is at $x_0 > 1$. With equality assigned to the active branch, define

$$\begin{aligned}\mathcal{A}_\Delta f(x) &= (1-x)f'(x) + \lambda(x)\mu_2[f(x + c(x)(1-\mu_2)) - f(x)] \\ &\quad + \lambda(x)(1-\mu_2)[f(x - c(x)\mu_2) - f(x)], \qquad x \le x_0, \\ \mathcal{A}_\Delta f(x) &= (-\mu_1/\Delta - x)f'(x) + \mu_2\lambda(x)[f(x + c(x)) - f(x)], \qquad x > x_0.\end{aligned} \tag{114}$$

The active downward landing remains positive because $q(x)\Delta x \le 2$ implies

$$x - c(x)\mu_2 \ge x(1 - \mu_2/2) > 0. \tag{115}$$

The centered active jumps have mean zero, while the inactive loss one jump compensates its drift. Consequently

$$\mathcal{A}_\Delta x = 1 - x \tag{116}$$

on both branches.

Lemma 6 (Well posedness and invariant uniqueness)**.** *Under* (21)*, the generator* (114) *has a conservative Borel right strong Markov realization on* $(0, \infty)$*. It is positive Harris recurrent and has a unique invariant probability* $\pi^{\mathrm{VA}}_{\Delta,L,U}$*.*

Proof. The active deterministic flow is $\phi_t(x) = 1 + (x-1)e^{-t}$ and remains below x_0. The inactive flow has velocity $-\mu_1/\Delta - x$ and crosses x_0 transversally from above. Thus the piecewise deterministic flow is unique. The condition $x_0 > 1$ matters: if the switch were below 1, the two velocities could point toward the switching surface and require an additional boundary convention.

For $r \geq 2$, let $V(x) = x + x^{-r}$. Near zero, $q(x) = U$ and the active centered jump cancels at first order. Since $\lambda(x)c(x)^2 = q(x)x^2/4$, Taylor's theorem gives

$$\mathcal{A}_\Delta x^{-r} \leq -rx^{-r-1} + C_{r,U}x^{-r}. \tag{117}$$

At infinity the process is inactive, its upward jump decreases x^{-r}, and (116) supplies the negative linear term. Hence, for a compact annulus K,

$$\mathcal{A}_\Delta V \leq aV + b\mathbf{1}_K. \tag{118}$$

Moreover, $\lambda(x) \leq C(1 + x^{-2}) \leq C'(1 + V(x))$. The stopped clock construction, Dynkin's formula, and Gronwall exclude accumulation of random jumps. The standard Davis construction therefore gives a conservative Borel right strong Markov PDMP; see the source map in Appendix J.

It remains to prove that the Foster compact is petite. Let $T_0(y) = y - \mu_2 c(y)$ be the active loss zero landing. Start from x near 1, condition on one loss zero jump at time $s < T$, and exclude other jumps before T. The endpoint is

$$F_x(s) = 1 + [T_0(\phi_s(x)) - 1]e^{-(T-s)}.$$

On either smooth side of a clip interface,

$$F'_x(s) = \mu_2 e^{-(T-s)}[(y-1)c'(y) - c(y)], \qquad y = \phi_s(x). \tag{119}$$

At $y = 1$, both one sided derivatives equal $-\mu_2 e^{-(T-s)}c(1) \neq 0$. A small starting interval C about 1 and a compact jump time window therefore have a Jacobian bounded above and away from zero. The channel intensity and the pre- and post jump survival probabilities are uniformly positive on this compact path tube. A change of variables gives an interval J and $\epsilon > 0$ such that

$$P_T(x, A) \geq \epsilon \,\mathrm{Leb}(A \cap J), \qquad x \in C. \tag{120}$$

If 1 is a clip interface, use one sided starting intervals. Equivalently, condition on a first loss zero jump below 1 and use the second jump time in a smooth subinterval; (119) supplies the same density tube with two jumps.

The set C is accessible from every state. Above 1, a no jump path moves toward C, crossing x_0 first if necessary. Below 1, finitely many rapid loss one jumps move the state above 1, after which a no jump path enters C. Each finite tube has positive probability. For every x in the Foster compact K, this construction provides a neighborhood of x, a bounded time window, and positive probability of reaching C. A finite subcover of K, composed with (120), gives a common resolvent minorization; hence K is petite. The same upward/downward jump controls show ψ-irreducibility for $\psi = \mathrm{Leb}\,|_J$. The Foster–petite criterion of Meyn & Tweedie (1993) now gives positive Harris recurrence and invariant uniqueness.

Notice that the point resolvent is not equivalent to Lebesgue measure. From $x = 1$, the process remains at 1 before its first jump, so the resolvent has an atom there. The small set argument needs only irreducibility and does not make this stronger, false claim. □

H.3 Threshold occupation and BKR identification

On the normalized state $X_{n-1} = D_{n-1}/[\Delta(n-1)]$, compact KKT asymptotics and (107) give

$$np_{n,2} \longrightarrow \lambda(x) = \frac{4}{q(x)\Delta^2 x^2} \tag{121}$$

uniformly away from the switch and clip interfaces. The clip interfaces do not create generator discontinuities because q, λ, and the landing maps are continuous there.

