Statistica Sinica: Supplement

INFINITE-ARMS BANDIT:
OPTIMALITY VIA CONFIDENCE BOUNDS

Hock Peng Chan and Shouri Hu

National University of Singapore

Supplementary Material

This document consists of four sections. In Section A we prove Lemma 2, in Section B we prove Theorem 1, in Section C we prove Theorem 2 and in Section D we provide details to the verifications of (A1), (B1) and (B2) in Examples 3–5.

A Proof of Lemma 2

Let the infinite arms bandit problem be labeled as Problem A, and let R_A be the smallest possible regret for this problem. We prove Lemma 2 by considering two related problems, Problems B and C.

PROOF OF LEMMA 2. Let Problem B be like Problem A except that when we observe the first positive loss from arm k, its mean μ_k is revealed.

Let R_B be the smallest regret for Problem B. Since in Problem B we have access to additional arm-mean information, $R_A \geq R_B$.

In Problem B the best solution involves an initial exploration phase in which we play K arms, each until its first positive loss. This is followed by an exploitation phase in which we play the best arm for the remaining $n-M$ trials, where M is the number of rewards in the exploration phase. It is always advantageous to experiment first because no information on arm mean is gained during exploitation. For continuous rewards $M = K$. Let $\mu_b (= \mu_{\text{best}}) = \min_{1 \leq k \leq K} \mu_k$.

In Problem C like in Problem B, μ_k is revealed upon the observation of its first positive X_{kt}. The difference is that instead of playing the best arm for $n - M$ additional trials, we play it for n additional trials, for a total of $n + M$ trials. Let R_C be the smallest regret of Problem C, the expected value of $\sum_{k=1}^K n_k \mu_k$, with $\sum_{k=1}^K n_k = n + M$. We can extend the optimal solution of Problem B to a (possibly non-optimal) solution of Problem C by simply playing the best arm with mean μ_b a further M times. Hence

$$[R_A + E(M\mu_b) \geq] R_B + E(M\mu_b) \geq R_C. \tag{A.1}$$

Lemma 2 follows from Lemmas 3 and 4 below. □

Lemma 3. *$R_C = n\zeta_n$ for ζ_n satisfying $v(\zeta_n) = \frac{\lambda}{n}$.*

Lemma 4. $E(M\mu_b) = o(n^{\frac{\beta}{\beta+1}})$.

Bonald and Proutière (2013) also referred to Problem B in their lower bounds for Bernoulli rewards. What is different in our proof of Lemma 2 is a further simplification by considering Problem C, in which the number of rewards in the exploitation phase is fixed to be n. We show in Lemma 3 that under Problem C the optimal regret has a simple expression $n\zeta_n$, and reduce the proof of Lemma 2 to showing Lemma 4.

Proof of Lemma 3. Let arm j be the best arm after k arms have been played in the experimentation phase, that is $\mu_j = \min_{1 \leq i \leq k} \mu_i$. Let ϕ_* be the strategy of trying out a new arm if and only if $nv(\mu_j) > \lambda$, or equivalently $\mu_j > \zeta_n$. Since we need on the average $\frac{1}{p(\zeta_n)}$ arms before achieving $\mu_j \leq \zeta_n$, and the exploration cost of each arm is λ, the regret of ϕ_* is

$$R_* = \frac{\lambda}{p(\zeta_n)} + nE_g(\mu|\mu \leq \zeta_n) = r_n(\zeta_n) = n\zeta_n, \tag{A.2}$$

see (5.1) and Lemma 1 in the main manuscript for the second and third equalities in (A.2).

Hence $R_C \leq n\zeta_n$ and to show Lemma 3, it remains to show that for any strategy ϕ, its regret R_ϕ is not less than R_*. Let K_* be the number of arms played by ϕ_* and K the number of arms played by ϕ. Let $\mu_* = \min_{1 \leq k \leq K_*} \mu_k$. Let $G_1 = \{K < K_*\} (= \{\min_{1 \leq k \leq K} \mu_k > \zeta_n\})$ and $G_2 =$

$\{K > K_*\}(= \{\mu_* \le \zeta_n, K > K_*\})$. Since

$$\begin{aligned} R_\phi &= \lambda E(K) + nE(\min_{1\le k\le K} \mu_k), \\ R_* &= \lambda E(K_*) + nE(\mu_*), \end{aligned}$$

we can express

$$R_\phi - R_* = \sum_{\ell=1}^{2} \Big\{\lambda E[(K-K_*)\mathbf{1}_{G_\ell}] + nE\Big[\Big(\min_{1\le k\le K} \mu_k - \mu_*\Big)\mathbf{1}_{G_\ell}\Big]\Big\}. \quad \text{(A.3)}$$

Under G_1, $\min_{1\le k\le K} \mu_k > \zeta_n$ and therefore by (A.2),

$$\begin{aligned} &\lambda E[(K-K_*)\mathbf{1}_{G_1}] + nE\Big[\Big(\min_{1\le k\le K} \mu_k - \mu_*\Big)\mathbf{1}_{G_1}\Big] \quad \text{(A.4)} \\ = \; & -\tfrac{\lambda P(G_1)}{p(\zeta_n)} + n\Big\{E\Big[\Big(\min_{1\le k\le K} \mu_k\Big)\mathbf{1}_{G_1}\Big] - P(G_1)E_g(\mu|\mu \le \zeta_n)\Big\} \\ \ge \; & P(G_1)\{-\tfrac{\lambda}{p(\zeta_n)} + n[\zeta_n - E_g(\mu|\mu\le\zeta_n)]\} = 0. \end{aligned}$$

The identity $E[(K_* - K)\mathbf{1}_{G_1}] = \frac{P(G_1)}{p(\zeta_n)}$ is due to $\min_{1\le k\le K} \mu_k > \zeta_n$ when there are K arms, and so an additional $\frac{1}{p(\zeta_n)}$ arms on average is required under strategy ϕ_*, to get an arm with μ_k not more than ζ_n. The identity

$$E(\mu_* \mathbf{1}_{G_1}) = P(G_1)E(\mu_*) = P(G_1)E_g(\mu|\mu\le\zeta_n)$$

is due to the independence between $\mathbf{1}_{G_1}$ and μ_*.

