A. Notation

Symbol	Definition
$\alpha(d)$	Attraction probability of item d
$\alpha_{\max}$	Highest attraction probability, $\alpha(1)$
A	Binary attraction vector, where $A(d)$ is the attraction indicator of item d
P_α	Distribution over binary attraction vectors
$\mathcal{A}$	Set of active batches
$b_{\max}$	Index of the last created batch
$\boldsymbol{B}_{b,\ell}$	Items in stage ℓ of batch b
$\boldsymbol{c}_t(k)$	Indicator of the click on position k at time t
$\boldsymbol{c}_{b,\ell}(d)$	Number of observed clicks on item d in stage ℓ of batch b
$\hat{\boldsymbol{c}}_{b,\ell}(d)$	Estimated probability of clicking on item d in stage ℓ of batch b
$\bar{\boldsymbol{c}}_{b,\ell}(d)$	Probability of clicking on item d in stage ℓ of batch b, $\mathbb{E}\left[\hat{\boldsymbol{c}}_{b,\ell}(d)\right]$
$\mathcal{D}$	Ground set of items $[L]$ such that $\alpha(1) \geq \ldots \geq \alpha(L)$
δ_T	$\log T + 3 \log\log T$
$\tilde{\Delta}_\ell$	$2^{-\ell}$
$\boldsymbol{I}_b$	Interval of positions in batch b
K	Number of positions to display items
$\mathrm{len}(b)$	Number of positions to display items in batch b
L	Number of items
$\boldsymbol{L}_{b,\ell}(d)$	Lower confidence bound of item d, in stage ℓ of batch b
n_ℓ	Number of times that each item is observed in stage ℓ
$\boldsymbol{n}_{b,\ell}$	Number of observations of item d in stage ℓ of batch b
$\Pi_K(\mathcal{D})$	Set of all K-tuples with distinct elements from $\mathcal{D}$
$r(\mathcal{R}, A, X)$	Reward of list $\mathcal{R}$, for attraction and examination indicators A and X
$r(\mathcal{R}, \alpha, \chi)$	Expected reward of list $\mathcal{R}$
$\mathcal{R} = (d_1, \ldots, d_K)$	List of K items, where d_k is the k-th item in $\mathcal{R}$
$\mathcal{R}^* = (1, \ldots, K)$	Optimal list of K items
$R(\mathcal{R}, A, X)$	Regret of list $\mathcal{R}$, for attraction and examination indicators A and X
$R(T)$	Expected cumulative regret in T steps
T	Horizon of the experiment
$\boldsymbol{U}_{b,\ell}(d)$	Upper confidence bound of item d, in stage ℓ of batch b
$\chi(\mathcal{R}, k)$	Examination probability of position k in list $\mathcal{R}$
$\chi^*(k)$	Examination probability of position k in the optimal list $\mathcal{R}^*$
X	Binary examination matrix, where $X(\mathcal{R}, k)$ is the examination indicator of position k in list $\mathcal{R}$
P_χ	Distribution over binary examination matrices

B. Proof of Theorem 1

Let $\boldsymbol{R}_{b,\ell}$ be the stochastic regret associated with stage ℓ of batch b. Then the expected T-step regret of `MergeRank` can be decomposed as

$$R(T) \leq \mathbb{E}\left[\sum_{b=1}^{2K}\sum_{\ell=0}^{T-1}\boldsymbol{R}_{b,\ell}\right]$$

because the maximum number of batches is $2K$. Let

$$\bar{\boldsymbol{\chi}}_{b,\ell}(d) = \frac{\bar{\boldsymbol{c}}_{b,\ell}(d)}{\alpha(d)} \tag{12}$$

