THE EXACT WORST CASE OF THOMPSON SAMPLING ON TWO BERNOULLI ARMS AT HORIZONS UP TO EIGHTEEN

KOREA SUPERINTELLIGENCE LABS

Abstract. Fix the algorithm: Thompson sampling with independent Beta$(1,1)$ priors on two Bernoulli arms, in the form of Agrawal and Goyal's Algorithm 1. Its expected regret at horizon T on the instance with means (p,q), $q \le p$, is a polynomial $R_T(p,q)$ with rational coefficients, so its worst case over the instance triangle $0 \le q \le p \le 1$ is a well-defined number W_T and its worst-case instances are well-defined points. We determine W_T exactly for $T \le 4$, enclose it to 10^{-6} for $5 \le T \le 17$, prove that the worst case is attained on the edge $q = 0$ at every horizon up to 17, and show that it leaves the edge at 18. (i) The deterministic instance $(1,0)$ is a worst case for every $T \le 4$ and is not one at $T = 5$: $W_1, \ldots, W_4 = 1/2,\ 5/6,\ 25/24,\ 281/240$, while $R_5(49/50, 0) > R_5(1,0)$. (ii) For every $T \le 17$ the worst case is attained on the edge $q = 0$ — every instance (p,q) is dominated by the edge instance $((p-q)/(1-q), 0)$ — whereas at $T = 18$ the interior instance $(29/50, 1/100)$ beats every instance of the edge; so 18 is the smallest horizon at which the worst case leaves the edge. The departure is a first-order event at the edge: the inward derivative $\partial_q R_T(q + (1-q)u,\ q)|_{q=0}$ is ≤ 0 at every edge point u for every $T \le 17$, and at $T = 18$ it is strictly positive at an edge point whose regret is within 10^{-6} of the maximum over the edge. (iii) For $5 \le T \le 17$ the value W_T is enclosed on both sides to 10^{-6}: $1.255002\ldots,\ 1.330657\ldots,\ \ldots,\ 1.937644\ldots$. (iv) Against explore-then-commit, whose one-round exploration rule has exact worst case $T^2/(8(T-2))$ for $T \ge 4$, horizon 11 is the smallest horizon ≥ 3 at which W_T is strictly below the worst case of every explore-then-commit rule. Every upper bound is a Bernstein certificate of polynomial nonnegativity on $[0,1]^2$, every lower bound is an exact rational evaluation, the derivatives are derivatives of the real polynomial with values enclosed in exact rational arithmetic, and every statement is machine-checked in Lean 4. Nothing here bounds Thompson sampling in general: the results concern one fixed algorithm on two arms at small horizons.

1. Introduction

The algorithm. Thompson sampling goes back to Thompson's 1933 paper [13]. We fix it in the form that Agrawal and Goyal state as Algorithm 1 of [2] (the same Algorithm 1 appears in [1]), specialised to two arms and independent uniform priors. Each arm $i \in \{1,2\}$ carries two counters (S_i, F_i), both 0 at time 0. In every round the learner draws $\theta_i \sim \mathrm{Beta}(S_i + 1, F_i + 1)$ independently for $i = 1, 2$, plays $\arg\max_i \theta_i$, observes a reward $r \in \{0,1\}$ distributed as Bernoulli(μ_i) for the arm i played, and adds 1 to S_i if $r = 1$ and to F_i otherwise. In the words of [2]: "For each arm $i = 1, \ldots, N$, sample $\theta_i(t)$ from the Beta$(S_i + 1, F_i + 1)$ distribution. Play arm $i(t) := \arg\max_i \theta_i(t)$ and observe reward r_t." Neither paper states a tie-breaking rule; none is needed, since the two draws come from atomless distributions and $\theta_1 = \theta_2$ has probability zero (Proposition 2.2 records the finite check that no tie mass is lost).

The quantity. Write $p = \mu_1$ and $q = \mu_2$ and assume $q \le p$, so that arm 2 is the worse arm; by the symmetry of the algorithm this loses nothing. The instance space is the triangle

$$\Omega = \{(p,q) \in \mathbb{R}^2 :\ 0 \le q \le p \le 1\},$$

and the expected regret at horizon T is

$$R_T(p,q) = (p-q)\,\mathbb{E}\big[N_2(T)\big],$$

where $N_2(T)$ is the number of times arm 2 is played in the first T rounds. Two elementary facts (Section 2) make R_T a polynomial in (p,q) with rational coefficients: the probability that one Beta variable with integer parameters exceeds another is rational, and the probability of standing in a given posterior state after t rounds is a rational multiple of $p^{S_1}(1-p)^{F_1}q^{S_2}(1-q)^{F_2}$. Consequently the *worst case*

$$W_T = \max_{(p,q)\in\Omega} R_T(p,q)$$

exists, and so do the *worst-case instances*, the points of Ω where it is attained. This paper is about these objects, for the fixed algorithm above, at small horizons.

What is in print. The finite-horizon literature on Thompson sampling is a literature of bounds. Agrawal and Goyal prove expected regret $O(\ln T/\Delta + 1/\Delta^3)$ on two arms [1] and $O(\sqrt{NT\ln T})$ on N arms [2]; Kaufmann, Korda and Munos give the first finite-time analysis matching the Lai–Robbins rate for Bernoulli rewards [8]. The worst case itself is the object that Jin, Xu, Shi, Xiao and Gu single out as open in the abstract of [6]: "Despite its popularity and empirical success, it has remained an open problem whether Thompson sampling can match the minimax lower bound $\Omega(\sqrt{KT})$ for K-armed bandit problems, where T is the total time horizon." Their own algorithm samples from clipped Gaussians, not from $\mathrm{Beta}(1,1)$ posteriors, and their introduction records that "the TS strategy with Gaussian priors has a worst-case regret $\Omega(\sqrt{KT\log K})$"; for the Beta-prior algorithm the finite-horizon results we know of are upper bounds. Kalvit and Zeevi [7] study the arm-sampling behaviour of UCB under what they call a "small" gap worst-case lens, in diffusion scaling, and contrast it with Thompson sampling: for the same two-armed Bernoulli algorithm as here (their Algorithm 2, sampling from $\mathrm{Beta}(S_i+1, F_i+1)$), their Theorem 3 gives the limit law of $N_1(n)/n$ as $n\to\infty$ on the two instances where both arms have mean 0 or both have mean 1 — an "incomplete learning" phenomenon — which is a statement about arm sampling in the limit on instances of zero regret, not about the regret at a fixed horizon. Jacko's multidisciplinary survey of the finite-horizon two-armed bandit problem with binary responses [5] casts the problem as a Markov decision process with Bernoulli responses and Beta updating and evaluates designs from several disciplines computationally, in Bayesian and in frequentist performance measures on a few fixed instances; it reports no worst case over instances. Lattimore [9] proves near-optimal frequentist regret guarantees for the finite-horizon Gittins index strategy with Gaussian noise and prior. None of these prints an exact worst-case value, or a worst-case instance, of the Beta-prior algorithm at a fixed horizon, and we are not aware of any such value or instance in print.

Results. All statements below are for two Bernoulli arms and independent $\mathrm{Beta}(1,1)$ priors.

(i) *The corner.* For every $T\le 4$ the deterministic instance $(1,0)$ is a worst case, so $W_1, W_2, W_3, W_4 = 1/2,\ 5/6,\ 25/24,\ 281/240$ exactly (Theorem 3.2); at $T=5$ it is not, since $R_5(49/50,0) = 141187542923/112500000000 > 2007/1600 = R_5(1,0)$ (Theorem 3.3). So 5 is the smallest horizon at which $(1,0)$ is not a worst case (Theorem 3.4). The mechanism is the sign of $\partial_p R_T(1,0)$, which is $31/120$ at $T=4$ and $-847/14400$ at $T=5$ (Proposition 3.5).

(ii) *The edge.* For every $T\le 17$ every instance $(p,q)\in\Omega$ is dominated by an instance of the edge $q=0$: $R_T(p,q)\le R_T(r,0)$ with $r=(p-q)/(1-q)$ (Theorem 4.1). At $T=18$ the interior instance $(29/50, 1/100)$ has larger regret than every instance of the edge (Theorem 4.2). Hence 18 is the smallest horizon at which the worst case is not attained on the edge (Theorem 4.3). The mechanism is again a derivative, now in the inward direction: along the segment $q\mapsto (q+(1-q)u,\, q)$ from the edge point $(u,0)$ the regret is differentiable in q (Lemma 4.6), its derivative at $q=0$ is ≤ 0 for every $u\in[0,1]$ and every $T\le 17$, and at $T=18$ it is strictly positive at the edge point $u=5771/10000$, whose regret lies within 10^{-6} of the maximum over the edge; so at $T=18$ arbitrarily small steps off the edge from that point already increase the regret (Theorem 4.8). At that point the derivative is enclosed to 10^{-6} for $T=14,\dots,19$, and its sign changes between 17 and 18 (Proposition 4.7).

(iii) *The values.* For $5\le T\le 17$ the worst case is enclosed to 10^{-6}, the upper bound by a certificate and the lower bound by an explicit rational edge instance (Theorems 3.6 and 4.4); at $T=18$ the same is done for the maximum over the edge, which is no longer the worst case.