In view that $(K-K_*)\mathbf{1}_{G_2} = \sum_{j=0}^{\infty} \mathbf{1}_{\{K>K_*+j\}}$ and

$$\Big(\min_{1\le k\le K} \mu_k - \mu_*\Big)\mathbf{1}_{G_2}$$

$$= \sum_{j=0}^{\infty} \Big(\min_{1 \le k \le K_*+j+1} \mu_k - \min_{1 \le k \le K_*+j} \mu_k \Big) \mathbf{1}_{\{K > K_*+j\}},$$

it follows that

$$\begin{aligned} & \lambda E[(K-K_*)\mathbf{1}_{G_2}] + nE\Big[\Big(\min_{1 \le k \le K} \mu_k - \mu_*\Big)\mathbf{1}_{G_2}\Big] \qquad \text{(A.5)} \\ = & \sum_{j=0}^{\infty} E\Big\{\Big[\lambda + n\Big(\min_{1 \le k \le K_*+j+1} \mu_k - \min_{1 \le k \le K_*+j} \mu_k\Big)\Big]\mathbf{1}_{\{K > K_*+j\}}\Big\} \\ = & \sum_{j=0}^{\infty} E\Big\{\Big[\lambda - nv\Big(\min_{1 \le k \le K_*+j} \mu_k\Big)\Big]\mathbf{1}_{\{K > K_*+j\}}\Big\} \ge 0. \end{aligned}$$

The second equality in (A.5) follows from

$$E\Big(\min_{1 \le k \le K_*+j} \mu_k - \min_{1 \le k \le K^*+j+1} \mu_k \Big| \min_{1 \le k \le K^*+j} \mu_k = x, K > K^* + j \Big) = v(x).$$

The inequality in (A.5) follows from

$$v\Big(\min_{1 \le k \le K_*+j} \mu_k\Big) \le v(\mu_*) \le v(\zeta_n) = \tfrac{\lambda}{n},$$

as v is monotone increasing. Lemma 3 follows from (A.2)–(A.5). □

Proof of Lemma 4. Let $\widehat{K} = \lfloor n\zeta_n (\log n)^{\beta+2} \rfloor$ for ζ_n satisfying $nv(\zeta_n) = \lambda$. Express $E(M\mu_b) = \sum_{i=1}^5 E(M\mu_b \mathbf{1}_{D_i})$, where

$$\begin{aligned} D_1 &= \{\mu_b \le \tfrac{\zeta_n}{\log n}\}, \\ D_2 &= \{\mu_b > \tfrac{\zeta_n}{\log n}, K > \widehat{K}\}, \\ D_3 &= \{\tfrac{\zeta_n}{\log n} < \mu_b \le \zeta_n (\log n)^{\beta+3}, K \le \widehat{K}\}, \\ D_4 &= \{\mu_b > \zeta_n (\log n)^{\beta+3}, K \le \widehat{K}, M > \tfrac{n}{2}\}, \end{aligned}$$

$$D_5 = \{\mu_b > \zeta_n (\log n)^{\beta+3}, K \leq \widehat{K}, M \leq \tfrac{n}{2}\}.$$

It suffices to show that for all i,

$$E(M\mu_b \mathbf{1}_{D_i}) = o(n^{\frac{\beta}{\beta+1}}). \tag{A.6}$$

Since $\zeta_n \sim Cn^{-\frac{1}{\beta+1}}$ [see (3.3) of the main manuscript], $\frac{M\zeta_n}{\log n} \leq \frac{n\zeta_n}{\log n} = o(n^{\frac{\beta}{\beta+1}})$ and (A.6) holds for $i = 1$.

Let $\widehat{\mu}_b = \min_{k \leq \hat{K}} \mu_k$. Since $M \leq n$, $\mu_b \leq \mu_1$ and $E(\mu_1) \leq \lambda$,

$$\begin{aligned} E(M\mu_b \mathbf{1}_{D_2}) &\leq nE(\mu_1 \mathbf{1}_{D_2}) \qquad (A.7) \\ &= nE(\mu_1 | \mu_1 > \tfrac{\zeta_n}{\log n}) P(D_2) \\ &\leq [\lambda + o(1)] n P(\widehat{\mu}_b > \tfrac{\zeta_n}{\log n}). \end{aligned}$$

By condition (A1), $p(\zeta) \sim \frac{\alpha}{\beta}\zeta^\beta$ as $\zeta \to 0$, hence substituting

$$P(\widehat{\mu}_b > \tfrac{\zeta_n}{\log n}) = [1 - p(\tfrac{\zeta_n}{\log n})]^{\hat{K}} = \exp\{-[1 + o(1)]\widehat{K}\tfrac{\alpha}{\beta}(\tfrac{\zeta_n}{\log n})^\beta] = O(n^{-1})$$

into (A.7) shows (A.6) for $i = 2$.

Let M_j be the number of plays of Π_j to the first positive X_{jt} (hence $M = \sum_{j=1}^K M_j$). It follows from condition (A2) that $E_\mu M_1 = \frac{1}{P_\mu(X_1 > 0)} \leq \frac{1}{a_1 \min(\mu, 1)}$, hence by $\mu_b \leq \zeta_n (\log n)^{\beta+3}$ under D_3,

$$\begin{aligned} E(M\mu_b \mathbf{1}_{D_3}) &\leq E(M_1 \mathbf{1}_{\{\mu_1 > \frac{\zeta_n}{\log n}\}}) \widehat{K} \zeta_n (\log n)^{\beta+3} \qquad (A.8) \\ &\leq \Big(\int_{\frac{\zeta_n}{\log n}}^\infty \tfrac{g(\mu)}{a_1 \min(\mu, 1)} d\mu \Big) n \zeta_n^2 (\log n)^{2\beta+5}. \end{aligned}$$

Substituting

$$\int_{\frac{\zeta_n}{\log n}}^{\infty} \frac{g(\mu)}{\mu} d\mu = \begin{cases} O(1) & \text{if } \beta > 1, \\ O(\log n) & \text{if } \beta = 1, \\ O((\frac{\zeta_n}{\log n})^{\beta-1}) & \text{if } \beta < 1, \end{cases}$$

into (A.8) shows (A.6) for $i = 3$.

If $\mu_j > \zeta_n (\log n)^{\beta+3}$, then by condition (A2), M_j is bounded above by a geometric random variable with mean ν^{-1}, where $\nu = a_1 \zeta_n (\log n)^{\beta+3}$. Hence for $0 < \theta < \log(\frac{1}{1-\nu})$,

$$E(e^{\theta M_j} \mathbf{1}_{\{\mu_j > \zeta_n (\log n)^{\beta+3}\}}) \leq \sum_{h=1}^{\infty} e^{\theta h} \nu (1-\nu)^{h-1} = \tfrac{\nu e^\theta}{1 - e^\theta (1-\nu)},$$

implying that

$$[e^{\frac{\theta n}{2}} P(D_4) \leq] E(e^{\theta M} \mathbf{1}_{D_4}) \leq (\tfrac{\nu e^\theta}{1-e^\theta(1-\nu)})^{\hat{K}}. \tag{A.9}$$

Consider θ such that $e^\theta = 1 + \frac{\nu}{2}$ and check that $e^\theta(1-\nu) \leq 1 - \frac{\nu}{2}$ [$\Rightarrow \theta < \log(\frac{1}{1-\nu})$]. It follows from (A.9) that

$$\begin{aligned} P(D_4) &\leq e^{-\frac{\theta n}{2}} (\tfrac{\nu e^\theta}{\nu/2})^{\hat{K}} = 2^{\hat{K}} e^{\theta(\hat{K} - \frac{n}{2})} \\ &= \exp[\widehat{K} \log 2 + [1 + o(1)] \tfrac{\nu}{2} (\widehat{K} - \tfrac{n}{2})] \\ &= \exp\{-[1+o(1)] \tfrac{n\nu}{4}\} = O(n^{-1}). \end{aligned}$$

Since $M \leq n$, $\mu_b \leq \mu_1$ and $E(\mu_1) \leq \lambda$,

$$E(M \mu_b \mathbf{1}_{D_4}) \leq n E[\mu_1 | \mu_1 > \zeta_n (\log n)^{\beta+3}] P(D_4) \leq n[\lambda + o(1)] P(D_4),$$

and (A.6) holds for $i = 4$.

