A. Notations

$[n]$	set $\{1, \cdots, n\}$ for any $n \in \mathbb{N}$
$[n]^{(m)}$	set of all m-permutations of $[n]$, i.e., all ordered m-subset of $[n]$ for any $m \leq n$
$[L]$	ground set of size L
$w(i)$	click probability of item $i \in [L]$
K	size of recommendation list/pulled arm
$[L]^{(K)}$	set of all K-permutations of $[L]$
S_t	recommendation list/pulled arm at time step t
i_i^t	i-th pulled item at time step t
$W_t(i)$	an r.v. that reflects whether the user clicks at item i at time step t
S_t^π	chosen arm at time step t by algorithm π
$\boldsymbol{O}_t^\pi$	stochastic outcome by pulling S_t^π at time step t by algorithm π
$\tilde{k}_t$	feedback from the user at time step t
$\bar{w}(i)$	equals to $1 - w(i)$, i.e., one minus the click probability
$\boldsymbol{w}$	the vector of click probabilities $w(i)$'s
w^*	maximum click probability
w'	minimum click probability
ϵ	tolerance parameter
K'_ϵ	number of ϵ-optimal items (abbreviated as K' for brevity)
S^*	optimal arm in $[K]^{(K)}$
π	deterministic and non-anticipatory algorithm
$\hat{S}^\pi$	output of algorithm π
$\mathcal{T}^\pi$	time complexity of algorithm π
ϕ^π	final recommendation rule of algorithm π
$\mathcal{F}_t$	observation history
δ	risk parameter (failure probability)
$\mathbb{T}^*(\boldsymbol{w}, \epsilon, \delta, K)$	optimal expected time complexity (abbreviated as $\mathbb{T}^*$)
D_t	survival set in Algorithm 1
A_t	accept set in Algorithm 1
R_t	reject set in Algorithm 1
$T_t(i)$	number of observations of item i by time step t
$\hat{w}_t(i)$	empirical mean of item i at time step t in Algorithm 1
k_t	number of ϵ-optimal items to identify at time step t in Algorithm 1
$C_t(i, \delta)$	confidence radius of item i at time step t in Algorithm 1
$U_t(i, \delta)$	upper confidence bound (UCB) of item i at time step t in Algorithm 1
$L_t(i, \delta)$	lower confidence bound (LCB) of item i at time step t in Algorithm 1

j^*	empirically $(K+1)$-th optimal item at time step t in Algorithm 1
j'	empirically K-th optimal item at time step t in Algorithm 1
$\rho(\delta)$	parameter used to define the confidence radius $C_t(i,\delta)$
$c_1, c_2, \ldots$	finite and positive universal constants whose values may vary from line to line
Δ_i	gap between the click probabilities
$\bar{\Delta}_i$	variation of Δ_i incurred by the tolerance parameter ϵ
$\bar{T}_{i,\delta}$	number of observations required to identify item i with fixed δ and ϵ
$\sigma(i)$	descending order of $\bar{\Delta}_i$ of ground items
$\hat{k}_t$	number of surviving items pulled at time step t during the proceeding of Algorithm 1
$X_{\hat{k}_t;t}$	number of observations of surviving items at time step t
$\mu(k,w)$	lower bound on $\mathbb{E}X_{k;t}$ (abbreviated as μ_k)
$v(k,w)$	upper bound on $\mathbb{E}X_{k;t}^2$ (abbreviated as v_k)
K_1, K_2, M_k	parameters used in Theorem 4.1
N_1, N_2, N_3	constituents in the upper bound established in Theorem 4.1
π_1	represents Algorithm 1 for brevity
$\tilde{\mu}(k,w)$	upper bound on $\mathbb{E}X_{k;t}$ (abbreviated as $\tilde{\mu}_k$)
$\mathcal{E}(i,\delta)$	"nice event" in the analysis of Algorithm 1
$w^{(\ell)}(\ell)$	click probability of item ℓ under instance ℓ $(1 \le \ell \le L)$
$S_t^{\pi,\ell}$	chosen arm at time step t by algorithm π under instance ℓ $(0 \le \ell \le L)$
$\boldsymbol{O}_t^{\pi,\ell}$	stochastic outcome by pulling S_t^{π} at time step t by algorithm π under instance ℓ $(0 \le \ell \le L)$

B. Useful definitions and theorems

Here are some basic facts from the literature that we will use:

Theorem B.1 (Azuma's Inequality for Martingales with Subgaussian Tails, implied by Shamir (2011))**.** *Let $\{(D_t, \mathcal{F}_t)\}_{t=1}^{\infty}$ be a martingale difference sequence, and suppose that for any $\lambda \le 0$, we have $\mathbb{E}[e^{\lambda D_t}|\mathcal{F}_{t-1}] \le e^{\lambda^2\omega^2/2}$ almost surely. Then for all $\omega \ge 0$,*

$$\Pr\left[\sum_{t=1}^{n} D_t \le -\omega\right] \le \exp\left(-\frac{\omega^2}{2\sum_{t=1}^{n} v_t^2}\right).$$

Theorem B.2 (Non-asymptotic law of the iterated logarithm (Jamieson et al., 2014; Jun et al., 2016))**.** *Let $X_1, X_2, \ldots$ be i.i.d. zero-mean sub-Gaussian random variables with scale $\sigma > 0$; i.e. $\mathbb{E}e^{\lambda X_i} \le e^{\frac{\lambda^2\sigma^2}{2}}$. Let $\omega \in (0, \sqrt{1/6})$. Then,*

$$\mathbb{P}\left(\forall\tau \ge 1, \left|\sum_{s=1}^{\tau} X_s\right| \le 4\sigma\sqrt{\frac{\log\left(\log_2(2\tau)/\omega\right)}{\tau}}\right) \ge 1 - 6\omega^2.$$

C. Influence of ϵ

In general, a larger ϵ indicates a smaller time complexity. Here are two explanations. (i) When ϵ grows, K'_ϵ, the number of ϵ-optimal items also grows. Then it should be easier to identify an ϵ-optimal arm. (ii) If ϵ is sufficiently large such that $K'_\epsilon \ge 2K-1$, then there are at least K items left in the survival set D_t before the algorithm stops. Otherwise, when $|D_t| < K$, the agent pulls $|D_t| < K$ surviving items at some steps and this results in a wastage in the number of time steps.

Proposition C.1. *Assume $K' \geq 2K-1$. With probability at least $1-\delta$, Algorithm 1 outputs an ϵ-optimal arm after at most $(c_1 N_1' + c_2 N_2')$ steps where*

$$N_1' = \frac{2v_K^2}{\mu_K^2}\log\left(\frac{2}{\delta}\right) = O\left(\frac{v_K^2}{\mu_K^2}\log\left(\frac{2}{\delta}\right)\right),$$

$$N_2' = \frac{2}{\mu_K}\left[\sum_{i=1}^{L-K'+K-1}\bar{T}_{\sigma(i)} + (K'-K+1)\bar{T}_{\sigma(L-K'+K)} + (K'-K)\right]$$

$$= O\left(\frac{1}{\mu_K}\left\{\sum_{i=1}^{L-K'+K-1}\bar{\Delta}_{\sigma(i)}^{-2}\log\left[\frac{L}{\delta}\log\left(\frac{L}{\delta\bar{\Delta}_{\sigma(i)}^2}\right)\right] + (K'-K+1)\bar{\Delta}_{\sigma(L-K'+K)}^{-2}\log\left[\frac{L}{\delta}\log\left(\frac{L}{\delta\bar{\Delta}_{\sigma(L-K'+K)}^2}\right)\right]\right\}\right).$$

D. Proofs of main results

In this Section, we provide proofs of Proposition 4.2, Corollary 4.3, Propositions 4.4, 4.6, 4.7, Corollary 4.9, Theorem 5.1, 5.4, Lemmas 5.2 – 5.11, and complete the proof of Theorems 4.1, 4.8, Proposition C.1 in this order.

D.1. Proof of Proposition 4.2

Proposition 4.2. *Assume $0 < w' < w^* \leq 1$. We have*

$$N_1 \leq \begin{cases} 4K\log\left(\frac{4K}{\delta}\right) & 0 < w^* \leq 1/K, \\ \frac{8Kw^{*2}}{w'^2}\log\left(\frac{8Kw^{*2}}{\delta w'^2}\right) & 1/K < w^* \leq 1. \end{cases}$$

Proof. According to Lemma 5.2 and Theorem 5.4,

$$\mu_k \geq \min\{k/2, 1/(2w^*)\}, \quad v_k = \min\{k, \sqrt{2}/w'\}.$$

We upper bound v_k/μ_k and k/μ_k in two cases:

(i): $0 < w^* \leq 1/K$: $0 < w' < w^* \leq 1/K$, $v_k = k$, $\mu_k \geq k/2$.

$$\frac{v_k}{\mu_k} \leq \frac{k}{k/2} = 2 \;\Rightarrow\; \sum_{k=1}^{K-K_2}\frac{v_{K-k+1}^2}{\mu_{K-k+1}^2}\log\left(\frac{1}{\delta}\sum_{j=1}^{K-K_2}\frac{v_{K-j+1}^2}{\mu_{K-j+1}^2}\right) \leq 4K\log\left(\frac{4K}{\delta}\right).$$

(ii): $1/K < w^* \leq 1$: $v_k \leq \sqrt{2}/w'$, $\mu_k \geq 1/(2w^*)$.

$$\frac{v_k}{\mu_k} \leq \frac{\sqrt{2}/w'}{1/(2w^*)} = \frac{2\sqrt{2}w^*}{w'} \;\Rightarrow\; \sum_{k=1}^{K-K_2}\frac{v_{K-k+1}^2}{\mu_{K-k+1}^2}\log\left(\frac{1}{\delta}\sum_{j=1}^{K-K_2}\frac{v_{K-j+1}^2}{\mu_{K-j+1}^2}\right) \leq \frac{8Kw^{*2}}{w'^2}\log\left(\frac{8Kw^{*2}}{\delta w'^2}\right).$$

□

D.2. Proof of Corollary 4.3

Corollary 4.3. *(i) If all $w(i)$'s are at most $1/K$, with probability at least $1-\delta$, Algorithm 1 outputs S^* after at most*

$$O\left(\frac{1}{K}\sum_{i=1}^{L-K}\bar{T}_{\sigma(i)} + \bar{T}_{\sigma(L-1)}\right)$$

$$= O\left(\frac{1}{K}\sum_{i=1}^{L-K}\Delta_{\sigma(i)}^{-2}\log\left[\frac{L}{\delta}\log\left(\frac{L}{\delta\Delta_{\sigma(i)}^2}\right)\right] + \Delta_{\sigma(L-1)}^{-2}\log\left[\frac{L}{\delta}\log\left(\frac{L}{\delta\Delta_{\sigma(L-1)}^2}\right)\right]\right)$$

steps; (ii) if all $w(i)$'s are at least $1/2$, with probability at least $1-\delta$, Algorithm 1 outputs S^* after at most

$$O\Big(\sum_{i=1}^{L-1}\bar{T}_{\sigma(i)}\Big)=O\Big(\sum_{i=1}^{L-1}\Delta_{\sigma(i)}^{-2}\log\Big[\frac{L}{\delta}\log\Big(\frac{L}{\delta\Delta_{\sigma(i)}^2}\Big)\Big]\Big)$$

steps.

Proof. According to Lemma 5.2 and Theorem 5.4,

$$\mu_k\geq\min\{k/2,1/(2w^*)\},\quad v_k=\min\{k,\sqrt{2}/w'\}.$$

We first upper bound v_k/μ_k and k/μ_k in two cases:

(i): $0<w^*\leq 1/K$: $0<w'<w^*\leq 1/K$, $v_k=k$, $\mu_k\geq k/2$.

$$\frac{v_k}{\mu_k}\leq\frac{k}{k/2}=2,\quad \frac{k}{\mu_k}\leq\frac{k}{k/2}=2.$$

(ii): $1/K<w^*\leq 1$: $v_k\leq\sqrt{2}/w'$, $\mu_k\geq 1/(2w^*)$.

$$\frac{v_k}{\mu_k}\leq\frac{\sqrt{2}/w'}{1/(2w^*)}=\frac{2\sqrt{2}w^*}{w'},\quad \frac{k}{\mu_k}\leq\frac{k}{1/(2w^*)}=2kw^*.$$

Next, we separate the upper bound in Theorem 4.1 into two parts and bound them separately:

$$(A)=\sum_{k=1}^{K-1}\frac{v_{K-k+1}^2}{\mu_{K-k+1}^2}\log\left(\frac{1}{\delta}\sum_{j=1}^{K-1}\frac{v_{K-j+1}^2}{\mu_{K-j+1}^2}\right),$$

$$(B)=\frac{1}{\mu_K}\sum_{i=1}^{L-K}\bar{T}_{\sigma(i)}+\sum_{k=1}^{K-K_1-1}\bar{T}_{\sigma(L-K+k)}\left(\frac{K-k+1}{\mu_{K-k+1}}-\frac{K-k}{\mu_{K-k}}\right)+\left(\frac{K_1+1}{\mu_{K_1+1}}-2\right)\bar{T}_{\sigma(L-K_1)}+2\bar{T}_{\sigma(L-K_2)}$$

with $K_1=\min\{\lfloor 1/w^*\rfloor,K-1\}$, $K_2=1$.

<u>Case 1</u>: All click probabilities $w(i)$ are at most $1/K$. $1/w^*\geq K$ and $v_k/\mu_k\leq 2$, $K_1=K-1$.

$$(A)\leq 4(K-1)\log\left(\frac{4(K-1)}{\delta}\right)=O\left(K\log\left(\frac{K}{\delta}\right)\right),$$

$$(B)\leq\frac{2}{K}\sum_{i=1}^{L-K}\bar{T}_{\sigma(i)}+\left(\frac{2}{\mu_2}-2\right)\bar{T}_{\sigma(L-K+1)}+2\bar{T}_{\sigma(L-1)}=O\left(\frac{1}{K}\sum_{i=1}^{L-K}\bar{T}_{\sigma(i)}+\bar{T}_{\sigma(L-1)}\right).$$

<u>Case 2</u>: All click probabilities $w(i)$ are at least $1/2$. $1/w^*\leq K$ implies $v_k/\mu_k\leq 4\sqrt{2}$, $K_1\geq 1$.

$$\begin{aligned}(A)&\leq 32(K-1)\log\left(\frac{32(K-1)}{\delta}\right)=O\left(K\log\left(\frac{K}{\delta}\right)\right),\\(B)&\leq 2\sum_{i=1}^{L-K}\bar{T}_{\sigma(i)}+\sum_{k=1}^{K-2}\bar{T}_{\sigma(L-K+k)}\left(\frac{K-k+1}{\mu_{K-k+1}}-\frac{K-k}{\mu_{K-k}}\right)+\frac{2}{\mu_2}\bar{T}_{\sigma(L-1)}\\&=2\sum_{i=1}^{L-K}\bar{T}_{\sigma(i)}+\sum_{k=0}^{K-3}\frac{K-k}{\mu_{K-k}}\bar{T}_{\sigma(L-K+k+1)}-\sum_{k=1}^{K-2}\frac{K-k}{\mu_{K-k}}\bar{T}_{\sigma(L-K+k)}+\frac{2}{\mu_2}\bar{T}_{\sigma(L-1)}\\&=2\sum_{i=1}^{L-K}\bar{T}_{\sigma(i)}+\frac{K}{\mu_K}\bar{T}_{\sigma(L-K+1)}+\sum_{k=0}^{K-2}\frac{K-k}{\mu_{K-k}}\left[\bar{T}_{\sigma(L-K+k+1)}-\bar{T}_{\sigma(L-K+k)}\right]\\&\leq 2\sum_{i=1}^{L-K}\bar{T}_{\sigma(i)}+\frac{K}{\mu_K}\bar{T}_{\sigma(L-K+1)}+\sum_{k=0}^{K-2}2(K-k)\cdot\left[\bar{T}_{\sigma(L-K+k+1)}-\bar{T}_{\sigma(L-K+k)}\right]\end{aligned}$$

$$\leq 2\sum_{i=1}^{L-K}\bar{T}_{\sigma(i)}+\sum_{k=1}^{K-2}\bar{T}_{\sigma(L-K+k)}\left[2(K-k+1)-2(K-k)\right]+4\bar{T}_{\sigma(L-1)}$$
$$=O\left(\sum_{i=1}^{L-1}\bar{T}_{\sigma(i)}\right).$$

Recall that when $\epsilon=0$,

$$\bar{T}_i=O\left(\Delta_i^{-2}\log\left(\frac{L}{\delta}\log\left(\frac{L}{\delta\Delta_i^2}\right)\right)\right)$$

for all $i\in[L]$. Altogether, we complete the proof.