(iv) *Explore-then-commit.* The rule that explores each arm once and then commits has exact worst case $T^2/(8(T-2))$ for $T\ge 4$ (Proposition 5.1), and 11 is the smallest horizon ≥ 3 at which W_T is strictly below the worst case of every explore-then-commit rule (Theorem 5.3); $T=2$ is a degenerate exception (Proposition 5.2).

What is *not* claimed deserves equal emphasis. Nothing here says anything about Thompson sampling with other priors, on more than two arms, or as $T \to \infty$; nothing bounds the algorithm's regret beyond horizon 18; the interior worst-case instance at $T = 18$ is located only to the extent that one explicit point beats the whole edge; and the positive inward derivative at $T = 18$ is proved at one rational edge point, not on a neighbourhood of it, while for $T \leq 17$ the derivative is shown to be ≤ 0, not strictly negative, on the edge.

Method. Every upper bound is a *Bernstein certificate*: a polynomial with integer coefficients is proved nonnegative on $[0,1]^2$ by exhibiting a binary subdivision of the square on whose pieces all Bernstein coefficients are nonnegative (Section 2.4). The polynomials are the regret in the coordinates $p = q + (1-q)u$, which map $[0,1]^2$ onto Ω, and the certificate polynomial is built from the list of posterior states by the certificate checker's own arithmetic, so its coefficients are an output of the pipeline rather than an input to be trusted. Every lower bound is an exact rational evaluation of R_T at an explicit instance. The two derivative statements (Propositions 3.5 and 4.7, Theorem 4.8) are derivatives of the real polynomial in the sense of Mathlib's `HasDerivAt`, assembled by the product rule from the same list of states; their values at a rational point are exact rationals compared with rational bounds, and the non-positivity of the inward derivative for $T \leq 17$ is read off the edge-dominance certificates with no further computation. Everything is checked in Lean 4 with Mathlib (Section 6); the finite searches that *found* the certificates and the witnesses are described with each proof.

Changes from version 1. Version 1 of this paper reported the sign change of the inward derivative at the edge between $T = 17$ and $T = 18$ as computed evidence (its Remark 4.6) and listed its proof as an open item. This version proves it: Lemma 4.6, Proposition 4.7, Theorem 4.8, Remark 4.9 and Proposition 4.10 are new, the abstract, the summary of results above, the verification table of Section 6 and Appendix A are extended accordingly, and item (4) of Section 7 is replaced by what remains open. Every statement of version 1 is unchanged.

2. Preliminaries

2.1. **The Beta comparison is rational.** For positive integers a, b the law $\mathrm{Beta}(a, b)$ is that of the a-th smallest of $a + b - 1$ independent uniform variables on $[0, 1]$.

Lemma 2.1. *Let $X \sim \mathrm{Beta}(a, b)$ and $Y \sim \mathrm{Beta}(c, d)$ be independent, with a, b, c, d positive integers, and put $n = a + b - 1$, $m = c + d - 1$. Then*

$$\mathbb{P}[X > Y] = \binom{n+m}{n}^{-1} \sum_{j=c}^{m} \binom{a+j-1}{j}\binom{b-1+m-j}{m-j}, \tag{1}$$

and equivalently

$$\mathbb{P}[X > Y] = \frac{1}{B(a,b)} \sum_{j=c}^{c+d-1} \binom{c+d-1}{j} B(a+j,\, b+c+d-1-j), \tag{2}$$

where B is Euler's Beta function, $B(a, b) = (a-1)!\,(b-1)!/(a+b-1)!$.

Proof. For (1), realise X as the a-th smallest of n uniforms and Y as the c-th smallest of m further uniforms. All $\binom{n+m}{n}$ interleavings of the merged order are equally likely, and the number of them in which exactly j of the second sample lie below the a-th element of the first is $\binom{a+j-1}{j}\binom{b-1+m-j}{m-j}$; $X > Y$ exactly when $j \geq c$. For (2), integrate the density of X against the distribution function of Y, which at integer parameters is the binomial tail $I_x(m', n' - m' + 1) = \sum_{j=m'}^{n'} \binom{n'}{j} x^j (1-x)^{n'-j}$ [3, eq. 8.17.5] with $m' = c$ and $n' = c + d - 1$. ☐

Both formulas are implemented, and the following finite check is part of the formal development. It is the reason the missing tie-breaking rule is harmless: no probability mass sits on $\theta_1 = \theta_2$.

Proposition 2.2. *Formulas* (1) *and* (2) *agree on all 10^4 quadruples $(a, b, c, d) \in \{1, \ldots, 10\}^4$. On all 8^4 quadruples in $\{1, \ldots, 8\}^4$, $\mathbb{P}[X > Y] + \mathbb{P}[Y > X] = 1$, and $\mathbb{P}[X > Y] = 1/2$ whenever $(a, b) = (c, d)$.*

Proof. Exhaustive evaluation in exact rational arithmetic over the stated ranges. ☐

2.2. **The regret is a polynomial carried by the posterior states.** A *state* is a quadruple $s = (S_1, F_1, S_2, F_2)$ of nonnegative integers; its *level* is $S_1+F_1+S_2+F_2$, the number of rounds played. Write $\pi(s) = \mathbb{P}[\theta_1 > \theta_2]$ for the probability, given by Lemma 2.1 with $(a,b,c,d) = (S_1+1, F_1+1, S_2+1, F_2+1)$, that the algorithm plays arm 1 in state s.

Lemma 2.3. *For every state s there is a rational number $c_s \geq 0$, independent of (p,q), such that the probability of being in state s after $t = \mathrm{level}(s)$ rounds on the instance (p,q) is $c_s\, p^{S_1}(1-p)^{F_1} q^{S_2}(1-q)^{F_2}$. The c_s are given by $c_{(0,0,0,0)} = 1$ and the forward recursion in which a state s contributes $c_s\pi(s)$ to each of its two arm-1 successors and $c_s(1-\pi(s))$ to each of its two arm-2 successors. Consequently, for $(p,q) \in \Omega$,*

$$R_T(p,q) = (p-q) \sum_{t<T} \sum_{\mathrm{level}(s)=t} c_s\,(1-\pi(s))\; p^{S_1}(1-p)^{F_1} q^{S_2}(1-q)^{F_2}. \tag{3}$$

Proof. Each transition from s multiplies the probability by $\pi(s)$ or $1-\pi(s)$ (the sampling step) and by exactly one of $p, 1-p, q, 1-q$ (the reward), and the exponents record which. Summing $1-\pi(s)$ against the state probabilities over the levels $t < T$ gives $\mathbb{E}[N_2(T)]$. □

Formula (3) is the definition of R_T used throughout, at every real (p,q); on Ω it is the expected regret. It is a polynomial of degree at most T in each variable, with rational coefficients. The number of states at levels $< T$ grows like $T^4/24$; there are 1001 terms at $T = 11$ and 5985 at $T = 18$.

2.3. **Box coordinates.** The substitution

$$p = q + (1-q)u, \qquad \text{so that} \qquad 1-p = (1-u)(1-q), \quad p-q = u(1-q), \tag{4}$$

maps the square $[0,1]^2 \ni (u,q)$ onto the triangle Ω (the edge $q = 1$ of the square collapses to the point $(1,1)$). The point $(u,0)$ of the square is the edge instance $(u,0)$ of Ω, and the line $q = \mathrm{const}$ of the square is a segment of Ω ending on the edge. Let D_T be the least common denominator of the coefficients $c_s(1-\pi(s))$ in (3). Then $D_T R_T(q+(1-q)u,\, q)$ is a polynomial in (u,q) with integer coefficients, and by (4) it is assembled from the four factors $u,\ 1-u,\ q,\ 1-q$ alone.

2.4. **Bernstein certificates.** Let $F \in \mathbb{Z}[x_0, \ldots, x_{n-1}]$ have degree at most d_j in x_j. Writing $u_j = x_j$, $v_j = 1 - x_j$, the homogenisation

$$\widehat{F}(u,v) = \prod_j (u_j+v_j)^{d_j}\ F\Big(\frac{u_0}{u_0+v_0}, \ldots, \frac{u_{n-1}}{u_{n-1}+v_{n-1}}\Big)$$

is a polynomial in the $2n$ variables u_j, v_j whose coefficient of $\prod_j u_j^{i_j} v_j^{d_j - i_j}$ is $b_i \prod_j \binom{d_j}{i_j}$, with b_i the Bernstein coefficient of F at multidegree d; and $\widehat{F}(x, 1-x) = F(x)$. If all coefficients of $\widehat{F}$ are nonnegative then $F \geq 0$ on $[0,1]^n$. When they are not, the square is subdivided: the substitutions $x_j \mapsto x_j/2$ and $x_j \mapsto (1+x_j)/2$ act on $\widehat{F}$ as $(u_j, v_j) \mapsto (u_j,\, u_j + 2v_j)$ and $(u_j, v_j) \mapsto (2u_j + v_j,\, v_j)$, integer substitutions with nonnegative coefficients. A *certificate* for $F \geq 0$ on $[0,1]^n$ is a binary tree each of whose internal nodes names an axis j to split at its midpoint, and each of whose leaves is a sub-box on which all coefficients of the transformed $\widehat{F}$ are nonnegative. We count the *nodes* of a tree as internal nodes plus leaves, so that the trivial certificate (no subdivision) has one node.