Under D_5 for n large, since $v(\zeta) \sim \frac{\alpha}{\beta(\beta+1)}\zeta^{\beta+1}$ as $\zeta \to 0$ and $\zeta_n \sim Cn^{-\frac{1}{\beta+1}}$,

$$(n-M)v(\mu_b)[> \tfrac{n}{2}v(\zeta_n(\log n)^{\beta+3})] > \lambda.$$

If we explore one more arm, then the additional exploration cost is not more than λ and reduction in exploitation cost is at least $(n-K)v(\mu_b)$. Hence D_5 is an event of zero probability, in view that we are looking at the optimal solution of Problem B. Therefore (A.6) holds for $i = 5$. □

B Proof of Theorem 1

We preface the proof of Theorem 1 with Lemmas 5–8. The lemmas are proved in Section B.1 and B.2. Consider $X_1, X_2, \dots$ i.i.d. F_μ. Let $S_t = \sum_{u=1}^t X_u$, $\bar{X}_t = \frac{S_t}{t}$ and $\widehat{\sigma}_t^2 = t^{-1}\sum_{u=1}^t (X_u - \bar{X}_t)^2$. Let

$$T_b = \inf\{t : S_t > b_n t \zeta_n\}, \tag{B.1}$$

$$T_c = \inf\{t : S_t > t\zeta_n + c_n \widehat{\sigma}_t \sqrt{t}\}, \tag{B.2}$$

with $b_n \to \infty$ and $c_n \to \infty$ such that $b_n + c_n = o(n^\delta)$ for all $\delta > 0$, and $\zeta_n \sim Cn^{-\frac{1}{\beta+1}}$ for $C = (\frac{\lambda\beta(\beta+1)}{\alpha})^{\frac{1}{\beta+1}}$. Let

$$d_n = n^{-\omega} \text{ for some } 0 < \omega < \tfrac{1}{\beta+1}. \tag{B.3}$$

Lemma 5. *As $n \to \infty$,*

$$\sup_{\mu \geq d_n} [\min(\mu, 1) E_\mu T_b] = O(1), \tag{B.4}$$

$$E_g(T_b \mu \mathbf{1}_{\{\mu \geq d_n\}}) \leq \lambda + o(1). \tag{B.5}$$

Lemma 6. *Let $\epsilon > 0$. As $n \to \infty$,*

$$\sup_{(1+\epsilon)\zeta_n \leq \mu \leq d_n} [\mu E_\mu (T_c \wedge n)] = O(c_n^3 + \log n), \tag{B.6}$$

$$E_g[(T_c \wedge n)\mu \mathbf{1}_{\{(1+\epsilon)\zeta_n \leq \mu \leq d_n\}}] \to 0. \tag{B.7}$$

Lemma 7. *Let $0 < \epsilon < 1$. As $n \to \infty$,*

$$\sup_{\mu \leq (1-\epsilon)\zeta_n} P_\mu(T_b < \infty) \to 0.$$

Lemma 8. *Let $0 < \epsilon < 1$. As $n \to \infty$,*

$$\sup_{\mu \leq (1-\epsilon)\zeta_n} P_\mu(T_c < \infty) \to 0.$$

The number of times an arm is played has distribution bounded above by $T := T_b \wedge T_c$. Lemmas 7 and 8 say that an arm with μ_k less than $(1-\epsilon)\zeta_n$ is unlikely to be rejected, whereas (B.5) and (B.7) say that the regret due to sampling from an arm with μ_k more than $(1+\epsilon)\zeta_n$ is asymptotically bounded by λ. The remaining (B.4) and (B.6) are technical relations used in the proof of Theorem 1.

Proof of Theorem 1. The number of times arm k is played is n_k, and it is distributed as $T_b \wedge T_c \wedge (n - \sum_{\ell=1}^{k-1} n_\ell)$. Let $0 < \epsilon < 1$. We can express

$$R_n - n\zeta_n = z_1 + z_2 + z_3 = z_1 + z_2 - |z_3|, \tag{B.8}$$

where $z_i = E[\sum_{k:\mu_k \in D_i} n_k(\mu_k - \zeta_n)]$ for

$$D_1 = [(1+\epsilon)\zeta_n, \infty), \quad D_2 = ((1-\epsilon)\zeta_n, (1+\epsilon)\zeta_n), \quad D_3 = (0, (1-\epsilon)\zeta_n].$$

It is easy to see that $z_2 \leq \epsilon n \zeta_n$. We shall show that

$$z_1 \leq \tfrac{\lambda + o(1)}{(1-\epsilon)^\beta p(\zeta_n)}, \tag{B.9}$$

$$|z_3| \geq [(\tfrac{1-\epsilon}{1+\epsilon})^\beta + o(1)][n\epsilon\zeta_n + \tfrac{(1-\epsilon)\lambda}{p(\zeta_n)}]. \tag{B.10}$$

We conclude Theorem 1 from (B.8)–(B.10) with $\epsilon \to 0$. □

Proof of (B.9). Since $T = T_b \wedge T_c$, by Lemmas 7 and 8,

$$\begin{aligned} q_n &:= \sup_{\mu \leq (1-\epsilon)\zeta_n} P_\mu(T < \infty) \qquad\qquad \text{(B.11)} \\ &\leq \sup_{\mu \leq (1-\epsilon)\zeta_n} [P_\mu(T_b < \infty) + P_\mu(T_c < \infty)] \to 0. \end{aligned}$$

That is an arm with μ_k less than $(1-\epsilon)\zeta_n$ is rejected with negligible probability for n large. Since the total number of played arms K is bounded above by a geometric random variable with mean $\frac{1}{P_g(T=\infty)}$, by (B.11) and $p(\zeta) \sim \frac{\alpha}{\beta}\zeta^\beta$ as $\zeta \to 0$,

$$EK \leq \tfrac{1}{P_g(T=\infty)} \leq \tfrac{1}{(1-q_n)p((1-\epsilon)\zeta_n)} \sim \tfrac{1}{(1-\epsilon)^\beta p(\zeta_n)}. \tag{B.12}$$

By (B.5) and (B.7),

$$E_g(n_1\mu_1 \mathbf{1}_{\{\mu_1 \geq (1+\epsilon)\zeta_n\}})$$

$$
\begin{aligned}
&= E_g(n_1\mu_1\mathbf{1}_{\{(1+\epsilon)\zeta_n\le\mu_1\le d_n\}}) + E_g(n_1\mu_1\mathbf{1}_{\{\mu_1\ge d_n\}}) \\
&\le E_g[(T_c\wedge n)\mu_1\mathbf{1}_{\{(1+\epsilon)\zeta_n\le\mu_1\le d_n\}}] + E_g(T_b\mu_1\mathbf{1}_{\{\mu_1\ge d_n\}}) \\
&\le \lambda + o(1),
\end{aligned}
$$

and (B.9) follows from (B.12) and $z_1 \le E_g(n_1\mu_1\mathbf{1}_{\{\mu_1\ge(1+\epsilon)\zeta_n\}})EK$. □