Only x_0 requires a source estimate. Since $r_n \le \kappa < 1/2$,

$$\frac{4r_n}{\Delta x_0} = 2Lr_n < L.$$

Thus $a_n = L$ exactly on a deterministic neighborhood of x_0, for all large n and every probe history. Take the decreasing $C^{1,1}$ ramp ϕ_ρ used in the proof of Lemma 2: its derivative is -1 on $[x_0-\rho, x_0+\rho]$, it vanishes outside the doubled interval, its range is $O(\rho)$, and $\mathrm{Lip}(\phi_\rho') = O(1/\rho)$.

For rare probability p, the active frequent arm cell satisfies

$$\begin{aligned} &n\mathbb{E}[(X_n - X_{n-1})\mathbf{1}\{I_n = 1\} \mid \mathcal{F}_{n-1}] \\ &\quad = \frac{(1-p)(b_{n,2} - b_{n,1}) - (\mu_1 - b_{n,1})}{\Delta} - (1-p)X_{n-1} + O(n^{-1}), \end{aligned} \tag{122}$$

which converges to $(1-p)(1-x)$. On the inactive side the corresponding quantity is

$$\frac{pb_{n,1} - \mu_1}{\Delta} - (1-p)X_{n-1} + O(n^{-1}). \tag{123}$$

Both are uniformly negative on a sufficiently small ramp because $x_0 > 1$. The active loss zero rare jump moves left and helps the decreasing ramp. A loss one rare jump has probability $O(1/n)$ and ramp cost $O(\rho)$; the frequent arm Taylor remainder is $O(1/(\rho n^2))$.

Probe errors must be charged by harmonic motion, not by the raw length of a probe block. Even on a miscentered history, the expected absolute ramp increment on the fixed-L threshold tube is bounded by

$$\mathbb{E}[|\phi_\rho(X_n) - \phi_\rho(X_{n-1})| \mid \mathcal{F}_{n-1}] \le C_{\rho,\Delta,L}/n. \tag{124}$$

Indeed, the frequent normalized step is $O(1/n)$, while a rare step changes the bounded ramp by $O(\rho)$ and occurs with probability $O(1/n)$. At probe count k, a fixed centering error has probability $O(e^{-ck})$, and its block has harmonic length $O(\log k)$. Therefore the total expected bad block charge is bounded by

$$\sum_{k\ge 1} O(e^{-ck}\log k) < \infty.$$

Telescoping the ramp yields

$$\limsup_{T\to\infty} \frac{1}{\log T} \sum_{n\le T} \frac{1}{n} \Pr(|X_{n-1} - x_0| \le \rho) \le C\rho. \tag{125}$$

The recovery bounds proved above make the positive harmonic occupations tight and exclude the zero boundary. Equations (121) and (125), together with the discrete martingale telescope, show that every subsequential limit annihilates $\mathcal{A}_\Delta$ on the compact test core. The stopped characteristic argument and the BKR invariance criterion of Bhatt et al. (2006) then make every limit invariant for the Borel right PDMP. Lemma 6 identifies it with $\pi^{\mathrm{VA}}_{\Delta,L,U}$.

The harmonic uniform integrability supplied by Theorem 4 upgrades the bounded test convergence to the unbounded intensity in (121). Combining this fact with (106) proves (22).

H.4 Small gap invariant law and clip removal

Fix an interior m and let $\Delta \downarrow 0$. On every positive compact, the branch is active for all sufficiently small gaps. With $v = m(1-m)$ and

$$q_v(x) = \mathrm{clip}\left(\frac{4}{vx}, L, U\right),$$

a second order expansion of the centered jumps gives

$$\mathcal{A}_\Delta f(x) \longrightarrow \mathcal{L}_{L,U,v} f(x) = (1-x)f'(x) + \frac{vq_v(x)x^2}{8} f''(x). \tag{126}$$

The first order jump term vanishes because the jump is centered, and

$$\frac{\lambda(x)}{2}\mathbb{E}[c(x)^2(Y_2 - \mu_2)^2] = \frac{v_2 q(x) x^2}{8}.$$

The inverse power drift (117) is uniform for small Δ, and (116) gives $\int x\, d\pi^{\mathrm{VA}}_{\Delta,L,U} = 1$. These bounds give tightness and inverse square uniform integrability. Every invariant subsequential limit annihilates (126). The one dimensional diffusion has a unique zero flux invariant law $\pi_{L,U,v}$, so

$$\pi^{\mathrm{VA}}_{\Delta,L,U} \Rightarrow \pi_{L,U,v}, \qquad \lim_{\Delta\downarrow 0} J_{\Delta,L,U} = J_{L,U,v}, \tag{127}$$

where

$$J_{L,U,v} = \int \frac{4}{q_v(x)x^2}\, \pi_{L,U,v}(dx).$$

For completeness, put

$$a = \frac{4}{Uv}, \quad b = \frac{4}{Lv}, \quad \xi_U = \frac{8}{Uv} = 2a, \quad \xi_L = \frac{8}{Lv} = 2b.$$

Up to normalization, the stationary density is

$$h_{L,U,v}(x) = \begin{cases} a^{3+\xi_U} e^{2-\xi_U} x^{-2-\xi_U} e^{-\xi_U/x}, & 0 < x < a, \\ xe^{-2x}, & a \le x \le b, \\ b^{3+\xi_L} e^{2-\xi_L} x^{-2-\xi_L} e^{-\xi_L/x}, & x > b. \end{cases} \tag{128}$$

In the middle region $q_v(x) = 4/(vx)$ and the normalized limiting density is $4xe^{-2x}$, the $\mathrm{Gamma}(2,2)$ law.