be the *average examination probability* of item d in stage ℓ of batch b. Let

$$\begin{aligned}\mathcal{E}_{b,\ell} = \Big\{&\text{Event 1: } \forall d \in \boldsymbol{B}_{b,\ell} : \bar{\boldsymbol{c}}_{b,\ell}(d) \in [\boldsymbol{L}_{b,\ell}(d), \boldsymbol{U}_{b,\ell}(d)]\,,\\ &\text{Event 2: } \forall \boldsymbol{I}_b \in [K]^2,\ d \in \boldsymbol{B}_{b,\ell},\ d^* \in \boldsymbol{B}_{b,\ell} \cap [K] \text{ s.t. } \Delta = \alpha(d^*) - \alpha(d) > 0 :\\ &\qquad n_\ell \geq \frac{16K}{\chi^*(\boldsymbol{I}_b(1))(1-\alpha_{\max})\Delta^2}\log T \implies \hat{\boldsymbol{c}}_{b,\ell}(d) \leq \bar{\boldsymbol{\chi}}_{b,\ell}(d)[\alpha(d) + \Delta/4]\,,\\ &\text{Event 3: } \forall \boldsymbol{I}_b \in [K]^2,\ d \in \boldsymbol{B}_{b,\ell},\ d^* \in \boldsymbol{B}_{b,\ell} \cap [K] \text{ s.t. } \Delta = \alpha(d^*) - \alpha(d) > 0 :\\ &\qquad n_\ell \geq \frac{16K}{\chi^*(\boldsymbol{I}_b(1))(1-\alpha_{\max})\Delta^2}\log T \implies \hat{\boldsymbol{c}}_{b,\ell}(d^*) \geq \bar{\boldsymbol{\chi}}_{b,\ell}(d^*)[\alpha(d^*) - \Delta/4]\Big\}\end{aligned}$$

be the "good event" in stage ℓ of batch b, where $\bar{\boldsymbol{c}}_{b,\ell}(d)$ is the probability of clicking on item d in stage ℓ of batch b, which is defined in (8); $\hat{\boldsymbol{c}}_{b,\ell}(d)$ is its estimate, which is defined in (7); and both χ^* and $\alpha_{\max}$ are defined in Section 5.3. Let $\overline{\mathcal{E}_{b,\ell}}$ be the complement of event $\mathcal{E}_{b,\ell}$. Let $\mathcal{E}$ be the "good event" that all events $\mathcal{E}_{b,\ell}$ happen; and $\overline{\mathcal{E}}$ be its complement, the "bad event" that at least one event $\mathcal{E}_{b,\ell}$ does not happen. Then the expected T-step regret can be bounded from above as

$$R(T) \leq \mathbb{E}\left[\sum_{b=1}^{2K}\sum_{\ell=0}^{T-1}\boldsymbol{R}_{b,\ell}\mathbb{1}\{\mathcal{E}\}\right] + TP(\overline{\mathcal{E}}) \leq \sum_{b=1}^{2K}\mathbb{E}\left[\sum_{\ell=0}^{T-1}\boldsymbol{R}_{b,\ell}\mathbb{1}\{\mathcal{E}\}\right] + 4KL(3e+K)\,,$$

where the second inequality is from Lemma 2. Now we apply Lemma 7 to each batch b and get that

$$\sum_{b=1}^{2K}\mathbb{E}\left[\sum_{\ell=0}^{T-1}\boldsymbol{R}_{b,\ell}\mathbb{1}\{\mathcal{E}\}\right] \leq \frac{192K^3L}{(1-\alpha_{\max})\Delta_{\min}}\log T\,.$$

This concludes our proof.

C. Upper Bound on the Probability of Bad Event $\overline{\mathcal{E}}$

Lemma 2. *Let $\overline{\mathcal{E}}$ be defined as in the proof of Theorem 1 and $T \geq 5$. Then*

$$P(\overline{\mathcal{E}}) \leq \frac{4KL(3e + K)}{T} \,.$$

Proof. By the union bound,

$$P(\overline{\mathcal{E}}) \leq \sum_{b=1}^{2K} \sum_{\ell=0}^{T-1} P(\overline{\mathcal{E}_{b,\ell}}) \,.$$

Now we bound the probability of each event in $\overline{\mathcal{E}_{b,\ell}}$ and then sum them up.