Proof of (B.10). Let ℓ be the first arm with mean not more than $(1-\epsilon)\zeta_n$. We have

$$
\begin{aligned}
|z_3| &= E\Big[\sum_{k:\mu_k\in D_3} n_k(\zeta_n-\mu_k)\Big] \qquad \text{(B.13)} \\
&\ge (En_\ell)\{\zeta_n - E_g[\mu|\mu\le(1-\epsilon)\zeta_n]\}.
\end{aligned}
$$

Since $v(\zeta_n)\sim\frac{\lambda}{n}$ and $p(\zeta)\sim\frac{\alpha}{\beta}\zeta^\beta$, $v(\zeta)\sim\frac{\alpha}{\beta(\beta+1)}\zeta^{\beta+1}$ as $\zeta\to 0$,

$$
\begin{aligned}
& \zeta_n - E_g[\mu|\mu\le(1-\epsilon)\zeta_n] \\
= \;& \zeta_n - \{(1-\epsilon)\zeta_n - E_g[(1-\epsilon)\zeta_n-\mu|\mu\le(1-\epsilon)\zeta_n]\} \\
= \;& \zeta_n - [(1-\epsilon)\zeta_n - \tfrac{v((1-\epsilon)\zeta_n)}{p((1-\epsilon)\zeta_n)}] \\
\sim \;& \epsilon\zeta_n + \tfrac{(1-\epsilon)v(\zeta_n)}{p(\zeta_n)} \sim \epsilon\zeta_n + \tfrac{(1-\epsilon)\lambda}{np(\zeta_n)},
\end{aligned}
$$

and (B.10) thus follows from (B.13) and

$$
En_\ell \ge [(\tfrac{1-\epsilon}{1+\epsilon})^\beta + o(1)]n. \qquad \text{(B.14)}
$$

Let j be the first arm with mean not more than $(1+\epsilon)\zeta_n$ and $M=\sum_{i=1}^{j-1} n_i$. We have

$$
En_\ell \ge (1-q_n)E(n-M)P(\ell=j).
$$

Since $q_n \to 0$ and $P(\ell = j) \to (\frac{1-\epsilon}{1+\epsilon})^\beta$, to show (B.14) it suffices to show that $EM = o(n)$.

Indeed by (B.4), (B.6) and $E_\mu n_1 \leq E_\mu(T \wedge n)$,

$$\begin{aligned} &\sup_{\mu \geq (1+\epsilon)\zeta_n} [\min(\mu, 1) E_\mu n_1] \\ \leq\ &\max\Big[\sup_{(1+\epsilon)\zeta_n \leq \mu \leq d_n} \mu E_\mu(T_c \wedge n), \sup_{\mu \geq d_n} \min(\mu, 1) E_\mu T_b\Big] = O(c_n^3 + \log n). \end{aligned}$$

Hence in view that $\frac{1}{p((1+\epsilon)\zeta_n)} = O(n^{\frac{\beta}{\beta+1}})$ and $P_g(\mu_1 > (1+\epsilon)\zeta_n) \to 1$ as $n \to \infty$,

$$\begin{aligned} EM &\leq \tfrac{1}{p((1+\epsilon)\zeta_n)} E_g(n_1 | \mu_1 > (1+\epsilon)\zeta_n) \\ &= O(n^{\frac{\beta}{\beta+1}}) E_g[\tfrac{c_n^3 + \log n}{\min(\mu_1, 1)} \big| \mu_1 > (1+\epsilon)\zeta_n] \\ &= O(n^{\frac{\beta}{\beta+1}}(c_n^3 + \log n)) \int_{(1+\epsilon)\zeta_n}^\infty \tfrac{g(\mu)}{\min(\mu,1)} d\mu \\ &= O(n^{\frac{\beta}{\beta+1}}(c_n^3 + \log n)) \max(n^{\frac{1-\beta}{\beta+1}}, \log n) = o(n). \end{aligned}$$

The first relation in the line above follows from

$$\int_{(1+\epsilon)\zeta_n}^\infty \frac{g(\mu)}{\min(\mu,1)} d\mu = \begin{cases} O(1) & \text{if } \beta > 1, \\ O(\log n) & \text{if } \beta = 1, \\ O(n^{\frac{1-\beta}{\beta+1}}) & \text{if } \beta < 1. \quad \square \end{cases}$$

B.1 Proofs of Lemmas 5–8 for discrete rewards

In the case of discrete rewards, one difficulty is that for μ_k small, there are potentially multiple plays on arm k before a positive X_{kt} is observed.

Condition (A2) is helpful in ensuring that the mean of this positive X_{kt} is not too large.

Recall that for integer-valued rewards we assume in condition (B1) that for $0 < \delta \leq 1$, there exists $\theta_\delta > 0$ such that for $\mu > 0$ and $0 \leq \theta \leq \theta_\delta$,

$$M_\mu(\theta) \leq e^{(1+\delta)\theta\mu}, \quad \text{(B.15)}$$

$$M_\mu(-\theta) \leq e^{-(1-\delta)\theta\mu}. \quad \text{(B.16)}$$

In addition,

$$P_\mu(X > 0) \leq a_2\mu \text{ for some } a_2 > 0, \quad \text{(B.17)}$$

$$E_\mu X^4 = O(\mu) \text{ as } \mu \to 0. \quad \text{(B.18)}$$

Proof of Lemma 5. Recall that

$$T_b = \inf\{t : S_t > b_n t\zeta_n\},$$

and that $d_n = n^{-\omega}$ for some $0 < \omega < \frac{1}{\beta+1}$. We shall show that

$$\sup_{\mu \geq d_n} [\min(\mu, 1) E_\mu T_b] = O(1), \quad \text{(B.19)}$$

$$E_g(T_b \mu \mathbf{1}_{\{\mu \geq d_n\}}) \leq \lambda + o(1). \quad \text{(B.20)}$$

Let $\theta = 2\omega \log n$. Since X_u is integer-valued, it follows from Markov's inequality that

$$P_\mu(S_t \leq b_n t\zeta_n) \leq [e^{\theta b_n \zeta_n} M_\mu(-\theta)]^t \leq \{e^{\theta b_n \zeta_n}[P_\mu(X = 0) + e^{-\theta}]\}^t. \quad \text{(B.21)}$$

By $P_\mu(X>0) \geq a_1 d_n$ for $\mu \geq d_n$ [see (A2)], $\theta b_n \zeta_n = o(d_n)$ [because θ and b_n are both sub-polynomial in n and $\zeta_n = O(n^{-\frac{1}{\beta+1}})$] and (B.21), uniformly over $\mu \geq d_n$,

$$\begin{aligned} E_\mu T_b &= 1 + \sum_{t=1}^{\infty} P_\mu(T_b > t) \qquad \text{(B.22)} \\ &\leq 1 + \sum_{t=1}^{\infty} P_\mu(S_t \leq b_n t \zeta_n) \\ &\leq \{1 - e^{\theta b_n \zeta_n}[P_\mu(X=0) + e^{-\theta}]\}^{-1} \\ &= \{1 - [1+o(d_n)][P_\mu(X=0) + d_n^2]\}^{-1} \\ &= [P_\mu(X>0) + o(d_n)]^{-1} \sim [P_\mu(X>0)]^{-1}. \end{aligned}$$