The checker takes a tree and a polynomial presented in the (u,v) form and returns a Boolean; its soundness — if it returns true then $F \geq 0$ on $[0,1]^n$ — is a theorem of the formal development. Two features matter for what follows. First, $\widehat{F}$ is *built* from the term list of (3) by the checker's own additions and multiplications of the four homogeneous factors $u,\ 1-u,\ q,\ 1-q$, and the identity between this build and $D_T R_T(q+(1-q)u, q)$ is proved by list induction; so the Bernstein coefficients are never typed in. Second, all upper bounds below are instances of three generic statements:

- (C1) if the checker accepts a tree for $M_n D_T - M_d\, D_T R_T(q+(1-q)u, q)$ with $M_d > 0$, then $R_T \leq M_n/M_d$ on Ω;
- (C2) if the checker accepts a tree for $D_T\big(R_T(u,0) - R_T(q+(1-q)u, q)\big)$, then for every $(p,q) \in \Omega$ there is $r \in [0,1]$ with $R_T(p,q) \leq R_T(r,0)$ (namely $r = (p-q)/(1-q)$ when $q < 1$);

(C3) if the checker accepts a tree for $M_n D_T - M_d\, D_T R_T(u,0)$ with $M_d > 0$, then $R_T(r,0) \le M_n/M_d$ for every $r \in [0,1]$.

The certificate polynomial in (C3) does not depend on q, so its trees only ever split the u axis; we call such certificates one-variable. The trees themselves were found by a greedy search in a Python model of the checker (split the axis and position that leaves the fewest negative coefficients, recurse), and only the tree is handed to the formal checker.

2.5. **Explore-then-commit.** For $m \ge 0$ with $2m \le T$, the rule ETC(m) plays arm 1 for m rounds and arm 2 for m rounds, and then plays for the remaining $T - 2m$ rounds the arm with the larger number of successes, choosing by a fair coin when the counts are equal. With $S_1 \sim \mathrm{Bin}(m,p)$ and $S_2 \sim \mathrm{Bin}(m,q)$ independent,

$$R_T^{\mathrm{ETC}(m)}(p,q) = (p-q)\Big(m + (T-2m)\big(\mathbb{P}[S_2 > S_1] + \tfrac{1}{2}\mathbb{P}[S_2 = S_1]\big)\Big), \tag{5}$$

again a polynomial with rational coefficients. Garivier, Kaufmann and Lattimore [4] print the same decomposition for two Gaussian arms — in the proof of their Theorem 1, "$R^n_\mu(T) = \Delta\,\mathbb{E}_\mu[N_2] = \Delta\big(n + (T-2n)\mathbb{P}_\mu(S_n \le 0)\big)$" with S_n Gaussian — where the commit probability is a Gaussian tail and ties have probability zero; over Bernoulli arms the commit probability is rational and the tie convention matters, which is why it is stated. For the rule in general see [10, Chapter 6]. ETC(0) commits by a coin flip and has regret $(p-q)T/2$.

3. The deterministic instance is extremal exactly up to horizon four

Proposition 3.1. *At the deterministic instance* $(1,0)$*,*

$$R_T(1,0) = \tfrac{1}{2},\ \tfrac{5}{6},\ \tfrac{25}{24},\ \tfrac{281}{240},\ \tfrac{2007}{1600},\ \tfrac{1983559}{1512000},\ \tfrac{12736343}{9408000},\ \tfrac{9855651817}{7112448000} \qquad (T = 1,\dots,8).$$

Proof. Exact rational evaluation of (3), carried out three times: by the state recursion of Lemma 2.3, by an explicit enumeration of all 4^T paths of arms and rewards, and by the formal development. All three agree. □

Theorem 3.2. *For every horizon $T \le 4$ and every $(p,q) \in \Omega$, $R_T(p,q) \le R_T(1,0)$. In particular $W_1 = 1/2$, $W_2 = 5/6$, $W_3 = 25/24$, $W_4 = 281/240$.*

Proof. $T = 0$ is trivial. For $T = 1,2,3,4$ apply (C1) with $M_n/M_d = R_T(1,0)$ from Proposition 3.1. At each of these horizons the certificate is the trivial tree: every coefficient of the homogenised difference is already nonnegative at the polynomial's own multidegree, with no subdivision at all. □

Theorem 3.3. *At horizon* 5 *the deterministic instance is not a worst case:*

$$R_5\big(\tfrac{49}{50},0\big) = \frac{141187542923}{112500000000} = 1.2550003\ldots\ >\ 1.254375 = \frac{2007}{1600} = R_5(1,0).$$

Proof. Two exact rational evaluations of (3). The witness was found by a grid search along the edge $q = 0$ followed by local refinement; any rational point of the edge with $R_5 > 2007/1600$ would do. □

Theorem 3.4. *Horizon* 5 *is the smallest horizon at which the deterministic instance* $(1,0)$ *is not a worst case of Thompson sampling on two Bernoulli arms: for every $T \le 4$, $R_T \le R_T(1,0)$ on Ω, and $R_5 \le R_5(1,0)$ fails on Ω.*

Proof. Theorems 3.2 and 3.3. □

The mechanism is visible in one derivative. Along the edge $q = 0$ the function $p \mapsto R_T(p,0)$ is a polynomial; if its derivative at $p = 1$ is negative, moving the good arm's mean slightly below 1 increases the regret and the corner cannot be a maximum.

Proposition 3.5. *The map $p \mapsto R_T(p,0)$ is differentiable at $p = 1$ with*

$$\partial_p R_T(1,0) = \tfrac{1}{2},\ \tfrac{2}{3},\ \tfrac{13}{24},\ \tfrac{31}{120},\ -\tfrac{847}{14400},\ -\tfrac{131167}{378000} \qquad (T = 1,\dots,6);$$

in particular $\partial_p R_4(1,0) = 31/120 > 0$ and $\partial_p R_5(1,0) = -847/14400 < 0$.

Proof. Differentiate (3) termwise at $q = 0$: only the states with $S_2 = 0$ survive, and at $p = 1$ a term $c\,p^{S_1}(1-p)^{F_1}$ contributes cS_1 if $F_1 = 0$, $-c$ if $F_1 = 1$ and 0 otherwise. The values are exact rational evaluations of this termwise derivative. In the formal development the statement for $T = 4, 5$ is a genuine derivative (`HasDerivAt`) of the real function, assembled from the same list of states, not a formal-derivative convention; the six values are stated for the rational termwise derivative. □

Past the transition no rational maximiser is known, and the worst case is enclosed rather than computed.

Theorem 3.6.

$$W_5 \in \big(1.255002,\ 1.255003\big], \qquad W_6 \in \big(1.330657,\ 1.330658\big].$$

Explicitly: $R_5 \le 1255003/10^6$ and $R_6 \le 1330658/10^6$ on Ω, while $R_5(4893/5000, 0) > 1255002/10^6$ and $R_6(8921/10000, 0) > 1330657/10^6$.

Proof. Upper bounds by (C1) with subdivision trees of 21 nodes ($T = 5$) and 23 nodes ($T = 6$), every split on the u axis. Lower bounds by exact evaluation:

$$R_5\big(\tfrac{4893}{5000}, 0\big) = \frac{313750732373183384923}{250000000000000000000} = 1.2550029\ldots,$$
$$R_6\big(\tfrac{8921}{10000}, 0\big) = \frac{745168464542115411884587831}{560000000000000000000000000} = 1.3306579\ldots$$

The witnesses come from a floating-point grid search with local refinement, which puts the maximisers near $(0.978638, 0)$ and $(0.892096, 0)$; those decimals are evidence about the location, not part of the theorem. □

Proposition 3.7 (controls)**.** *(a) $R_5 \le 2007/1600$ fails on Ω, so the bound of Theorem 3.2 does not extend to $T = 5$. (b) No $(p, q) \in \Omega$ has $R_5(p, q) > 1255004/10^6$, so the enclosure of Theorem 3.6 cannot be widened upward. (c) Some $(p, q) \in \Omega$ has $R_5(p, q) > 0$, and $R_5(p, p) = 0$ for every real p. (d) $R_1(p, q) = (p - q)/2$ for all real p, q.*

Proof. (a) is Theorem 3.3; (b) follows from the upper bound of Theorem 3.6; (c) from Proposition 3.1 and the factor $p - q$ in (3); (d) because the single term at level 0 has weight $1/2$. □

Remark 3.8. Combining Proposition 3.1 with the lower bounds of Theorems 3.6 and 4.4 shows that $(1, 0)$ is not a worst case at $T = 6, 7, 8$ either: $R_6(1, 0) = 1.3118\ldots < 1.330657$, $R_7(1, 0) = 1.3537\ldots < 1.404396$ and $R_8(1, 0) = 1.3856\ldots < 1.474045$. We have not recorded $R_T(1, 0)$ beyond $T = 8$, so nothing in this form is asserted for larger T; the floating-point edge maximisers of Remark 4.11 lie well inside the edge.

4. The worst case stays on the edge $q = 0$ exactly up to horizon seventeen

Theorem 3.6 enclosed the worst case at $T = 5, 6$, with lower-bound witnesses in the interior of the edge $q = 0$. The next question is whether the worst case is attained on that edge. Say that the worst case at horizon T is *attained on the edge* if every $(p, q) \in \Omega$ is dominated by some edge instance, i.e. $R_T(p, q) \le R_T(r, 0)$ for some $r \in [0, 1]$; equivalently, if $W_T = \max_{r \in [0,1]} R_T(r, 0)$.