□

D.3. Proof of Proposition 4.4

Proposition 4.4. *(i) If all $w(i)$'s are at most $1/K$,*

$$\mathbb{E}\mathcal{T}^{\pi_1}\leq c_1\log\left(\frac{1}{\delta}\right)\cdot\left\{\frac{1}{K}\sum_{i=1}^{L-K}\Delta_{\sigma(i)}^{-2}\log\left[L\log\left(\frac{L}{\Delta_{\sigma(i)}^2}\right)\right]\right.$$
$$\left.+\Delta_{\sigma(L-1)}^{-2}\log\left[L\log\left(\frac{L}{\Delta_{\sigma(L-1)}^2}\right)\right]\right\};$$

(ii) if all $w(i)$'s are at least $1/2$,

$$\mathbb{E}\mathcal{T}^{\pi_1}\leq c_2\sum_{i=1}^{L-1}\Delta_{\sigma(i)}^{-2}\log\left[L\log\left(\frac{L}{\Delta_{\sigma(i)}^2}\right)\right]\log\left(\frac{1}{\delta}\right).$$

Proof. **(i) Consider all click probabilities $w(i)$'s are at most $1/K$.** For any $0<\delta'\leq\delta$, revisit the proof and result of Theorem 4.1. First, Lemma 5.7 implies that $\mathbb{P}\left(\bigcap_{i=1}^{L}\mathcal{E}(\epsilon,\delta')\right)\geq 1-\delta'/2$. Assume $\bigcap_{i=1}^{L}\mathcal{E}(\epsilon,\delta')$ holds from now on. Secondly, Lemma 5.8 indicates that Algorithm 1 can correctly identify item i after $\bar{T}_{i,\delta}$ observations. Thirdly, similar to the discussion in Section 5.2, we set $\sum_{k=1}^{K-1}\delta_k\leq\delta'/2$. Additionally applying the analysis in Proposition 4.2 and Corollary 4.3, with probability at least $1-\delta'$, we can bound the time complexity by

$$\sum_{k=1}^{K-1}\frac{v_{K-k+1}^2}{\mu_{K-k+1}^2}\log\left(\frac{1}{\delta'}\sum_{j=1}^{K-K_2}\frac{v_{K-j+1}^2}{\mu_{K-j+1}^2}\right)+K+\frac{8}{K}\sum_{i=1}^{L-K}\bar{T}_{\sigma(i)}+8\bar{T}_{\sigma(L-1)}$$
$$\leq 4K\log\left(\frac{4K}{\delta'}\right)+K+\frac{8}{K}\sum_{i=1}^{L-K}\frac{217}{\Delta_{\sigma(i)}^2}\log\left[\frac{24L}{\delta'}\log_2\left(\frac{648\times 12L}{\delta'\Delta_{\sigma(i)}^2}\right)\right]+\frac{7136}{\Delta_{\sigma(L-1)}^2}\log\left[\frac{24L}{\delta'}\log_2\left(\frac{648\times 12L}{\delta'\Delta_{\sigma(L-1)}^2}\right)\right]$$
$$\leq 5K\log\left(\frac{4K}{\delta'}\right)+\frac{1}{K}\sum_{i=1}^{L-K}\frac{10600}{\Delta_{\sigma(i)}^2}\log\left[\frac{L}{\delta'}\log_2\left(\frac{L}{\delta'\Delta_{\sigma(i)}^2}\right)\right]+\frac{10600}{\Delta_{\sigma(L-1)}^2}\log\left[\frac{L}{\delta'}\log_2\left(\frac{L}{\delta'\Delta_{\sigma(L-1)}^2}\right)\right]$$
$$\leq 10610\log\left(\frac{1}{\delta'^2}\right)\cdot\left\{\frac{1}{K}\sum_{i=1}^{L-K}\Delta_{\sigma(i)}^{-2}\log\left[L\log\left(\frac{L}{\Delta_{\sigma(i)}^2}\right)\right]+\Delta_{\sigma(L-1)}^{-2}\log\left[L\log\left(\frac{L}{\Delta_{\sigma(L-1)}^2}\right)\right]\right\}.$$

In short, set

$$A=21220\left\{\frac{1}{K}\sum_{i=1}^{L-K}\Delta_{\sigma(i)}^{-2}\log\left[L\log\left(\frac{L}{\Delta_{\sigma(i)}^2}\right)\right]+\Delta_{\sigma(L-1)}^{-2}\log\left[L\log\left(\frac{L}{\Delta_{\sigma(L-1)}^2}\right)\right]\right\},$$

then $\Pr(\mathcal{T}>-A\log\delta')<\delta'$ for any $0\leq\delta'\leq\delta$.

Meanwhile, Tonelli's Theorem implies that

$$\mathbb{E}\mathcal{T} = \mathbb{E}\left[\int_0^{\mathcal{T}} 1 \,\mathrm{d}x\right] = \mathbb{E}\left[\int_0^{+\infty} \mathbb{I}(\mathcal{T} > x) \,\mathrm{d}x\right] = \int_0^{+\infty} \mathbb{E}\left[\mathbb{I}(\mathcal{T} > x)\right] \,\mathrm{d}x = \int_0^{+\infty} \mathbb{P}(\mathcal{T} > x) \,\mathrm{d}x.$$

Since $x = -A\log\delta$ implies $\delta = e^{-x/A}$ and $\int_0^{+\infty} e^{-x/A} \,\mathrm{d}x = Ae^{-x/A}|_{x=+\infty}^{0} = A$,

$$\begin{aligned}
\mathbb{E}\mathcal{T} &\le \int_0^{-A\log\delta} 1 \,\mathrm{d}x + \int_{-A\log\delta}^{+\infty} \mathbb{P}(\mathcal{T} > x) \,\mathrm{d}x \le -A\log\delta + \int_0^{+\infty} \mathbb{P}(\mathcal{T} > x) \,\mathrm{d}x = -A\log\delta + A \\
&\le 42440 \log\left(\frac{1}{\delta}\right) \cdot \left\{\frac{1}{K}\sum_{i=1}^{L-K} \Delta_{\sigma(i)}^{-2} \log\left[L\log\left(\frac{L}{\Delta_{\sigma(i)}^2}\right)\right] + \Delta_{\sigma(L-1)}^{-2} \log\left[L\log\left(\frac{L}{\Delta_{\sigma(L-1)}^2}\right)\right]\right\}.
\end{aligned}$$

(ii) Consider all click probabilities $w(i)$'s are at least $1/2$. The analysis is similar to that in Case (i). With the analysis in Theorem 4.1 and results in Proposition 4.2 and Corollary 4.3, for any $0 < \delta' \le \delta$, with probability at least $1 - \delta'$, the time complexity is at most

$$\begin{aligned}
&\sum_{k=1}^{K-1} \frac{v_{K-k+1}^2}{\mu_{K-k+1}^2} \log\left(\frac{1}{\delta'}\sum_{j=1}^{K-1} \frac{v_{K-j+1}^2}{\mu_{K-j+1}^2}\right) + K + 8\sum_{i=1}^{L-1} \bar{T}_{\sigma(i)} \\
&\le \frac{8Kw^{*2}}{w'^2} \log\left(\frac{8Kw^{*2}}{\delta w'^2}\right) + K + \sum_{i=1}^{L-1} \frac{8 \times 217}{\Delta_{\sigma(i)}^2} \log\left[\frac{24L}{\delta'} \log_2\left(\frac{648 \times 12L}{\delta'\Delta_{\sigma(i)}^2}\right)\right] \\
&\le 32K \log\left(\frac{32K}{\delta}\right) + K + \sum_{i=1}^{L-1} \frac{10598}{\Delta_{\sigma(i)}^2} \log\left[\frac{L}{\delta'} \log_2\left(\frac{L}{\delta'\Delta_{\sigma(i)}^2}\right)\right] \\
&\le 10630 \sum_{i=1}^{L-1} \Delta_{\sigma(i)}^{-2} \log\left[L\log\left(\frac{L}{\Delta_{\sigma(i)}^2}\right)\right] \log\left(\frac{1}{\delta'^2}\right).
\end{aligned}$$

Set $A = 21260\sum_{i=1}^{L-1} \Delta_{\sigma(i)}^{-2} \log\left[L\log\left(\frac{L}{\Delta_{\sigma(i)}^2}\right)\right]$, then $\Pr(\mathcal{T} > -A\log\delta') < \delta'$ for any $0 \le \delta' \le \delta$. Lastly,

$$\mathbb{E}\mathcal{T} \le 42520 \sum_{i=1}^{L-1} \Delta_{\sigma(i)}^{-2} \log\left[L\log\left(\frac{L}{\Delta_{\sigma(i)}^2}\right)\right] \log\left(\frac{1}{\delta}\right).$$

□

D.4. Proof of Proposition 4.6

Proposition 4.6. *Under Assumption 4.5, (i) if $0 < w^* \le 1/K$, with probability at least $1 - \delta$, Algorithm 1 outputs S^* after at most*

$$O\left(\frac{L}{K(w^* - w')^2} \log\left[\frac{L}{\delta}\log\left(\frac{L}{\delta(w^* - w')^2}\right)\right]\right)$$

steps; (ii) if $1/K < w^* \le 1$, with probability at least $1 - \delta$, Algorithm 1 outputs S^* after at most

$$O\left(\frac{w^*L}{(w^* - w')^2}\log\left[\frac{L}{\delta}\log\left(\frac{L}{\delta(w^* - w')^2}\right)\right] + \frac{w^{*2}}{w'^2}\log\left(\frac{1}{\delta}\right)\right)$$

steps.

Proof. **We first remind ourselves how the algorithm proceeds.** In this instance, $\epsilon = 0$ yields $K^* = 1$. For any item $i \in [L]$, $\bar{\Delta}_i = \Delta_i = w^* - w'$. And according to Lemma 5.8, item i will be correctly classified with high probability after $\bar{T}_i$ observations where $\rho = \delta/(12L)$,

$$\bar{T}_{i,\delta} = \bar{T}_{(w,\delta)} = \bar{T}_{(w)} = 1 + \left\lfloor \frac{216}{(w^* - w')^2} \log\left(\frac{2}{\rho}\log_2\left(\frac{648}{\rho(w^* - w')^2}\right)\right)\right\rfloor$$

$$= O\left(\frac{1}{(w^*-w')^2}\log\left[\frac{L}{\delta}\log_2\left(\frac{L}{\delta(w^*-w')^2}\right)\right]\right).$$

This implies that each item requires the same number of observations to be correctly identified. According to the design of algorithm, $T_t(j)-1 \le T_t(i) \le T_t(j)+1$ for any remaining items $i \neq j$. Therefore, the worst case is as follows:

- the agent observes one item for $\bar{T}_{(w)}$ times and the others for $\bar{T}_{(w)}-1$ times after t' steps, and identifies one item per step for the subsequent $L-2$ steps.

Therefore, we now turn to upper bounding the number of steps required to eliminate an item for the first time. According to Lemma 5.6, we set $\delta_0=\delta/2$, $k=K$, $n=t'$, $\omega'_K=-\sqrt{-2t'v_K^2\log(\delta/2)}$. Then the total number of observations during t' steps should be larger than $t'\mu_K+\omega'_K$ with probability at least $1-\delta/2$. The number of observations can be upper bounded as follows:

$$t'\mu_K+\omega'_K \le \bar{T}_{(w)}+(L-1)[\bar{T}_{(w)}-1] = L\cdot\bar{T}_{(w)}-L+1.$$

Then with Lemma 5.7 and its ensuing discussion in Section 5.2, with probability at least $1-\delta$, the time complexity is upper bounded by

$$\frac{2(L\cdot\bar{T}_{(w)}-L+1)}{\mu_K}+\frac{2v_K^2}{\mu_K^2}\log\left(\frac{2}{\delta}\right).$$

Next, we consider how the values of w^* and w' affect the bound. According to Lemma 5.2 and Theorem 5.4,

$$\mu_k \ge \min\{k/2, 1/(2w^*)\}, \quad v_k=\min\{k,\sqrt{2}/w'\}.$$

We discuss two cases separately:

Case 1: $0<w^*\le 1/K$: $0<w'<w^*\le 1/K$, $v_K=K$, $\mu_K\ge K/2$. The upper bound becomes:

$$\frac{4(L\cdot\bar{T}_{(w)}-L+1)}{K}+\frac{8K^2}{K^2}\log\left(\frac{2}{\delta}\right)=O\left(\frac{L}{K(w^*-w')^2}\log\left[\frac{L}{\delta}\log_2\left(\frac{L}{\delta(w^*-w')^2}\right)\right]\right).$$

Case 2: $1/K<w^*\le 1$: $v_k\le\sqrt{2}/w'$, $\mu_k\ge 1/(2w^*)$. The bound becomes

$$\frac{2(L\cdot\bar{T}_{(w)}-L+1)}{1/(2w^*)}+\frac{4/w'^2}{1/(2w^*)^2}\log\left(\frac{2}{\delta}\right)=O\left(\frac{w^*L}{(w^*-w')^2}\log\left[\frac{L}{\delta}\log_2\left(\frac{L}{\delta(w^*-w')^2}\right)\right]+\left(\frac{w^*}{w'}\right)^2\cdot\log\left(\frac{1}{\delta}\right)\right).$$

□

D.5. Proof of Proposition 4.7

Proposition 4.7. *Under Assumption 4.5, (i) if* $0<w^*\le 1/K$,

$$\mathbb{E}\mathcal{T}^{\pi_1}\le\frac{c_1L}{K(w^*-w')^2}\log\left[L\log\left(\frac{L}{(w^*-w')^2}\right)\right]\log\left(\frac{1}{\delta}\right);$$

(ii) if $w'\ge 1/2$ *or* $w^*/w'\le 2$,

$$\mathbb{E}\mathcal{T}^{\pi_1}\le\frac{c_2w^*L}{(w^*-w')^2}\log\left[L\log\left(\frac{L}{(w^*-w')^2}\right)\right]\log\left(\frac{1}{\delta}\right).$$

Proof. For any $0<\delta'\le\delta$, we revisit the proof of Proposition 4.6. Firstly, Lemma 5.7 implies that $\mathbb{P}\left(\bigcap_{i=1}^L\mathcal{E}(\epsilon,\delta')\right)\ge 1-\delta'/2$. Assume $\bigcap_{i=1}^L\mathcal{E}(\epsilon,\delta')$ holds from now on. Secondly, Lemma 5.8 implies that the agent can identify any item correctly after

$$\bar{T}_{(w,\delta')}=1+\left\lfloor\frac{216}{(w^*-w')^2}\log\left(\frac{24L}{\delta'}\log_2\left(\frac{648*24L}{\delta'(w^*-w')^2}\right)\right)\right\rfloor\le\frac{1320}{(w^*-w')^2}\log\left[\frac{L}{\delta'}\log\left(\frac{L}{\delta'(w^*-w')^2}\right)\right]$$

observations. Then with analysis similar to Appendix D.4, we can upper bound the time complexity of Algorithm 1 with probability $1-\delta'$.

Case 1: $0 < w^* \leq 1/K$: with probability at least $1-\delta'$, the time complexity is upper bounded by

$$\begin{aligned}&\frac{4L\cdot\bar{T}_{(w,\delta')}}{K}+8\log\left(\frac{2}{\delta'}\right)\leq\frac{5288L}{K(w^*-w')^2}\log\left[\frac{L}{\delta'}\log\left(\frac{L}{\delta'(w^*-w')^2}\right)\right]\\&\leq\frac{10576L}{K(w^*-w')^2}\log\left[L\log\left(\frac{L}{(w^*-w')^2}\right)\right]\log(\frac{1}{\delta'}):=-A\log\delta'.\end{aligned}$$

Then for any $0<\delta'\leq\delta$, $\Pr(\mathcal{T}>-A\log\delta')<\delta'$. Meanwhile, Tonelli's Theorem implies that

$$\mathbb{E}\mathcal{T}=\mathbb{E}\left[\int_0^{\mathcal{T}}1\,\mathrm{d}x\right]=\mathbb{E}\left[\int_0^{+\infty}\mathbb{I}(\mathcal{T}>x)\,\mathrm{d}x\right]=\int_0^{+\infty}\mathbb{E}\left[\mathbb{I}(\mathcal{T}>x)\right]\,\mathrm{d}x=\int_0^{+\infty}\mathbb{P}(\mathcal{T}>x)\,\mathrm{d}x.$$

Since $x=-A\log\delta$ implies $\delta=e^{-x/A}$ and $\int_0^{+\infty}e^{-x/A}\,\mathrm{d}x=Ae^{-x/A}|_{x=+\infty}^0=A$,

$$\begin{aligned}&\mathbb{E}\mathcal{T}\leq\int_0^{-A\log\delta}1\,\mathrm{d}x+\int_{-A\log\delta}^{+\infty}\mathbb{P}(\mathcal{T}>x)\,\mathrm{d}x\leq-A\log\delta+\int_0^{+\infty}\mathbb{P}(\mathcal{T}>x)\,\mathrm{d}x=-A\log\delta+A\\&\leq\frac{21152L}{K(w^*-w')^2}\log\left[L\log\left(\frac{L}{(w^*-w')^2}\right)\right]\log(\frac{1}{\delta}).\end{aligned}$$

Case 2: $1/2\leq w'<1$ or $w^*/w'\leq 2$: with a similar analysis, for any $0<\delta\leq\delta'$, with

$$\begin{aligned}&4w^*L\bar{T}_{(w,\delta')}+16\left(\frac{w^*}{w'}\right)^2\log\left(\frac{2}{\delta'}\right)\leq\frac{5280w^*L}{(w^*-w')^2}\log\left[\frac{L}{\delta'}\log\left(\frac{L}{\delta'(w^*-w')^2}\right)\right]+64\log\left(\frac{1}{\delta'}\right)\\&\leq\frac{10624w^*L}{(w^*-w')^2}\log\left[L\log\left(\frac{L}{(w^*-w')^2}\right)\right]\log\left(\frac{1}{\delta'}\right):=-A\log\delta'\end{aligned}$$

$\Pr(\mathcal{T}>-A\log\delta')<\delta'$. Lastly,

$$\mathbb{E}\mathcal{T}\leq\frac{21248w^*L}{(w^*-w')^2}\log\left[L\log\left(\frac{L}{(w^*-w')^2}\right)\right]\log\left(\frac{1}{\delta}\right).$$

□

D.6. Proof of Corollary 4.9

Corollary 4.9. *Under Assumption 4.5, we have*

$$\begin{aligned}\mathbb{T}^*&\geq\frac{\mathrm{KL}(1-\delta,\delta)}{\tilde{\mu}_K}\cdot\left[\frac{K}{\mathrm{KL}(w^*,w')}+\frac{L-K}{\mathrm{KL}(w',w^*)}\right]\\&=\Omega\Big(\min\{w',1-w^*\}\cdot\frac{Lw'}{(w^*-w')^2}\log\left[\frac{1}{\delta}\right]\Big).\end{aligned}$$

where $\tilde{\mu}_K=[1-(1-w')^K]/w'\leq 1/w'$.

