A LEMMAS

Lemma 3. *Let $\mathcal{R}$ be any list over $[K]$. Let*

$$\Delta(\mathcal{R}) = \sum_{k=1}^{K-1} \mathbb{1}\{\alpha(\mathcal{R}(k+1)) - \alpha(\mathcal{R}(k)) > 0\} \times (\alpha(\mathcal{R}(k+1)) - \alpha(\mathcal{R}(k))) \tag{8}$$

be the attraction gap *of list $\mathcal{R}$. Then the expected regret of $\mathcal{R}$ is bounded as*

$$\sum_{k=1}^{K} (\chi(\mathcal{R}^*, k)\alpha(k) - \chi(\mathcal{R}, k)\alpha(\mathcal{R}(k))) \leq K\chi_{\max}\Delta(\mathcal{R})\,.$$

Proof. Fix position $k \in [K]$. Then

$$\begin{aligned}\chi(\mathcal{R}^*, k)\alpha(k) - \chi(\mathcal{R}, k)\alpha(\mathcal{R}(k)) &\leq \chi(\mathcal{R}^*, k)(\alpha(k) - \alpha(\mathcal{R}(k))) \\ &\leq \chi_{\max}(\alpha(k) - \alpha(\mathcal{R}(k)))\,,\end{aligned}$$

where the first inequality follows from the fact that the examination probability of any position is the lowest in the optimal list (Assumption A5) and the second inequality follows from the definition of $\chi_{\max}$. In the rest of the proof, we bound $\alpha(k) - \alpha(\mathcal{R}(k))$. We consider three cases. First, let $\alpha(\mathcal{R}(k)) \geq \alpha(k)$. Then $\alpha(k) - \alpha(\mathcal{R}(k)) \leq 0$ and can be trivially bounded by $\Delta(\mathcal{R})$. Second, let $\alpha(\mathcal{R}(k)) < \alpha(k)$ and $\pi(k) > k$, where $\pi(k)$ is the position of item k in list $\mathcal{R}$. Then

$$\begin{aligned}\alpha(k) - \alpha(\mathcal{R}(k)) &= \alpha(\mathcal{R}(\pi(k))) - \alpha(\mathcal{R}(k)) \\ &\leq \sum_{i=k}^{\pi(k)-1} \mathbb{1}\{\alpha(\mathcal{R}(i+1)) - \alpha(\mathcal{R}(i)) > 0\}\, \alpha(\mathcal{R}(i+1)) - \alpha(\mathcal{R}(i))).\end{aligned}$$

From the definition of $\Delta(\mathcal{R})$, this quantity is bounded from above by $\Delta(\mathcal{R})$. Finally, let $\alpha(\mathcal{R}(k)) < \alpha(k)$ and $\pi(k) < k$. This implies that there exists an item at a lower position than k, $j > k$, such that $\alpha(\mathcal{R}(j)) \geq \alpha(k)$. Then

$$\begin{aligned}\alpha(k) - \alpha(\mathcal{R}(k)) &\leq \alpha(\mathcal{R}(j)) - \alpha(\mathcal{R}(k)) \\ &\leq \sum_{i=k}^{j-1} \mathbb{1}\{\alpha(\mathcal{R}(i+1)) - \alpha(\mathcal{R}(i)) > 0\}\, (\alpha(\mathcal{R}(i+1)) - \alpha(\mathcal{R}(i)))\,.\end{aligned}$$

From the definition of $\Delta(\mathcal{R})$, this quantity is bounded from above by $\Delta(\mathcal{R})$. This concludes the proof. □

Lemma 4. *Let*

$$\boldsymbol{\mathcal{P}}_t = \left\{(i,j) \in [K]^2 : i < j,\ \left|\bar{\boldsymbol{\mathcal{R}}}_t^{-1}(i) - \bar{\boldsymbol{\mathcal{R}}}_t^{-1}(j)\right| = 1, \boldsymbol{s}_{t-1}(i,j) \leq 2\sqrt{\boldsymbol{n}_{t-1}(i,j)\log(1/\delta)}\right\}$$

be the set of potentially randomized item pairs at time t and $\boldsymbol{\Delta}_t = \max_{\boldsymbol{\mathcal{R}}_t} \Delta(\boldsymbol{\mathcal{R}}_t)$ be the maximum attraction gap *of any list $\boldsymbol{\mathcal{R}}_t$, where $\Delta(\boldsymbol{\mathcal{R}}_t)$ is defined in (8). Then on event $\mathcal{E}$ in Lemma 9,*

$$\boldsymbol{\Delta}_t \leq 3\sum_{i=1}^{K}\sum_{j=i+1}^{K} \mathbb{1}\{(i,j) \in \boldsymbol{\mathcal{P}}_t\}\,(\alpha(i) - \alpha(j))$$

holds at any time $t \in [n]$.