Event 1

The probability that event 1 in $\mathcal{E}_{b,\ell}$ does not happen is bounded as follows. Fix $\boldsymbol{I}_b$ and $\boldsymbol{B}_{b,\ell}$. For any $d \in \boldsymbol{B}_{b,\ell}$,

$$\begin{aligned} P(\bar{\boldsymbol{c}}_{b,\ell}(d) \notin [\boldsymbol{L}_{b,\ell}(d), \boldsymbol{U}_{b,\ell}(d)]) &\leq P(\bar{\boldsymbol{c}}_{b,\ell}(d) < \boldsymbol{L}_{b,\ell}(d)) + P(\bar{\boldsymbol{c}}_{b,\ell}(d) > \boldsymbol{U}_{b,\ell}(d)) \\ &\leq \frac{2e \left\lceil \log(T \log^3 T) \log n_\ell \right\rceil}{T \log^3 T} \\ &\leq \frac{2e \left\lceil \log^2 T + \log(\log^3 T) \log T \right\rceil}{T \log^3 T} \\ &\leq \frac{2e \left\lceil 2 \log^2 T \right\rceil}{T \log^3 T} \\ &\leq \frac{6e}{T \log T} \,, \end{aligned}$$

where the second inequality is by Theorem 10 of Garivier & Cappe (2011), the third inequality is from $T \geq n_\ell$, the fourth inequality is from $\log(\log^3 T) \leq \log T$ for $T \geq 5$, and the last inequality is from $\lceil 2 \log^2 T \rceil \leq 3 \log^2 T$ for $T \geq 3$. By the union bound,

$$P(\exists d \in \boldsymbol{B}_{b,\ell} \text{ s.t. } \bar{\boldsymbol{c}}_{b,\ell}(d) \notin [\boldsymbol{L}_{b,\ell}(d), \boldsymbol{U}_{b,\ell}(d)]) \leq \frac{6eL}{T \log T}$$

for any $\boldsymbol{B}_{b,\ell}$. Finally, since the above inequality holds for any $\boldsymbol{B}_{b,\ell}$, the probability that event 1 in $\mathcal{E}_{b,\ell}$ does not happen is bounded as above.

Event 2

The probability that event 2 in $\mathcal{E}_{b,\ell}$ does not happen is bounded as follows. Fix $\boldsymbol{I}_b$ and $\boldsymbol{B}_{b,\ell}$, and let $k = \boldsymbol{I}_b(1)$. If the event does not happen for items d and d^*, then it must be true that

$$n_\ell \geq \frac{16K}{\chi^*(k)(1 - \alpha_{\max})\Delta^2} \log T \,, \quad \hat{\boldsymbol{c}}_{b,\ell}(d) > \bar{\boldsymbol{\chi}}_{b,\ell}(d)[\alpha(d) + \Delta/4] \,.$$

From the definition of the average examination probability in (12) and a variant of Hoeffding's inequality in Lemma 8, we have that

$$P\left(\hat{\boldsymbol{c}}_{b,\ell}(d) > \bar{\boldsymbol{\chi}}_{b,\ell}(d)[\alpha(d) + \Delta/4]\right) \leq \exp\left[-n_\ell D_{\mathrm{KL}}(\bar{\boldsymbol{\chi}}_{b,\ell}(d)[\alpha(d) + \Delta/4] \,\|\, \bar{\boldsymbol{c}}_{b,\ell}(d))\right] \,.$$

From Lemma 9, $\bar{\boldsymbol{\chi}}_{b,\ell}(d) \geq \chi^*(k)/K$ (Lemma 3), and Pinsker's inequality, we have that

$$\begin{aligned} \exp\left[-n_\ell D_{\mathrm{KL}}(\bar{\boldsymbol{\chi}}_{b,\ell}(d)[\alpha(d) + \Delta/4] \,\|\, \bar{\boldsymbol{c}}_{b,\ell}(d))\right] &\leq \exp\left[-n_\ell \bar{\boldsymbol{\chi}}_{b,\ell}(d)(1 - \alpha_{\max}) D_{\mathrm{KL}}(\alpha(d) + \Delta/4 \,\|\, \alpha(d))\right] \\ &\leq \exp\left[-n_\ell \frac{\chi^*(k)(1 - \alpha_{\max})\Delta^2}{8K}\right] \,. \end{aligned}$$

From our assumption on n_ℓ, we conclude that

$$\exp\left[-n_\ell \frac{\chi^*(k)(1-\alpha_{\max})\Delta^2}{8K}\right] \leq \exp[-2\log T] = \frac{1}{T^2}\,.$$

Finally, we chain all above inequalities and get that event 2 in $\mathcal{E}_{b,\ell}$ does not happen for any fixed $\boldsymbol{I}_b$, $\boldsymbol{B}_{b,\ell}$, d, and d^* with probability of at most T^{-2}. Since the maximum numbers of items d and d^* are L and K, respectively, the event does not happen for any fixed $\boldsymbol{I}_b$ and $\boldsymbol{B}_{b,\ell}$ with probability of at most KLT^{-2}. In turn, the probability that event 2 in $\mathcal{E}_{b,\ell}$ does not happen is bounded by KLT^{-2}.