The term inside $\{\cdot\}$ in (B.21) is not more than $[1+o(d_n)](1-a_1 d_n + d_n^2) < 1$ for n large and this gives us the second inequality in (B.22). We conclude (B.19) from (B.22) and (A2). By (B.22),

$$\begin{aligned} E_g[T_b \mu \mathbf{1}_{\{\mu \geq d_n\}}] &= \int_{d_n}^{\infty} E_\mu(T_b) \mu g(\mu) d\mu \\ &\leq [1+o(1)] \int_{d_n}^{\infty} \tfrac{E_\mu(X)}{P_\mu(X>0)} g(\mu) d\mu \\ &= [1+o(1)] \int_{d_n}^{\infty} E_\mu(X|X>0) g(\mu) d\mu \to \lambda, \end{aligned}$$

hence (B.20) holds. $\square$

Proof of Lemma 6. Recall that $T_c = \inf\{t : S_t > t\zeta_n + c_n \widehat{\sigma}_t \sqrt{t}\}$ and let $\epsilon > 0$. We want to show that

$$\begin{aligned} \sup_{(1+\epsilon)\zeta_n \leq \mu \leq d_n} \mu E_\mu(T_c \wedge n) &= O(c_n^3 + \log n), \quad &\text{(B.23)} \\ E_g[(T_c \wedge n)\mu \mathbf{1}_{\{(1+\epsilon)\zeta_n \leq \mu \leq d_n\}}] &\to 0. \quad &\text{(B.24)} \end{aligned}$$

We first show that there exists $\kappa > 0$ such that as $n \to \infty$,

$$\sup_{\mu \leq d_n} \Big[\mu \sum_{t=1}^n P_\mu(\widehat{\sigma}_t^2 \geq \kappa\mu) \Big] = O(\log n). \quad \text{(B.25)}$$

Since X is non-negative integer-valued, $X^2 \leq X^4$. Indeed by (B.18), there exists $\kappa > 0$ such that $\rho_\mu := E_\mu X^2 \leq \frac{\kappa\mu}{2}$ for $\mu \leq d_n$ and n large, therefore by (B.18) again and Chebyshev's inequality,

$$\begin{aligned} P_\mu(\widehat{\sigma}_t^2 \geq \kappa\mu) &\leq P_\mu\Big(\sum_{u=1}^t X_u^2 \geq t\kappa\mu \Big) \\ &\leq P_\mu\Big(\sum_{u=1}^t (X_u^2 - \rho_\mu) \geq \tfrac{t\kappa\mu}{2} \Big) \\ &\leq \tfrac{t\mathrm{Var}_\mu(X^2)}{(t\kappa\mu/2)^2} = O((t\mu)^{-1}), \end{aligned}$$

and (B.25) holds.

By (B.25), uniformly over $(1+\epsilon)\zeta_n \leq \mu \leq d_n$,

$$\begin{aligned} E_\mu(T_c \wedge n) &= 1 + \sum_{t=1}^{n-1} P_\mu(T_c > t) \quad \text{(B.26)} \\ &\leq 1 + \sum_{t=1}^{n-1} P_\mu(S_t \leq t\zeta_n + c_n \widehat{\sigma}_t \sqrt{t}) \\ &\leq 1 + \sum_{t=1}^{n-1} P_\mu(S_t \leq t\zeta_n + c_n \sqrt{\kappa\mu t}) + O(\tfrac{\log n}{\mu}). \end{aligned}$$

Let $0 < \delta < \frac{1}{2}$ to be further specified. Uniformly over $t \geq c_n^3 \mu^{-1}$, $\mu t/(c_n \sqrt{\kappa \mu t}) \to \infty$ and therefore by (B.16), $\mu \geq (1+\epsilon)\zeta_n$ and Markov's inequality, for n large,

$$\begin{aligned} P_\mu(S_t \leq t\zeta_n + c_n\sqrt{\kappa\mu t}) &\leq P_\mu(S_t \leq t(\zeta_n + \delta\mu)) \\ &\leq e^{\theta_\delta t(\zeta_n+\delta\mu)} M_\mu^t(-\theta_\delta) \\ &\leq e^{t\theta_\delta[\zeta_n - (1-2\delta)\mu]} \leq e^{-\eta t\theta_\delta \mu}, \end{aligned} \tag{B.27}$$

where $\eta = 1 - 2\delta - \frac{1}{1+\epsilon} > 0$ (for δ chosen small). Since $1 - e^{-\eta\theta_\delta\mu} \sim \eta\theta_\delta\mu$ as $\mu \to 0$,

$$\sum_{t=1}^{n-1} e^{-\eta t\theta_\delta\mu} \leq c_n^3\mu^{-1} + \sum_{t \geq c_n^3\mu^{-1}} e^{-\eta t\theta_\delta\mu} = O(c_n^3\mu^{-1}), \tag{B.28}$$

and substituting (B.27) into (B.26) gives us (B.23). By (B.23),

$$\begin{aligned} E_g[(T_c \wedge n)\mu \mathbf{1}_{\{(1+\epsilon)\zeta_n \leq \mu \leq d_n\}}] &= P_g((1+\epsilon)\zeta_n \leq \mu \leq d_n) O(c_n^3 + \log n) \\ &= O(d_n^\beta(c_n^3 + \log n)), \end{aligned}$$

and (B.24) holds since c_n is sub-polynomial in n. $\square$

Proof of Lemma 7. We want to show that

$$P_\mu(S_t > tb_n\zeta_n \text{ for some } t \geq 1) \to 0 \tag{B.29}$$

uniformly over $\mu \leq (1-\epsilon)\zeta_n$.

By (B.17) and Bonferroni's inequality,

$$\begin{aligned} & P_\mu(S_t > tb_n\zeta_n \text{ for some } t \leq \tfrac{1}{\sqrt{b_n}\zeta_n}) & \text{(B.30)} \\ \leq \quad & P_\mu(X_t > 0 \text{ for some } t \leq \tfrac{1}{\sqrt{b_n}\zeta_n}) \leq \tfrac{a_2\mu}{\sqrt{b_n}\zeta_n} \rightarrow 0. \end{aligned}$$

By (B.15) and Markov's inequality, for n large,

$$\begin{aligned} & P_\mu(S_t > tb_n\zeta_n \text{ for some } t > \tfrac{1}{\sqrt{b_n}\zeta_n}) & \text{(B.31)} \\ \leq \quad & \sup_{t > \frac{1}{\sqrt{b_n}\zeta_n}} [e^{-\theta_1 b_n\zeta_n} M_\mu(\theta_1)]^t \leq e^{-\theta_1(b_n\zeta_n - 2\mu)/(\zeta_n\sqrt{b_n})} \rightarrow 0. \end{aligned}$$

To see the first inequality of (B.31), let f_μ be the density of X_1 with respect to some σ-finite measure, and let $E_\mu^{\theta_1}(P_\mu^{\theta_1})$ denote expectation (probability) with respect to density

$$f_\mu^{\theta_1}(x) := [M_\mu(\theta_1)]^{-1} e^{\theta_1 x} f_\mu(x).$$

Let $T = \inf\{t > \frac{1}{\sqrt{b_n}\zeta_n} : S_t > tb_n\zeta_n\}$. It follows from Markov's inequality that

$$\begin{aligned} P_\mu(T = t) \quad &= \quad M_\mu^t(\theta_1) E_\mu^{\theta_1}(e^{-\theta_1 S_t} \mathbf{1}_{\{T=t\}}) & \text{(B.32)} \\ &\leq \quad [e^{-\theta_1 b_n\zeta_n} M_\mu(\theta_1)]^t P_\mu^{\theta_1}(T = t), \end{aligned}$$

and the first inequality of (B.31) follows from summing (B.32) over $t > \frac{1}{\sqrt{b_n}\zeta_n}$. □

Proof of Lemma 8. We want to show that

$$P_\mu(S_t > t\zeta_n + c_n\widehat{\sigma}_t\sqrt{t} \text{ for some } t \geq 1) \rightarrow 0 \qquad \text{(B.33)}$$

uniformly over $\mu \le (1-\epsilon)\zeta_n$.