Proof. Fix list $\boldsymbol{\mathcal{R}}_t$ and position $k \in [K-1]$. Let i', i, j, j' be items at positions $k-1, k, k+1, k+2$ in $\bar{\boldsymbol{\mathcal{R}}}_t$. If $k = 1$, let $i' = i$; and if $k = K-1$, let $j' = j$. We consider two cases.

First, suppose that the permutation at time t is such that i and j could be exchanged. Then

$$\alpha(\boldsymbol{\mathcal{R}}_t^{-1}(k+1)) - \alpha(\boldsymbol{\mathcal{R}}_t^{-1}(k)) \leq \mathbb{1}\{(\min\{i,j\}, \max\{i,j\}) \in \boldsymbol{\mathcal{P}}_t\}\,(\alpha(\min\{i,j\}) - \alpha(\max\{i,j\}))$$

holds on event $\mathcal{E}$ by the design of `BubbleRank`. More specifically, $(\min\{i,j\}, \max\{i,j\}) \notin \boldsymbol{\mathcal{P}}_t$ implies that $\alpha(\boldsymbol{\mathcal{R}}_t^{-1}(k+1)) - \alpha(\boldsymbol{\mathcal{R}}_t^{-1}(k)) \leq 0$.

Second, suppose that the permutation at time t is such that i and i' could be exchanged, j and j' could be exchanged, or both. Then

$$\begin{aligned}\alpha(\boldsymbol{\mathcal{R}}_t^{-1}(k+1)) - \alpha(\boldsymbol{\mathcal{R}}_t^{-1}(k)) \leq{} & \mathbb{1}\{(\min\{i,i'\}, \max\{i,i'\}) \in \boldsymbol{\mathcal{P}}_t\}\,(\alpha(\min\{i,i'\}) - \alpha(\max\{i,i'\})) + \\ & \alpha(j) - \alpha(i) + \\ & \mathbb{1}\{(\min\{j,j'\}, \max\{j,j'\}) \in \boldsymbol{\mathcal{P}}_t\}\,(\alpha(\min\{j,j'\}) - \alpha(\max\{j,j'\}))\end{aligned}$$

holds by the same argument as in the first case. Also note that

$$\alpha(j) - \alpha(i) \leq \mathbb{1}\{(\min\{i,j\}, \max\{i,j\}) \in \boldsymbol{\mathcal{P}}_t\}\,(\alpha(\min\{i,j\}) - \alpha(\max\{i,j\}))$$

holds on event $\mathcal{E}$ by the design of `BubbleRank`. Therefore, for any position $k \in [K-1]$ in both above cases,

$$\begin{aligned}\alpha(\boldsymbol{\mathcal{R}}_t^{-1}(k+1)) - \alpha(\boldsymbol{\mathcal{R}}_t^{-1}(k)) \leq \sum_{\ell=k-1}^{k+1} \mathbb{1}\left\{\left(\min\left\{\bar{\boldsymbol{\mathcal{R}}}_t^{-1}(\ell), \bar{\boldsymbol{\mathcal{R}}}_t^{-1}(\ell+1)\right\}, \max\left\{\bar{\boldsymbol{\mathcal{R}}}_t^{-1}(\ell), \bar{\boldsymbol{\mathcal{R}}}_t^{-1}(\ell+1)\right\}\right) \in \boldsymbol{\mathcal{P}}_t\right\} \times \\ \left(\alpha\left(\min\left\{\bar{\boldsymbol{\mathcal{R}}}_t^{-1}(\ell), \bar{\boldsymbol{\mathcal{R}}}_t^{-1}(\ell+1)\right\}\right) - \alpha\left(\max\left\{\bar{\boldsymbol{\mathcal{R}}}_t^{-1}(\ell), \bar{\boldsymbol{\mathcal{R}}}_t^{-1}(\ell+1)\right\}\right)\right).\end{aligned}$$