Event 3

This bound is analogous to that of event 2.

Total probability

The maximum number of stages in any batch in `BatchRank` is $\log T$ and the maximum number of batches is $2K$. Hence, by the union bound,

$$P(\overline{\mathcal{E}}) \leq \left(\frac{6eL}{T\log T} + \frac{KL}{T^2} + \frac{KL}{T^2}\right)(2K\log T) \leq \frac{4KL(3e+K)}{T}\,.$$

This concludes our proof. ■

D. Upper Bound on the Regret in Individual Batches

Lemma 3. *For any batch b, positions $\boldsymbol{I}_b$, stage ℓ, set $\boldsymbol{B}_{b,\ell}$, and item $d \in \boldsymbol{B}_{b,\ell}$,*

$$\frac{\chi^*(k)}{K} \leq \bar{\boldsymbol{\chi}}_{b,\ell}(d)\,,$$

where $k = \boldsymbol{I}_b(1)$ is the highest position in batch b.

Proof. The proof follows from two observations. First, by Assumption 6, $\chi^*(k)$ is the lowest examination probability of position k. Second, by the design of `DisplayBatch`, item d is placed at position k with probability of at least $1/K$. ■

Lemma 4. *Let event $\mathcal{E}$ happen and $T \geq 5$. For any batch b, positions $\boldsymbol{I}_b$, set $\boldsymbol{B}_{b,0}$, item $d \in \boldsymbol{B}_{b,0}$, and item $d^* \in \boldsymbol{B}_{b,0} \cap [K]$ such that $\Delta = \alpha(d^*) - \alpha(d) > 0$, let $k = \boldsymbol{I}_b(1)$ be the highest position in batch b and ℓ be the first stage where*

$$\tilde{\Delta}_\ell < \sqrt{\frac{\chi^*(k)(1-\alpha_{\max})}{K}}\Delta\,.$$

Then $\boldsymbol{U}_{b,\ell}(d) < \boldsymbol{L}_{b,\ell}(d^*)$.

Proof. From the definition of n_ℓ in `BatchRank` and our assumption on $\tilde{\Delta}_\ell$,

$$n_\ell \geq \frac{16}{\tilde{\Delta}_\ell^2}\log T > \frac{16K}{\chi^*(k)(1-\alpha_{\max})\Delta^2}\log T\,. \tag{13}$$

Let $\mu = \bar{\boldsymbol{\chi}}_{b,\ell}(d)$ and suppose that $\boldsymbol{U}_{b,\ell}(d) \geq \mu[\alpha(d) + \Delta/2]$ holds. Then from this assumption, the definition of $\boldsymbol{U}_{b,\ell}(d)$, and event 2 in $\mathcal{E}_{b,\ell}$,

$$\begin{aligned}D_{\mathrm{KL}}(\hat{\boldsymbol{c}}_{b,\ell}(d) \,\|\, \boldsymbol{U}_{b,\ell}(d)) &\geq D_{\mathrm{KL}}(\hat{\boldsymbol{c}}_{b,\ell}(d) \,\|\, \mu[\alpha(d) + \Delta/2])\,\mathbb{1}\{\hat{\boldsymbol{c}}_{b,\ell}(d) \leq \mu[\alpha(d) + \Delta/2]\}\\ &\geq D_{\mathrm{KL}}(\mu[\alpha(d) + \Delta/4] \,\|\, \mu[\alpha(d) + \Delta/2])\,.\end{aligned}$$

From Lemma 9, $\mu \geq \chi^*(k)/K$ (Lemma 3), and Pinsker's inequality, we have that

$$\begin{aligned}D_{\mathrm{KL}}(\mu[\alpha(d) + \Delta/4] \,\|\, \mu[\alpha(d) + \Delta/2]) &\geq \mu(1-\alpha_{\max})D_{\mathrm{KL}}(\alpha(d) + \Delta/4 \,\|\, \alpha(d) + \Delta/2)\\ &\geq \frac{\chi^*(k)(1-\alpha_{\max})\Delta^2}{8K}\,.\end{aligned}$$