Proof. First, by setting $w(i)=w^*$ for all $1\leq i\leq K$ and $w(j)=w'$ for all $k<j\leq L$, the result in Theorem 4.8 becomes

$$\frac{\mathrm{KL}(1-\delta,\delta)}{\tilde{\mu_K}}\cdot\left[\frac{K}{\mathrm{KL}(w^*,w')}+\frac{L-K}{\mathrm{KL}(w',w^*)}\right]\geq\frac{\log(1/2.4\delta)}{\tilde{\mu_K}}\cdot\left[\frac{K}{\mathrm{KL}(w^*,w')}+\frac{L-K}{\mathrm{KL}(w',w^*)}\right].$$

Next, according to Pinsker's and reverse Pinsker's inequality for any two distributions P and Q defined in the same finite space X we have

$$\delta(P,Q)^2\leq\frac{1}{2}\mathrm{KL}(P,Q)\leq\frac{1}{\alpha_Q}\delta(P,Q)^2$$

where $\delta(P,Q) = \sup\{|P(A) - Q(A)| \big| A \subset X\}$ and $\alpha_Q = \min_{x \in X: Q(x) > 0} Q(x)$. In our case, set $\delta(w^*, w') = (w^* - w')^2$ and $\alpha = \min\{w', w^*, 1 - w^*, 1 - w'\} = \min\{w', 1 - w^*\}$, we have

$$(w^* - w')^2 \leq \frac{1}{2}\mathrm{KL}(w^*, w') \leq \frac{1}{\alpha}(w^* - w')^2 = \frac{1}{\min\{w', 1 - w^*\}}(w^* - w')^2,$$
$$(w^* - w')^2 \leq \frac{1}{2}\mathrm{KL}(w', w^*) \leq \frac{1}{\alpha}(w^* - w')^2 = \frac{1}{\min\{w', 1 - w^*\}}(w^* - w')^2.$$

Further since $\tilde{\mu}_K \leq 1/w'$ as stated by Lemma 5.2, the lower bound becomes

$$\Omega\left(\min\{w', 1 - w^*\} \cdot \frac{Lw'}{(w^* - w')^2} \log\left[\frac{1}{\delta}\right]\right).$$

□

D.7. Proof of Theorem 5.1

Theorem 5.1. *Consider a set of items with weights $\boldsymbol{u} = (u_1, \ldots, u_k)$ such that $u_1 \geq \ldots \geq u_k$, and let $\mu_k(\boldsymbol{u}, I)$ be the expected number of observations when items are placed with order I. Let $I_{dec} = (1, \ldots, k)$, $I_{inc} = (k, \ldots, 1)$, and I be any order, then*
(i) boundedness: $\mu_k(\boldsymbol{u}, I_{dec}) \leq \mu_k(\boldsymbol{u}, I) \leq \mu_k(\boldsymbol{u}, I_{inc})$;
(ii) monotonicity: let $\boldsymbol{v} = (v_1, \ldots, v_k)$ be another vector of weights, then $\mu_k(\boldsymbol{u}, I) \geq \mu_k(\boldsymbol{v}, I)$ if $u_i \leq v_i$ for all $i \in [k]$.

Proof. (i) Consider any ordered set $I = (i_1^I, \ldots, i_k^I)$. To show $\mu_k(\boldsymbol{u}, I_{\text{dec}}) \leq \mu_k(\boldsymbol{u}, I) \leq \mu_k(\boldsymbol{u}, I_{\text{inc}})$, it is sufficient to show the following:

$(*)$: If there exists $1 \leq m < k$ such that $u_{i_m^I} < u_{i_{m+1}^I}$, we can change their positions to get I' and have $\mu_k(\boldsymbol{u}, I) > \mu_k(\boldsymbol{u}, I')$.

The proof of $(*)$ is as follows:

if $1 \leq m < k - 1$,

$$\begin{aligned}\mu_k(\boldsymbol{u}, I) - \mu_k(\boldsymbol{u}, I') &= m \cdot \prod_{j=1}^{m-1}(1 - u_{i_j^I})(u_{i_m^I} - u_{i_{m+1}^I}) + (m+1) \cdot \prod_{j=1}^{m-1}(1 - u_{i_j^I})[u_{i_{m+1}^I}(1 - u_{i_m^I}) - u_{i_m^I}(1 - u_{i_{m+1}^I})] \\ &= -\prod_{j=1}^{m-1}(1 - u_{i_j^I})(u_{i_m^I} - u_{i_{m+1}^I}) > 0;\end{aligned}$$

if $m = k - 1$,

$$\begin{aligned}\mu_k(\boldsymbol{u}, I) - \mu_k(\boldsymbol{u}, I') &= m \cdot \prod_{j=1}^{m-1}(1 - u_{i_j^I})(u_{i_m^I} - u_{i_{m+1}^I}) + (m+1) \cdot \prod_{j=1}^{m-1}(1 - u_{i_j^I})[(1 - u_{i_m^I}) - (1 - u_{i_{m+1}^I})] \\ &= -\prod_{j=1}^{m-1}(1 - u_{i_j^I})(u_{i_m^I} - u_{i_{m+1}^I}) > 0.\end{aligned}$$

(ii) To show the monotonicity, it is sufficient to show the following:

$(\#)$: Set two sets of click probabilities u, v such that $v_{i_m^I} > u_{i_m^I}$ for some $1 \leq m \leq k$ and $v_{i_j^I} = u_{i_j^I}$ for $j \neq m$. Then we have $\mu_k(\boldsymbol{u}, I) \geq \mu_k(\boldsymbol{v}, I)$.

Here is the proof of $(\#)$. If $m = k$, then obviously we have $\mu_k(\boldsymbol{u}, I) = \mu_k(\boldsymbol{v}, I)$. If $1 \leq m < k$, we exchange positions of the m-th and $(m+1)$-th item to get a new ordered set I_1, then

$$\mu_k(\boldsymbol{u}, I) - \mu_k(\boldsymbol{u}, I_1) = -\prod_{j=1}^{m-1}(1 - u_{i_j^I})(u_{i_m^I} - u_{i_{m+1}^I}), \;\; \mu_k(\boldsymbol{v}, I) - \mu_k(\boldsymbol{v}, I_1) = -\prod_{j=1}^{m-1}(1 - u_{i_j^I})(v_{i_m^I} - u_{i_{m+1}^I}).$$

Hence

$$\begin{aligned}\mu_k(\boldsymbol{u}, I) - \mu_k(\boldsymbol{v}, I) &= [\mu_k(\boldsymbol{u}, I) - \mu_k(\boldsymbol{u}, I_1)] - [\mu_k(\boldsymbol{v}, I) - \mu_k(\boldsymbol{v}, I_1)] + \mu_k(\boldsymbol{u}, I_1) - \mu_k(\boldsymbol{v}, I_1) \\ &= -\prod_{j=1}^{m-1}(1 - u_{i_j^I})(u_{i_m^I} - v_{i_m^I}) + \mu_k(\boldsymbol{u}, I_1) - \mu_k(\boldsymbol{v}, I_1) \\ &> \mu_k(\boldsymbol{u}, I_1) - \mu_k(\boldsymbol{v}, I_1).\end{aligned}$$

If $m+1 < k$, note that the only difference between $(\boldsymbol{u}, I_1)$ and $(\boldsymbol{v}, I_1)$ now lies in the click probability of the $(m+1)$-th item. In detail,

$$u_{i_{m+1}^{I_1}} = u_{i_m^I},\ v_{i_{m+1}^{I_1}} = v_{i_m^I} \text{ and } u_{i_j^{I_1}} = v_{i_j^{I_1}},\ \forall j \neq m+1.$$

We exchange positions of the $(m+1)$-th and $(m+2)$-th item in I_1 to get a new ordered set I_2. Similarly we have

$$\begin{aligned}\mu_k(\boldsymbol{u}, I_1) - \mu_k(\boldsymbol{v}, I_1) &= [\mu_k(\boldsymbol{u}, I_1) - \mu_k(\boldsymbol{u}, I_2)] - [\mu_k(\boldsymbol{v}, I_1) - \mu_k(\boldsymbol{v}, I_2)] + \mu_k(\boldsymbol{u}, I_2) - \mu_k(\boldsymbol{v}, I_2) \\ &= -\prod_{j=1}^{m}(1 - u_{i_j^{I_1}})(u_{i_{m+1}^{I_1}} - v_{i_{m+1}^{I_1}}) + \mu_k(\boldsymbol{u}, I_2) - \mu_k(\boldsymbol{v}, I_2) \\ &= -\prod_{j=1}^{m}(1 - u_{i_j^{I_1}})(u_{i_m^I} - v_{i_m^I}) + \mu_k(\boldsymbol{u}, I_2) - \mu_k(\boldsymbol{v}, I_2) \\ &> \mu_k(\boldsymbol{u}, I_2) - \mu_k(\boldsymbol{v}, I_2).\end{aligned}$$

We can continue this operation for $n = k - m$ times and get I_n. Iteratively, we have $\mu_k(\boldsymbol{u}, I) - \mu_k(\boldsymbol{v}, I) \geq \mu_k(\boldsymbol{u}, I_n) - \mu_k(\boldsymbol{v}, I_n)$. Besides, the only difference between $(\boldsymbol{u}, I_n)$ and $(\boldsymbol{v}, I_n)$ now lies in the click probability of the k-th item:

$$u_{i_k^{I_n}} = u_{i_m^I},\ v_{i_k^{I_n}} = v_{i_m^I} \text{ and } u_{i_j^{I_n}} = v_{i_j^{I_n}},\ \forall j \neq k.$$

Since $\mu_k(\boldsymbol{u}, I_n) = \mu_k(\boldsymbol{v}, I_n)$, $\mu_k(\boldsymbol{u}, I) \geq \mu_k(\boldsymbol{v}, I)$.

□

D.8. Proof of Lemma 5.2

Lemma 5.2. *For any k, t,*

$$\min\left\{\frac{k}{2}, \frac{1}{2w^*}\right\} \leq \mu_k \leq \mathbb{E}X_{k;t} \leq \tilde{\mu}_k \leq \min\left\{\frac{1}{w'}, k\right\}.$$

Proof. **Lower bound.** According to Lemma 5.1, the expectation of observations attains its minimum when we pull an ordered set $\{1, 2, \ldots, k\}$, and attains its maximum when we pull an ordered set $\{L-k+1, L-k+2, \ldots, L\}$. In other words, depending on the instance, the expectation of observations can be lower bounded as follows:

$$\mu_k = \mu(k, w) = \sum_{i=1}^{k-1} i \cdot w(i) \cdot \prod_{j=1}^{i-1}[1 - w(j)] + k \cdot \prod_{j=1}^{k-1}[1 - w(j)].$$

Moreover, since the lower bound μ_k is larger than the expectation of observations when $w(i) = w^*$ for all $1 \leq i \leq k$ or we pull item 1 for K times (note that this is not allowed in Algorithm 1), we can utilize only w^* to lower bound the expectation:

$$\mu_k \geq \sum_{i=1}^{k-1} i \cdot w^*(1 - w^*)^{i-1} + k(1 - w^*)^{k-1} := g(w^*)$$

then

$$g(w) = w + 2w(1-w) + \ldots + (k-1)w(1-w)^{k-2} + k(1-w)^{k-1}$$

$$
\begin{aligned}
(1-w)\cdot g(w) &= w(1-w) + 2w(1-w)^2 + \ldots + (k-1)w(1-w)^{k-1} + k(1-w)^k \\
w\cdot g(w) &= w + w(1-w) + w(1-w)^2 + \ldots + w(1-w)^{k-2} + [k-(k-1)w](1-w)^{k-1} - k(1-w)^k \\
w\cdot g(w) &= w + w(1-w) + w(1-w)^2 + \ldots + w(1-w)^{k-2} + (k-kw+w-k+kw)(1-w)^{k-1} \\
w\cdot g(w) &= w\cdot \frac{1-(1-w)^k}{w} \\
g(w) &= \frac{1-(1-w)^k}{w}.
\end{aligned}
$$

Let $w^* = k^{-\beta} \in [0,1]$, then $\beta \geq 0$. Since $(1-1/x)^x$ is a nondecreasing function of x and $\lim_{x\to\infty}(1-1/x)^x = 1/e$, $k^{1-\beta} \geq 0$,

$$
g(w^*) = \frac{1-(1-w^*)^k}{w^*} = \frac{1-(1-k^{-\beta})^{k^\beta\cdot k^{1-\beta}}}{k^{-\beta}} \geq k^\beta\cdot\left(1-e^{-k^{1-\beta}}\right).
$$

If $\beta \geq 1$, let $f(x) = e^{-x}$, then $f^{(n)}(x) = (-1)^n\cdot e^{-x}$. For any $x \geq 0$, there exists $y \in [0,x]$ such that

$$
f(x) = f(0) + f'(0)\cdot x + \frac{1}{2!}f^{(2)}(0)\cdot x^2 + \frac{1}{3!}f^{(3)}(0)\cdot y^3 = 1 - x + \frac{x^2}{2} - \frac{y^3}{3} \leq 1 - x + \frac{x^2}{2}.
$$

This leads to $1-e^{-x} \geq x(1-x/2)$ and

$$
g(w^*) \geq k^\beta\cdot k^{1-\beta}(1-k^{1-\beta}/2) \geq k(1-1/2) = k/2.
$$

Otherwise, $0 \leq \beta < 1$. Since

$$
\beta\nearrow \;\Rightarrow\; 1-\beta\searrow \;\Rightarrow\; k^{1-\beta}\searrow \;\Rightarrow\; -k^{1-\beta}\nearrow \;\Rightarrow\; e^{-k^{1-\beta}}\nearrow \;\Rightarrow\; 1-e^{-k^{1-\beta}}\searrow,
$$

$1-e^{-k^{1-\beta}}$ decreases as β increases. Then,

$$
g(w^*) \geq k^\beta\cdot(1-e^{-k^{1-1}}) = k^\beta\cdot(1-e^{-1}) \geq k^\beta\cdot(1-1/2) = k^\beta/2.
$$

Altogether, $\mu_k \geq \min\{k/2, k^\beta/2\} = \min\{k/2, 1/(2w^*)\}$.