Now we sum over all positions and note that each pair of $\bar{\boldsymbol{\mathcal{R}}}_t^{-1}(\ell)$ and $\bar{\boldsymbol{\mathcal{R}}}_t^{-1}(\ell+1)$ appears on the right-hand side at most three times, in any list $\boldsymbol{\mathcal{R}}_t$. This concludes our proof. □

Lemma 5. *Let $\boldsymbol{\mathcal{P}}_t$ be defined as in Lemma 4. Then on event $\mathcal{E}$ in Lemma 9,*

$$\sum_{k=1}^{K}(\chi(\mathcal{R}^*, k)\alpha(k) - \chi(\boldsymbol{\mathcal{R}}_t, k)\alpha(\boldsymbol{\mathcal{R}}_t(k))) \leq 3K\chi_{\max}\sum_{i=1}^{K}\sum_{j=i+1}^{K}\mathbb{1}\{(i,j) \in \boldsymbol{\mathcal{P}}_t\}\,(\alpha(i) - \alpha(j))$$

holds at any time $t \in [n]$.

Proof. A direct consequence of Lemmas 3 and 4. □

Lemma 6. *Let $\boldsymbol{\mathcal{P}}_t$ be defined as in Lemma 4, $\boldsymbol{\mathcal{P}} = \bigcup_{t=1}^{n}\boldsymbol{\mathcal{P}}_t$, and $\mathcal{V}_0$ be defined as in (6). Then on event $\mathcal{E}$ in Lemma 9,*

$$|\boldsymbol{\mathcal{P}}| \leq K - 1 + 2\,|\mathcal{V}_0|\ .$$

Proof. From the design of `BubbleRank`, $|\boldsymbol{\mathcal{P}}_1| = K-1$. The set of randomized item pairs grows only if the base list in `BubbleRank` changes. When this happens, the number of incorrectly-ordered item pairs decreases by one, on event $\mathcal{E}$, and the set of randomized item pairs increases by at most two pairs. This event occurs at most $|\mathcal{V}_0|$ times. This concludes our proof. □

Lemma 7. *For any items i and j such that $i < j$,*

$$\boldsymbol{s}_n(i,j) \leq 15\frac{\alpha(i)+\alpha(j)}{\alpha(i)-\alpha(j)}\log(1/\delta)$$

on event $\mathcal{E}$ in Lemma 9.

Proof. To simplify notation, let $\boldsymbol{s}_t = \boldsymbol{s}_t(i,j)$ and $\boldsymbol{n}_t = \boldsymbol{n}_t(i,j)$. The proof has two parts. First, suppose that $\boldsymbol{s}_t \leq 2\sqrt{\boldsymbol{n}_t\log(1/\delta)}$ holds at all times $t \in [n]$. Then from this assumption and on event $\mathcal{E}$ in Lemma 9,

$$\frac{\alpha(i)-\alpha(j)}{\alpha(i)+\alpha(j)}\boldsymbol{n}_t - 2\sqrt{\boldsymbol{n}_t\log(1/\delta)} \leq \boldsymbol{s}_t \leq 2\sqrt{\boldsymbol{n}_t\log(1/\delta)}\,.$$

This implies that

$$\boldsymbol{n}_t \leq \left[4\frac{\alpha(i) + \alpha(j)}{\alpha(i) - \alpha(j)}\right]^2 \log(1/\delta)$$

at any time t, and in turn that

$$\boldsymbol{s}_t \leq 2\sqrt{\boldsymbol{n}_t \log(1/\delta)} \leq 8\frac{\alpha(i) + \alpha(j)}{\alpha(i) - \alpha(j)} \log(1/\delta)$$

at any time t. Our claim follows from setting $t = n$.

Now suppose that $\boldsymbol{s}_t \leq 2\sqrt{\boldsymbol{n}_t \log(1/\delta)}$ does not hold at all times $t \in [n]$. Let τ be the first time when $\boldsymbol{s}_\tau > 2\sqrt{\boldsymbol{n}_\tau \log(1/\delta)}$. Then from the definition of τ and on event $\mathcal{E}$ in Lemma 9,

$$\begin{aligned}\frac{\alpha(i) - \alpha(j)}{\alpha(i) + \alpha(j)}\boldsymbol{n}_\tau - 2\sqrt{\boldsymbol{n}_\tau \log(1/\delta)} &\leq \boldsymbol{s}_\tau \leq \boldsymbol{s}_{\tau-1} + 1 \\ &\leq 2\sqrt{\boldsymbol{n}_\tau \log(1/\delta)} + 1 \\ &\leq 3\sqrt{\boldsymbol{n}_\tau \log(1/\delta)}\,,\end{aligned}$$