Theorem 4.1. *For every horizon $T \le 17$ and every $(p, q) \in \Omega$ there is $r \in [0, 1]$ with $R_T(p, q) \le R_T(r, 0)$; one may take $r = (p - q)/(1 - q)$ when $q < 1$. So the worst case is attained on the edge $q = 0$ at every horizon up to 17.*

Proof. Apply (C2) at each $T \le 17$. In the box coordinates (4) the inequality $R_T(u, 0) \ge R_T(q + (1-q)u, q)$ on $[0, 1]^2$ says precisely that sliding an instance along its segment $q = \text{const}$ down to the edge never lowers the regret. The formal statement is existential in r; its proof supplies $r = (p-q)/(1-q)$ for $q < 1$ and $r = 0$ for $q = 1$, where $(p, q) = (1, 1)$ and the regret is 0. The certificates are strikingly cheap: for every $T \le 16$ the trivial tree suffices — all Bernstein coefficients of the homogenised difference are already nonnegative at its own multidegree — and at $T = 17$ a tree of 5 nodes (two splits of the u axis) suffices. □

T	$W_T \in (v_T,\ v_T + 10^{-6}]$	edge witness r_T	nodes
7	(1.404396, 1.404397]	1663/2000	17
8	(1.474045, 1.474046]	7853/10000	17
9	(1.539283, 1.539284]	7481/10000	19
10	(1.600311, 1.600312]	7173/10000	19
11	(1.657475, 1.657476]	6911/10000	21
12	(1.711140, 1.711141]	1337/2000	17
13	(1.761644, 1.761645]	6487/10000	21
14	(1.809293, 1.809294]	789/1250	19
15	(1.854356, 1.854357]	1231/2000	19
16	(1.897070, 1.897071]	1203/2000	15
17	(1.937644, 1.937645]	5887/10000	21
18	edge only: (1.976265, 1.976266]	5771/10000	19

TABLE 1. The certified worst case of Thompson sampling at horizons 7 to 17 (Theorem 4.4), and the maximum over the edge $q = 0$ at horizon 18 (Theorem 4.2). "Nodes" is the size of the one-variable subdivision tree certifying the upper bound; the lower bound is the exact value at $(r_T, 0)$. For $T \le 4$ the worst case is the exact rational of Theorem 3.2, and for $T = 5, 6$ the enclosures are those of Theorem 3.6.

Theorem 4.2. *At horizon 18 the interior instance $(29/50, 1/100)$, with $q = 1/100 > 0$, has strictly larger regret than every instance of the edge: $R_{18}(r, 0) < R_{18}(29/50, 1/100)$ for all $r \in [0, 1]$. Quantitatively,*

$$\max_{r\in[0,1]} R_{18}(r,0) \in \big(1.976265,\ 1.976266\big] \qquad \text{and} \qquad R_{18}\big(\tfrac{29}{50}, \tfrac{1}{100}\big) > 1.976266.$$

Proof. The edge maximum is bounded above by (C3) with a one-variable tree of 19 nodes, and below by the exact evaluation $R_{18}(5771/10000, 0) > 1976265/10^6$. The interior value is an exact evaluation of (3) at a rational point: a fraction with an 89-digit numerator and an 89-digit denominator, equal to $1.9764401426391\ldots$, which exceeds $1976266/10^6$. The interior witness was found by a floating-point search over the triangle; the exact comparison is what is proved. □

Theorem 4.3. *Horizon 18 is the smallest horizon at which the worst case of Thompson sampling on two Bernoulli arms is not attained on the edge $q = 0$: for every $T \le 17$ every instance is dominated by an edge instance, and at $T = 18$ some instance with $q > 0$ dominates every edge instance.*

Proof. Theorems 4.1 and 4.2. □

Because the worst case is attained on the edge for $T \le 17$, a one-variable certificate on the edge polynomial bounds the whole triangle, and the worst case is enclosed at every horizon up to the departure.

Theorem 4.4. *For $7 \le T \le 17$ the worst case W_T lies in the interval $(v_T,\ v_T + 10^{-6}]$ with v_T as in Table 1; explicitly $R_T \le v_T + 10^{-6}$ on Ω and $R_T(r_T, 0) > v_T$ for the rational r_T listed there. At $T = 18$ the maximum of R_{18} over the edge lies in $(1.976265, 1.976266]$.*

Proof. Upper bounds: (C3) with the trees of Table 1, chained with Theorem 4.1, which turns an edge bound into a bound on Ω. Lower bounds: exact evaluation at $(r_T, 0)$. The witnesses come from the same floating-point search as before; the independent floating-point maxima (1.404396158, 1.474045260, 1.539283150, 1.600311377, 1.657475885, 1.711140469 at $T = 7, \ldots, 12$) fall inside the enclosures. □

Proposition 4.5 (controls). *(a) The conclusion of Theorem 4.1 fails at $T = 18$: it is not the case that every $(p, q) \in \Omega$ is dominated by an edge instance. (b) Some $r \in [0, 1]$ has $R_{18}(r, 0) > 1.976265$, so the edge enclosure at $T = 18$ cannot be pushed down; and no $(p, q) \in \Omega$ has $R_{12}(p, q) > 1.711142$, so the $T = 12$ enclosure cannot be widened upward. (c) $0 < 1/100 \le 29/50 \le 1$, so the witness of Theorem 4.2 is a genuine interior point of Ω; and the certificate that proves edge dominance at $T = 17$ (the 5-node tree) is rejected by the checker at $T = 18$.*

Proof. (a) is Theorem 4.2; (b) is the lower half of Theorem 4.2 and the upper half of Theorem 4.4; (c) is arithmetic plus one Boolean evaluation of the checker, which returns false, as it must, since the statement it would certify is false. ☐

The mechanism: the departure is first-order. The analogue of Proposition 3.5 for the departure from the edge is the derivative of the regret in the inward direction. Fix $u \in [0,1]$. As q runs from 0 to 1 the point $(q+(1-q)u,\, q)$ runs from the edge point $(u,0)$ along the segment $u = \text{const}$ of the box coordinates (4) into the interior of Ω, and

$$\ell_{T,u}(q) = R_T\big(q+(1-q)u,\, q\big) \tag{6}$$

is a polynomial in q. Its derivative at $q=0$,

$$\delta_T(u) = \ell'_{T,u}(0) = \partial_q\big[R_T(q+(1-q)u,\, q)\big]_{q=0},$$

is the rate at which the regret changes when an instance leaves the edge along its own segment — precisely the direction in which the certificates of Theorem 4.1 control it.

Lemma 4.6. *For every horizon T, every real u and every real x, the function $\ell_{T,u}$ is differentiable at x, and its derivative is the product rule applied termwise to* (3) *along the segment, where $p = q+(1-q)u$ has $p' = 1-u$.*

Proof. By (4) a term of (3) restricted to the segment is $w_s\, u(1-u)^{F_1}\,(q+(1-q)u)^{S_1}\, q^{S_2}\,(1-q)^{1+F_1+F_2}$ with $w_s = c_s(1-\pi(s))$; each factor is a polynomial in q and the product rule applies. In the formal development the derivative is a `HasDerivAt` of the real function at every real x, assembled from the list of states by Mathlib's product rule. At $q=0$ the factor q^{S_2} leaves only the states with $S_2 \in \{0,1\}$, and the derivative collapses to the polynomial in u

$$\delta_T(u) = \sum_{S_2=0} w_s\Big(S_1\, u^{S_1}(1-u)^{F_1+1} - (1+F_1+F_2)\, u^{S_1+1}(1-u)^{F_1}\Big) + \sum_{S_2=1} w_s\, u^{S_1+1}(1-u)^{F_1}, \tag{7}$$

the sums over the states of level $< T$ with the indicated S_2. Formula (7) is not part of the formal development, which evaluates the uncollapsed termwise derivative at $q=0$; it is the form used by the independent computation in the next proof. ☐

The edge point at which the derivative is evaluated below is $u_0 = 5771/10000$, the rational witness of the lower bound on the edge maximum at $T = 18$ in Theorem 4.2.

Proposition 4.7. *At the edge point $u_0 = 5771/10000$,*

T	$\delta_T(u_0) \in$
14	$(-0.260867,\ -0.260866)$
15	$(-0.183010,\ -0.183009)$
16	$(-0.107288,\ -0.107287)$
17	$(-0.034413,\ -0.034412)$
18	$(+0.035163,\ +0.035164)$
19	$(+0.101198,\ +0.101199)$

In particular $\delta_{17}(u_0) < 0 < \delta_{18}(u_0)$: the sign of the inward derivative at u_0 changes exactly between horizons 17 *and* 18.