Upper bound. Similarly we can see that the expectation of observations attains its maximum when we pull an ordered set $\{L, L-1, \ldots, L-k+1\}$, and therefore upper bounded by

$$
\tilde{\mu}_k = \tilde{\mu}(k,w) = \sum_{i=1}^{k-1} i\cdot w(L+1-i)\cdot\prod_{j=1}^{i-1}[1-w(L+1-j)] + k\cdot\prod_{j=1}^{k-1}[1-w(L+1-j)].
$$

Furthermore, the upper bound $\tilde{\mu}_k$ is smaller than the expectation of observations when $w(j) = w'$ for all $L-k+1 \leq j \leq L$ or we pull item L for K times (again note that this is not allowed in Algorithm 1):

$$
\tilde{\mu}_k \leq \sum_{i=1}^{k-1} i\cdot w'(1-w')^{i-1} + k(1-w')^{k-1} = g(w') \leq \frac{1}{w'}.
$$

□

D.9. Proof of Theorem 5.4

Theorem 5.4. *Let X be an almost surely bounded nonnegative r.v.. If $\mathbb{E}X^2 \leq v^2$, then X is v-LSG.*

Proof. Set $\mathbb{E}X = \mu$ and $0 \leq X \leq M$ a.s., then $M \geq 0$ and $0 \leq \mu \leq M$. It is equivalent to show that for any $v \geq \mathbb{E}X^2$, $\lambda \leq 0$,

$$
\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{v^2\lambda^2}{2} + \lambda\mu\right).
$$

Set

$$f(\lambda) := \frac{v^2\lambda^2}{2} + \lambda\mu - \log \mathbb{E}[\exp(\lambda X)],$$

it is further equivalent to show $f(\lambda) \geq 0$. Then since $0 \leq X \leq M$ a.s., for any $\lambda \leq 0$, by Bounded Convergence Theorem,

$$\begin{aligned}
&\mathbb{E}[\exp(\lambda X)] \leq 1, \quad \left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\exp(\lambda X)\right| = |X\exp(\lambda X)| \leq M \ a.s., \\
\Rightarrow\ &\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}[\exp(\lambda X)] = \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}\lambda}\exp(\lambda X)\right] = \mathbb{E}[X\exp(\lambda X)] \leq M, \quad \left|\frac{\mathrm{d}}{\mathrm{d}\lambda}X\exp(\lambda X)\right| = |X^2\exp(\lambda X)| \leq M^2 \ a.s., \\
\Rightarrow\ &\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}[X\exp(\lambda X)] = \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}\lambda}X\exp(\lambda X)\right] = \mathbb{E}[X^2\exp(\lambda X)] \leq M^2, \\
&\left|\frac{\mathrm{d}}{\mathrm{d}\lambda}X^2\exp(\lambda X)\right| = |X^3\exp(\lambda X)| \leq M^3 \ a.s., \\
\Rightarrow\ &\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}[X^2\exp(\lambda X)] = \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}\lambda}X^2\exp(\lambda X)\right] = \mathbb{E}[X^3\exp(\lambda X)].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&f(0) = 0, \\
&f'(\lambda) = v^2\lambda + \mu - \frac{\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}[\exp(\lambda X)]}{\mathbb{E}[\exp(\lambda X)]} = v^2\lambda + \mu - \frac{\mathbb{E}[X\exp(\lambda X)]}{\mathbb{E}[\exp(\lambda X)]}, \\
&f'(0) = \mu - \mathbb{E}X = 0, \\
&f''(\lambda) = v^2 - \frac{\mathbb{E}[X^2\exp(\lambda X)]\mathbb{E}[\exp(\lambda X)] - (\mathbb{E}[X\exp(\lambda X)])^2}{(\mathbb{E}[\exp(\lambda X)])^2} \geq v^2 - \frac{\mathbb{E}[X^2\exp(\lambda X)]}{\mathbb{E}[\exp(\lambda X)]} := g(\lambda), \\
&g'(\lambda) = \frac{-\mathbb{E}[X^3\exp(\lambda X)]\mathbb{E}[\exp(\lambda X)] + \mathbb{E}[X^2\exp(\lambda X)]\mathbb{E}[X\exp(\lambda X)]}{\mathbb{E}[\exp(\lambda X)])^2}.
\end{aligned}$$

Let μ be the probability measure of X on $\mathbb{R}$. Since $0 \leq X \leq M$ a.s., $\mu([0, M]) = 1$ and

$$\begin{aligned}
&-\mathbb{E}[X^3\exp(\lambda X)]\mathbb{E}[\exp(\lambda X)] + \mathbb{E}[X^2\exp(\lambda X)]\mathbb{E}[X\exp(\lambda X)] \\
&= -\int_{[0,M]}\int_{[0,M]} x^3 e^{\lambda x+\lambda y}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) + \int_{[0,M]}\int_{[0,M]} x^2 y e^{\lambda x+\lambda y}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \\
&= \frac{1}{2}\int_{[0,M]}\int_{[0,M]} e^{\lambda x+\lambda y}(-x^3 - y^3 + x^2 y + xy^2)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \\
&= \frac{1}{2}\int_{[0,M]}\int_{[0,M]} e^{\lambda x+\lambda y}[-x^2(x-y) - y^2(y-x)]\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \\
&= \frac{1}{2}\int_{[0,M]}\int_{[0,M]} e^{\lambda x+\lambda y}(x-y)(y^2-x^2)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \\
&= -\frac{1}{2}\int_{[0,M]}\int_{[0,M]} e^{\lambda x+\lambda y}(x-y)^2(x+y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \leq 0.
\end{aligned}$$

Since $\mathbb{E}[\exp(\lambda X)])^2 > 0$, $g'(\lambda) \leq 0$. Hence $g(\lambda)$ is monotonically decreasing on $\mathbb{R}$. Further, for any $\lambda \leq 0$, since $v^2 \geq \mathbb{E}X^2$

$$\begin{aligned}
&f''(\lambda) \geq g(\lambda) \geq g(0) = v^2 - \mathbb{E}X^2 \geq 0 \ \Rightarrow\ f'(\lambda) \text{ is monotonically increasing} \\
\Rightarrow\ &f'(\lambda) \leq f'(0) = 0 \qquad\qquad \Rightarrow\ f(\lambda) \text{ is monotonically decreasing} \ \Rightarrow\ f(\lambda) \geq f(0) = 0.
\end{aligned}$$

□

Given $v^2 \geq \mathbb{E}X^2$, it is more challenging to show X is v-SG than to show X is v-LSG. By revisiting the proof above, we see that given X is v-LSG, to show X is v-SG suffices to show $f(\lambda) \geq 0$ for any $\lambda \geq 0$. Since it is hard to directly tell whether the inequality above holds for any $\lambda \geq 0$, it is natural to look at how $f(\lambda)$ grows in $\mathbb{R}$.

Fix any $M_0 > 0$. For any $\lambda \in [0, M_0]$, again, applying the Bounded Convergence Theorem, we have

$$\begin{aligned} & \frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}[\exp(\lambda X)] = \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}\lambda}\exp(\lambda X)\right], \ \frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}[X\exp(\lambda X)] = \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}\lambda}X\exp(\lambda X)\right], \\ \Rightarrow \ & f'(\lambda) = v^2\lambda + \mu - \frac{\mathbb{E}[X\exp(\lambda X)]}{\mathbb{E}[\exp(\lambda X)]}, \\ & f''(\lambda) = v^2 - \frac{\mathbb{E}[X^2\exp(\lambda X)]\mathbb{E}[\exp(\lambda X)] - (\mathbb{E}[X\exp(\lambda X)])^2}{(\mathbb{E}[\exp(\lambda X)])^2}. \end{aligned}$$

Since $f(0) = 0$ and f' is differentiable on $\mathbb{R}$, it requires at least $r > 0$ such that $f'(\lambda) \geq 0$ for any $\lambda \in [0, r]$. Furthermore, since $f'(0) = 0$, one may consider showing that $f''(\lambda) \geq 0$ on $[0, r]$.

In the proof above, we define a function g to show that $f''(\lambda) \geq g(\lambda) \geq 0$ on $(-\infty, 0]$. However, since $g(\lambda) \leq 0$ on $[0, +\infty)$, this cannot help to show $f''(\lambda) \geq 0$ on $[0, r]$.

The discussion above indicates that showing X is v-SG is more challenging than showing X is v-LSG.

D.10. Proof of Lemma 5.5

Lemma 5.5. *For any k, t, $\mathbb{E}X_{k;t}^2 \leq v_k^2 = \min\{k^2, 2/w'^2\}$.*

Proof. Recall w' is the minimum click probability. We abbreviate $X_{k;t}$ as X. Firstly, since $X \in [1, k]$, $\mathbb{E}X^2 \leq k^2$. Next, note that $\mathbb{E}X^2$ increases when the click probabilities decrease or k increases. Set Y as a random variable drawn from a geometric distribution with parameter w', then $\mathbb{E}X^2 \leq \mathbb{E}Y^2$. Since $\mathbb{E}Y^2 = 2/w'^2 - 1/w'$, $\mathbb{E}X^2 \leq 2/w'^2$. □

D.11. Proof of Lemma 5.6

Lemma 5.6. *For any $k, t, \delta > 0$, set*

$$\mathcal{E}^* := \left\{ \sum_{t=1}^{n} X_{k;t} \leq n\mu_k - \sqrt{2nv_k^2 \log\left(\frac{1}{\delta}\right)} \right\},$$

then $\Pr(\mathcal{E}^*) \leq \delta$. Further when $\overline{\mathcal{E}^*}$ holds, for any $T > 0$, $\sum_{t=1}^{n} X_{k;t} \leq T$ implies that $n \leq 2T/\mu_k + 2\log(1/\delta)v_k^2/\mu_k^2$.

Proof. We abbreviate $X_{k;t}$ as X_t (the number of observations of surviving items at step t when pulling k surviving items), and set $D_t = X_t - \mathbb{E}X_t$, $\mathcal{F}_t$ denote the decisions and observations up to step t. Besides, let S_t be the set to pull at step t, then S_t is determined by $\mathcal{F}_{t-1}$, and X_t depends on S_t. Since

$$\mathbb{E}[D_t|\mathcal{F}_{t-1}] = \mathbb{E}[\ \mathbb{E}[X_t - \mathbb{E}X_t|S_t] \mid \mathcal{F}_{t-1}\ ] = 0,$$

$D_1, \ldots, D_t$ i s a martingale difference sequence adapted to $\mathcal{F} = (\mathcal{F}_t)_t$. Besides, according to Theorem 5.4, for any t, any $\lambda \leq 0$, $v_k^2 \geq \mathbb{E}X^2$ yields $\mathbb{E}[e^{\lambda D_t}|\mathcal{F}_{t-1}] \leq e^{\lambda^2 v^2/2}$. Then for any $\omega > 0$,

$$\Pr\left[\sum_{t=1}^{n}(X_t - \mathbb{E}X_t) \leq -\omega\right] = \Pr\left[\sum_{t=1}^{n} D_t \leq -\omega\right] \leq \exp\left(-\frac{\omega^2}{2nv_k^2}\right).$$

Let the probability bound in the right hand side be δ, then

$$\delta = \exp\left(-\frac{\omega^2}{2nv_k^2}\right) \ \Rightarrow \ \omega = \sqrt{2nv_k^2 \log\left(\frac{1}{\delta}\right)}.$$

Note that $\mathbb{E}X_t \geq \mu_k$ for any t,

$$\delta \geq \Pr\left(\sum_{t=1}^{n}(X_t - \mathbb{E}X_t) \leq -\omega\right) = \Pr\left(\sum_{t=1}^{n} X_t \leq \sum_{t=1}^{n}\mathbb{E}X_t - \omega\right)$$
$$\geq \Pr\left(\sum_{t=1}^{n} X_t \leq n\mu_k - \omega\right) \geq \Pr\left(\sum_{t=1}^{n} X_t \leq n\mu_k - \sqrt{2nv_k^2\log\left(\frac{1}{\delta}\right)}\right).$$

Next, for any $T > 0$, consider

$$n\mu_k - \sqrt{2nv_k^2\log\left(\frac{1}{\delta}\right)} \leq T.$$

Set

$$a_0 = \frac{1}{\mu_k}\sqrt{2v_k^2\log\left(\frac{1}{\delta}\right)},\ b_0 = \frac{T}{\mu_k},\ x = \sqrt{n},$$

then $x \geq 0$ and $x^2 - a_0 x - b_0 \leq 0$. Note that $(p+q)^2 \leq 2(p^2+q^2)$,

$$x \leq \frac{a_0 + \sqrt{a_0^2 + 4b_0}}{2}$$
$$\Rightarrow n \leq \left(\frac{a_0 + \sqrt{a_0^2 + 4b_0}}{2}\right)^2 \leq a_0^2 + 2b_0 = \frac{2T}{\mu_k} + \frac{2v_k^2}{\mu_k^2}\log\left(\frac{1}{\delta}\right).$$

□

D.12. Proof of Lemma 5.7

Lemma 5.7. *For any $\delta \in [0,1]$, $\mathbb{P}\left(\bigcap_{i=1}^{L}\mathcal{E}(i,\delta)\right) \geq 1 - \delta/2$.*

Remark D.1 (Sub-Gaussian property)**.** *Define $\eta_t(i) = W_t(i) - w(i)$, then $\eta_t(i)$ is $1/2$-sub-Guassian.*

Proof of Remark D.1. Any non-negative random variable bounded in $[a,b]$ a.s. is sub-Gaussian with parameter $(b-a)/2$. Meanwhile, $W_t(i) \in [0,1]$ yields that $\eta_t(i) \in [w(i)-1, w(i)]$. $[w(i) - (w(i)-1)]/2 = 1/2$. □

Proof. For all $i \in [L]$, $\mathcal{E}(i,\delta) = \{\forall t \geq 1, |\hat{w}_t(i) - w(i)| \leq C_t(i,\delta)\}$. Recall that

$$C_t(i,\delta) = \tilde{C}\left(T_t(i),\rho\right), \quad \tilde{C}(\tau,\rho) = 4\sqrt{\frac{\log\left(\log_2(2\tau)/\rho\right)}{\tau+1}}, \quad \rho(\delta) = \sqrt{\delta/(12L)},$$

then according to Theorem B.2,

$$\mathbb{P}\left(\mathcal{E}(i,\delta)\right) \geq 1 - 6\rho(\delta)^2 = 1 - \frac{\delta}{2L} \ \Rightarrow\ \mathbb{P}\left(\bar{\mathcal{E}}(i,\delta)\right) \leq \frac{\delta}{2L},$$
$$\Rightarrow \mathbb{P}\left(\bigcap_{i=1}^{L}\mathcal{E}(i,\delta)\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{L}\bar{\mathcal{E}}(i,\delta)\right) \geq 1 - \sum_{i=1}^{L}\mathbb{P}\left(\bar{\mathcal{E}}(i,\delta)\right) \geq 1 - L\cdot\frac{\delta}{2L} = 1 - \delta/2.$$

□

D.13. Proof of Lemma 5.8

Lemma 5.8. *Fix any $0 < \delta' \leq \delta$, assume $\bigcap_{i=1}^{L}\mathcal{E}(i,\delta')$ holds. Set $T_t' := \min_{i\in D_t} T_t(i)$, then for any time step t,*

$$\forall i \leq K', T'(t) \geq \bar{T}_{i,\delta'} \Rightarrow L_t(i,\delta) > U_t(j^*,\delta) - \epsilon \Rightarrow i \in A_t,$$
$$\forall i > K', T'(t) \geq \bar{T}_{i,\delta'} \Rightarrow U_t(i,\delta) < L_t(j',\delta) - \epsilon \Rightarrow i \in R_t.$$

Preliminary. Since we use the UCB of the empirical top-(k_t+1) item to accept ϵ-optimal items, hopefully it should be close to the true click probability of item (k_t+1); likewise, the LCB of the empirical top-(k_t) item should be close to the true click probability of item (k_t). This is stated in Lemma D.2.

Lemma D.2 (Jun et al. (2016, Lemma 3))**.** *Denote by $\hat{i}$ the index of the item with empirical mean is i-th largest: i.e., $\hat{w}(\hat{1}) \geq \ldots \geq \hat{w}(\hat{L})$. Assume that the empirical means of the arms are controlled by ϵ : i.e., $\forall i, |\hat{w}(i) - w(i)| < \epsilon$. Then,*

$$\forall i, w(i) - \epsilon \leq \hat{w}(\hat{i}) \leq w(i) + \epsilon.$$

After that, Lemma 5.8 shows that the agent will correctly classify the items after a sufficient number of observations, and also show what is the sufficient number of observations for each item.

Proof. Recall

$$k_t = K - |A_t|,\ \rho(\delta') = \delta'/(12L),\ \bar{T}_{i,\delta'} = 1 + \left\lfloor \frac{216}{\bar{\Delta}_i^2} \log\left(\frac{2}{\rho(\delta')} \log_2\left(\frac{648}{\rho(\delta')\bar{\Delta}_i^2}\right)\right)\right\rfloor.$$

And We use ρ and ρ' as abbreviations for $\rho(\delta)$ and $\rho(\delta')$ respectively.

It suffices to show for the case where A_t and R_t are empty since otherwise the problem is equivalent to removing rejected or accepted arms from consideration and starting a new problem with $L_{\text{new}} = L - |A_t| - |R_t|$ and $K_{\text{new}} = K - |A_t|$ while maintaining the observations so far. Note that when A_t is empty, $k_t = K$.