where the last inequality holds for any $\delta \leq 1/e$. This implies that

$$\boldsymbol{n}_\tau \leq \left[5\frac{\alpha(i) + \alpha(j)}{\alpha(i) - \alpha(j)}\right]^2 \log(1/\delta)\,,$$

and in turn that

$$\boldsymbol{s}_\tau \leq 3\sqrt{\boldsymbol{n}_\tau \log(1/\delta)} \leq 15\frac{\alpha(i) + \alpha(j)}{\alpha(i) - \alpha(j)} \log(1/\delta)\,.$$

Now note that $\boldsymbol{s}_t = \boldsymbol{s}_\tau$ for any $t > \tau$, from the design of `BubbleRank`. This concludes our proof. □

For some $\mathcal{F}_t = \sigma(\boldsymbol{\mathcal{R}}_1, \boldsymbol{c}_1, \dots, \boldsymbol{\mathcal{R}}_t, \boldsymbol{c}_t)$-measurable event A, let $\mathbb{P}_t(A) = \mathbb{P}(A \mid \mathcal{F}_t)$ be the conditional probability of A given history $\boldsymbol{\mathcal{R}}_1, \boldsymbol{c}_1, \dots, \boldsymbol{\mathcal{R}}_t, \boldsymbol{c}_t$. Let the corresponding conditional expectation operator be $\mathbb{E}_t\left[\cdot\right]$. Note that $\bar{\boldsymbol{\mathcal{R}}}_t$ is $\mathcal{F}_{t-1}$-measurable.

Lemma 8. *Let $i, j \in [K]$ be any items at consecutive positions in $\bar{\boldsymbol{\mathcal{R}}}_t$ and*

$$\boldsymbol{z} = \boldsymbol{c}_t(\boldsymbol{\mathcal{R}}_t^{-1}(i)) - \boldsymbol{c}_t(\boldsymbol{\mathcal{R}}_t^{-1}(j))\,.$$

Then, on the event that i and j are subject to randomization at time t,

$$\mathbb{E}_{t-1}\left[\boldsymbol{z} \mid \boldsymbol{z} \neq 0\right] \geq \frac{\alpha(i) - \alpha(j)}{\alpha(i) + \alpha(j)}$$

when $\alpha(i) > \alpha(j)$, and $\mathbb{E}_{t-1}\left[-\boldsymbol{z} \mid \boldsymbol{z} \neq 0\right] \leq 0$ when $\alpha(i) < \alpha(j)$.

Proof. The first claim is proved as follows. From the definition of expectation and $\boldsymbol{z} \in \{-1, 0, 1\}$,

$$\begin{aligned}\mathbb{E}_{t-1}\left[\boldsymbol{z} \mid \boldsymbol{z} \neq 0\right] &= \frac{\mathbb{P}_{t-1}(\boldsymbol{z} = 1, \boldsymbol{z} \neq 0) - \mathbb{P}_{t-1}(\boldsymbol{z} = -1, \boldsymbol{z} \neq 0)}{\mathbb{P}_{t-1}(\boldsymbol{z} \neq 0)} \\ &= \frac{\mathbb{P}_{t-1}(\boldsymbol{z} = 1) - \mathbb{P}_{t-1}(\boldsymbol{z} = -1)}{\mathbb{P}_{t-1}(\boldsymbol{z} \neq 0)} \\ &= \frac{\mathbb{E}_{t-1}\left[\boldsymbol{z}\right]}{\mathbb{P}_{t-1}(\boldsymbol{z} \neq 0)}\,,\end{aligned}$$

where the last equality is a consequence of $\boldsymbol{z} = 1 \implies \boldsymbol{z} \neq 0$ and that $\boldsymbol{z} = -1 \implies \boldsymbol{z} \neq 0$.