Proof. By Lemma 4.6, $\delta_T(u_0)$ is an exact rational number: the termwise derivative of (3) evaluated over $\mathbb{Q}$ at $q=0$, $u=u_0$, identified with the real derivative by a cast lemma. Each entry of the table is one comparison of that rational with two rational bounds, evaluated exactly. The values were computed first, independently of the formal development, in two ways — from the collapsed form (7), and as an exact-rational central difference of R_T along the segment with step 10^{-6} — which agree to 6×10^{-13} (the truncation error of the difference quotient) and give $-0.260866546\ldots$, $-0.183009585\ldots$, $-0.107287265\ldots$, $-0.034412967\ldots$, $+0.035163848\ldots$, $+0.101198975\ldots$. The step from one horizon to the next lies between 0.066 and 0.078 throughout the window, so the sign change between 17 and 18 is transversal, more than four orders of magnitude wider than the enclosures. ☐

Theorem 4.8. (a) *For every horizon $T \le 17$ and every $u \in [0,1]$, $\delta_T(u) \le 0$: no infinitesimal step off the edge, from any point of the edge, increases the regret.*

(b) *At horizon 18 and the edge point $u_0 = 5771/10000$, $\delta_{18}(u_0) > 0$; consequently for every $\varepsilon > 0$ there is $q \in (0, \varepsilon)$ with*

$$R_{18}\big(q + (1-q)u_0,\, q\big) > R_{18}(u_0, 0).$$

(c) *The point is, in value, within 10^{-6} of the maximum over the edge: $R_{18}(u_0, 0) > 1.976265$, while $R_{18}(r, 0) \le 1.976266$ for every $r \in [0,1]$.*

So the departure of the worst case from the edge at $T = 18$ is a first-order sign change at the edge, at an edge point whose regret is within 10^{-6} of the edge maximum.

Proof. (a) The certificate of Theorem 4.1 says, in box coordinates, that for $T \le 17$ and all $(u,q) \in [0,1]^2$, $R_T(q + (1-q)u, q) \le R_T(u, 0)$, i.e. $\ell_{T,u}(q) \le \ell_{T,u}(0)$ for $0 \le q \le 1$. A function differentiable at 0 whose value at 0 is a maximum over $[0,1)$ has derivative ≤ 0 there, as the limit of difference quotients from the right that are all ≤ 0. This half reuses the eighteen certificates of Theorem 4.1 unchanged and involves no new computation.

(b) $\delta_{18}(u_0) > 0.035163$ by Proposition 4.7, and a function differentiable at 0 with positive derivative exceeds its value at 0 at every sufficiently small positive argument; so every $\varepsilon > 0$ admits such a $q \in (0, \varepsilon)$.

(c) is the pair of bounds already used in the proof of Theorem 4.2: the exact evaluation $R_{18}(u_0, 0) = 1.9762656\ldots > 1976265/10^6$ and the one-variable certificate (C3) with the 19-node tree. □

Remark 4.9 (what Theorem 4.8 adds to Theorems 4.1–4.3). The two descriptions of the departure are logically independent, in both directions. Theorems 4.2 and 4.3 locate the *global* maximum at $T = 18$ off the edge; they say nothing about any derivative on the edge, and are compatible with $\delta_{18}(u) \le 0$ at every $u \in [0,1]$: a maximum can leave a boundary through a second, distant bump while the boundary itself stays first-order stable. Theorem 4.8(b) rules this out at $T = 18$ and puts the instability at an edge point within 10^{-6}, in value, of the edge maximum. Conversely, (b) does not imply Theorem 4.2: a positive derivative at one edge point yields interior points that beat *that* edge point, by an amount that is not quantified and, for small q, smaller than the 10^{-6} slack in (c); it does not yield a point that beats *every* edge instance. What (b) does yield is a second proof that the inequality certified in Theorem 4.1 — $R_T(q + (1-q)u, q) \le R_T(u, 0)$ on $[0,1]^2$ — fails at $T = 18$, from the sign of one rational number, without the Bernstein control of Proposition 4.5(c) or the 89-digit witness of Theorem 4.2 (Proposition 4.10(e)). Half (a) is a corollary of Theorem 4.1 and is stated as such; the new content is (b), (c) and their conjunction with (a), which says *why* the departure horizon is 18 — the q-direction analogue of the sign change at the corner in Proposition 3.5.

Proposition 4.10 (controls). *With $u_0 = 5771/10000$: (a) it is not the case that $\delta_{18}(u_0) \ge 0.036$, and (b) it is not the case that $\delta_{17}(u_0) \le -0.035$, so a claim above the enclosure at 18 and a claim below the enclosure at 17 are both refuted. (c) Statement (a) of Theorem 4.8 with the hypothesis $T \le 17$ dropped is false: some T and some $u \in [0,1]$ have $\delta_T(u) > 0$. (d) Some $u \in [0,1]$ has $\delta_{17}(u) < 0$ strictly, so the bound of Theorem 4.8(a) is met strictly at an explicit edge point and is not vacuous. (e) It is not the case that $R_{18}(q + (1-q)u, q) \le R_{18}(u, 0)$ for all $(u,q) \in [0,1]^2$ — the box-coordinate inequality of Theorem 4.1 fails at $T = 18$ — proved from Theorem 4.8(b) alone, with no certificate and no interior witness.*

Proof. (a) and (b) from the enclosures of Proposition 4.7 and the uniqueness of the derivative; (c) and (d) from the sign change at u_0; (e) from Theorem 4.8(b) with $\varepsilon = 1$, whose point (u_0, q) violates the inequality. □

Remark 4.11 (computed evidence, not a theorem). Theorem 4.8 fixes one rational edge point for the whole ladder. A floating-point search over the edge puts the edge maximiser u_T^* at 0.615544, 0.601466, 0.588712, 0.577085, 0.566428 for $T = 15, \ldots, 19$, so $u_0 = 0.5771$ lies within 2×10^{-5} of u_{18}^*; the derivative (7) evaluated in exact rational arithmetic at these six-place points and rounded to six places is -0.148879, -0.089379, -0.027887, $+0.035158$, $+0.099385$; and on a grid of 2001 points of the edge it is ≤ 0 everywhere for $T \le 17$ and positive near u_T^* from $T = 18$ on — consistent with Theorem 4.8, which proves the first fact at every edge point and the second at u_0. The same search gives the edge maximisers $u_5^* \approx 0.978638$, $u_6^* \approx 0.892096$ and, after the departure, interior maximisers with $q \approx 0.011,\ 0.029,\ 0.045$ at $T = 18, 19, 20$. These are computations outside the formal development; no theorem of this paper depends on them.

5. Explore-then-commit is beaten first at horizon eleven

Proposition 5.1. *For every horizon $T \geq 4$ and every $(p,q) \in \Omega$,*

$$R_T^{\mathrm{ETC}(1)}(p,q) \leq \frac{T^2}{8(T-2)},$$

with equality at $\left(\frac{T}{2(T-2)}, 0\right)$. So the worst case of ETC(1) *is exactly $T^2/(8(T-2))$, attained wherever $p - q = T/(2(T-2))$.*

Proof. For $m = 1$ the commit probability in (5) collapses to $q(1-p) + \frac{1}{2}\big(pq + (1-p)(1-q)\big) = (1-p+q)/2$, so with $\Delta = p - q$, $R_T^{\mathrm{ETC}(1)} = \Delta\big(1 + (T-2)(1-\Delta)/2\big)$, and one checks the identity

$$\frac{T^2}{8(T-2)} - R_T^{\mathrm{ETC}(1)}(p,q) = \frac{\big(2(T-2)\Delta - T\big)^2}{8(T-2)}.$$

The square vanishes at $\Delta = T/(2(T-2))$, which lies in $[0,1]$ for $T \geq 4$. $\square$

Proposition 5.2. *At horizon* 2 *both admissible rules* ETC(0) *and* ETC(1) *have worst case exactly* 1*, attained at* $(1,0)$*, while $W_2 = 5/6$: for each $m \in \{0,1\}$ there is an instance on which* ETC(m) *suffers strictly more than Thompson sampling does anywhere on Ω.*

Proof. At $T = 2$, ETC(0) is a coin flip and ETC(1) plays each arm once with no commit round; both have regret $p - q$, which is 1 at $(1,0)$. Theorem 3.2 gives $R_2 \leq 5/6 < 1$. $\square$

For a horizon T write $W_T^{(m)} = \max_\Omega R_T^{\mathrm{ETC}(m)}$ for the worst case of ETC(m), $2m \leq T$.

Theorem 5.3. *Horizon* 11 *is the smallest horizon ≥ 3 at which the worst case of Thompson sampling is strictly below the worst case of every explore-then-commit rule:*

(a) *$W_{11} < W_{11}^{(m)}$ for every $m \in \{0,1,\dots,5\}$; indeed for each such m there is an instance (p,q) with $R_{11}(p',q') < R_{11}^{\mathrm{ETC}(m)}(p,q)$ for all $(p',q') \in \Omega$;*

(b) *for every $T \leq 10$ with $T \neq 2$ there is an admissible m and an instance (p,q) with $R_T^{\mathrm{ETC}(m)}(p',q') \leq R_T(p,q)$ for all $(p',q') \in \Omega$; in particular $W_T^{(m)} \leq W_T$.*

Proof. (a) First, $R_{11} \leq 1.658$ on Ω, by (C1) with a 9-node tree (the true value is $1.65747\ldots$ by Theorem 4.4, but 1.658 is enough). Then, for each m, one instance: ETC(0) at $(1,0)$ costs $11/2$; ETC(1) at $(11/18, 0)$ costs $121/72 = 1.6805\ldots$, its worst case by Proposition 5.1; and ETC(m) for $m = 2,3,4,5$ at $(1,0)$ commits correctly and costs exactly $m \geq 2$. Each exceeds 1.658.