First of all, $T_t(i) \geq T_t'$ implies that

$$C_t(i,\delta) = \tilde{C}\left(T_t(i), \rho\right) \leq \tilde{C}\left(T_t', \rho\right),\ C_t(i,\delta') = \tilde{C}\left(T_t(i), \rho'\right) \leq \tilde{C}\left(T_t', \rho'\right). \tag{D.1}$$

Then since $\bigcap_{i=1}^{L} \mathcal{E}(i,\delta')$ holds, $|\hat{w}_t(i) - w(i)| \leq \tilde{C}\left(T_t(i), \rho'\right) \leq \tilde{C}\left(T_t', \rho'\right)$ for all $i \in D_t$. Combining this with Lemma D.2, we have

$$w(i) + \tilde{C}\left(T_t', \rho'\right) \leq \hat{w}_t(i) \leq w(i) + \tilde{C}\left(T_t', \rho'\right)\ \forall i \in D_t. \tag{D.2}$$

We first prove that for any $i \leq K'$,

$$T'(t) \geq \bar{T}_{i,\delta'} \ \Rightarrow\ L_t(i,\delta) > U_t(j^*,\delta) - \epsilon \text{ where } j^* = \operatorname*{arg\,max}_{j \in D_t}^{(k_t+1)} \hat{w}_t \ \Rightarrow\ i \in A_t.$$

For clarity, we write $j^* = \widehat{K+1}$, which is the item with the $(K+1)^{\text{st}}$ largest empirical mean at the t-th step. We assume the contrary: $L_t(i,\delta) \leq U_t(\widehat{K+1},\delta) - \epsilon$. We can apply (D.1) and (D.2) to obtain

$$\begin{aligned}
&L_t(i,\delta) \geq \hat{w}_t(i) - \tilde{C}\left(T_t', \rho\right) \geq w(i) - \tilde{C}\left(T_t', \rho\right) - \tilde{C}\left(T_t', \rho'\right),\\
&U_t(\widehat{K+1}) - \epsilon \leq \hat{w}_t(\widehat{K+1}) + \tilde{C}\left(T_t', \rho\right) - \epsilon \leq w(K+1) + \tilde{C}\left(T_t', \rho\right) + \tilde{C}\left(T_t', \rho'\right) - \epsilon.
\end{aligned}$$

Next,

$$\begin{aligned}
&w(i) - \tilde{C}\left(T_t', \rho\right) - \tilde{C}\left(T_t', \rho'\right) \leq w(K+1) + \tilde{C}\left(T_t', \rho\right) + \tilde{C}\left(T_t', \rho'\right) - \epsilon,\\
\Rightarrow\ &0 \overset{(a)}{<} w(i) - w(K+1) + \epsilon \leq 2\tilde{C}\left(T_t', \rho\right) + 2\tilde{C}\left(T_t', \rho'\right) \leq 4\tilde{C}\left(T_t', \rho'\right) = 16\sqrt{\frac{\log\left(\log_2(2T_t')/\rho'\right)}{T_t'}},\\
\Rightarrow\ &T_t' \leq \frac{216}{([w(i) - w(K+1) + \epsilon]^2} \log\left(\log_2(2T_t')/\rho'\right).
\end{aligned}$$

Part (a) of the second line above follows from: (i) if $i \leq K$, $w(i) - w(K+1) + \epsilon = \Delta_i + \epsilon > 0$; (ii) else, $K < i \leq K'$, since $w(i) \geq w(K) - \epsilon$, we have $w(i) - w(K+1) + \epsilon = w(i) - w(K) + w(K) - w(K+1) + \epsilon = \Delta_K - \Delta_i + \epsilon \geq \Delta_K > 0$. Then invert to the third line using

$$\tau \leq c \log\left(\frac{\log_2 2\tau}{\rho'}\right) \Rightarrow \tau \leq c \log\left(\frac{2}{\rho'} \log_2\left(\frac{3c}{\rho'}\right)\right)$$

with $c = 216[w(i) - w(K+1) + \epsilon]^{-2}$ to have

$$T'_t \le \frac{216}{[w(i) - w(K+1) + \epsilon]^2} \log\left(\frac{2}{\rho'} \log_2\left(\frac{648}{\rho'[w(i) - w(K+1) + \epsilon]^2}\right)\right)$$
$$< 1 + \left\lfloor \frac{216}{[w(i) - w(K+1) + \epsilon]^2} \log\left(\frac{2}{\rho'} \log_2\left(\frac{648}{\rho'[w(i) - w(K+1) + \epsilon]^2}\right)\right)\right\rfloor = \bar{T}_{i,\delta'}.$$

Therefore, $\bar{T}'_t \ge \bar{T}_{i,\delta'}$ implies that $L_t(i,\delta) > U_t(j^*,\delta) - \epsilon$ where $j^* = \arg\max_{j \in D_t}^{(k_t+1)} \hat{w}_t$. Then $i \in A_t$ is accepted.

Subsequently, we prove that for any $i > K'$,

$$T'(t) \ge \bar{T}_{i,\delta'} \;\Rightarrow\; U_t(i,\delta) < L_t(j',\delta) - \epsilon \text{ where } j' = \arg\max_{j \in D_t}^{(k_t)} \hat{w}_t \;\Rightarrow\; i \in R_t.$$

Again for brevity, we write $\hat{K} = j'$, the item with the K^{th} largest empirical mean at the t-th step. We assume the contrary: $U_t(i,\delta) \ge L_t(\hat{K},\delta) - \epsilon$. Again applying (D.1) and (D.2), we have

$$U_t(i,\delta) \le \hat{w}_t(i) + \tilde{C}\left(T'_t,\rho\right) \le w(i) + \tilde{C}\left(T'_t,\rho\right) + \tilde{C}\left(T'_t,\rho'\right),$$
$$L_t(\hat{K},\delta) - \epsilon \ge \hat{w}_t(\hat{K}) - \tilde{C}\left(T'_t,\rho\right) - \epsilon \ge w(K) - \tilde{C}\left(T'_t,\rho\right) - \tilde{C}\left(T'_t,\rho'\right) - \epsilon.$$

Next,

$$w(i) + \tilde{C}\left(T'_t,\rho\right) + \tilde{C}\left(T'_t,\rho'\right) \ge w(K) - \tilde{C}\left(T'_t,\rho\right) - \tilde{C}\left(T'_t,\rho'\right) - \epsilon,$$
$$\Rightarrow\; 0 < w(K) - w(i) - \epsilon \le 2\tilde{C}\left(T'_t,\rho\right) + 2\tilde{C}\left(T'_t,\rho'\right) \le 4\tilde{C}\left(T'_t,\rho'\right) = 16\sqrt{\frac{\log\left(\log_2(2T'_t)/\rho'\right)}{T'_t}}.$$

Similar to the first case, with

$$\bar{T}_{i,\delta'} = 1 + \left\lfloor \frac{216}{(w(K) - w(i) - \epsilon)^2} \log\left(\frac{2}{\rho'} \log_2\left(\frac{648}{\rho'(w(K) - w(i) - \epsilon)^2}\right)\right)\right\rfloor$$

we obtain that $\bar{T}'_t \ge \bar{T}_{i,\delta'}$ implies $U_t(i,\delta) < L_t(j',\delta) - \epsilon$ where $j' = \arg\max_{j \in D_t}^{(k_t)} \hat{w}_t$. Then $i \in R_t$ is rejected.

□

D.14. Proof of Lemma 5.9

Lemma 5.9. *Assume $\bigcap_{i=1}^{L} \mathcal{E}(i,\delta)$ holds. Algorithm 1 stops after identifying at most $L - \max\{K' - K, 1\}$ items.*

Proof. Assume $\bigcap_{i=1}^{L} \mathcal{E}(i,\delta)$ holds.

Case (i): $K' = K$. In the worst case, the algorithm does not terminate before identifying the $(L-1)$-th one. In this case, after identifying the $(L-1)$-th one with sufficient observations, either the accept set or the reject set is full, i.e., $|A_t| = K$ or $|R_t| = L - K$, the the agent can just place the remaining item in the unfilled set.

Hence, the algorithm terminates after sufficiently observing and identifying at most $L - 1 = L - \max\{K' + K, 1\}$ items.

Case (ii): $K' > K$. The algorithm classify all items correctly according to Lemma 5.8. since the number of ϵ-optimal items is $K' = \max\{i : w(i) \ge w(K) - \epsilon\} \ge K$, the number of suboptimal items is $L - K' \le L - K$. Hence, $|R_t| \le L - K'$. Besides, $|A_t| \le K$ according to the design of the algorithm. Therefore,

$$|A_t| + |R_t| \le L - K' + K.$$

In other words, the algorithm terminates after sufficiently observing and identifying at most $L - K' + K = L - \max\{K' + K, 1\}$ items. □

D.15. Proof of Lemma 5.11

Lemma 5.11. *For any* $1 \le \ell \le L$,

$$\begin{aligned}&\mathrm{KL}\big(\ \{S_t^{\pi,0}, \boldsymbol{O}_t^{\pi,0}\}_{t=1}^{\mathcal{T}}, \{S_t^{\pi,\ell}, \boldsymbol{O}_t^{\pi,\ell}\}_{t=1}^{\mathcal{T}}\ \big)\\ &= \mathbb{E}[T_{\mathcal{T}}(\ell)] \cdot \mathrm{KL}\big(w(\ell), w^{(\ell)}(\ell)\big) \ge \sup_{\mathcal{E}\in\mathcal{T}} \mathrm{KL}\big(\mathbb{P}_0(\mathcal{E}), \mathbb{P}_\ell(\mathcal{E})\big).\end{aligned}$$

To manifest the difference between instance ℓ and other instances, with $w^{(0)}(i) = w(i)$ for all $i \in [L]$ we write

- $\{w^{(0)}(1), w^{(0)}(2), \ldots, w^{(0)}(L)\}$ under instance 0;
- $\{w^{(0)}(1), w^{(0)}(2), \ldots, w^{(0)}(\ell-1), w^{(\ell)}(\ell), w^{(0)}(\ell+1), \ldots, w^{(0)}(L)\}$ under instance ℓ.

We combine Lemma 5.10 and a result from Kaufmann et al. (2016) to relate the time complexity and KL divergence together.

Lemma D.3 ((Kaufmann et al., 2016, Lemma 19))**.** *Let $\mathcal{T}$ be any almost surely finite stopping time with respect to $\mathcal{F}_t$. For every event $\mathcal{E} \in \mathcal{F}_{\mathcal{T}}$, instance $1 \le \ell \le L$,*

$$\mathrm{KL}\left(\{S_t^{\pi,0}, \boldsymbol{O}_t^{\pi,0}\}_{t=1}^{\mathcal{T}}, \{S_t^{\pi,\ell}, \boldsymbol{O}_t^{\pi,\ell}\}_{t=1}^{\mathcal{T}}\right) \ge \mathrm{KL}(\mathbb{P}_0(\mathcal{E}), \mathbb{P}_\ell(\mathcal{E})).$$

Notations. Before presenting the proof, we remind the reader of the definition of the KL divergence (Cover & Thomas, 2012). For two discrete random variables X and Y with common support $\mathcal{A}$,

$$\mathrm{KL}(X, Y) = \sum_{x\in\mathcal{A}} P_X(x) \log \frac{P_X(x)}{P_Y(x)}$$

denotes the KL divergence between probability mass functions of X and Y. Next, we also use $\mathrm{KL}(P_X \| P_Y)$ to also signify this KL divergence. Lastly, when a and b are two real numbers between 0 and 1, $\mathrm{KL}(a, b) = \mathrm{KL}\left(\mathrm{Bern}(a) \| \mathrm{Bern}(b)\right)$, i.e., $\mathrm{KL}(a, b)$ denotes the KL divergence between $\mathrm{Bern}(a)$ and $\mathrm{Bern}(b)$.

Proof. For any certain s_t, we can observe that the KL divergence $\mathrm{KL}\left(P_{\boldsymbol{O}_t^{\pi,0}|S_t^{\pi,0}}(\cdot \mid s_t) \,\middle\|\, P_{\boldsymbol{O}_t^{\pi,\ell}|S_t^{\pi,\ell}}(\cdot \mid s_t)\right)$ should grow with the KL divergence of observed items. Further, for each observed item i, there is a KL divergence of $\mathrm{KL}\left(w^{(0)}(i), w^{(\ell)}(i)\right)$. Whenever $S_t^{\pi,0} = s_t$, we have

$$\mathrm{KL}\left(P_{\boldsymbol{O}_t^{\pi,0}|S_t^{\pi,0}}(\cdot \mid s_t) \,\middle\|\, P_{\boldsymbol{O}_t^{\pi,\ell}|S_t^{\pi,\ell}}(\cdot \mid s_t)\right) = \sum_{i\in s_t} \mathbb{E}_0\left[1\{i \text{ is observed at time } t\}\right] \cdot \mathrm{KL}\left(w^{(0)}(i), w^{(\ell)}(i)\right).$$

Then according to Lemma 5.10,

$$\begin{aligned}&\mathrm{KL}\left(\{S_t^{\pi,0}, \boldsymbol{O}_t^{\pi,0}\}_{t=1}^{\mathcal{T}}, \{S_t^{\pi,\ell}, \boldsymbol{O}_t^{\pi,\ell}\}_{t=1}^{T}\right)\\ &= \sum_{t=1}^{\mathcal{T}} \sum_{s_t\in[L]^{(K)}} \Pr[S_t^{\pi,0} = s_t] \cdot \mathrm{KL}\left(P_{\boldsymbol{O}_t^{\pi,0}|S_t^{\pi,0}}(\cdot \mid s_t) \,\middle\|\, P_{\boldsymbol{O}_t^{\pi,\ell}|S_t^{\pi,\ell}}(\cdot \mid s_t)\right)\\ &= \sum_{t=1}^{\mathcal{T}} \sum_{s_t\in[L]^{(K)}} \Pr[S_t^{\pi,0} = s_t] \cdot \sum_{i\in s_t} \mathbb{E}_0\left[1\{i \text{ is observed at time } t\}\right] \cdot \mathrm{KL}\left(w^{(0)}(i), w^{(\ell)}(i)\right)\\ &= \sum_{i=1}^{L} \sum_{t=1}^{\mathcal{T}} \sum_{s_t\in[L]^{(K)}} \mathbb{E}_0\left[1\{S_t^{\pi,0} = s_t,\ i \in s_t,\ i \text{ is observed at time } t\}\right] \cdot \mathrm{KL}\left(w^{(0)}(i), w^{(\ell)}(i)\right)\\ &= \sum_{i=1}^{L} \mathbb{E}[T_{\mathcal{T}}(i)] \cdot \mathrm{KL}\left(w^{(0)}(i), w^{(\ell)}(i)\right)\\ &= \mathbb{E}[T_{\mathcal{T}}(\ell)] \cdot \mathrm{KL}\left(w^{(0)}(\ell), w^{(\ell)}(\ell)\right).\end{aligned}$$

□

D.16. Proof of Theorem 4.1

Preliminary. Recall that $\bar{\Delta}_{\sigma(1)} \geq \bar{\Delta}_{\sigma(2)} \geq \ldots \geq \bar{\Delta}_{\sigma(L)}$, and $T_t(i)$ counts the number of observations of item i up to the t-th step. The worst case is that the algorithm eliminates $\sigma(1), \sigma(2), \ldots$ in order, and the algorithm eliminates at most 1 item at one time step. Besides, the design of Algorithm 1 implies that

$$T_t(j) - 1 \leq T_t(i) \leq T_t(j) + 1, \ \forall i \neq j \in D_t. \tag{D.3}$$

In the following discussion,we assume $\bigcap_{i=1}^{L} \mathcal{E}(i,\delta)$ holds and $K' < 2K-1$ (discussion on $K' \geq 2K-1$ is in Appendix C). Note that Lemma 5.7 implies that $\mathbb{P}\left(\bigcap_{i=1}^{L} \mathcal{E}(i,\delta)\right) \geq 1 - \delta/2$. Besides, we write $\mu(k,w)$ as μ_k, $v(k,w)$ as v_k, $\bar{T}_{i,\delta}$ as $\bar{T}_i$, $\rho(\delta)$ as ρ for simplicity.