Let $\chi_i = \mathbb{E}_{t-1}\left[\chi(\boldsymbol{\mathcal{R}}_t, \boldsymbol{\mathcal{R}}_t^{-1}(i))\right]$ and $\chi_j = \mathbb{E}_{t-1}\left[\chi(\boldsymbol{\mathcal{R}}_t, \boldsymbol{\mathcal{R}}_t^{-1}(j))\right]$ denote the average examination probabilities of the positions with items i and j, respectively, in $\boldsymbol{\mathcal{R}}_t$; and consider the event that i and j are subject to randomization at time t. By Assumption A2, the values of χ_i and χ_j do not depend on the randomization of other parts of $\bar{\boldsymbol{\mathcal{R}}}_t$, only on the positions of i and j. Then $\chi_i \geq \chi_j$; from $\alpha(i) > \alpha(j)$ and Assumption A4. Based on this fact, $\mathbb{E}_{t-1}\left[\boldsymbol{z}\right]$ is bounded from below as

$$\mathbb{E}_{t-1}\left[\boldsymbol{z}\right] = \chi_i \alpha(i) - \chi_j \alpha(j) \geq \chi_i(\alpha(i) - \alpha(j))\,,$$

where the inequality is from $\chi_i \geq \chi_j$. Moreover, $\mathbb{P}_{t-1}(\boldsymbol{z} \neq 0)$ is bounded from above as

$$\begin{aligned}\mathbb{P}_{t-1}(\boldsymbol{z} \neq 0) &= \mathbb{P}_{t-1}(\boldsymbol{z} = 1) + \mathbb{P}_{t-1}(\boldsymbol{z} = -1)\\ &\leq \chi_i \alpha(i) + \chi_j \alpha(j)\\ &\leq \chi_i(\alpha(i) + \alpha(j))\,,\end{aligned}$$

where the first inequality is from inequalities $\mathbb{P}_{t-1}(\boldsymbol{z} = 1) \leq \chi_i \alpha(i)$ and $\mathbb{P}_{t-1}(\boldsymbol{z} = -1) \leq \chi_j \alpha(j)$, and the last inequality is from $\chi_i \geq \chi_j$.

Finally, we chain all above inequalities and get our first claim. The second claim follows from the observation that $\mathbb{E}_{t-1}\left[-\boldsymbol{z} \mid \boldsymbol{z} \neq 0\right] = -\mathbb{E}_{t-1}\left[\boldsymbol{z} \mid \boldsymbol{z} \neq 0\right]$. ☐

Lemma 9. *Let $S_1 = \left\{(i, j) \in [K]^2 : i < j\right\}$ and $S_2 = \left\{(i, j) \in [K]^2 : i > j\right\}$. Let*

$$\begin{aligned}\mathcal{E}_{t,1} &= \left\{\forall (i, j) \in S_1 : \frac{\alpha(i) - \alpha(j)}{\alpha(i) + \alpha(j)} \boldsymbol{n}_t(i, j) - 2\sqrt{\boldsymbol{n}_t(i, j) \log(1/\delta)} \leq \boldsymbol{s}_t(i, j)\right\},\\ \mathcal{E}_{t,2} &= \left\{\forall (i, j) \in S_2 : \boldsymbol{s}_t(i, j) \leq 2\sqrt{\boldsymbol{n}_t(i, j) \log(1/\delta)}\right\}.\end{aligned}$$

Let $\mathcal{E} = \bigcap_{t \in [n]} (\mathcal{E}_{t,1} \cap \mathcal{E}_{t,2})$ and $\overline{\mathcal{E}}$ be the complement of $\mathcal{E}$. Then $\mathbb{P}(\overline{\mathcal{E}}) \leq \delta^{\frac{1}{2}} K^2 n$.

Proof. First, we bound $\mathbb{P}(\overline{\mathcal{E}_{t,1}})$. Fix $(i, j) \in S_1$, $t \in [n]$, and $(\boldsymbol{n}_\ell(i, j))_{\ell=1}^t$. Let $\tau(m)$ be the time of observing item pair (i, j) for the m-th time, $\tau(m) = \min\left\{\ell \in [t] : \boldsymbol{n}_\ell(i, j) = m\right\}$ for $m \in [\boldsymbol{n}_t(i, j)]$. Let $\boldsymbol{z}_\ell = \boldsymbol{c}_\ell(\boldsymbol{\mathcal{R}}_\ell^{-1}(i)) - \boldsymbol{c}_\ell(\boldsymbol{\mathcal{R}}_\ell^{-1}(j))$. Since $(\boldsymbol{n}_\ell(i, j))_{\ell=1}^t$ is fixed, note that $\boldsymbol{z}_\ell \neq 0$ if $\ell = \tau(m)$ for some $m \in [\boldsymbol{n}_t(i, j)]$. Let $\boldsymbol{X}_0 = 0$ and