By (B.17) and Bonferroni's inequality,

$$\begin{aligned} & P_\mu(S_t > t\zeta_n + c_n \widehat{\sigma}_t \sqrt{t} \text{ for some } t \le \tfrac{1}{c_n\mu}) \\ \le \ & P_\mu(X_t > 0 \text{ for some } t \le \tfrac{1}{c_n\mu}) \le \tfrac{a_2}{c_n} \to 0. \end{aligned} \tag{B.34}$$

Moreover

$$P_\mu(S_t > t\zeta_n + c_n \widehat{\sigma}_t \sqrt{t} \text{ for some } t > \tfrac{1}{c_n\mu}) \le \text{(I)} + \text{(II)}, \tag{B.35}$$

$$\begin{aligned} \text{where (I)} &= P_\mu(S_t > t\zeta_n + c_n(\mu t/2)^{\frac{1}{2}} \text{ for some } t > \tfrac{1}{c_n\mu}), \\ \text{(II)} &= P_\mu(\widehat{\sigma}_t^2 \le \tfrac{\mu}{2} \text{ and } S_t \ge t\zeta_n \text{ for some } t > \tfrac{1}{c_n\mu}). \end{aligned}$$

By (B.34) and (B.35), to show (B.33), it suffices to show that (I)$\to 0$ and (II)$\to 0$.

Let $0 < \delta \le 1$ be such that $1+\delta < (1-\epsilon)^{-1}$. Hence $\mu \le (1-\epsilon)\zeta_n$ implies $\zeta_n \ge (1+\delta)\mu$. It follows from (B.15) and Markov's inequality [see (B.31) and (B.32)] that

$$\begin{aligned} \text{(I)} &\le \sup_{t > \frac{1}{c_n\mu}} [e^{-\theta_\delta[t\zeta_n + c_n(\mu t/2)^{\frac{1}{2}}]} M_\mu^t(\theta_\delta)] \\ &\le \exp\{-\theta_\delta[\zeta_n - (1+\delta)\mu]/(c_n\mu) - \theta_\delta(c_n/2)^{\frac{1}{2}}\} \\ &\le \exp\{-\theta_\delta(c_n/2)^{\frac{1}{2}}\} \to 0. \end{aligned}$$

Since $X_u^2 \ge X_u$, the inequality $S_t \ge t\zeta_n(\ge t\mu)$ implies $\sum_{u=1}^t X_u^2 \ge t\mu$, and this, together with $\widehat{\sigma}_t^2 \le \frac{\mu}{2}$ implies that $\bar{X}_t^2 \ge \frac{\mu}{2}$. Hence by (B.15) and

Markov's inequality argument, for n large,

$$\begin{aligned}
\text{(II)} &\leq P_\mu\Big(\bar{X}_t \geq \sqrt{\tfrac{\mu}{2}} \text{ for some } t > \tfrac{1}{c_n\mu}\Big) \\
&\leq \sup_{t > \frac{1}{c_n\mu}} [e^{-\theta_1\sqrt{\mu/2}} M_\mu(\theta_1)]^t \\
&\leq \exp\{-\theta_1[\sqrt{\tfrac{\mu}{2}} - 2\mu]/(c_n\mu)\} \\
&\leq \exp\Big\{-\theta_1\Big[\tfrac{1}{c_n\sqrt{2(1-\epsilon)\zeta_n}} - \tfrac{2}{c_n}\Big]\Big\} \rightarrow 0. \quad \square
\end{aligned}$$

B.2 Proofs of Lemmas 5–8 for continuous rewards

In the case of continuous rewards, the proofs are simpler due to positive X_{kt}, in particular $\lambda = E_g\mu$. Recall that for continuous rewards, we assume in condition (B2) that

$$\sup_{\mu>0} P_\mu(X \leq \gamma\mu) \rightarrow 0 \text{ as } \gamma \rightarrow 0. \tag{B.36}$$

Moreover (B.18) holds and for $0 < \delta \leq 1$, there exists $\tau_\delta > 0$ such that for $0 < \theta\mu \leq \tau_\delta$,

$$M_\mu(\theta) \leq e^{(1+\delta)\theta\mu}, \tag{B.37}$$

$$M_\mu(-\theta) \leq e^{-(1-\delta)\theta\mu}. \tag{B.38}$$

In addition for each $t \geq 1$, there exists $\xi_t > 0$ such that

$$\sup_{\mu \leq \xi_t} P_\mu(\widehat{\sigma}_t^2 \leq \gamma\mu^2) \rightarrow 0 \text{ as } \gamma \rightarrow 0, \tag{B.39}$$

where $\widehat{\sigma}_t^2 = t^{-1}\sum_{u=1}^t (X_u - \bar{X}_t)^2$ and $\bar{X}_t = t^{-1}\sum_{u=1}^t X_u$ for i.i.d. $X_u \overset{d}{\sim} F_\mu$.

Proof of Lemma 5. To show (B.4) and (B.5), it suffices to show that

$$\sup_{\mu \geq d_n} E_\mu T_b \leq 1 + o(1). \tag{B.40}$$

Let $\theta > 0$ to be further specified. By Markov's inequality,

$$P_\mu(S_t \leq b_n t \zeta_n) \leq [e^{\theta b_n \zeta_n} M_\mu(-\theta)]^t.$$

Moreover, for any $\gamma > 0$,

$$M_\mu(-\theta) \leq P_\mu(X \leq \gamma\mu) + e^{-\gamma\theta\mu},$$

hence

$$\begin{aligned} E_\mu T_b &\leq 1 + \sum_{t=1}^\infty P_\mu(S_t \leq b_n t \zeta_n) \qquad \text{(B.41)} \\ &\leq \{1 - e^{\theta b_n \zeta_n}[P_\mu(X \leq \gamma\mu) + e^{-\gamma\theta\mu}]\}^{-1}. \end{aligned}$$

Let $\gamma = \frac{1}{\log n}$ and $\theta = n^\eta$ for some $\omega < \eta < \frac{1}{\beta+1}$. By (B.36), b_n is sub-polynomial in n, and $d_n = n^{-\omega}$, for $\mu \geq d_n$,

$$e^{\theta b_n \zeta_n} \to 1, \quad e^{-\gamma\theta\mu} \to 0, \quad P_\mu(X \leq \gamma\mu) \to 0,$$

and (B.40) follows from (B.41). □

Proof of Lemma 6. By (B.18), for μ small,

$$\begin{aligned} \rho_\mu := E_\mu X^2 &= E_\mu(X^2 \mathbf{1}_{\{X<1\}}) + E_\mu(X^2 \mathbf{1}_{\{X \geq 1\}}) \\ &\leq E_\mu X + E_\mu X^4 = O(\mu). \end{aligned}$$