Bound the number of observations per phrase. Observe that there are less than K surviving items remaining in the survival set D_t at some steps before the algorithm terminates, we separate the steps into several phrases:

(i) During the first phrase, the agent eliminates $L-K+1$ items within t_1 steps. According to Lemma 5.8 and Line (D.3),

$$\begin{aligned} &T_{t_1}(\sigma(j)) \leq \bar{T}_{\sigma(j)}, && \forall 1 \leq j \leq L-K+1; \\ &T_{t_1}(\sigma(i)) \leq \bar{T}_{\sigma(L-K+1)} + 1, && \forall L-K+1 < i \leq L. \end{aligned}$$

Then the total number of observations of surviving items in D_t within this phrase can be bounded as follows:

$$\sum_{i=1}^{L-K+1} \bar{T}_{\sigma(i)} + \sum_{i=L-K+2}^{L} T_{t_1}(\sigma(i)) \leq \sum_{i=1}^{L-K} \bar{T}_{\sigma(i)} + K\bar{T}_{\sigma(L-K+1)} + (K-1) := \tilde{T}_1.$$

(ii) During the k-th phrase for any $2 \leq k \leq K - \max\{K'-K, 1\}$, the agent eliminates the $L-K+k$-th item within t_k steps. Again apply Lemma 5.8 and Line (D.3):

$$\begin{aligned} &T_{\sum_{j=1}^{k} t_j}(\sigma(L-K+k)) \leq \bar{T}_{\sigma(L-K+k)}; \\ &T_{\sum_{j=1}^{k} t_j}(\sigma(i)) \leq \bar{T}_{\sigma(L-K+k)} + 1, && \forall L-K+k+1 \leq i \leq L; \\ &T_{\sum_{j=1}^{k-1} t_j}(\sigma(i)) \geq \bar{T}_{\sigma(L-K+k-1)} - 1, && \forall L-K+k \leq i \leq L. \end{aligned}$$

Then the total number of observations of surviving items in D_t within this phrase can also be bounded:

$$\begin{aligned} &\bar{T}_{\sigma(L-K+k)} + \sum_{i=L-K+k+1}^{L} T_{\sum_{i=j}^{k} t_j}(\sigma(i)) - \sum_{i=L-K+k}^{L} T_{\sum_{j=1}^{k-1} t_j}(\sigma(i)) \\ &\leq (K-k+1)[\bar{T}_{\sigma(L-K+k)} - \bar{T}_{\sigma(L-K+k-1)}] + 2(K-k) + 1 := \tilde{T}_k. \end{aligned}$$

Bound the number of time steps per phrase. Recall that the k-th ($1 \leq k \leq K - \max\{K'-K, 1\}$) phrase consist of t_k time steps. Let Z_k be the total number of observations within the t_k steps. Lemma 5.6 indicates that

$$\mathbb{P}(Z_k \geq t_k \mu_{K+1-k} \omega_{K+1-k}) \geq 1 - \delta_k \quad \text{with} \quad \omega_{K+1-k} = -\sqrt{-2t_k v_{K+1-k}^2 \log \delta_k}.$$

Then according to Lemma 5.6, for any k ($1 \leq k \leq K - \max\{K'-K, 1\}$), with probability at least $1-\delta_k$,

$$t_k \leq \frac{2\tilde{T}_k}{\mu_{K-k+1}} - \frac{2v_{K+1-k}^2}{\mu_{K-k+1}^2} \cdot \log \delta_k.$$

Bound the time complexity. Altogether, we would have $\sum_{k=1}^{K-\max\{K'-K,1\}} t_k$ as the time complexity. Besides, we bound the total error incurred by partial observation by $\delta/2$. In other words,

$$\mathcal{T} \leq \sum_{k=1}^{K-\max\{K'-K,1\}} \left(-\frac{2v_{K-k+1}^2}{\mu_{K-k+1}^2} \cdot \log \delta_k + 2 \sum_{k=1}^{2K-K'} \frac{\tilde{T}_k}{\mu_{K-k+1}} \right) \quad \text{where} \sum_{k=1}^{K-\max\{K'-K,1\}} \delta_k \leq \delta/2.$$

Depending on the value of $K'-K$, there are two cases:

Case 1: $K' - K \geq 1$, i.e., $K - \max\{K' - K, 1\} = 2K - K'$;

Case 2: $K' = K$, i.e., $K - \max\{K' - K, 1\} = K - 1$.

For brevity, we only go through the remaining analysis for the first case, the analysis for the second one is similar.

Since the second term in the bound on $\mathcal{T}$ merely depends on the problem, we turn to analyze the first term. Since the first term holds for any values of δ_k's such that $\sum_{k=1}^{2K-K'} \delta_k \leq \delta/2$, we minimize the first term with the method of Lagrange multiplier. Set $c_k = \frac{2v_{K-k+1}^2}{\mu_{K-k+1}^2}$, the problem turns to

$$(\blacktriangle) = \max_{\delta_k : 1 \leq k \leq 2K-K'} \sum_{k=1}^{2K-K'} c_k \log \delta_k \quad \text{s.t.} \quad \sum_{k=1}^{2K-K'} \delta_k \leq \delta/2.$$

Let

$$L\left(\{\delta_k\}_{k=1}^{2K-K'}, \{c_k\}_{k=1}^{2K-K'-1}, \lambda\right) = \sum_{k=1}^{2K-K'} c_k \log \delta_k + \lambda \left(\sum_{k=1}^{2K-K'} \delta_k - \delta/2 \right),$$

then for all $1 \leq k \leq 2K - K'$,

$$\frac{\partial L}{\partial \delta_k} = \frac{c_k}{\delta_k} + \lambda = 0 \;\Rightarrow\; \delta_k^* = \frac{c_k \delta}{2 \sum_{j=1}^{2K-K'} c_j}.$$

$(\blacktriangle)$ attains its maximum when $\delta_k = \delta_k^*$ for all $1 \leq k \leq 2K - K'$. Hence

$$\begin{aligned}
\sum_{k=1}^{2K-K'} t_k &\leq - \sum_{k=1}^{2K-K'} c_k \log \delta_k^* + 2 \sum_{k=1}^{2K-K'} \frac{\tilde{T}_k}{\mu_{K-k+1}} \\
&= \sum_{k=1}^{2K-K'} c_k \log \left(\frac{2 \sum_{j=1}^{2K-K'} c_j}{c_k \delta} \right) + 2 \sum_{k=1}^{2K-K'} \frac{\tilde{T}_k}{\mu_{K-k+1}} \\
&= \underbrace{\sum_{k=1}^{2K-K'} c_k \log \left(\frac{2}{\delta} \sum_{j=1}^{2K-K'} c_j \right)}_{(\spadesuit)} + \underbrace{\sum_{k=1}^{2K-K'} c_k \log \left(\frac{1}{c_k} \right)}_{(\heartsuit)} + \underbrace{2 \sum_{k=1}^{2K-K'} \frac{\tilde{T}_k}{\mu_{K-k+1}}}_{(\clubsuit)}.
\end{aligned}$$

Now we bound $(\spadesuit)$, $(\heartsuit)$, $(\clubsuit)$ individually.

Bounding $(\spadesuit)$: note that $\mu_{K+1-k} \geq 2$ for all $1 \leq k \leq 2K - K'$, $K' \geq K$ and $c_k = \frac{2v_{K-k+1}^2}{\mu_{K-k+1}^2}$,

$$(\spadesuit) = \sum_{k=1}^{2K-K'} c_k \log \left(\frac{2}{\delta} \sum_{j=1}^{2K-K'} c_j \right) = \sum_{k=1}^{2K-K'} \frac{2v_{K-k+1}^2}{\mu_{K-k+1}^2} \log \left(\frac{2}{\delta} \sum_{j=1}^{2K-K'} \frac{2v_{K-j+1}^2}{\mu_{K-j+1}^2} \right).$$

Bounding $(\heartsuit)$: Let $g(x) = \frac{\log x}{x}$ for $x > 0$, then $g'(x) = \frac{1 - \log x}{x^2}$. Since $g'(x) > 0$ when $x \in (0, e)$, $g'(e) = 0$, $g'(x) < 0$ when $x \in (e, +\infty)$, $g(x)$ is increasing on $(0, e)$, is decreasing on $(e, +\infty)$ and attains its global maximum $g(e) = \frac{1}{e}$ at $x = e$. Hence,

$$(\heartsuit) = \sum_{k=1}^{2K-K'} g\left(\frac{1}{c_k} \right) \leq \frac{2K - K'}{e} \leq K.$$

Bounding $(\clubsuit)$: We first rewrite this term according to the definition of $\tilde{T}_k$'s:

$$\tilde{T}_1 = \sum_{i=1}^{L-K} \bar{T}_{\sigma(i)} + K \bar{T}_{\sigma(L-K+1)} + (K - 1),$$

$$\tilde{T}_k = (K-k+1)[\bar{T}_{\sigma(L-K+k)} - \bar{T}_{\sigma(L-K+k-1)}] + 2(K-k) + 1 \quad \forall 2 \le k \le 2K - K' - 1,$$

$$\Rightarrow (\clubsuit) \le \frac{2}{\mu_K}\left[\sum_{i=1}^{L-K} \bar{T}_{\sigma(i)} + K\bar{T}_{\sigma(L-K+1)} + K\right] + \sum_{k=2}^{2K-K'} \frac{2(K-k+1)}{\mu_{K-k+1}}\left[\bar{T}_{\sigma(L-K+k)} - \bar{T}_{\sigma(L-K+k-1)} + 3\right]$$

$$\Rightarrow (\clubsuit)/4 \le \frac{1}{\mu_K}\left[\sum_{i=1}^{L-K} \bar{T}_{\sigma(i)} + K\bar{T}_{\sigma(L-K+1)}\right] + \sum_{k=2}^{2K-K'} \frac{K-k+1}{\mu_{K-k+1}}\left[\bar{T}_{\sigma(L-K+k)} - \bar{T}_{\sigma(L-K+k-1)}\right].$$

Next, since $\mu_k \ge \min\{k/2, 1/(2w^*)\}$ as shown in Lemma 5.2, when $K-k+1 \le 1/w^*$,

$$k \ge K + 1 - \lfloor 1/w^* \rfloor,\ \mu_{K-k+1} \ge \frac{K-k+1}{2},\ \frac{K-k+1}{\mu_{K-k+1}} \le 2.$$

Hence with $K_0 = \max\{\min\{2K-K', K - \lfloor 1/w^* \rfloor\}, 1\}$,

$$\begin{aligned}
&\frac{K}{\mu_K}\bar{T}_{\sigma(L-K+1)} + \sum_{k=2}^{K_1} \frac{K-k+1}{\mu_{K-k+1}}\left[\bar{T}_{\sigma(L-K+k)} - \bar{T}_{\sigma(L-K+k-1)}\right] \\
&= \frac{K\bar{T}_{\sigma(L-K+1)}}{\mu_K} + \sum_{k=2}^{K_0} \frac{(K-k+1)\bar{T}_{\sigma(L-K+k)}}{\mu_{K-k+1}} - \sum_{k=1}^{K_0-1} \frac{(K-k)\bar{T}_{\sigma(L-K+k)}}{\mu_{K-k}} \\
&= \frac{(K-K_0+1)\bar{T}_{\sigma(L-K+K_0)}}{\mu_{K-K_0+1}} + \sum_{k=1}^{K_0-1} \bar{T}_{\sigma(L-K+k)}\left(\frac{K-k+1}{\mu_{K-k+1}} - \frac{K-k}{\mu_{K-k}}\right),
\end{aligned}$$

and

$$\begin{aligned}
\sum_{k=K_0+1}^{2K-K'} \frac{K-k+1}{\mu_{K-k+1}}\left[\bar{T}_{\sigma(L-K+k)} - \bar{T}_{\sigma(L-K+k-1)}\right] &\le 2\sum_{k=K_0+1}^{2K-K'} \bar{T}_{\sigma(L-K+k)} - \bar{T}_{\sigma(L-K+k-1)} \\
&= 2\bar{T}_{\sigma(L+K-K')} - 2\bar{T}_{\sigma(L-K+K_0)}.
\end{aligned}$$

Further,

$$(\clubsuit)/4 \le \frac{1}{\mu_K}\sum_{i=1}^{L-K} \bar{T}_{\sigma(i)} + \sum_{k=1}^{K_0-1} \bar{T}_{\sigma(L-K+k)}\left(\frac{K-k+1}{\mu_{K-k+1}} - \frac{K-k}{\mu_{K-k}}\right) + \left(\frac{K-K_0+1}{\mu_{K-K_0+1}} - 2\right)\bar{T}_{\sigma(L-K+K_0)} + 2\bar{T}_{\sigma(L+K-K')}.$$

Summation of (♠), (♡), (♣). Recall $\rho = \delta/(12L)$ and

$$\bar{T}_i = 1 + \left\lfloor \frac{216}{\bar{\Delta}_i^2} \log\left(\frac{2}{\rho}\log_2\left(\frac{648}{\rho\bar{\Delta}_i^2}\right)\right)\right\rfloor.$$

The time complexity is upper bounded by

$$c_1 \sum_{k=1}^{2K-K'} \frac{v_{K-k+1}^2}{\mu_{K-k+1}^2} \log\left(\frac{1}{\delta}\sum_{j=1}^{2K-K'} \frac{v_{K-j+1}^2}{\mu_{K-j+1}^2}\right) + c_2\frac{1}{\mu_K}\sum_{i=1}^{L-K} \bar{T}_{\sigma(i)} + c_3 \sum_{k=2}^{2K-K'} \frac{K-k+1}{\mu_{K-k+1}}\left[\bar{T}_{\sigma(L-K+k)} - \bar{T}_{\sigma(L-K+k-1)}\right]$$

where

$$\frac{1}{\mu_K}\sum_{i=1}^{L-K} \bar{T}_{\sigma(i)} = O\left(\frac{1}{\mu_K}\sum_{i=1}^{L-K} \bar{\Delta}_{\sigma(i)}^{-2} \log\left[\frac{L}{\delta}\log\left(\frac{L}{\delta\bar{\Delta}_{\sigma(i)}^2}\right)\right]\right),$$

$$\sum_{k=2}^{2K-K'} \frac{K-k+1}{\mu_{K-k+1}}\left[\bar{T}_{\sigma(L-K+k)} - \bar{T}_{\sigma(L-K+k-1)}\right]$$

$$
\begin{aligned}
&= \sum_{k=1}^{K_0-1} \bar{T}_{\sigma(L-K+k)} \left(\frac{K-k+1}{\mu_{K-k+1}} - \frac{K-k}{\mu_{K-k}} \right) + \left(\frac{K-K_0+1}{\mu_{K-K_0+1}} - 2 \right) \bar{T}_{\sigma(L-K+K_0)} + 2\bar{T}_{\sigma(L+K-K')} \\
&= c_4 \sum_{k=1}^{K_0-1} \bar{\Delta}^{-2}_{\sigma(L-K+k)} \log \left[\frac{L}{\delta} \log \left(\frac{L}{\delta \bar{\Delta}^2_{\sigma(L-K+k)}} \right) \right] \cdot \left(\frac{K+1-k}{\mu_{K+1-k}} - \frac{K-k}{\mu_{K-k}} \right) \\
&\quad + c_5 \left(\frac{K-K_0+1}{\mu_{K-K_0+1}} - 2 \right) \bar{\Delta}^{-2}_{\sigma(L-K+K_0)} \log \left[\frac{L}{\delta} \log \left(\frac{L}{\delta \bar{\Delta}^2_{\sigma(L-K+K_0)}} \right) \right] + c_6 \bar{\Delta}^{-2}_{\sigma(L+K-K')} \log \left[\frac{L}{\delta} \log \left(\frac{L}{\delta \bar{\Delta}^2_{\sigma(L+K-K')}} \right) \right].
\end{aligned}
$$

D.17. Proof of Theorem 4.8

Recall that $\boldsymbol{O}_t^\pi$ is a vector in $\{0,1,\star\}^K$, where $0,1,\star$ represents observing no click, observing a click and no observation respectively. For example, when $S_t^\pi = (2,1,5,4)$ and $\boldsymbol{O}_t^\pi = (0,0,1,\star)$, items $2,1,5,4$ are listed in the displayed order; items $2,1$ are not clicked, item 5 is clicked, and the response to item 4 is not observed. By the definition of the cascading model, the outcome $\boldsymbol{O}_t^\pi = (0,0,1,\star)$ is in general a (possibly emtpy) string of 0s, followed by a 1 (if the realized reward is 1), and then followed by a possibly empty string of $\star$s. Clearly, $S_t^{\pi,\ell}$, $\boldsymbol{O}_t^{\pi,\ell}$ are random variables with distribution depending on $w^{(\ell)}$ (hence these random variables distribute differently for different ℓ), albeit a possibly complicated dependence on $w^{(\ell)}$.

With the analysis in Section 5.3, according to Lemma 5.11 and the definition of the instance ℓ, one obtains for $i \in \{1,\dots,K\}$ or $j \in \{K+1,\dots,L\}$ respectively,

$$
\mathbb{E}[T_{\mathcal{T}}(i)] \geq \frac{\mathrm{KL}(1-\delta,\delta)}{\mathrm{KL}\left(w(i), w(K+1)\right) + \alpha}, \quad \mathbb{E}[T_{\mathcal{T}}(j)] \geq \frac{\mathrm{KL}(1-\delta,\delta)}{\mathrm{KL}\left(w(j), w(K)\right) + \alpha}.
$$

Let Y_t denote the number of observations of items at time step t. By revisiting the definition of $X_{k;t}$ in Section 4.1, we see that $X_{K;t}$ actually counts the observation of all pulled items at time step t. Hence, $Y_t \leq X_{K;t}$. Setting $\alpha \to 0$ and summing over the items yields a bound on the expected number of total observations $\mathbb{E}\left[\sum_{t=1}^{\mathcal{T}} Y_t\right] = \sum_{i=1}^{L} \mathbb{E}[T_{\mathcal{T}}(i)]$. Meanwhile, an upper bound of $\mathbb{E}X_{K;t}$ as stated in Lemma 5.2 and tower property indicates that

$$
\mathbb{E}\left[\sum_{t=1}^{\mathcal{T}} Y_t\right] = \mathbb{E}\left[\ \mathbb{E}\left[\sum_{t=1}^{T} Y_t \mid \mathcal{T}=T\right]\ \right] \leq \mathbb{E}\left[\ \mathbb{E}\left[\sum_{t=1}^{T} \tilde{\mu}_K \,\middle|\, \mathcal{T}=T\right]\ \right] = \mathbb{E}\left[\ \tilde{\mu}_K \cdot \mathcal{T}\ \right] = \tilde{\mu}_K \cdot \mathbb{E}[\mathcal{T}].
$$

Note that $\mathrm{KL}(x,1-x) \geq \log(1/2.4x)$ for any $x \in [0,1]$, we complete the proof of Theorem 4.8.