Stationarity applied to truncated approximations of $\log x$ gives

$$\mathbb{E}[X^{-1}] = 1 + \frac{v}{8}\mathbb{E}q_v(X).$$

Cauchy–Schwarz consequently yields

$$J_{L,U,v} \ge \frac{4(\mathbb{E}X^{-1})^2}{\mathbb{E}q_v(X)} \ge 2v. \tag{129}$$

Matching the three density pieces and substituting $y = \xi_U/x$ and $y = \xi_L/x$ in the tails gives

$$J_{L,U,v} = 2v + O(U^{-1}) + O\left(L^{-1}e^{-8/(vL)}\right). \tag{130}$$

Thus $U \uparrow \infty$ followed by $L \downarrow 0$ sends the cost to $2v$. Combining (22), (127), and (130) proves (23).

I Fixed gap occupation with moving clips

We now define the moving clip learner. Lemma 7 identifies its limiting law for fixed clips, Proposition 2 proves the occupation identity after the horizon limit, and a diagonal argument for the singular cost proves Theorem 3. This construction uses the central reduced variance estimator and is separate from Theorem 2. The order $T \to \infty$ followed by $\Delta \downarrow 0$ is retained throughout.

Consider two Bernoulli loss arms with $1/2 \le \mu_1 < \mu_2 \le 1$, gap $\Delta = \mu_2 - \mu_1 < 1/8$, and deterministic probe times $\tau_k = \lceil k^k \rceil$, $k \ge 2$. At τ_k probe arm 1 when k is even and arm 2 when k is odd. Probe observations do not update the FTRL state. Immediately after probe k, let

$$N_i(k) = \sum_{j=2}^{k} \mathbf{1}\{i_j = i\}, \qquad r_{i,k} = \sqrt{\frac{M \log k}{2N_i(k)}}$$

and let $\bar{\mu}^P_{i,k}$ be the corresponding post probe empirical mean. When $N_i(k) = 0$, set $r_{i,k} = +\infty$ and $\bar{\mu}^P_{i,k} = 1/2$; equivalently use $g_k = 1$ and $\widehat{\Delta}_{L,k} = 0$ until both counts are positive. The finitely many ordinary rounds before probe 3 use $L = U = 1$ and $\widehat{\Delta}_L = 0$. Put $\rho_k = r_{1,k} + r_{2,k}$ and define

$$g_k = \min\{1, |\bar{\mu}^P_{2,k} - \bar{\mu}^P_{1,k}| + \rho_k\}, \tag{131}$$

$$\widehat{\Delta}_{L,k} = \left(|\bar{\mu}^P_{2,k} - \bar{\mu}^P_{1,k}| - \rho_k\right)_+, \tag{132}$$

$$U_k = \min\{k^\alpha, g_k^{-\alpha}\}, \qquad L_k = \max\{k^{-\alpha}, g_k^{\alpha}\}. \tag{133}$$

The simultaneous post probe event is

$$E_k = \bigcap_{i=1}^{2} \left\{ |\bar{\mu}_{i,k}^{P} - \mu_i| \le r_{i,k} \right\}.$$

Hoeffding's inequality gives $\Pr(E_k^c) \le 4k^{-M}$. For ordinary update n, let $K(n)$ be the index of the last completed probe; wall clock time and regret horizon follow Section 5. On an ordinary round after probe $k = K(n)$ and before probe $k+1$, set

$$a_n = \mathrm{clip}\left(\frac{16n\widehat{\Delta}_{L,k}}{|D_{n-1}|}, L_k, U_k \right), \qquad \eta_n = \sqrt{a_n/n}, \tag{134}$$

with value L_k when the ratio is undefined. The learner uses the exact two arm $1/2$-Tsallis KKT map and the predictable central estimator

$$B_{n,i} = \frac{1}{2}\mathbf{1}\{p_{n,i} \ge a_n/n\}, \qquad \widehat{\ell}_{n,i} = B_{n,i} + \frac{\mathbf{1}\{I_n = i\}}{p_{n,i}}(\ell_{n,i} - B_{n,i}). \tag{135}$$

For fixed $0 < L < U < \infty$, define

$$\theta = \frac{1/2 - \mu_1}{\Delta}, \quad a = \frac{\mu_1}{\Delta}, \quad q_{L,U}^{\mathrm{MC}}(x) = \mathrm{clip}(16/x, L, U),$$

$$x_0 = \frac{2}{L\Delta}, \qquad \lambda(x) = \frac{4}{q_{L,U}^{\mathrm{MC}}(x)\Delta^2 x^2}, \qquad j(x) = \frac{q_{L,U}^{\mathrm{MC}}(x)\Delta x^2}{8}.$$

The generator of the limiting MC process with fixed clips is

$$\begin{aligned} \mathcal{A}_{\Delta,\mu_1,L,U}^{\mathrm{MC}} f(x) &= (\theta - x) f'(x) \\ &\quad + \lambda(x)\{\mu_2[f(x + j(x)) - f(x)] + (1-\mu_2)[f(x - j(x)) - f(x)]\}, \quad x \le x_0, \\ &= (-a - x) f'(x) \\ &\quad + \mu_2 \lambda(x)[f(x + 2j(x)) - f(x)], \quad x > x_0. \end{aligned} \tag{136}$$

Lemma 7 (Fixed limit MC invariant law)**.** *For every instance and clips above,* (136) *has a conservative Borel right strong Markov realization, is positive Harris recurrent, and has a unique invariant probability* $\pi_{\Delta,\mu_1,L,U}^{\mathrm{MC}}$.