Hence to show (B.6) and (B.7), we proceed as in the proof of Lemma 6 for discrete rewards, applying (B.38) in place of (B.16), with any fixed $\theta > 0$ in place of θ_δ in (B.27) and (B.28). □

PROOF OF LEMMA 7. It follows from (B.37) with $\theta = \frac{\tau_1}{\mu}$ and Markov's inequality [see (B.31) and (B.32)] that for n large,

$$\begin{aligned} & P_\mu(S_t > t b_n \zeta_n \text{ for some } t \geq 1) \\ \leq \; & \sup_{t \geq 1} [e^{-\theta b_n \zeta_n} M_\mu(\theta)]^t \leq e^{-\theta(b_n \zeta_n - 2\mu)} \to 0. \quad \square \end{aligned}$$

PROOF OF LEMMA 8. Let $\eta > 0$ and choose $\delta > 0$ such that $(1+\delta)(1-\epsilon) < 1$. It follows from (B.37) with $\theta = \frac{\tau_\delta}{\mu}$ and Markov's inequality that for u large,

$$\begin{aligned} & P_\mu(S_t \geq t\zeta_n + c_n \widehat{\sigma}_t \sqrt{t} \text{ for some } t > u) \qquad \text{(B.42)} \\ \leq \; & P_\mu(S_t \geq t\zeta_n \text{ for some } t > u) \\ \leq \; & \sup_{t > u} [e^{-\theta \zeta_n} M_\mu(\theta)]^t \leq e^{-u\theta[\zeta_n - (1+\delta)\mu]} \leq e^{-u\tau_\delta[(1-\epsilon)^{-1} - (1+\delta)]} \leq \eta. \end{aligned}$$

By (B.39), we can select $\gamma > 0$ such that for n large (so that $\mu \leq (1-\epsilon)\zeta_n \leq \min_{1 \leq t \leq u} \xi_t$),

$$\sum_{t=1}^{u} P_\mu(\widehat{\sigma}_t^2 \leq \gamma \mu^2) \leq \eta. \qquad \text{(B.43)}$$

Let $\theta = \frac{\tau_1}{\mu}$. By (B.37), (B.43) and Bonferroni's inequality,

$$
\begin{aligned}
& P_\mu(S_t > t\zeta + c_n \widehat{\sigma}_t \sqrt{t} \text{ for some } t \leq u) \qquad \text{(B.44)} \\
\leq \; & P_\mu(S_t \geq c_n \widehat{\sigma}_t \sqrt{t} \text{ for some } t \leq u) \\
\leq \; & \eta + \sum_{t=1}^{u} P_\mu(S_t \geq c_n \mu \sqrt{\gamma t}) \\
\leq \; & \eta + \sum_{t=1}^{u} e^{-\theta c_n \mu \sqrt{\gamma t}} M_\mu^t(\theta) \\
\leq \; & \eta + \sum_{t=1}^{u} e^{-\tau_1 (c_n \sqrt{\gamma t} - 2t)} \rightarrow \eta.
\end{aligned}
$$

Lemma 8 follows from (B.42) and (B.44) since η can be chosen arbitrarily small. □

C Proof of Theorem 2

The idealized algorithm in the beginning of Section 5.1 of the main manuscript captures the essence of how CBT behaves. We reveal μ_k when the first positive loss of arm k appears. If $\mu_k > \zeta$ [with optimality when $\zeta = \zeta_n$, see (5.3) of the main manuscript] then we stop sampling from arm k and sample the next arm $k+1$. If $\mu_k \leq \zeta$ then we exploit arm k a further n times before stopping.

In the idealized version of empirical CBT, we reveal μ_k when the first positive loss of arm k appears and stop exploring the arm. Since the first positive loss of an arm has mean λ, the sum of losses after k arms have been played has mean $k\lambda$. When $\min_{1 \leq i \leq k} \mu_i \leq \widehat{\zeta}_k (:= \frac{k\lambda}{n})$ we stop exploring, and exploit the best arm a further n times. More specifically:

Idealized empirical CBT

1. For $k = 1, 2, \ldots$: Draw n_k rewards from arm k, where

$$n_k = \inf\{t \geq 1 : X_{kt} > 0\}.$$

2. Stop when there are K arms, where

$$K = \inf\left\{k \geq 1 : \min_{1 \leq i \leq k} \mu_i \leq \tfrac{k\lambda}{n}\right\}.$$

3. Draw n additional rewards from arm j satisfying $\mu_j = \min_{1 \leq k \leq K} \mu_k$.

The regret of this algorithm is $R'_n = \lambda EK + nE(\min_{1 \leq k \leq K} \mu_k)$.

Theorem 2. *The idealized empirical CBT has regret* $R'_n \sim C I_\beta n^{\frac{\beta}{\beta+1}}$, *where* $C = (\frac{\lambda\beta(\beta+1)}{\alpha})^{\frac{1}{\beta+1}}$ *and* $I_\beta = (\frac{1}{\beta+1})^{\frac{1}{\beta+1}}(2 - \frac{1}{(\beta+1)^2})\Gamma(2 - \frac{\beta}{\beta+1})$.

Proof. We stop exploring after K arms, where

$$K = \inf\{k : \min_{1 \leq j \leq k} \mu_j \leq \widehat{\zeta}_k\}, \quad \widehat{\zeta}_k = \tfrac{k\lambda}{n}. \tag{C.1}$$

Let

$$D_k^1 = \{\widehat{\zeta}_k - \tfrac{\lambda}{n} < \min_{1 \le j \le k-1} \mu_j \le \widehat{\zeta}_k\}, \quad D_k^2 = \{\min_{1 \le j \le k-1} \mu_j > \widehat{\zeta}_k, \mu_k \le \widehat{\zeta}_k\}.$$

We check that D_k^1, D_k^2 are disjoint, and that $D_k^1 \cup D_k^2 = \{K = k\}$. Essentially D_k^1 is the event that $K = k$ and the best arm is not k, and D_k^2 the event that $K = k$ and the best arm is k.