From the definition of $\boldsymbol{U}_{b,\ell}(d)$, $T \geq 5$, and above inequalities,

$$n_\ell = \frac{\log T + 3\log\log T}{D_{\mathrm{KL}}(\hat{\boldsymbol{c}}_{b,\ell}(d) \,\|\, \boldsymbol{U}_{b,\ell}(d))} \leq \frac{2\log T}{D_{\mathrm{KL}}(\hat{\boldsymbol{c}}_{b,\ell}(d) \,\|\, \boldsymbol{U}_{b,\ell}(d))} \leq \frac{16K\log T}{\chi^*(k)(1-\alpha_{\max})\Delta^2}\,.$$

This contradicts to (13), and therefore it must be true that $\boldsymbol{U}_{b,\ell}(d) < \mu[\alpha(d) + \Delta/2]$ holds.

On the other hand, let $\mu^* = \bar{\boldsymbol{\chi}}_{b,\ell}(d^*)$ and suppose that $\boldsymbol{L}_{b,\ell}(d^*) \leq \mu^*[\alpha(d^*) - \Delta/2]$ holds. Then from this assumption, the definition of $\boldsymbol{L}_{b,\ell}(d^*)$, and event 3 in $\mathcal{E}_{b,\ell}$,

$$\begin{aligned}D_{\mathrm{KL}}(\hat{\boldsymbol{c}}_{b,\ell}(d^*) \,\|\, \boldsymbol{L}_{b,\ell}(d^*)) &\geq D_{\mathrm{KL}}(\hat{\boldsymbol{c}}_{b,\ell}(d^*) \,\|\, \mu^*[\alpha(d^*) - \Delta/2])\,\mathbb{1}\{\hat{\boldsymbol{c}}_{b,\ell}(d^*) \geq \mu^*[\alpha(d^*) - \Delta/2]\}\\ &\geq D_{\mathrm{KL}}(\mu^*[\alpha(d^*) - \Delta/4] \,\|\, \mu^*[\alpha(d^*) - \Delta/2])\,.\end{aligned}$$

From Lemma 9, $\mu^* \geq \chi^*(k)/K$ (Lemma 3), and Pinsker's inequality, we have that

$$\begin{aligned}D_{\mathrm{KL}}(\mu^*[\alpha(d^*) - \Delta/4] \,\|\, \mu^*[\alpha(d^*) - \Delta/2]) &\geq \mu^*(1-\alpha_{\max})D_{\mathrm{KL}}(\alpha(d^*) - \Delta/4 \,\|\, \alpha(d^*) - \Delta/2)\\ &\geq \frac{\chi^*(k)(1-\alpha_{\max})\Delta^2}{8K}\,.\end{aligned}$$

From the definition of $\boldsymbol{L}_{b,\ell}(d^*)$, $T \geq 5$, and above inequalities,

$$n_\ell = \frac{\log T + 3\log\log T}{D_{\mathrm{KL}}(\hat{\boldsymbol{c}}_{b,\ell}(d) \,\|\, \boldsymbol{L}_{b,\ell}(d^*))} \leq \frac{2\log T}{D_{\mathrm{KL}}(\hat{\boldsymbol{c}}_{b,\ell}(d^*) \,\|\, \boldsymbol{L}_{b,\ell}(d^*))} \leq \frac{16K\log T}{\chi^*(k)(1-\alpha_{\max})\Delta^2}\,.$$

This contradicts to (13), and therefore it must be true that $\boldsymbol{L}_{b,\ell}(d^*) > \mu^*[\alpha(d^*) - \Delta/2]$ holds.