$$\boldsymbol{X}_\ell = \sum_{\ell'=1}^{\ell} \mathbb{E}_{\tau(\ell')-1}\left[\boldsymbol{z}_{\tau(\ell')} \,\middle|\, \boldsymbol{z}_{\tau(\ell')} \neq 0\right] - \boldsymbol{s}_{\tau(\ell)}(i, j)$$

for $\ell \in [\boldsymbol{n}_t(i, j)]$. Then $(\boldsymbol{X}_\ell)_{\ell=1}^{\boldsymbol{n}_t(i,j)}$ is a martingale, because

$$\begin{aligned}\boldsymbol{X}_\ell - \boldsymbol{X}_{\ell-1} &= \mathbb{E}_{\tau(\ell)-1}\left[\boldsymbol{z}_{\tau(\ell)} \,\middle|\, \boldsymbol{z}_{\tau(\ell)} \neq 0\right] - (\boldsymbol{s}_{\tau(\ell)}(i, j) - \boldsymbol{s}_{\tau(\ell-1)}(i, j))\\ &= \mathbb{E}_{\tau(\ell)-1}\left[\boldsymbol{z}_{\tau(\ell)} \,\middle|\, \boldsymbol{z}_{\tau(\ell)} \neq 0\right] - \boldsymbol{z}_{\tau(\ell)}\,,\end{aligned}$$

where the last equality follows from the definition of $\boldsymbol{s}_{\tau(\ell)}(i, j) - \boldsymbol{s}_{\tau(\ell-1)}(i, j)$. Now we apply the Azuma-Hoeffding inequality and get that

$$P\left(\boldsymbol{X}_{\boldsymbol{n}_t(i,j)} - \boldsymbol{X}_0 \geq 2\sqrt{\boldsymbol{n}_t(i, j) \log(1/\delta)}\right) \leq \delta^{\frac{1}{2}}\,.$$

Moreover, from the definitions of $\boldsymbol{X}_0$ and $\boldsymbol{X}_{\boldsymbol{n}_t(i,j)}$, and by Lemma 8, we have that

$$\begin{aligned}\delta^{\frac{1}{2}} &\geq P\left(\boldsymbol{X}_{\boldsymbol{n}_t(i,j)} - \boldsymbol{X}_0 \geq 2\sqrt{\boldsymbol{n}_t(i, j) \log(1/\delta)}\right)\\ &= P\left(\sum_{\ell'=1}^{\boldsymbol{n}_t(i,j)} \mathbb{E}_{\tau(\ell')-1}\left[\boldsymbol{z}_{\tau(\ell')} \,\middle|\, \boldsymbol{z}_{\tau(\ell')} \neq 0\right] - \boldsymbol{s}_t(i, j) \geq 2\sqrt{\boldsymbol{n}_t(i, j) \log(1/\delta)}\right)\\ &\geq P\left(\frac{\alpha(i) - \alpha(j)}{\alpha(i) + \alpha(j)} \boldsymbol{n}_t(i, j) - \boldsymbol{s}_t(i, j) \geq 2\sqrt{\boldsymbol{n}_t(i, j) \log(1/\delta)}\right)\\ &= P\left(\frac{\alpha(i) - \alpha(j)}{\alpha(i) + \alpha(j)} \boldsymbol{n}_t(i, j) - 2\sqrt{\boldsymbol{n}_t(i, j) \log(1/\delta)} \geq \boldsymbol{s}_t(i, j)\right).\end{aligned}$$

The above inequality holds for any $(\boldsymbol{n}_\ell(i,j))_{\ell=1}^t$, and therefore also in expectation over $(\boldsymbol{n}_\ell(i,j))_{\ell=1}^t$. From the definition of $\mathcal{E}_{t,1}$ and the union bound, we have $\mathbb{P}(\overline{\mathcal{E}_{t,1}}) \leq \frac{1}{2}\delta^{\frac{1}{2}}K(K-1)$.