Proof. The state dependent $q_{L,U}^{\mathrm{MC}}$ is continuous, bounded above and away from zero, and piecewise smooth. The Davis construction used in Lemma 1 therefore applies with the same active and inactive flows. Since $\theta \le 0$, the integrated active jump rate diverges before a flow can reach zero. Also $j(x)/x \le 1/4$ on the active branch, so downward landings remain positive. Direct substitution gives $\mathcal{A}_{\Delta,\mu_1,L,U}^{\mathrm{MC}} x = 1 - x$. The stopped Taylor calculation for x^{-r}, $r > 2$, is unchanged up to constants depending on L, U and yields a Foster drift for $x + x^{-r}$. Finally, choose a compact active interval inside a constant clip region and condition on one upward jump in a compact time window. The endpoint Jacobian is bounded above and away from zero, giving the one jump minorization used in Lemma 1. Accessibility and the Foster–petite criterion give positive Harris recurrence and uniqueness. □

Proposition 2 (Exact fixed gap coefficient for moving clips)**.** *Let* $0 < \alpha < 1$ *and* $M > 6\alpha + 1$. *For every fixed Bernoulli instance with* $1/2 \le \mu_1 < \mu_2 \le 1$ *and* $0 < \Delta < 1/8$*, the learner in* (133)–(135) *satisfies*

$$\lim_{T\to\infty} \frac{R_T}{\log T} = \frac{1}{\Delta} \int_0^\infty \frac{4}{q_\infty(x) x^2}\, \pi_{\Delta,\mu_1,L_\infty,U_\infty}^{\mathrm{MC}}(dx), \tag{137}$$

where $L_\infty = \Delta^\alpha$, $U_\infty = \Delta^{-\alpha}$, $q_\infty(x) = \mathrm{clip}(16/x, L_\infty, U_\infty)$, *and* $\pi_{\Delta,\mu_1,L_\infty,U_\infty}^{\mathrm{MC}}$ *is the invariant law in Lemma 7.*

Theorem 3 states the resulting central small gap equality. Its horizon then gap order is part of the statement.

Global recovery with a moving strip. Let $J_b(t)$ be a deterministic nondecreasing bound on the one step increment inside $0 \le D_t \le b\sqrt{t}$. The global high and low episodes used in Appendix E need not be restarted when a probe changes the clips. The same stopped exponential supermartingale proof remains valid if

$$\sum_{t \ge t_0} t^v \exp\left\{-c\frac{\sqrt{t}}{J_b(4t)^2}\right\} < \infty \quad \text{for every } c, v > 0. \tag{138}$$

Indeed, phase zero retains its order-$\sqrt{s}$ separation from the lower barrier. A later phase-r success is charged to survival through phase $r-1$, rather than to a restarted barrier estimate. Summing over starts and dyadic phases introduces at most a logarithmic multiplicity, absorbed by (138). This proves finite expected low duration and summable polynomial moments of global episode entry potentials. Here $J_b(t) = O((\log t/\log\log t)^\alpha)$, so the condition holds.

Lemma 8 (Moving strip envelope and inverse cube drift)**.** *Let $0 < \alpha < 1$ and $M > 6\alpha + 1$.*

1. *On any ordinary round with $0 \le D_n \le b\sqrt{n}$,*

$$|D_{n+1} - D_n| \le J_b(n) := 1 + \tfrac{1}{4}\big(b\sqrt{U_{K(n)}} + 2\sqrt{2}\big)^2, \qquad J_b(n) = O((\log n/\log\log n)^\alpha)\,.$$

2. *Condition* (138) *holds.*

3. *Fix $b_0 > 0$, let $h_n = b_0\sqrt{n+1}$, and define global high episodes to start on reaching $D_n \ge 2h_n$ and end on reaching $D_n \le h_n$; let H_n be their predictable round indicator. Then on high rounds $V_n = n^2/D_n^2$ obeys* (139)*.*

Proof. (i) The estimator (135) changes each coordinate by at most 1 through its baseline and by $|\ell - B|/p_{n,I_n} \le 1/p_{n,I_n}$ through the played arm. The exact two arm KKT map solves $4/w^2 + 4/(s+w)^2 = 1$ with $s = \eta_n|D_n| \ge 0$; since $4/w^2 \le 1$ and $1 \le 8/w^2$, the leader root satisfies $2 \le w \le 2\sqrt{2}$, so the lagging probability obeys $p_{\mathrm{lag}} = 4/(s+w)^2 \ge 4/(s+2\sqrt{2})^2$. With $\eta_n = \sqrt{a_n/n}$, $a_n \le U_{K(n)}$ and $|D_n| \le b\sqrt{n}$ we get $s \le b\sqrt{U_{K(n)}}$, which gives the displayed envelope. Because $\tau_k = \lceil k^k \rceil$, the index $K(n)$ satisfies $K(n)\log K(n) \le \log n$, hence $K(n) \le (1+o(1))\log n/\log\log n$, and $U_k \le k^\alpha$ gives the stated growth.