(b) $T = 0$ is trivial and at $T = 1$ the coin flip ETC(0) costs $(p-q)/2 \leq 1/2 = R_1(1,0)$. At $T = 3$, ETC(1) costs $\Delta(3-\Delta)/2 \leq 1 < 25/24 = R_3(1,0)$. For $4 \leq T \leq 8$, Proposition 5.1 bounds ETC(1) by $T^2/(8(T-2))$, which is at most $R_T(1,0)$ of Proposition 3.1 ($1,\ 1.0416\ldots,\ 1.125,\ 1.225,\ 1.3333\ldots$ against $1.1708\ldots,\ 1.254375,\ 1.3118\ldots,\ 1.3537\ldots,\ 1.3856\ldots$). At $T = 9$ and $T = 10$ the corner no longer suffices and the Thompson-sampling witnesses are interior edge points:

$$R_9\big(\tfrac{3}{4}, 0\big) = \frac{15067274420293667}{9788549234688000} = 1.5392\ldots \ \geq\ \tfrac{81}{56} = 1.4464\ldots,$$
$$R_{10}\big(\tfrac{7}{10}, 0\big) = \frac{147769604615202264401}{92378880000000000000} = 1.5996\ldots \ \geq\ \tfrac{25}{16} = 1.5625.$$

Since R_T and each $R_T^{\mathrm{ETC}(m)}$ are polynomials on the compact set Ω, their maxima are attained, and (a), (b) are the statements about W_T and $W_T^{(m)}$ claimed; the formal statements are the witness forms, which avoid suprema altogether. $\square$

Remark 5.4. For $3 \leq T \leq 13$ the best explore-then-commit rule is ETC(1): ETC(0) has worst case $T/2$, and ETC(m) with $m \geq 2$ costs exactly $m \geq 2$ at $(1,0)$, whereas $T^2/(8(T-2)) < 2$ for $T \leq 13$ (and at $T = 3$ the worst case of ETC(1) is 1). This elementary comparison is not part of the formal development beyond the pieces used in Theorem 5.3. With it, the two worst cases read as follows, every entry exact or certified:

T	1	2	3	4	5	6	7	8	9	10	11
W_T	$\frac{1}{2}$	$\frac{5}{6}$	$\frac{25}{24}$	$\frac{281}{240}$	1.255002^+	1.330657^+	1.404396^+	1.474045^+	1.539283^+	1.600311^+	1.657475^+
$\min_m W_T^{(m)}$	$\frac{1}{2}$	1	1	1	$\frac{25}{24}$	$\frac{9}{8}$	$\frac{49}{40}$	$\frac{4}{3}$	$\frac{81}{56}$	$\frac{25}{16}$	$\frac{121}{72}$

where v^+ abbreviates the interval $(v, v+10^{-6}]$. Thompson sampling is strictly better at $T = 2$ (Proposition 5.2), no better at $3 \le T \le 10$, and strictly better again from $T = 11$; whether it stays strictly better for every $T \ge 12$ is not addressed here.

6. Verification

Every theorem and proposition above is a statement about the polynomial (3) (and (5)) over the real numbers, formalised and checked in Lean 4 [11] with Mathlib [12]; Appendix A reproduces the formal statements verbatim. The development has no `sorry` and no axiom beyond Lean's standard three (`propext`, `Classical.choice`, `Quot.sound`) and the axioms introduced by `native_decide`, which is used only to evaluate a Boolean on finite data: the certificate checker on a tree, an exact rational comparison, or an exhaustive check over a finite range. The reasoning layer — the soundness of the checker, the identity between the certificate polynomial and $D_T R_T$ in box coordinates, the three generic statements (C1)–(C3), the derivatives of Proposition 3.5 and Lemma 4.6 as `HasDerivAt`s, the two elementary facts about a derivative at an endpoint maximum and at a point of positive derivative used in Theorem 4.8, and the transfer of rational evaluations to $\mathbb{R}$ — uses only the standard three axioms. Running `#print axioms` on the headline declarations gives, beyond the standard three, exactly the following numbers of named `native_decide` axioms (one per evaluated Boolean):

declaration	statement	axioms
`corner_is_max_upto_four`	Theorem 3.2	20
`corner_not_max_at_five`	Theorem 3.3	9
`first_non_corner_horizon_is_five`	Theorem 3.4	21
`worst_case_five_enclosure`	Theorem 3.6	4
`worst_case_six_enclosure`	Theorem 3.6	4
`edge_dominates_upto_seventeen`	Theorem 4.1	18
`edge_not_max_at_eighteen`	Theorem 4.2	2
`edge_departure_horizon_is_eighteen`	Theorem 4.3	20
`worst_case_enclosures_seven_to_twelve`	Theorem 4.4	18
`worst_case_enclosures_thirteen_to_seventeen`	Theorem 4.4	15
`ts_below_every_etc_at_eleven`	Theorem 5.3(a)	9
`etc_at_most_ts_below_eleven`	Theorem 5.3(b)	12
`etc_one_worst`, `etc_one_attained`	Proposition 5.1	1, 1
`ts_below_every_etc_at_two`	Proposition 5.2	22
`edge_derivative_sign_flip`	Proposition 3.5	2
`edge_derivative_ladder`	Proposition 3.5	6
`edge_q_derivative_ladder`	Proposition 4.7	6
`edge_q_derivative_sign_flip`	Proposition 4.7	6
`edge_q_deriv_nonpos_upto_seventeen`	Theorem 4.8(a)	18
`interior_step_increases_at_eighteen`	Theorem 4.8(b)	6
`deriv_point_is_near_edge_maximiser`	Theorem 4.8(c)	2
`edge_departure_is_first_order`	Theorem 4.8(a),(b)	24
`corner_ladder`	Proposition 3.1	8
`beta_comparison_consistent`	Proposition 2.2	3
`tsRegret_le_of_cert`	(C1)	0
`tsRegret_le_edge_of_cert`	(C2)	0
`tsRegret_le_edge_box_of_cert`	(C2) in box coordinates	0
`tsRegret_edge_le_of_cert`	(C3)	0
`hasDerivAt_tsRegret_edge`	Proposition 3.5, real derivative	0
`hasDerivAt_lineRegret`	Lemma 4.6, real derivative	0

The six axioms of `edge_q_derivative_ladder` are the six rational comparisons of Proposition 4.7; every other statement of the first-order picture is derived from them, from the eighteen edge-dominance certificates and from the two $T = 18$ edge bounds of Theorem 4.2. `edge_departure_is_first_order` counts $18+6 = 24$, and each of the five controls of Proposition 4.10 depends on the same six and nothing else. The forward recursion of Lemma 2.3, the two formulas of Lemma 2.1, the polynomial build in box coordinates and all 44 Boolean evaluations behind the certificates and witnesses of Theorems 4.1–4.4 and Proposition 4.5 live in modules that import nothing from Mathlib, so that the expensive evaluations run in about 0.8 GB; the six comparisons of Proposition 4.7 are the only Boolean evaluations in a module that imports Mathlib, and that module checks in 33 s of wall time (38 s of CPU, peak 7.1 GB); a clean check of the whole development takes about six minutes of CPU time, the single largest item being the $T = 18$ certificates. The exact values were computed independently, before formalisation, by two Python implementations (the state recursion and an enumeration of all 4^T paths) which agree with each other and with the formal evaluations at every point checked, including $R_{18}(29/50, 1/100)$ and $R_{18}(5771/10000, 0)$; the six derivative values of Proposition 4.7 were likewise computed twice (collapsed form and central difference) before the formal comparisons were run. Repository reference to be added at submission.

7. Questions

(1) For $T \geq 18$ the worst-case instance is interior, and enclosing W_T needs a genuinely two-variable subdivision; its cost has not been measured. The evidence of Remark 4.11 suggests q grows slowly with T after the departure.

(2) The certified values give $W_T/\sqrt{T} = 0.601,\ 0.585,\ 0.561,\ \ldots,\ 0.470$ at $T = 3, 4, 5, \ldots, 17$, decreasing throughout the certified range. Whether $W_T/\sqrt{T}$ converges, and to what, is not addressed by anything here; the comparison of interest is with the Gaussian-prior lower bound $\Omega(\sqrt{KT\log K})$ quoted in [6].

(3) With priors $\mathrm{Beta}(\alpha, \beta)$, α, β positive integers, every probability stays rational and the same pipeline applies unchanged. Whether the two transition horizons (5 for the corner, 18 for the edge) move with the prior is a well-posed question that we have not asked; the inward derivative of Section 4 is prior-agnostic and is the cheapest way to *find* a candidate departure horizon for another prior, one rational sign per horizon.

(4) Theorem 4.8(a) gives $\delta_T(u) \leq 0$ on the whole edge for $T \leq 17$ from the edge-dominance certificates, but not a strict bound: $\delta_T(u) < 0$ for u in a closed sub-interval of the open edge would need a one-variable certificate on the derivative polynomial (7) itself, and so would positivity of δ_{18} on a whole neighbourhood of u_0, which the grid of Remark 4.11 suggests. Neither has been attempted.