Finally, based on inequality (11),

$$\mu^* = \frac{\bar{\boldsymbol{c}}_{b,\ell}(d^*)}{\alpha(d^*)} \geq \frac{\bar{\boldsymbol{c}}_{b,\ell}(d)}{\alpha(d)} = \mu\,,$$

and item d is guaranteed to be eliminated by the end of stage ℓ because

$$\begin{aligned}
\boldsymbol{U}_{b,\ell}(d) &< \mu[\alpha(d) + \Delta/2] \\
&\leq \mu\alpha(d) + \frac{\mu^*\alpha(d^*) - \mu\alpha(d)}{2} \\
&= \mu^*\alpha(d^*) - \frac{\mu^*\alpha(d^*) - \mu\alpha(d)}{2} \\
&\leq \mu^*[\alpha(d^*) - \Delta/2] \\
&< \boldsymbol{L}_{b,\ell}(d^*)\,.
\end{aligned}$$

This concludes our proof. ■

Lemma 5. *Let event $\mathcal{E}$ happen and $T \geq 5$. For any batch b, positions $\boldsymbol{I}_b$ where $\boldsymbol{I}_b(2) = K$, set $\boldsymbol{B}_{b,0}$, and item $d \in \boldsymbol{B}_{b,0}$ such that $d > K$, let $k = \boldsymbol{I}_b(1)$ be the highest position in batch b and ℓ be the first stage where*

$$\tilde{\Delta}_\ell < \sqrt{\frac{\chi^*(k)(1 - \alpha_{\max})}{K}}\Delta$$

for $\Delta = \alpha(K) - \alpha(d)$. Then item d is eliminated by the end of stage ℓ.

Proof. Let $B^+ = \{k, \dots, K\}$. Now note that $\alpha(d^*) - \alpha(d) \geq \Delta$ for any $d^* \in B^+$. By Lemma 4, $\boldsymbol{L}_{b,\ell}(d^*) > \boldsymbol{U}_{b,\ell}(d)$ for any $d^* \in B^+$; and therefore item d is eliminated by the end of stage ℓ. ■

Lemma 6. *Let $\mathcal{E}$ happen and $T \geq 5$. For any batch b, positions $\boldsymbol{I}_b$, and set $\boldsymbol{B}_{b,0}$, let $k = \boldsymbol{I}_b(1)$ be the highest position in batch b and ℓ be the first stage where*

$$\tilde{\Delta}_\ell < \sqrt{\frac{\chi^*(k)(1 - \alpha_{\max})}{K}}\Delta_{\max}$$

for $\Delta_{\max} = \alpha(s) - \alpha(s+1)$ and $s = \operatorname*{arg\,max}_{d \in \{\boldsymbol{I}_b(1), \dots, \boldsymbol{I}_b(2)-1\}} [\alpha(d) - \alpha(d+1)]$. Then batch b is split by the end of stage ℓ.

Proof. Let $B^+ = \{k, \dots, s\}$ and $B^- = \boldsymbol{B}_{b,0} \setminus B^+$. Now note that $\alpha(d^*) - \alpha(d) \geq \Delta_{\max}$ for any $(d^*, d) \in B^+ \times B^-$. By Lemma 4, $\boldsymbol{L}_{b,\ell}(d^*) > \boldsymbol{U}_{b,\ell}(d)$ for any $(d^*, d) \in B^+ \times B^-$; and therefore batch b is split by the end of stage ℓ. ■

Lemma 7. *Let event $\mathcal{E}$ happen and $T \geq 5$. Then the expected T-step regret in any batch b is bounded as*

$$\mathbb{E}\left[\sum_{\ell=0}^{T-1} \boldsymbol{R}_{b,\ell}\right] \leq \frac{96K^2L}{(1 - \alpha_{\max})\Delta_{\max}} \log T\,.$$

Proof. Let $k = \boldsymbol{I}_b(1)$ be the highest position in batch b. Choose any item $d \in \boldsymbol{B}_{b,0}$ and let $\Delta = \alpha(k) - \alpha(d)$.

First, we show that the expected per-step regret of any item d is bounded by $\chi^*(k)\Delta$ when event $\mathcal{E}$ happens. Since event $\mathcal{E}$ happens, all eliminations and splits up to any stage ℓ of batch b are correct. Therefore, items $1, \dots, k-1$ are at positions $1, \dots, k-1$; and position k is examined with probability $\chi^*(k)$. Note that this is the highest examination probability in batch b (Assumption 4). Our upper bound follows from the fact that the reward is linear in individual items (Section 3.1).