(ii) With $J_b(4t)^2 \le C(\log t)^{2\alpha}$ the exponent in (138) is at most $-c'\sqrt{t}/(\log t)^{2\alpha}$, and $\sqrt{t}/(\log t)^{2\alpha} \ge t^{1/4}$ for large t, so $t^v \exp\{-c't^{1/4}\}$ is summable for every v.

(iii) Write $d = D_n$ and $Z_n = D_{n+1} - D_n$. The conditional drift is $\mathbb{E}[Z_n \mid \mathcal{F}_n] = \Delta$, which does not depend on the clip. The stopped Taylor bound for $d \mapsto (n+1)^2/d^2$ gives

$$\mathbb{E}[V_{n+1} - V_n \mid \mathcal{F}_n] \le \frac{2n+1}{d^2} - \frac{2(n+1)^2\Delta}{d^3} + \frac{3(n+1)^2\,\mathbb{E}[Z_n^2 \mid \mathcal{F}_n]}{(\gamma d)^4},$$

where γd is the no skipping floor for the interpolation interval. Exact KKT increments give $\mathbb{E}[Z_n^2 \mid \mathcal{F}_n] \le C_{2,n}(1 + d^2/n)$ with $C_{2,n} \le C(1 + U_{K(n)})$, so the right side is at most

$$C(1 + C_{2,n})\left(\frac{n}{d^2} + \frac{n^2}{d^4}\right) - \kappa_0\frac{n^2}{d^3}, \qquad \kappa_0 = 2\Delta.$$

Put $u = n/d$, so that $n/d^2 = u^2/n$ and $n^2/d^3 = u^3/n$. On a high round $d \ge h_n$, hence $n/d^2 \le 1/b_0^2$ and therefore $n^2/d^4 = (u^2/n)(n/d^2) \le u^2/(b_0^2 n)$. Both positive terms are thus of the form u^2/n, and

$$\mathbb{E}[V_{n+1} - V_n \mid \mathcal{F}_n] \le \frac{C'(1+C_{2,n})u^2 - \kappa_0 u^3}{n} \le \frac{C'(1+C_{2,n})u^2 - \frac{\kappa_0}{2}u^3}{n} - \frac{\kappa_0}{2}\frac{n^2}{d^3}.$$

The function $u \mapsto Au^2 - Bu^3$ attains $4A^3/(27B^2)$ on $u \ge 0$, so the first term is at most $C(1 + C_{2,n})^3/n$, which is (139) with $\kappa = \kappa_0/2$. The cube is exactly this maximization and is not an artifact of the bound. □

With the episodes of Lemma 8(iii), on high rounds $V_n = n^2/D_n^2$ obeys

$$\mathbb{E}[V_{n+1} - V_n \mid \mathcal{F}_n] \le \frac{C(1 + C_{2,n}^3)}{n} - \kappa \frac{n^2}{D_n^3}. \tag{139}$$

Calibrated epochs have $C_{2,n} \le C_\Delta$; on E_k^c it is at most Ck^α. The harmonic length of epoch k is $O(\log k)$, so $M > 3\alpha + 1$ makes the positive remainders summable. Telescoping over truncated high episodes, dropping the nonnegative terminal potentials and using that an entry potential is at most $(n+1)/(4b_0^2)$ with summable expectation by the global entry moments of Appendix E, this yields the unweighted inverse cube bound.

A predictable compensation for the cost weight. The cost also contains $L_k^{-3/2}$, and multiplying (139) directly by this random epoch weight creates an uncontrolled boundary term. Let

$$S = \sup_{k \ge k_0} L_k^{-3/2}, \qquad Q = \sup_{k \ge k_0} U_k, \qquad \bar{A}_k = \mathbb{E}[S \mid \mathcal{G}_k], \tag{140}$$

where $\mathcal{G}_k$ is generated by the probe observations through probe k. From $\Pr(E_k^c) \le 4k^{-M}$,

$$S \in L^2, \qquad Q \in L^6, \qquad \mathbb{E}[S(1 + Q^3)] < \infty \tag{141}$$

when $M > 6\alpha + 1$. Use $A_n = \bar{A}_k$ on ordinary rounds in epoch k. It is predictable, dominates $L_k^{-3/2}$, and its probe increments have conditional mean zero. For every localized high episode entry time σ,

$$\mathbb{E}[A_\sigma V_\sigma] = \mathbb{E}[S V_\sigma]. \tag{142}$$

Lemma 9 (Probe filtration and weighted stopping)**.** *Let $\mathcal{F}_n$ contain the ordinary observations through update n and the probe observations completed by that time, before the next ordinary arm draw. Realize the probe and ordinary Bernoulli samples by independent reward arrays. Then $A_n = \bar{A}_{K(n)}$ is $\mathcal{F}_n$-measurable, is predictable for update $n+1$, and, for every bounded localized $\mathcal{F}_n$-stopping time σ and every nonnegative $\mathcal{F}_\sigma$-measurable Z_σ,*

$$\mathbb{E}[A_\sigma Z_\sigma] = \mathbb{E}[S Z_\sigma].$$

Moreover, (141) *holds when $M > 6\alpha + 1$.*

Proof. On $\{K(\sigma) = k\}$, the stopped ordinary history uses only the probe observations through k and the independent ordinary reward array. Future probe observations are conditionally independent of that history given $\mathcal{G}_k$. Hence $\mathbb{E}[S \mid \mathcal{F}_\sigma] = \mathbb{E}[S \mid \mathcal{G}_k] = \bar{A}_k$, which proves the projection identity by conditioning and summing over k. Between probes A_n is constant. At probe $k+1$ its increment is the martingale difference $\mathbb{E}[S \mid \mathcal{G}_{k+1}] - \mathbb{E}[S \mid \mathcal{G}_k]$, so its conditional contribution in a stopped telescope is zero.