For any fixed $k \in \mathbf{Z}^+$,

$$\begin{aligned} P(D_k^1) &= [1 - p(\widehat{\zeta}_k - \tfrac{\lambda}{n})]^{k-1} - [1 - p(\widehat{\zeta}_k)]^{k-1} \quad &(C.2) \\ &= \{1 - p(\widehat{\zeta}_k) + [1 + o(1)]\tfrac{\lambda}{n} g(\widehat{\zeta}_k)\}^{k-1} - [1 - p(\widehat{\zeta}_k)]^{k-1} \\ &\sim \{[1 - p(\widehat{\zeta}_k)]^{k-1}\} \tfrac{k\lambda}{n} g(\widehat{\zeta}_k) \\ &\sim \exp(-\tfrac{\alpha\lambda^\beta}{\beta} k^{\beta+1} n^{-\beta}) \alpha \lambda^\beta k^\beta n^{-\beta}. \end{aligned}$$

Moreover

$$E(R_n' | D_k^1) \sim k\lambda + n(\tfrac{k\lambda}{n}) = 2k\lambda. \quad (C.3)$$

Likewise,

$$\begin{aligned} P(D_k^2) &= \{[1 - p(\widehat{\zeta}_k)]^{k-1}\} p(\widehat{\zeta}_k) \quad &(C.4) \\ &\sim \exp(-\tfrac{\alpha\lambda^\beta}{\beta} k^{\beta+1} n^{-\beta}) \tfrac{\alpha\lambda^\beta}{\beta} k^\beta n^{-\beta}, \\ E(R_n' | D_k^2) &= k\lambda + nE(\mu | \mu \le \widehat{\zeta}_k) \quad &(C.5) \\ &= 2k\lambda - \tfrac{nv(\hat{\zeta}_k)}{p(\hat{\zeta}_k)} \sim (2 - \tfrac{1}{\beta+1}) k\lambda. \end{aligned}$$

Combining (C.2)–(C.5) gives us

$$\begin{aligned} R'_n &= \sum_{k=1}^{\infty}[E(R'_C|D_k^1)P(D_k^1)+E(R'_C|D_k^2)P(D_k^2)] \\ &\sim \sum_{k=1}^{\infty}\exp(-\tfrac{\alpha\lambda^\beta}{\beta}k^{\beta+1}n^{-\beta})(\tfrac{\alpha\lambda^{\beta+1}}{\beta}k^{\beta+1}n^{-\beta})(2\beta+2-\tfrac{1}{\beta+1}), \end{aligned} \tag{C.6}$$

It follows from (C.6) and a change of variables $x=\frac{\alpha\lambda^\beta}{\beta}k^{\beta+1}n^{-\beta}$ that

$$\begin{aligned} R'_n &\sim (2\beta+2-\tfrac{1}{\beta+1})\int_0^\infty \exp(-\tfrac{\alpha\lambda^\beta}{\beta}k^{\beta+1}n^{-\beta})(\tfrac{\alpha\lambda^{\beta+1}}{\beta}k^{\beta+1}n^{-\beta})dk \\ &= (2\beta+2-\tfrac{1}{\beta+1})\int_0^\infty \tfrac{1}{\beta+1}(\tfrac{\lambda\beta}{\alpha})^{\frac{1}{\beta+1}}n^{\frac{\beta}{\beta+1}}\exp(-x)x^{\frac{1}{\beta+1}}dx \\ &= (2-\tfrac{1}{(\beta+1)^2})(\tfrac{\lambda\beta}{\alpha})^{\frac{1}{\beta+1}}\Gamma(2-\tfrac{\beta}{\beta+1})n^{\frac{\beta}{\beta+1}}, \end{aligned}$$

and Theorem 2 holds. □

D Verifications of (A2), (B1) and (B2)

The optimality of CBT in Theorem 1 holds under the assumption:

(A2) There exists $a_1>0$ such that $P_\mu(X>0)\geq a_1\min(\mu,1)$ for all μ.

In addition, optimality for discrete rewards requires assumption (B1) [i.e. (B.15)–(B.18)] and optimality for continuous rewards requires assumption (B2) [i.e. (B.36)–(B.39)]. In the following examples we check that these assumptions hold in specific discrete and continuous distributions.

Example 3. Let F_μ be a distribution with support on $0,\dots,I$ for some positive integer $I > 1$ and having mean μ. Let $p_i = P_\mu(X = i)$. We check that $P_\mu(X > 0) \geq \mu I^{-1}$ and therefore (A2) holds with $a_1 = I^{-1}$.

Let $\theta_\delta > 0$ be such that

$$e^{i\theta} - 1 \leq i\theta(1+\delta) \text{ and } e^{-i\theta} - 1 \leq -i\theta(1-\delta) \text{ for } 0 \leq i\theta \leq I\theta_\delta. \qquad \text{(D.1)}$$

By (D.1) for $0 \leq \theta \leq \theta_\delta$,

$$\begin{aligned} M_\mu(\theta) &= \sum_{i=0}^{I} p_i e^{i\theta} \leq 1 + (1+\delta)\mu\theta, \\ M_\mu(-\theta) &= \sum_{i=0}^{I} p_i e^{-i\theta} \leq 1 - (1-\delta)\mu\theta, \end{aligned}$$

and (B.15), (B.16) follow from $1 + x \leq e^x$. Moreover (B.17) holds with $a_2 = 1$ and (B.18) holds because $E_\mu X^4 = \sum_{i=0}^{I} p_i i^4 \leq I^3\mu$.

Example 4. If $X \overset{d}{\sim}$ Poisson(μ), then

$$M_\mu(\theta) = \exp[\mu(e^\theta - 1)],$$

and both (B.15) and (B.16) hold for $\theta_\delta > 0$ satisfying

$$e^{\theta_\delta} - 1 \leq \theta_\delta(1+\delta) \text{ and } e^{-\theta_\delta} - 1 \leq -\theta_\delta(1-\delta).$$

Since $P_\mu(X > 0) = 1 - e^{-\mu}$, (A2) holds with $a_1 = 1 - e^{-1}$, and (B.17) holds with $a_2 = 1$. The relation in (B.18) holds because

$$E_\mu X^4 = \sum_{k=1}^{\infty} \tfrac{k^4\mu^k e^{-\mu}}{k!} = \mu e^{-\mu} + e^{-\mu} O\Big(\sum_{k=2}^{\infty} \mu^k\Big).$$

EXAMPLE 5. Let Z be a continuous non-negative random variable with mean 1, and with $Ee^{\tau_0 Z} < \infty$ for some $\tau_0 > 0$. Consider X distributed as μZ. Condition (A2) holds with $a_1 = 1$. We conclude (B.36) from

$$\sup_{\mu>0} P_\mu(X \leq \gamma\mu) = P(Z \leq \gamma) \to 0 \text{ as } \gamma \to 0.$$

Let $0 < \delta \leq 1$. Since $\lim_{\tau\to 0} \tau^{-1} \log Ee^{\tau Z} = EZ = 1$, there exists $\tau_\delta > 0$ such that for $0 < \tau \leq \tau_\delta$,

$$Ee^{\tau Z} \leq e^{(1+\delta)\tau} \text{ and } Ee^{-\tau Z} \leq e^{-(1-\delta)\tau}. \tag{D.2}$$

Since $M_\mu(\theta) = E_\mu e^{\theta X} = Ee^{\theta\mu Z}$ and $M_\mu(-\theta) = Ee^{-\theta\mu Z}$, we conclude (B.37) and (B.38) from (D.2) with $\tau = \theta\mu$. We conclude (B.18) from $E_\mu X^4 = \mu^4 EZ^4$, and (B.39), for arbitrary $\xi_t > 0$, from

$$P_\mu(\widehat{\sigma}_t^2 \leq \gamma\mu^2) = P(\widehat{\sigma}_{tZ}^2 \leq \gamma) \to 0 \text{ as } \gamma \to 0,$$

where $\widehat{\sigma}_{tZ}^2 = t^{-1} \sum_{u=1}^t (Z_u - \bar{Z}_t)^2$, for i.i.d. Z and Z_u.