We analyze two cases. First, suppose that $\Delta \leq 2K\Delta_{\max}$ for $\Delta_{\max}$ in Lemma 6. Then by Lemma 6, batch b splits when the number of steps in a stage is at most

$$\frac{16K}{\chi^*(k)(1 - \alpha_{\max})\Delta_{\max}^2} \log T\,.$$

By the design of `DisplayBatch`, any item in stage ℓ of batch b is displayed at most $2n_\ell$ times. Therefore, the maximum regret due to item d in the last stage before the split is

$$\frac{32K\chi^*(k)\Delta}{\chi^*(k)(1-\alpha_{\max})\Delta_{\max}^2}\log T \leq \frac{64K^2\Delta_{\max}}{(1-\alpha_{\max})\Delta_{\max}^2}\log T = \frac{64K^2}{(1-\alpha_{\max})\Delta_{\max}}\log T\,.$$

Now suppose that $\Delta > 2K\Delta_{\max}$. This implies that item d is easy to distinguish from item K. In particular,

$$\alpha(K) - \alpha(d) = \Delta - (\alpha(k) - \alpha(K)) \geq \Delta - K\Delta_{\max} \geq \frac{\Delta}{2}\,,$$

where the equality is from the identity

$$\Delta = \alpha(k) - \alpha(d) = \alpha(k) - \alpha(K) + \alpha(K) - \alpha(d)\,;$$

the first inequality is from $\alpha(k) - \alpha(K) \leq K\Delta_{\max}$, which follows from the definition of $\Delta_{\max}$ and $k \in [K]$; and the last inequality is from our assumption that $K\Delta_{\max} < \Delta/2$. Now we apply the derived inequality and, by Lemma 5 and from the design of `DisplayBatch`, the maximum regret due to item d in the stage where that item is eliminated is

$$\frac{32K\chi^*(k)\Delta}{\chi^*(k)(1-\alpha_{\max})(\alpha(K) - \alpha(d))^2}\log T \leq \frac{128K}{(1-\alpha_{\max})\Delta}\log T \leq \frac{64}{(1-\alpha_{\max})\Delta_{\max}}\log T\,.$$

The last inequality is from our assumption that $\Delta > 2K\Delta_{\max}$.

Because the lengths of the stages quadruple and `BatchRank` resets all click estimators at the beginning of each stage, the maximum expected regret due to any item d in batch b is at most 1.5 times higher than that in the last stage, and hence

$$\mathbb{E}\left[\sum_{\ell=0}^{T-1}\boldsymbol{R}_{b,\ell}\right] \leq \frac{96K^2\left|\boldsymbol{B}_{b,0}\right|}{(1-\alpha_{\max})\Delta_{\max}}\log T\,.$$

This concludes our proof. ■

E. Technical Lemmas

Lemma 8. *Let $(\boldsymbol{X}_1)_{i=1}^n$ be n i.i.d. Bernoulli random variables, $\bar{\boldsymbol{\mu}} = \sum_{i=1}^n \boldsymbol{X}_i$, and $\mu = \mathbb{E}[\bar{\boldsymbol{\mu}}]$. Then*

$$P(\bar{\boldsymbol{\mu}} \geq \mu + \varepsilon) \leq \exp[-n D_{\mathrm{KL}}(\mu + \varepsilon \,\|\, \mu)]$$

for any $\varepsilon \in [0, 1-\mu]$, and

$$P(\bar{\boldsymbol{\mu}} \leq \mu - \varepsilon) \leq \exp[-n D_{\mathrm{KL}}(\mu - \varepsilon \,\|\, \mu)]$$

for any $\varepsilon \in [0, \mu]$.

Proof. We only prove the first claim. The other claim follows from symmetry.

From inequality (2.1) of Hoeffding (1963), we have that

$$P(\bar{\boldsymbol{\mu}} \geq \mu + \varepsilon) \leq \left[\left(\frac{\mu}{\mu+\varepsilon}\right)^{\mu+\varepsilon} \left(\frac{1-\mu}{1-(\mu+\varepsilon)}\right)^{1-(\mu+\varepsilon)}\right]^n$$

for any $\varepsilon \in [0, 1-\mu]$. Now note that

$$\begin{aligned}
\left[\left(\frac{\mu}{\mu+\varepsilon}\right)^{\mu+\varepsilon} \left(\frac{1-\mu}{1-(\mu+\varepsilon)}\right)^{1-(\mu+\varepsilon)}\right]^n &= \exp\left[n\left[(\mu+\varepsilon)\log\frac{\mu}{\mu+\varepsilon} + (1-(\mu+\varepsilon))\log\frac{1-\mu}{1-(\mu+\varepsilon)}\right]\right] \\
&= \exp\left[-n\left[(\mu+\varepsilon)\log\frac{\mu+\varepsilon}{\mu} + (1-(\mu+\varepsilon))\log\frac{1-(\mu+\varepsilon)}{1-\mu}\right]\right] \\
&= \exp[-n D_{\mathrm{KL}}(\mu+\varepsilon \,\|\, \mu)]\,.
\end{aligned}$$