Appendix A. The formal statements

The declarations below are reproduced verbatim from the formal development (proofs omitted). Here `tsRegret T p q` is $R_T(p, q)$ of (3) over $\mathbb{R}$, defined at every real (p, q); `tsRegretQ` is the same over $\mathbb{Q}$; `etcRegret T m p q` is $R_T^{\mathrm{ETC}(m)}(p, q)$ of (5); `edgeDerivQ T` is the rational termwise derivative of $p \mapsto R_T(p, 0)$ at $p = 1$; `betaCross`, `betaSym` and `betaDiag` are the three exhaustive checks of Proposition 2.2; `edgeDomOK T tr` is the checker on the edge-dominance polynomial of (C2) with the tree `tr`, and `domTree17` is the 5-node tree of Theorem 4.1; `boxP u q` is $q + u(1-q)$, so that `fun q => tsRegret T (boxP u q) q` is $\ell_{T,u}$ of (6), and `sumValR`, `dqSumR` are the state sum of (3) and its termwise q-derivative along the segment, over $\mathbb{R}$. The generic certificate theorems and the two real derivatives are listed last.

```
theorem corner_ladder :
    tsRegret 1 1 0 = 1/2 ∧ tsRegret 2 1 0 = 5/6 ∧ tsRegret 3 1 0 = 25/24 ∧
      tsRegret 4 1 0 = 281/240 ∧ tsRegret 5 1 0 = 2007/1600 ∧
      tsRegret 6 1 0 = 1983559/1512000 ∧ tsRegret 7 1 0 = 12736343/9408000 ∧
      tsRegret 8 1 0 = 9855651817/7112448000

theorem corner_is_max_upto_four (T : ℕ) (hT : T ≤ 4) :
    ∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret T p q ≤ tsRegret T 1 0
```

```
theorem corner_not_max_at_five :
    ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ tsRegret 5 1 0 < tsRegret 5 p q

theorem first_non_corner_horizon_is_five :
    (∀ T : ℕ, T ≤ 4 → ∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret T p q ≤ tsRegret T 1 0)
      ∧ ¬ (∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 5 p q ≤ tsRegret 5 1 0)

theorem worst_case_five_enclosure :
    (∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 5 p q ≤ 1255003/1000000)
      ∧ (∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1255002/1000000 : ℝ) < tsRegret 5 p q)

theorem worst_case_six_enclosure :
    (∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 6 p q ≤ 1330658/1000000)
      ∧ (∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1330657/1000000 : ℝ) < tsRegret 6 p q)

theorem beta_comparison_consistent :
    betaCross 10 = true ∧ betaSym 8 = true ∧ betaDiag 8 = true

theorem control_corner_bound_fails_at_five :
    ¬ (∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 5 p q ≤ 2007/1600)

theorem control_no_instance_above_enclosure :
    ¬ (∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1255004/1000000 : ℝ) < tsRegret 5 p q)

theorem control_nonvacuous :
    (∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ 0 < tsRegret 5 p q)
      ∧ (∀ p : ℝ, tsRegret 5 p p = 0)

theorem control_horizon_one (p q : ℝ) : tsRegret 1 p q = (p - q) / 2

theorem edge_derivative_sign_flip :
    HasDerivAt (fun p : ℝ => tsRegret 4 p 0) (31/120) 1
      ∧ HasDerivAt (fun p : ℝ => tsRegret 5 p 0) (-(847/14400)) 1

theorem edge_derivative_ladder :
    edgeDerivQ 1 = 1/2 ∧ edgeDerivQ 2 = 2/3 ∧ edgeDerivQ 3 = 13/24 ∧ edgeDerivQ 4 = 31/120
      ∧ edgeDerivQ 5 = -(847/14400) ∧ edgeDerivQ 6 = -(131167/378000)

theorem edge_dominates_upto_seventeen (T : ℕ) (hT : T ≤ 17) :
    ∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 →
      ∃ r : ℝ, 0 ≤ r ∧ r ≤ 1 ∧ tsRegret T p q ≤ tsRegret T r 0

theorem edge_not_max_at_eighteen :
    ∃ p q : ℝ, 0 < q ∧ q ≤ p ∧ p ≤ 1 ∧
      ∀ r : ℝ, 0 ≤ r → r ≤ 1 → tsRegret 18 r 0 < tsRegret 18 p q

theorem edge_departure_horizon_is_eighteen :
    (∀ T : ℕ, T ≤ 17 → ∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 →
        ∃ r : ℝ, 0 ≤ r ∧ r ≤ 1 ∧ tsRegret T p q ≤ tsRegret T r 0)
      ∧ (∃ p q : ℝ, 0 < q ∧ q ≤ p ∧ p ≤ 1 ∧
          ∀ r : ℝ, 0 ≤ r → r ≤ 1 → tsRegret 18 r 0 < tsRegret 18 p q)

theorem worst_case_enclosures_seven_to_twelve :
    ((∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 7 p q ≤ 1404397/1000000)
      ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1404396/1000000 : ℝ) < tsRegret 7 p q)
    ∧ ((∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 8 p q ≤ 1474046/1000000)
      ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1474045/1000000 : ℝ) < tsRegret 8 p q)
    ∧ ((∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 9 p q ≤ 1539284/1000000)
      ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1539283/1000000 : ℝ) < tsRegret 9 p q)
    ∧ ((∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 10 p q ≤ 1600312/1000000)
      ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1600311/1000000 : ℝ) < tsRegret 10 p q)
    ∧ ((∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 11 p q ≤ 1657476/1000000)
      ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1657475/1000000 : ℝ) < tsRegret 11 p q)
    ∧ ((∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 12 p q ≤ 1711141/1000000)
      ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1711140/1000000 : ℝ) < tsRegret 12 p q)

theorem worst_case_enclosures_thirteen_to_seventeen :
    ((∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 13 p q ≤ 1761645/1000000)
      ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1761644/1000000 : ℝ) < tsRegret 13 p q)
    ∧ ((∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 14 p q ≤ 1809294/1000000)
      ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1809293/1000000 : ℝ) < tsRegret 14 p q)
    ∧ ((∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 15 p q ≤ 1854357/1000000)
      ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1854356/1000000 : ℝ) < tsRegret 15 p q)
    ∧ ((∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 16 p q ≤ 1897071/1000000)
      ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1897070/1000000 : ℝ) < tsRegret 16 p q)
    ∧ ((∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret 17 p q ≤ 1937645/1000000)
      ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1937644/1000000 : ℝ) < tsRegret 17 p q)
```

```
theorem control_edge_dominance_fails_at_eighteen :
    ¬ (∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 →
        ∃ r : ℝ, 0 ≤ r ∧ r ≤ 1 ∧ tsRegret 18 p q ≤ tsRegret 18 r 0)

theorem control_enclosures_are_tight :
    (∃ r : ℝ, 0 ≤ r ∧ r ≤ 1 ∧ (1976265/1000000 : ℝ) < tsRegret 18 r 0)
      ∧ ¬ (∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧ (1711142/1000000 : ℝ) < tsRegret 12 p q)

theorem control_departure_nonvacuous :
    (0 : ℝ) < 1/100 ∧ (1/100 : ℝ) ≤ 29/50 ∧ (29/50 : ℝ) ≤ 1
      ∧ edgeDomOK 18 domTree17 = false

theorem edge_q_derivative_ladder :
    (∃ d : ℝ, HasDerivAt (fun q : ℝ => tsRegret 14 (boxP (5771/10000) q) q) d 0
        ∧ -(260867/1000000) < d ∧ d < -(260866/1000000))
      ∧ (∃ d : ℝ, HasDerivAt (fun q : ℝ => tsRegret 15 (boxP (5771/10000) q) q) d 0
        ∧ -(183010/1000000) < d ∧ d < -(183009/1000000))
      ∧ (∃ d : ℝ, HasDerivAt (fun q : ℝ => tsRegret 16 (boxP (5771/10000) q) q) d 0
        ∧ -(107288/1000000) < d ∧ d < -(107287/1000000))
      ∧ (∃ d : ℝ, HasDerivAt (fun q : ℝ => tsRegret 17 (boxP (5771/10000) q) q) d 0
        ∧ -(34413/1000000) < d ∧ d < -(34412/1000000))
      ∧ (∃ d : ℝ, HasDerivAt (fun q : ℝ => tsRegret 18 (boxP (5771/10000) q) q) d 0
        ∧ (35163/1000000) < d ∧ d < (35164/1000000))
      ∧ (∃ d : ℝ, HasDerivAt (fun q : ℝ => tsRegret 19 (boxP (5771/10000) q) q) d 0
        ∧ (101198/1000000) < d ∧ d < (101199/1000000))

theorem edge_q_derivative_sign_flip :
    (∃ d : ℝ, HasDerivAt (fun q : ℝ => tsRegret 17 (boxP (5771/10000) q) q) d 0 ∧ d < 0)
      ∧ (∃ d : ℝ, HasDerivAt (fun q : ℝ => tsRegret 18 (boxP (5771/10000) q) q) d 0
          ∧ 0 < d)

theorem deriv_point_is_near_edge_maximiser :
    (1976265/1000000 : ℝ) < tsRegret 18 (5771/10000) 0
      ∧ ∀ r : ℝ, 0 ≤ r → r ≤ 1 → tsRegret 18 r 0 ≤ 1976266/1000000

theorem edge_q_deriv_nonpos_upto_seventeen (T : ℕ) (hT : T ≤ 17) (u d : ℝ)
    (hu0 : 0 ≤ u) (hu1 : u ≤ 1)
    (hd : HasDerivAt (fun q : ℝ => tsRegret T (boxP u q) q) d 0) : d ≤ 0

theorem interior_step_increases_at_eighteen (ε : ℝ) (hε : 0 < ε) :
    ∃ q : ℝ, 0 < q ∧ q < ε ∧
      tsRegret 18 (5771/10000) 0 < tsRegret 18 (boxP (5771/10000) q) q

theorem edge_departure_is_first_order :
    (∀ T : ℕ, T ≤ 17 → ∀ u d : ℝ, 0 ≤ u → u ≤ 1 →
        HasDerivAt (fun q : ℝ => tsRegret T (boxP u q) q) d 0 → d ≤ 0)
      ∧ (∀ ε : ℝ, 0 < ε → ∃ q : ℝ, 0 < q ∧ q < ε ∧
          tsRegret 18 (5771/10000) 0 < tsRegret 18 (boxP (5771/10000) q) q)

theorem control_edge_dominance_fails_at_eighteen_by_derivative :
    ¬ (∀ u q : ℝ, 0 ≤ u → u ≤ 1 → 0 ≤ q → q ≤ 1 →
        tsRegret 18 (boxP u q) q ≤ tsRegret 18 u 0)

theorem control_eighteen_derivative_not_too_large :
    ¬ (∃ d : ℝ, HasDerivAt (fun q : ℝ => tsRegret 18 (boxP (5771/10000) q) q) d 0
        ∧ (36/1000 : ℝ) ≤ d)

theorem control_seventeen_derivative_not_too_small :
    ¬ (∃ d : ℝ, HasDerivAt (fun q : ℝ => tsRegret 17 (boxP (5771/10000) q) q) d 0
        ∧ d ≤ -(35/1000 : ℝ))

theorem control_horizon_hypothesis_not_idle :
    ¬ (∀ T : ℕ, ∀ u d : ℝ, 0 ≤ u → u ≤ 1 →
        HasDerivAt (fun q : ℝ => tsRegret T (boxP u q) q) d 0 → d ≤ 0)

theorem control_derivative_nonvacuous :
    ∃ u d : ℝ, 0 ≤ u ∧ u ≤ 1 ∧
      HasDerivAt (fun q : ℝ => tsRegret 17 (boxP u q) q) d 0 ∧ d < 0

theorem etc_one_worst (T : ℕ) (hT : 4 ≤ T) (p q : ℝ)
    (h0 : 0 ≤ q) (h1 : q ≤ p) (h2 : p ≤ 1) :
    etcRegret T 1 p q ≤ (T : ℝ)^2 / (8 * ((T : ℝ) - 2))

theorem etc_one_attained (T : ℕ) (hT : 4 ≤ T) :
    etcRegret T 1 ((T : ℝ)/(2 * ((T : ℝ) - 2))) 0 = (T : ℝ)^2 / (8 * ((T : ℝ) - 2))

theorem ts_eleven_upper (p q : ℝ) (h0 : 0 ≤ q) (h1 : q ≤ p) (h2 : p ≤ 1) :
    tsRegret 11 p q ≤ 1658/1000
```

```
theorem ts_below_every_etc_at_eleven (m : ℕ) (hm : 2 * m ≤ 11) :
    ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧
      ∀ p' q' : ℝ, 0 ≤ q' → q' ≤ p' → p' ≤ 1 → tsRegret 11 p' q' < etcRegret 11 m p q

theorem ts_below_every_etc_at_two (m : ℕ) (hm : 2 * m ≤ 2) :
    ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧
      ∀ p' q' : ℝ, 0 ≤ q' → q' ≤ p' → p' ≤ 1 → tsRegret 2 p' q' < etcRegret 2 m p q

theorem etc_at_most_ts_below_eleven (T : ℕ) (hT : T ≤ 10) (hT2 : T ≠ 2) :
    ∃ m : ℕ, 2 * m ≤ T ∧ ∃ p q : ℝ, 0 ≤ q ∧ q ≤ p ∧ p ≤ 1 ∧
      ∀ p' q' : ℝ, 0 ≤ q' → q' ≤ p' → p' ≤ 1 → etcRegret T m p' q' ≤ tsRegret T p q

theorem tsRegret_le_of_cert (T : ℕ) (Mn Md : ℤ) (tr : SubTree)
    (hden : denOK T = true) (hD : 0 < tsDen T) (hMd : 0 < Md)
    (hc : checkHP tr (targetHP T Mn Md) = true) :
    ∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 → tsRegret T p q ≤ (Mn : ℝ) / (Md : ℝ)

theorem tsRegret_le_edge_of_cert (T : ℕ) (tr : SubTree) (hc : edgeDomOK T tr = true) :
    ∀ p q : ℝ, 0 ≤ q → q ≤ p → p ≤ 1 →
      ∃ r : ℝ, 0 ≤ r ∧ r ≤ 1 ∧ tsRegret T p q ≤ tsRegret T r 0

theorem tsRegret_edge_le_of_cert (T : ℕ) (Mn Md : ℤ) (tr : SubTree) (hMd : 0 < Md)
    (hc : edgeUpOK T Mn Md tr = true) :
    ∀ r : ℝ, 0 ≤ r → r ≤ 1 → tsRegret T r 0 ≤ (Mn : ℝ) / (Md : ℝ)

theorem hasDerivAt_tsRegret_edge (T : ℕ) (x : ℝ) :
    HasDerivAt (fun p : ℝ => tsRegret T p 0)
      (sumValR x 0 (tsTerms T) + x * dSumR x (tsTerms T)) x

theorem tsRegret_le_edge_box_of_cert (T : ℕ) (tr : SubTree) (hc : edgeDomOK T tr = true) :
    ∀ u q : ℝ, 0 ≤ u → u ≤ 1 → 0 ≤ q → q ≤ 1 → tsRegret T (boxP u q) q ≤ tsRegret T u 0

theorem hasDerivAt_lineRegret (T : ℕ) (u x : ℝ) :
    HasDerivAt (fun q : ℝ => tsRegret T (boxP u q) q)
      (-u * sumValR (boxP u x) x (tsTerms T) + (boxP u x - x) * dqSumR u x (tsTerms T)) x
```