For the moments, good epochs eventually keep g_k between fixed positive multiples of Δ. A bad epoch has $L_k^{-3/2} \le k^{3\alpha/2}$ and $U_k \le k^\alpha$. Thus, for every $r > 0$, the tail-sum bound and $\Pr(E_k^c) \le 4k^{-M}$ give $\mathbb{E}[\sup_{k:E_k^c} k^{r\alpha}] < \infty$ whenever $M > r\alpha + 1$. Taking $r = 3$ gives $S \in L^2$, and taking $r = 6$ gives $Q \in L^6$. Cauchy–Schwarz then yields $\mathbb{E}[SQ^3] < \infty$. □

Apply the lemma with $Z_\sigma = V_\sigma$. Conditioning on the complete probe sequence makes the clip path deterministic, and the global entry estimate is uniform over that path. The stopped episode telescope, the zero-mean probe increments, (141), and monotone localization therefore give

$$\sum_{n \le T} \mathbb{E}\left[H_n L_{K(n)}^{-3/2} \frac{n^2}{D_n^3} \right] \le C_\Delta (1 + \log T). \tag{143}$$

The KKT inequality $p_{n+1,2} \le 4n/(L_{K(n)} D_n^2)$ then proves harmonic uniform integrability of $np_{n+1,2}$.

Identification. Almost surely $(L_k, U_k, \widehat{\Delta}_{L,k}/\Delta)$ converges to $(L_\infty, U_\infty, 1)$. A decreasing $C^{1,1}$ ramp around the limiting switch $x_\infty = 2/(L_\infty \Delta)$ gives, on all sufficiently late calibrated epochs, a uniform inward source inequality. Choose its width $0 < \rho \le \rho_0(\Delta, L_\infty) < x_\infty/8$ so that the lower clip margin covers the entire outer ramp layer. On bad epochs, the scale free compact test source is bounded, and $\sum_k k^{-M} \log k < \infty$ pays their total contribution. The normalized occupation measure consequently assigns no mass to the switch.

Recovery gives tightness away from zero and infinity; local KKT asymptotics identify every subsequential occupation limit with the limit with fixed clips martingale problem. Lemma 7 identifies that limit as $\pi^{\mathrm{MC}}_{\Delta,\mu_1,L_\infty,U_\infty}$. Finally, (143) transfers the unbounded sampling cost and the probe contribution is $o(\log T)$, proving (137).

The diagonal invariant law. It remains to evaluate the fixed gap formula along the controller's own moving clips. Write $\delta = \Delta$,

$$L_\delta = \delta^\alpha, \qquad U_\delta = \delta^{-\alpha}, \qquad q_\delta(x) = \mathrm{clip}(16/x, L_\delta, U_\delta),$$

let $\pi_\delta = \pi^{\mathrm{MC}}_{\delta,1/2,L_\delta,U_\delta}$, and set

$$J_\delta = \int_0^\infty \frac{4}{q_\delta(x)x^2}\, \pi_\delta(dx).$$

We prove

$$\pi_\delta \Rightarrow \pi_0(dx) = 4xe^{-2x}\, dx, \qquad J_\delta \longrightarrow \frac{1}{2}. \tag{144}$$

The invariant statement itself holds for every fixed $\alpha > 0$; the restriction $\alpha < 1$ enters only through the controller composition in Proposition 2.

Put

$$a_\delta = 16/U_\delta, \quad b_\delta = 16/L_\delta, \quad x_{0,\delta} = 2/(L_\delta \delta),$$

and, with $\mu_+ = 1/2 + \delta$,

$$\lambda_\delta(x) = \frac{4}{q_\delta(x)\delta^2 x^2}, \qquad j_\delta(x) = \frac{q_\delta(x)\delta x^2}{8}.$$

On the active branch $x \le x_{0,\delta}$ the limit with fixed clips generator is

$$\mathcal{A}_\delta f(x) = -xf'(x) + \lambda_\delta(x)\left\{\mu_+[f(x + j_\delta(x)) - f(x)] + (1-\mu_+)[f(x - j_\delta(x)) - f(x)]\right\}. \tag{145}$$

On the inactive branch it is

$$\mathcal{A}_\delta f(x) = \left(-\frac{1}{2\delta} - x\right) f'(x) + \mu_+\lambda_\delta(x)[f(x + 2j_\delta(x)) - f(x)]. \tag{146}$$

Direct substitution gives the same identity on both branches,

$$\mathcal{A}_\delta x = 1 - x. \tag{147}$$

The fixed box Foster construction justifies bounded truncations and their monotone removal, so $\int x\, d\pi_\delta = 1$.