The claim that $\mathbb{P}(\overline{\mathcal{E}_{t,2}}) \leq \frac{1}{2}\delta^{\frac{1}{2}}K(K-1)$ is proved similarly, except that we use $\mathbb{E}_{\tau(\ell)-1}\left[\boldsymbol{z}_{\tau(\ell)} \,\middle|\, \boldsymbol{z}_{\tau(\ell)} \neq 0\right] \leq 0$. From the definition of $\overline{\mathcal{E}}$ and the union bound,

$$\mathbb{P}(\overline{\mathcal{E}}) \leq \sum_{t=1}^{n} \mathbb{P}(\overline{\mathcal{E}_{t,1}}) + \sum_{t=1}^{n} \mathbb{P}(\overline{\mathcal{E}_{t,2}}) \leq \delta^{\frac{1}{2}} K^2 n\,.$$

This completes our proof. $\square$

B RESULTS WITH NDCG

In this section, we report the NDCG of compared algorithms, which measures the quality of displayed lists. Since `CascadeKL-UCB` fails in the PBM and we focus on learning from all types of click feedback, we leave out `CascadeKL-UCB` from this section.

In the first two experiments, we evaluate algorithms by their regret in (4) and safety constraint violation in (7). Neither of these metrics measure the quality of ranked lists directly. In this experiment, we report the per-step NDCG@5 of `BubbleRank`, `BatchRank`, `TopRank`, and `Baseline` (Figure 4), which directly measures the quality of ranked lists and is widely used in the LTR literature [12, 2]. Since the *Yandex* dataset does not contain relevance scores for all query-item pairs, we take the attraction probability of the item in its learned click model as a proxy to its relevance score. This substitution is natural since our goal is to rank items in the descending order of their attraction probabilities [6]. We compute the NDCG@5 of a ranked list $\mathcal{R}$ as

$$NDCG@5(\mathcal{R}) = \frac{DCG@5(\mathcal{R})}{DCG@5(\mathcal{R}^*)}\,, \quad DCG@5(\mathcal{R}) = \sum_{k=1}^{5} \frac{\alpha(\mathcal{R}(k))}{\log_2(k+1)}\,,$$

where $\mathcal{R}^*$ is the optimal list and $\alpha(\mathcal{R}(k)$ is the attraction probability of the k-th item in list $\mathcal{R}$. This is a standard evaluation metric, and is used in TREC evaluation benchmarks [2], for instance. It measures the discounted gain over the attraction probabilities of the 5 highest ranked items in list $\mathcal{R}$, which is normalized by the DCG@5 of $\mathcal{R}^*$.

In Figure 4, we observe that `Baseline` has good NDCG@5 scores in all click models. Yet there is still room for improvement. `BubbleRank`, `BatchRank`, and `TopRank` have similar NDCG@5 scores after 5 million steps. But `BubbleRank` starts with NDCG@5 close to that of `Baseline`, while `BatchRank` and `TopRank` start with lists with very low NDCG@5.

These results validate our earlier findings. As in Section 5.2, we observe that `BubbleRank` converges to the optimal list in hindsight, since its NDCG@5 approaches 1. As in Section 5.3, we observe that `BubbleRank` is safe, since its NDCG@5 is never much worse than that of `Baseline`.

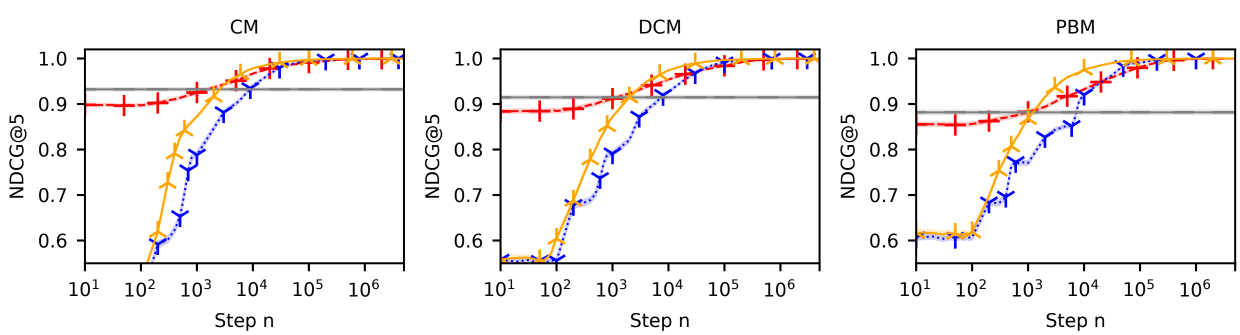

Figure 4: The per-step NDCG@5 of `BubbleRank` (red), `BatchRank` (blue), `TopRank` (orange), and `Baseline` (grey) in the CM, DCM, and PBM in up to 5 million steps. Higher is better. The shaded regions represent standard errors of our estimates.