References

[1] S. Agrawal and N. Goyal, *Analysis of Thompson Sampling for the multi-armed bandit problem*, arXiv:1111.1797v3 (2012); published in: Proceedings of the 25th Annual Conference on Learning Theory (COLT 2012), Proceedings of Machine Learning Research **23** (2012), pp. 39.1–39.26.

[2] S. Agrawal and N. Goyal, *Further Optimal Regret Bounds for Thompson Sampling*, arXiv:1209.3353v1 (2012); published in: Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2013), Proceedings of Machine Learning Research **31** (2013), pp. 99–107.

[3] NIST Digital Library of Mathematical Functions, §8.17, eq. 8.17.5, `https://dlmf.nist.gov/8.17`.

[4] A. Garivier, E. Kaufmann and T. Lattimore, *On Explore-Then-Commit Strategies*, arXiv:1605.08988v2 (2016); published in: Advances in Neural Information Processing Systems **29** (NIPS 2016), pp. 784–792.

[5] P. Jacko, *The Finite-Horizon Two-Armed Bandit Problem with Binary Responses: A Multidisciplinary Survey of the History, State of the Art, and Myths*, arXiv:1906.10173v1 (2019).

[6] T. Jin, P. Xu, J. Shi, X. Xiao and Q. Gu, *MOTS: Minimax Optimal Thompson Sampling*, arXiv:2003.01803v3 (2020); published in: Proceedings of the 38th International Conference on Machine Learning (ICML 2021), Proceedings of Machine Learning Research **139** (2021), pp. 5074–5083.

[7] A. Kalvit and A. Zeevi, *A Closer Look at the Worst-case Behavior of Multi-armed Bandit Algorithms*, Advances in Neural Information Processing Systems **34** (NeurIPS 2021), pp. 8807–8819; arXiv:2106.02126v3.

[8] E. Kaufmann, N. Korda and R. Munos, *Thompson Sampling: An Asymptotically Optimal Finite-Time Analysis*, in: Algorithmic Learning Theory (ALT 2012), Lecture Notes in Computer Science **7568**, Springer, 2012, pp. 199–213; also arXiv:1205.4217; `https://doi.org/10.1007/978-3-642-34106-9_18`.

[9] T. Lattimore, *Regret Analysis of the Finite-Horizon Gittins Index Strategy for Multi-Armed Bandits*, 29th Annual Conference on Learning Theory (COLT 2016), Proceedings of Machine Learning Research **49** (2016), pp. 1214–1245; arXiv:1511.06014v3.

[10] T. Lattimore and Cs. Szepesvári, *Bandit Algorithms*, Cambridge University Press, 2020. Chapter 6, "The Explore-Then-Commit Algorithm".

[11] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction — CADE 28, Lecture Notes in Computer Science **12699**, Springer, 2021, pp. 625–635; `https://doi.org/10.1007/978-3-030-79876-5_37`.

[12] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, pp. 367–381; `https://doi.org/10.1145/3372885.3373824`.

[13] W. R. Thompson, *On the likelihood that one unknown probability exceeds another in view of the evidence of two samples*, Biometrika **25** (1933), no. 3–4, 285–294.