D.18. Proof of Proposition C.1

Proposition C.1. *Assume $K' \geq 2K-1$. With probability at least $1-\delta$, Algorithm 1 outputs an ϵ-optimal arm after at most $(c_1 N_1' + c_2 N_2')$ steps where*

$$
\begin{aligned}
N_1' &= \frac{2v_K^2}{\mu_K^2} \log\left(\frac{2}{\delta}\right) = O\left(\frac{v_K^2}{\mu_K^2} \log\left(\frac{2}{\delta}\right)\right), \\
N_2' &= \frac{2}{\mu_K} \left[\sum_{i=1}^{L-K'+K-1} \bar{T}_{\sigma(i)} + (K'-K+1)\bar{T}_{\sigma(L-K'+K)} + (K'-K) \right] \\
&= O\left(\frac{1}{\mu_K} \left\{ \sum_{i=1}^{L-K'+K-1} \bar{\Delta}^{-2}_{\sigma(i)} \log\left[\frac{L}{\delta} \log\left(\frac{L}{\delta \bar{\Delta}^2_{\sigma(i)}} \right) \right] + (K'-K+1)\bar{\Delta}^{-2}_{\sigma(L-K'+K)} \log\left[\frac{L}{\delta} \log\left(\frac{L}{\delta \bar{\Delta}^2_{\sigma(L-K'+K)}} \right) \right] \right\} \right).
\end{aligned}
$$

Proof. Consider $K' \geq 2K-1$, i.e, $K'-K \geq K-1$. According to Lemma 5.9, there are at least $K'-K+1 \geq K$ items in the survival set D_t before the algorithm terminates, so the algorithm pulls K items from the surviving set D_t at each time step. And for simplicity, we again write $\mu(k,w)$ as μ_k, $v(k,w)$ as v_k, $\bar{T}_{i,\delta}$ as $\bar{T}_i$, $\rho(\delta)$ as ρ.

Recall Lemma 5.6, we set $\delta_0 = \delta/2$, $k=K$, $n=t_0'$, $\rho' = -\sqrt{-2t_0' v_K^2 \log(\delta/2)}$. Then the total number of observations during t_0' steps should be larger than $t_0'\mu_K + \rho'$ with probability at least $1-\delta/2$. And since the number of observations can

be upper bounded, we consider

$$t_0'\mu_K + \rho' \leq \sum_{i=1}^{L-K'+K} \bar{T}_{\sigma(i)} + \sum_{i=L-K'+K+1}^{L} T_{t_0'}(\sigma(i)) \leq \sum_{i=1}^{L-K'+K} \bar{T}_{\sigma(i)} + (K'-K)\left(\bar{T}_{\sigma(L-K'+K)} + 1\right)$$
$$= \sum_{i=1}^{L-K'+K-1} \bar{T}_{\sigma(i)} + (K'-K+1)\bar{T}_{\sigma(L-K'+K)} + (K'-K) := \tilde{T}_0.$$

Lastly, with Lemma 5.6, 5.7 and 5.8, we obtain that with probability at least $1-\delta$, Algorithm 1 stops after at most

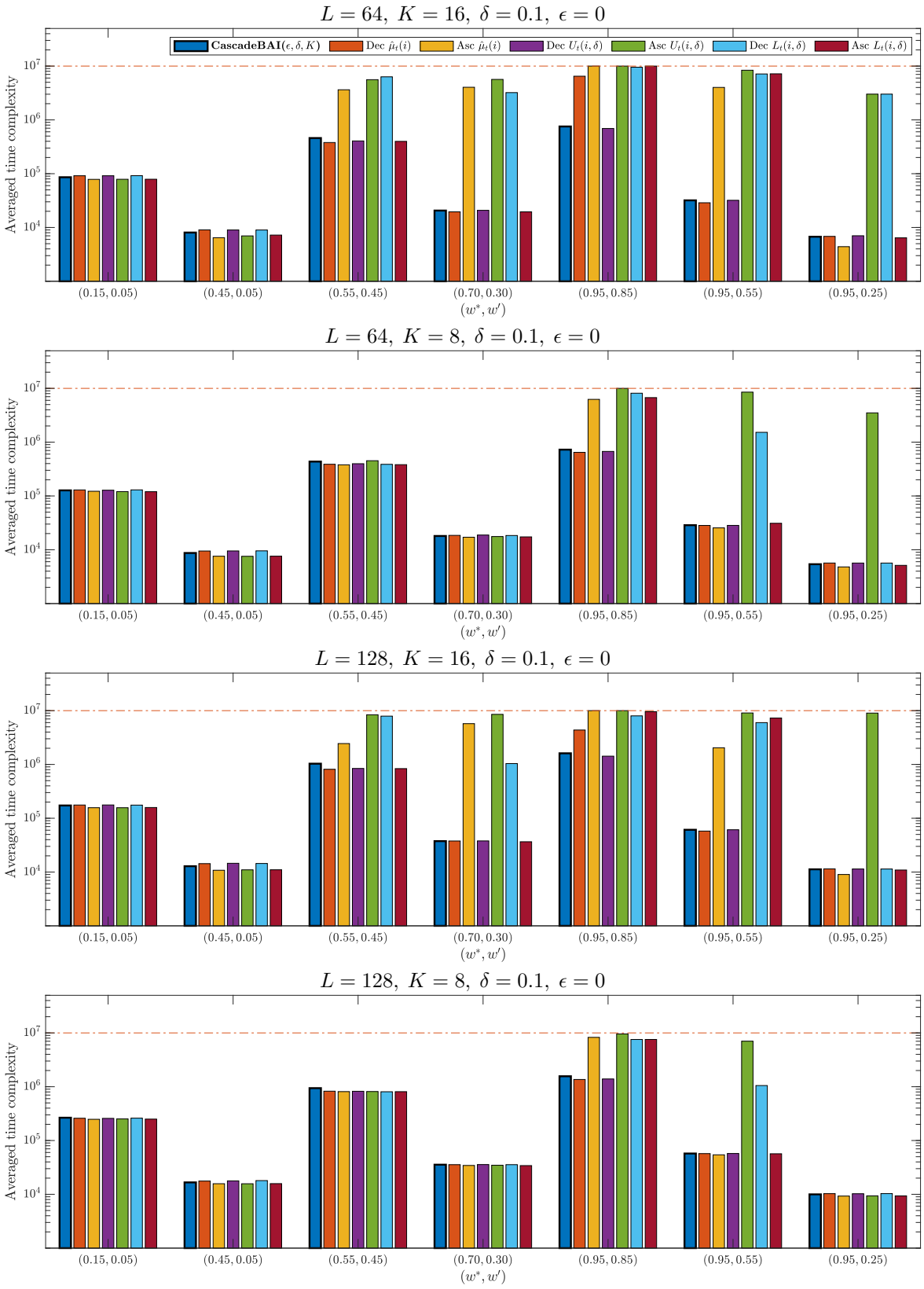
$$\frac{2\tilde{T}_0}{\mu_K} + \frac{2v_K^2}{\mu_K^2}\log\left(\frac{2}{\delta}\right) = N_2' + N_1'$$

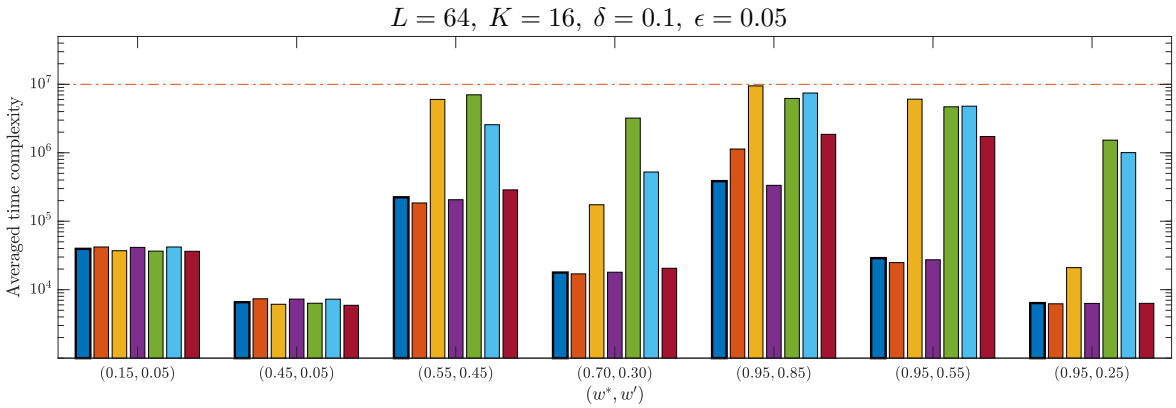
steps.

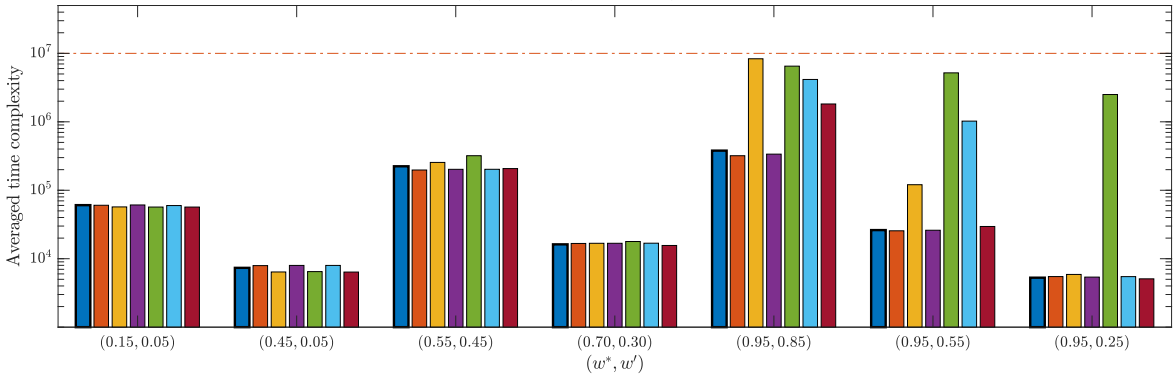
□

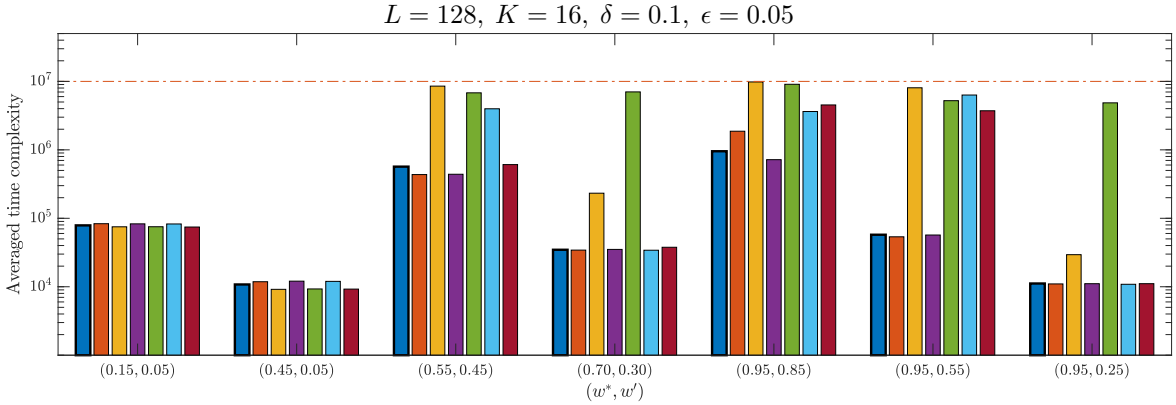
E. Additional numerical results

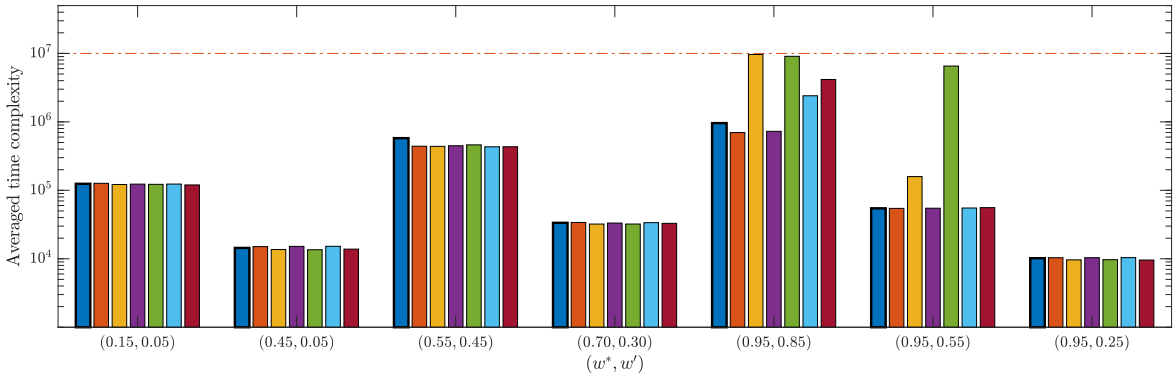
E.1. Order of pulled items

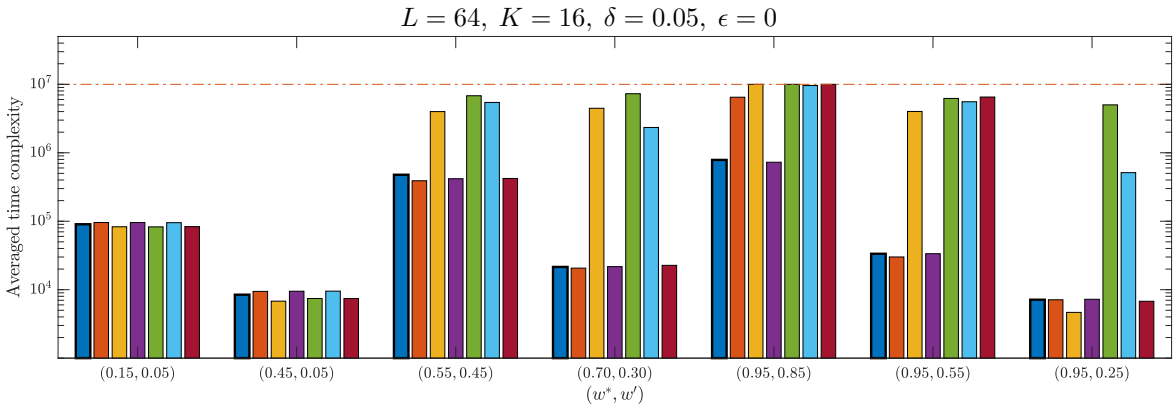


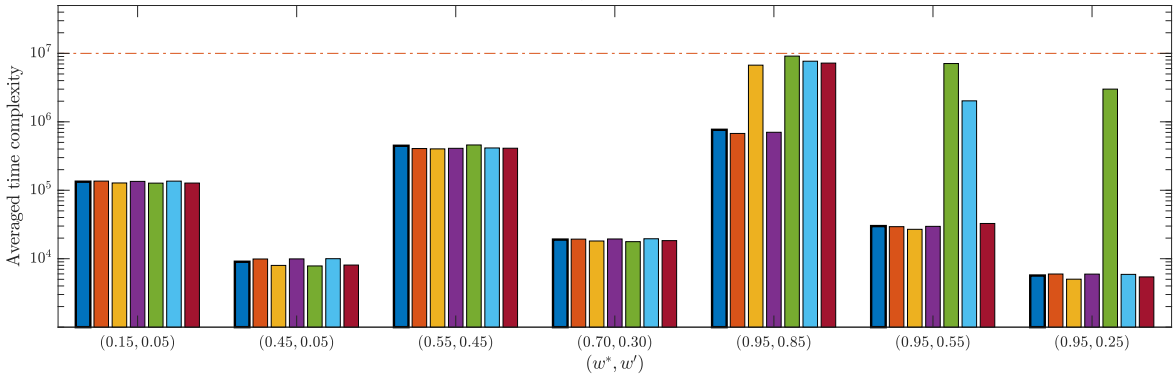

$L = 64,\ K = 16,\ \delta = 0.1,\ \epsilon = 0.05$
Averaged time complexity
10^4
10^5
10^6
10^7
$(0.15, 0.05)$
$(0.45, 0.05)$
$(0.55, 0.45)$
$(0.70, 0.30)$
$(0.95, 0.85)$
$(0.95, 0.55)$
$(0.95, 0.25)$
(w^*, w')

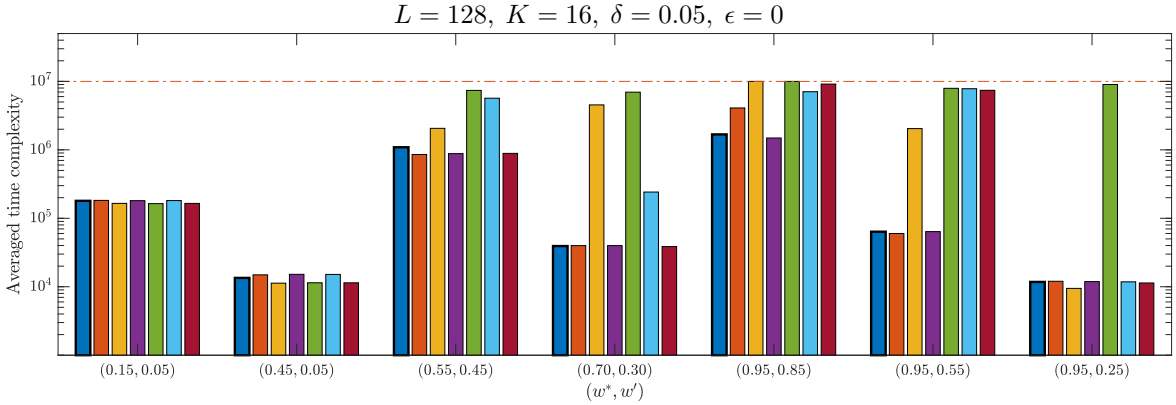

$L = 64,\ K = 8,\ \delta = 0.1,\ \epsilon = 0.05$
Averaged time complexity
10^4
10^5
10^6
10^7
$(0.15, 0.05)$
$(0.45, 0.05)$
$(0.55, 0.45)$
$(0.70, 0.30)$
$(0.95, 0.85)$
$(0.95, 0.55)$
$(0.95, 0.25)$
(w^*, w')