A uniform lower boundary moment. Fix $p \in (0,1)$ and put $s = p + 1 \in (1,2)$. On the lower and middle regions, $r = j_\delta(x)/x = q_\delta(x)\delta x/8 \le 2\delta$. Taylor expansion of x^{-p}, together with the exact identities $\lambda_\delta(2\delta r) = x^{-1}$ and $\lambda_\delta r^2 = q_\delta(x)/16$, yields

$$\mathcal{A}_\delta x^{-p} \le x^{-p-1}\left[-p + px + \frac{p(p+1)q_\delta(x)x}{32} + C_p\delta\right].$$

Since $q_\delta(x)x \le 16$, the constant part is $-p(1-p)/2 < 0$. Thus, for fixed small $\bar{x}$ and all sufficiently small δ,

$$\mathcal{A}_\delta x^{-p} \le -c_p x^{-p-1}, \qquad 0 < x \le \bar{x}.$$

Apply invariance to smooth versions of $\min\{x^{-p}, \varepsilon^{-p}\}$ that are constant above $2\bar{x}$. Below ε an upward jump can only decrease the capped function, and a downward jump crossing the cap has

no larger positive contribution than for x^{-p}. On the fixed cutoff annulus, the first order jump term cancels the flow and the remainder is uniformly bounded. Letting $\varepsilon \downarrow 0$ gives

$$\sup_{0<\delta\le\delta_0} \int x^{-s}\, \pi_\delta(dx) < \infty \quad \text{for every fixed } 1 < s < 2. \tag{148}$$

Together with the first moment, this gives tightness away from both boundaries.

On every compact subset of $(0,\infty)$, eventually $q_\delta(x) = 16/x$, $j_\delta(x) = 2\delta x$, and $\lambda_\delta(x) = 1/(4\delta^2 x)$. Hence, uniformly on compacts,

$$\mathcal{A}_\delta f(x) = (1-x)f'(x) + \frac{x}{2} f''(x) + O_{f,K}(\delta), \qquad f \in C_c^3(0,\infty). \tag{149}$$

Active jump landing multipliers lie in $[3/4, 5/4]$, while the inactive branch has only upward jumps. Thus every compactly supported test has one fixed compact source hull, and stationarity passes to subsequential limits. The limiting CIR generator $\mathcal{L}_0 f = (1-x)f' + xf''/2$ has the unique invariant probability $4xe^{-2x}dx$. This proves the first part of (144).

The singular cost. A uniform inverse square moment would be false because the limiting $\Gamma(2,2)$ law has an infinite one. Instead, on the active branch an exact calculation gives

$$\mathcal{A}_\delta x^{-1} = \frac{1}{x} - \frac{1 - q_\delta(x)x/16}{x^2(1-r^2)}. \tag{150}$$

On the inactive branch, with $k = L_\delta \delta x/4$,

$$\mathcal{A}_\delta x^{-1} = \frac{1}{x} + \frac{k/2 - \delta}{\delta x^2(1+k)}.$$

Localizing x^{-1} first gives a one sided stationary inequality; the uniform x^{-s} moment then makes every term integrable and permits removal of the localization. If

$$N_\delta = \int_{x<a_\delta} \frac{1 - U_\delta x/16}{x^2(1-r^2)}\, \pi_\delta(dx),$$

the positive upper region terms are $O(L_\delta^2)$ and $O(L_\delta^2\delta)$, so $\sup_{\delta\le\delta_0} N_\delta < \infty$. Consequently the deep lower clip cost is

$$\int_{x\le 8/U_\delta} \frac{4}{U_\delta x^2}\, d\pi_\delta \le \frac{8N_\delta}{U_\delta} = O(U_\delta^{-1}).$$

On the remaining lower shell, (148) gives $O(U_\delta^{1-s})$; on $x \ge 16/L_\delta$ the cost is at most $L_\delta/64$. All clipped contributions therefore vanish.

In the middle region the cost is $1/(4x)$. Equation (148) supplies uniform integrability at zero, and $x^{-1}\mathbf{1}\{x > 16/L_\delta\} \le L_\delta/16$ handles infinity. Weak convergence now yields

$$J_\delta \longrightarrow \frac{1}{4}\int_0^\infty x^{-1} 4xe^{-2x}\, dx = \frac{1}{2}.$$

This proves (144). Combining it with Proposition 2 proves Theorem 3.

J Sources for the stochastic process arguments

The arguments in Appendices A, H, and I use standard results on martingale problems, Borel right PDMPs, stopped extended generators, and Harris recurrence. We specify the imported conclusions here because two boundaries matter for the proof. First, Jacod–Shiryaev identifies the stopped semimartingale characteristics, but does not provide the conditional time shift argument used for existence and uniqueness. Second, Davis's extended generator is defined through a local martingale, whereas the Foster–Lyapunov criteria of Meyn and Tweedie use true martingales for the stopped processes.