This concludes the proof. ■

Lemma 9. *For any $c, p, q \in [0, 1]$,*

$$c(1 - \max\{p, q\}) D_{\mathrm{KL}}(p \,\|\, q) \leq D_{\mathrm{KL}}(cp \,\|\, cq) \leq c D_{\mathrm{KL}}(p \,\|\, q)\ . \tag{14}$$

Proof. The proof is based on differentiation. The first two derivatives of $D_{\mathrm{KL}}(cp \,\|\, cq)$ with respect to q are

$$\frac{\partial}{\partial q} D_{\mathrm{KL}}(cp \,\|\, cq) = \frac{c(q-p)}{q(1-cq)}\,, \quad \frac{\partial^2}{\partial q^2} D_{\mathrm{KL}}(cp \,\|\, cq) = \frac{c^2(q-p)^2 + cp(1-cp)}{q^2(1-cq)^2}\,;$$

and the first two derivatives of $c D_{\mathrm{KL}}(p \,\|\, q)$ with respect to q are

$$\frac{\partial}{\partial q}[c D_{\mathrm{KL}}(p \,\|\, q)] = \frac{c(q-p)}{q(1-q)}\,, \quad \frac{\partial^2}{\partial q^2}[c D_{\mathrm{KL}}(p \,\|\, q)] = \frac{c(q-p)^2 + cp(1-p)}{q^2(1-q)^2}\,.$$

The second derivatives show that $D_{\mathrm{KL}}(cp \,\|\, cq)$ and $c D_{\mathrm{KL}}(p \,\|\, q)$ are convex in q for any p. Their minima are at $q = p$.

Now we fix p and c, and prove (14) for any q. The upper bound is derived as follows. Since

$$D_{\mathrm{KL}}(cp \,\|\, cx) = c D_{\mathrm{KL}}(p \,\|\, x) = 0$$

when $x = p$, the upper bound holds when $c D_{\mathrm{KL}}(p \,\|\, x)$ increases faster than $D_{\mathrm{KL}}(cp \,\|\, cx)$ for any $p < x \leq q$, and when $c D_{\mathrm{KL}}(p \,\|\, x)$ decreases faster than $D_{\mathrm{KL}}(cp \,\|\, cx)$ for any $q \leq x < p$. This follows from the definitions of $\frac{\partial}{\partial x} D_{\mathrm{KL}}(cp \,\|\, cx)$ and $\frac{\partial}{\partial x}[c D_{\mathrm{KL}}(p \,\|\, x)]$. In particular, both derivatives have the same sign and $\left|\frac{\partial}{\partial x} D_{\mathrm{KL}}(cp \,\|\, cx)\right| \leq \left|\frac{\partial}{\partial x}[c D_{\mathrm{KL}}(p \,\|\, x)]\right|$ for any feasible $x \in [\min\{p, q\}, \max\{p, q\}]$.

The lower bound is derived as follows. The ratio of $\frac{\partial}{\partial x}[c D_{\mathrm{KL}}(p \,\|\, x)]$ and $\frac{\partial}{\partial x} D_{\mathrm{KL}}(cp \,\|\, cx)$ is bounded from above as

$$\frac{\frac{\partial}{\partial x}[c D_{\mathrm{KL}}(p \,\|\, x)]}{\frac{\partial}{\partial x} D_{\mathrm{KL}}(cp \,\|\, cx)} = \frac{1-cx}{1-x} \leq \frac{1}{1-x} \leq \frac{1}{1-\max\{p, q\}}$$

for any $x \in [\min\{p, q\}, \max\{p, q\}]$. Therefore, we get a lower bound on $D_{\mathrm{KL}}(cp \,\|\, cx)$ when we multiply $c D_{\mathrm{KL}}(p \,\|\, x)$ by $1 - \max\{p, q\}$. ■