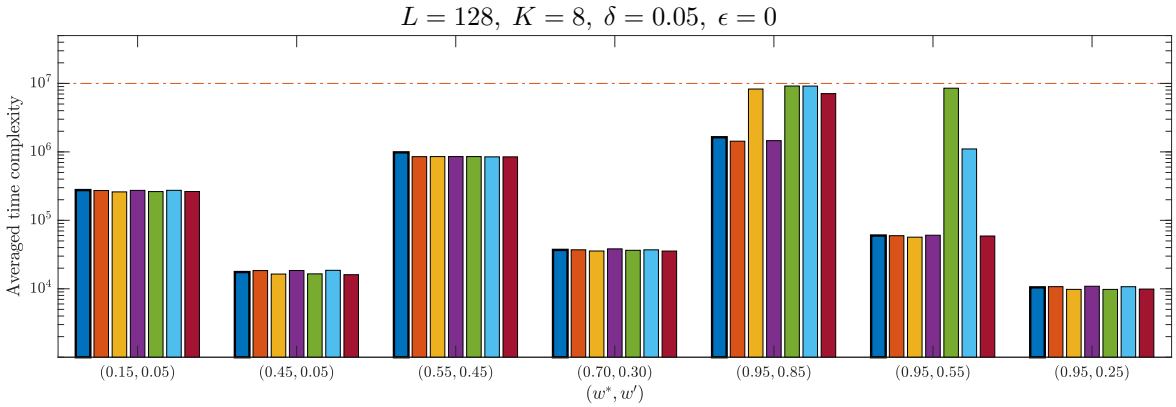

$L = 128,\ K = 16,\ \delta = 0.1,\ \epsilon = 0.05$
Averaged time complexity
10^4
10^5
10^6
10^7
$(0.15, 0.05)$
$(0.45, 0.05)$
$(0.55, 0.45)$
$(0.70, 0.30)$
$(0.95, 0.85)$
$(0.95, 0.55)$
$(0.95, 0.25)$
(w^*, w')


$L = 128,\ K = 8,\ \delta = 0.1,\ \epsilon = 0.05$
Averaged time complexity
10^4
10^5
10^6
10^7
$(0.15, 0.05)$
$(0.45, 0.05)$
$(0.55, 0.45)$
$(0.70, 0.30)$
$(0.95, 0.85)$
$(0.95, 0.55)$
$(0.95, 0.25)$
(w^*, w')


$L = 64,\ K = 16,\ \delta = 0.05,\ \epsilon = 0$
Averaged time complexity
10^4
10^5
10^6
10^7
$(0.15, 0.05)$
$(0.45, 0.05)$
$(0.55, 0.45)$
$(0.70, 0.30)$
$(0.95, 0.85)$
$(0.95, 0.55)$
$(0.95, 0.25)$
(w^*, w')


$L = 64,\ K = 8,\ \delta = 0.05,\ \epsilon = 0$
Averaged time complexity
10^4
10^5
10^6
10^7
$(0.15, 0.05)$
$(0.45, 0.05)$
$(0.55, 0.45)$
$(0.70, 0.30)$
$(0.95, 0.85)$
$(0.95, 0.55)$
$(0.95, 0.25)$
(w^*, w')


$L = 128,\ K = 16,\ \delta = 0.05,\ \epsilon = 0$
Averaged time complexity
10^4
10^5
10^6
10^7
$(0.15, 0.05)$
$(0.45, 0.05)$
$(0.55, 0.45)$
$(0.70, 0.30)$
$(0.95, 0.85)$
$(0.95, 0.55)$
$(0.95, 0.25)$
(w^*, w')


$L = 128,\ K = 8,\ \delta = 0.05,\ \epsilon = 0$
Averaged time complexity
10^4
10^5
10^6
10^7
$(0.15, 0.05)$
$(0.45, 0.05)$
$(0.55, 0.45)$
$(0.70, 0.30)$
$(0.95, 0.85)$
$(0.95, 0.55)$
$(0.95, 0.25)$
(w^*, w')

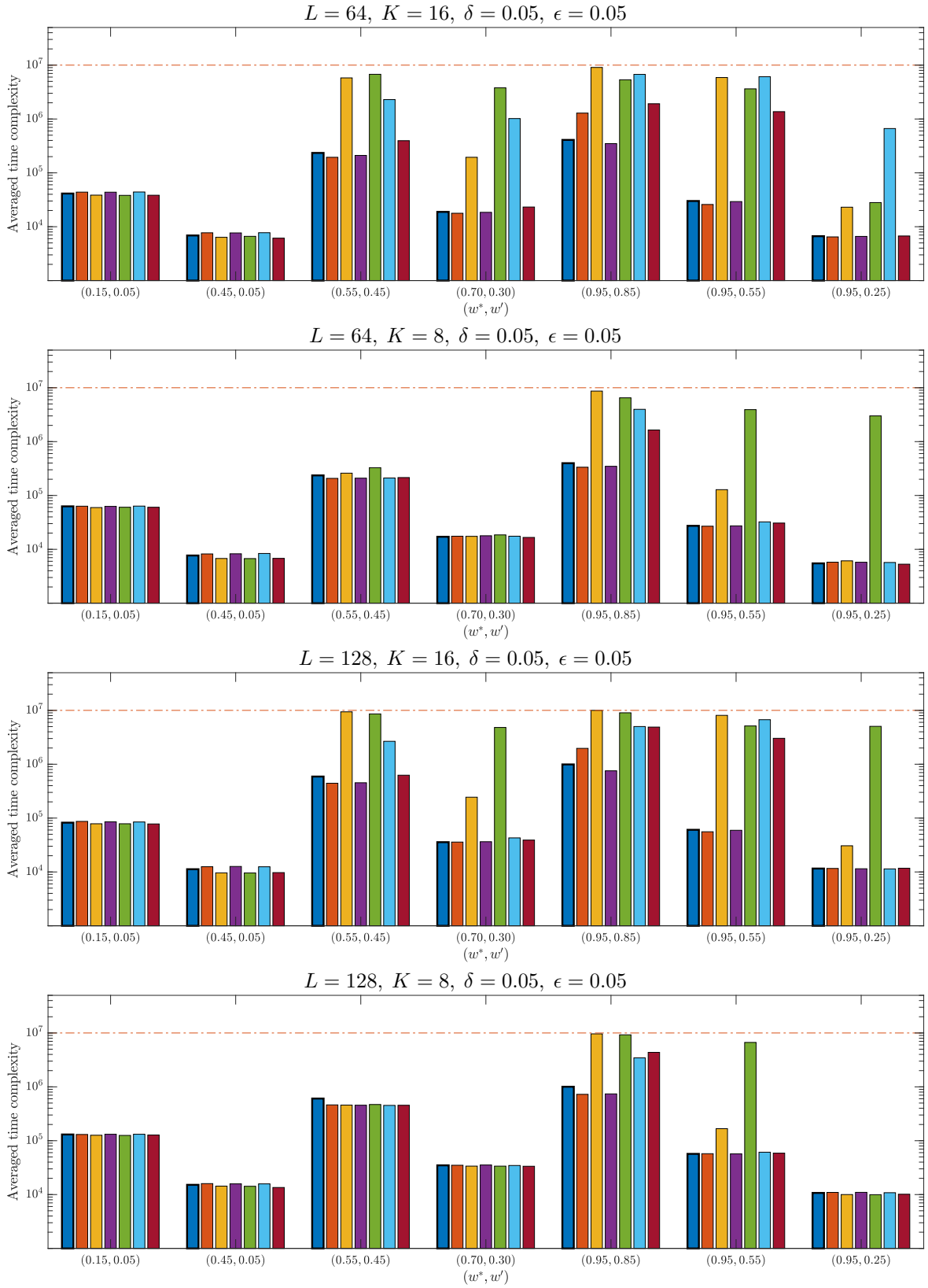


Figure E.1: Average time complexity incurred by different sorting order of S_t: ascending order of $T_i(t)$ (Algorithm 1), ascending/descending order of $\hat{\mu}_t(i)/U_t(i)/L_t(i)$ in the cascading bandits.

After a large amount of observations, it is likely that the empirical mean $\hat{w}_t(i)$ approaches the true weight $w(i)$, and $w(i)$ lies between the confidence bounds $U_t(i,\delta)$ and $L_t(i,\delta)$ with high probability. Therefore, one may consider to sort S_t in the descending or ascending order of $\hat{w}_t(i)$'s, $U_t(i,\delta)$'s or $L_t(i,\delta)$'s (the difference to Algorithm 1 reveals in Line 5–9). Diving into the numerical results, we found an algorithm always manages to find an ϵ-optimal arm provided that it is not terminated by the limit of 10^7 steps. Hence, we focus on the comparison of averaged stopping time.

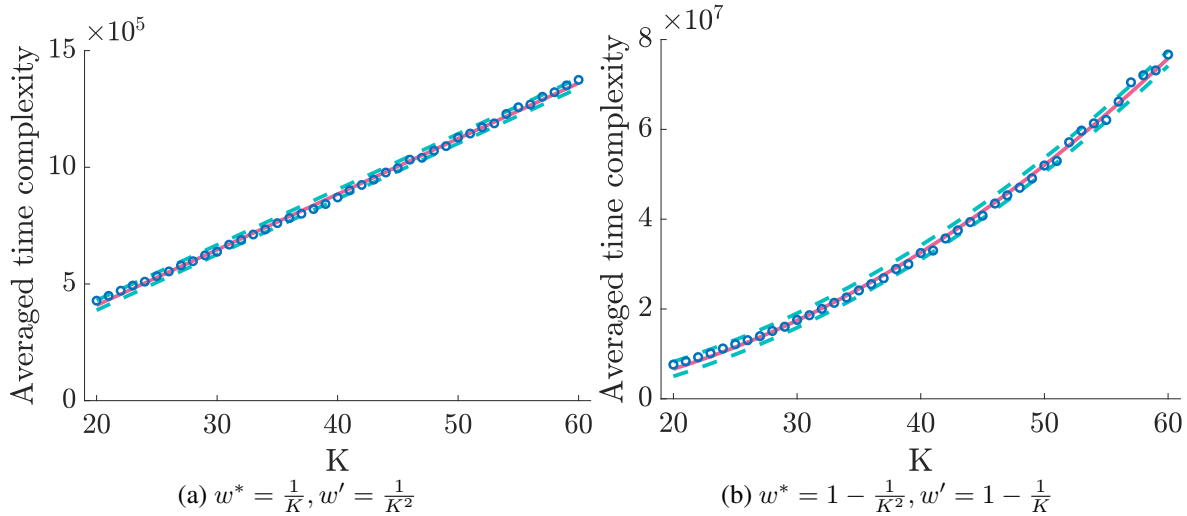
In Figure E.1, we can see that sorting S_t in the ascending order of $\hat{\mu}_t(i)$ or $U_t(i)$, especially the latter one, incurs an apparently larger averaged stopping time than other methods in most cases. Next, the descending order of $\hat{\mu}_t(i)$ does not work well in some cases. Thirdly, the ascending order of $L_t(i)$ performs almost the same as our algorithm in most cases but there are several cases where it performs much worse and does not terminate even after 10^7 iterations. Lastly, the descending order of $U_t(i)$ works almost as well as Algorithm 1 empirically but is in lack of theoretical guarantee on time complexity. Meanwhile, the standard deviation of the stopping time of our algorithm is negligible comparing to the average value. For instance, in the left-most case of Figure 6.1, the standard deviation is about 22318.54 when the average is about 754140.65.

E.2. Further empirical evidence

Table E.1: Fitted results of upper bounds on the stopping time $\mathcal{T}$ of Algorithm 1 with $\epsilon = 0$ (Proposition 4.6).

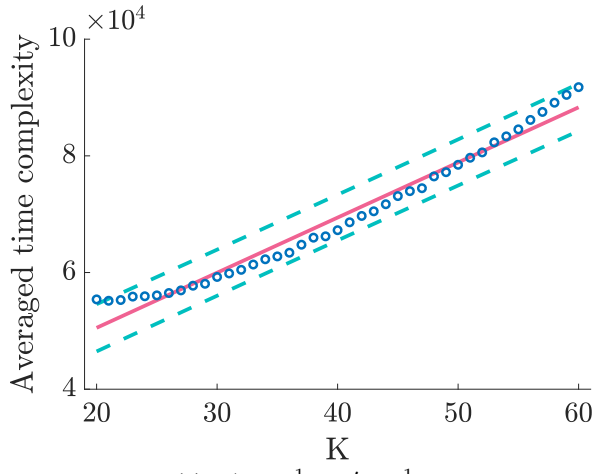
No.	w^*	w'	Fitting model	c_1	c_2	R^2-statistic	p-value
1	$1/K$	$1/K^2$	$c_1K + c_2$	23802.95	67400.19	0.9988	1.39×10^{-58}
2	$1 - 1/K^2$	$1 - 1/K$	$c_1K^2 + c_2$	21615.50	2007597.07	0.9987	1.29×10^{-57}
3	$1/\sqrt{K}$	$1/K$	$c_1K + c_2$	944.82	31626.49	0.9729	3.58×10^{-32}
4	$1 - 1/K$	$1 - 1/\sqrt{K}$	$c_1K + c_2$	23343.29	8823.27	0.9995	3.00×10^{-65}
5	$1 - 1/K$	$1/K$	$c_1K^2 + c_2$	1.22	3414.56	0.9917	3.03×10^{-42}

As shown in Table E.1, p-value is the probability that we reject the assumption of our fitting model versus a constant model (Glantz et al., 1990). Hence, the small p-values indicates that our fitting models are reasonable. Next, all c_1's are positive, implying all averaged stopping time grows with K, which corroborates our theoretical results.

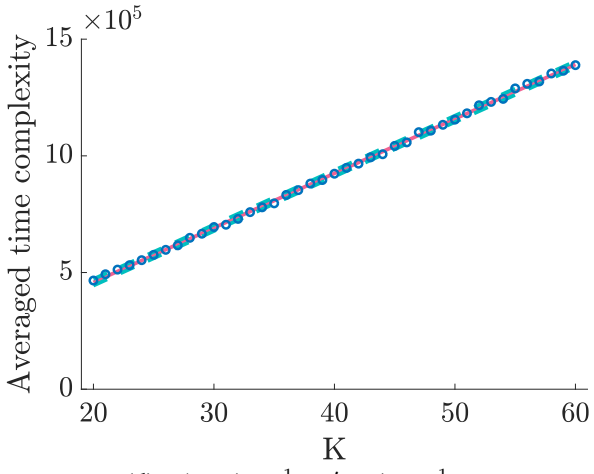


(a) $w^* = \frac{1}{K}, w' = \frac{1}{K^2}$

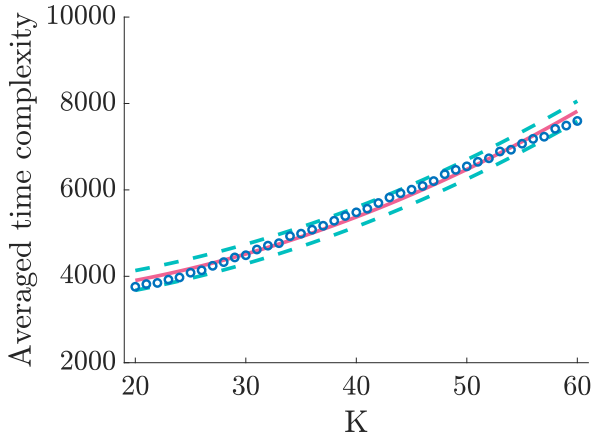
(b) $w^* = 1 - \frac{1}{K^2}, w' = 1 - \frac{1}{K}$



(c) $w^* = \frac{1}{\sqrt{K}}, w' = \frac{1}{K}$



(d) $w^* = 1 - \frac{1}{K}, w' = 1 - \frac{1}{\sqrt{K}}$



(e) $w^* = 1 - \frac{1}{K}, w' = \frac{1}{K}$

Figure E.2: Fit the averaged stopping time with functions of K for each case in order. Fix $L = 128$, $\delta = 0.1$, $\epsilon = 0$. Blue dots are the averaged stopping time, red line is the fitted curve, and cyan dashed lines show the 95% confidence interval.