J.1 Stopped characteristics and the BKR martingale problem

Fix a compact exit time τ_N and a bounded truncation function κ that is the identity near the origin. The coordinate martingales first show directly that Y^{τ_N} is a special semimartingale. We then propose the predictable triplet

$$B_t^N = \int_0^{t\wedge\tau_N} \left[b(Y_{s-}) + \int \kappa(z) K(Y_{s-}, dz) \right] ds, \tag{151}$$

$$C_t^N = 0, \tag{152}$$

$$\nu^N(ds, dz) = \mathbf{1}_{\{s\le\tau_N\}} K(Y_{s-}, dz)\, ds. \tag{153}$$

On the compact landing hull, the processes in (153) are predictable, B^N has locally finite variation, C^N is continuous and nonnegative, and $\int (1 \wedge |z|^2)\nu^N(ds, dz)$ is finite on bounded intervals. The cutoff sine and cosine functions are legal martingale problem tests for every $u \in \mathbb{R}^{K-1}$ and give the exponential local martingale identity for this candidate triplet. The implication (b)⇒(a) in Jacod and Shiryaev's Theorem II.2.42 therefore identifies (153) as the characteristics of Y^{τ_N} (Jacod & Shiryaev, 2003, pp. 86–88). These are the results used directly below. Proposition II.2.9 supplies the compatible supporting fact that a semimartingale's characteristics admit one predictable increasing dominating clock and predictable differential characteristics (Jacod & Shiryaev, 2003, pp. 77–78); no separate common clock comparison is needed.

To obtain the conditional time shift, we use regular conditional shifted laws, Galmarino measurability, optional sampling at bounded stopping times, and an argument on a common null set. These steps do not follow from the cited Jacod–Shiryaev results. We prove them on canonical Skorokhod space using BKR's countable graph core (A2), rational times, a countable bounded algebra, conditional Φ-integrability, and monotone class closure. The continuous compensator in (153) then determines each first jump survival law and channel mark. Recursion proves uniqueness for each initial point (BKR A3), and disintegration at Y_0 proves uniqueness for a general initial law with $h_1 = 0$ (BKR A6). BKR Theorem 4.2 and Remark 4.3 are invoked only after A1–A3, A6, A8–A11, the finite Φ moment, and $\int A f \, d\Gamma = 0$ hold for the process (Bhatt et al., 2006).

J.2 The scalar PDMP and the right process bridge

For one coordinate with $\theta \ge 0$, the Davis lift uses $y = \log x$ and the disjoint mode space

$$\widetilde{E} = (\{A\} \times (-\infty, y_0]) \,\dot\cup\, (\{I\} \times (y_0, \infty)), \qquad y_0 = \log x_0,$$

with (I, y_0) an outgoing boundary point whose kernel is concentrated at (A, y_0). For $\theta < 0$ we instead use the x-coordinate lift. Its active flow has a finite outgoing endpoint at zero, but the integrated rate diverges there, so the endpoint is almost surely preempted and its kernel need not be specified, exactly as in Davis's Example 24.7 (Davis, 1993, pp. 61–62). These lifts satisfy all four standard conditions in Davis (24.8): the mode fields are locally Lipschitz and have no finite ODE explosion whenever the boundary time is infinite, the rate is Borel and locally integrable along flows, the kernel is measurable with no self jump from a state, and the separately proved estimate for the number of events gives $\mathbb{E}_z N_t < \infty$. Theorem 25.5 may therefore be used for the strong Markov property, and Theorem 27.8 for the Borel right property (Davis, 1993, pp. 62, 64, 77–78). Both are invoked only after all parts of condition (24.8), including (24.8)(4), have been established.

For the scalar process, the coordinate map is a Borel bijection with Borel inverse, so the natural filtrations coincide. This transfers the strong Markov identity and, by pullback of excessive functions, Borel rightness. Davis does not assert that a Borel projection preserves these properties. Nor does Theorem 27.8 assert left limits. The lifted path is only right continuous at the forced mode switch, while the scalar path is proved càdlàg directly from the flow and jump construction.

J.3 Local and stopped extended generator domains

For $V_q(x) = x + x^{-q}$, the three hypotheses of Davis's Theorem 26.14 are checked explicitly: V_q is absolutely continuous along each flow, its two boundary values obey the mean identity at the outgoing boundary, and its jump increments are locally integrable. The theorem then places V_q in Davis's *local*-martingale extended generator domain and gives the usual drift plus jump formula (Davis,

1993, pp. 66, 69–70). Theorem 26.14 does not yield an untruncated true martingale. Remark 26.16 separates that stronger conclusion through its additional absolute jump integrability condition, which we do not claim globally.

Instead take $O_N = (1/N, N)$, let τ_N be its exit time, and set $X_t^N = X_{t\wedge\tau_N}$. Before τ_N, the coefficients and jump increments are bounded, and the exit landing remains in a bounded interval. Direct compensation hence proves that

$$V_q(X_t^N) - V_q(X_0) - \int_0^{t\wedge\tau_N} \mathcal{A}V_q(X_s)\,ds$$

is an integrable true martingale and that the absolute drift integral is integrable. Thus $V_q \in D(\mathcal{A}^{(N)})$ for every N. This is exactly the stopped domain setup used by Meyn and Tweedie's CD0 and CD2 conditions. Their Theorem 2.1 applies to CD0, and their Theorem 4.2 applies to the already nonexplosive right process together with CD2, the closed petite drift set, and boundedness of V_q on that set (Meyn & Tweedie, 1993, pp. 521–525, 529–530). We make no claim that $V_q \in D(\mathcal{A})$ for the untruncated process.