7 Supplementary

Proof of Lemma 1: We restate the lemma before giving the proof, for ease of reading:

Lemma 1: Let $F : \mathbb{R}^m \mapsto \mathbb{R}^a$ be a vector valued function, where $m \geq 1$, $a \geq 1$. For a fixed $x \in \mathbb{R}^m$, let k entries of x be observed at random. That is, for a fixed probability distribution $\mathbb{P}$ and some random $\sigma \sim \mathbb{P}(S_m)$, observed tuple is $\{\sigma, x_{\sigma(1)}, \ldots, x_{\sigma(k)}\}$. The necessary condition for existence of an unbiased estimator of $F(x)$, that can be constructed from $\{\sigma, x_{\sigma(1)}, \ldots, x_{\sigma(k)}\}$, is that it should be possible to decompose $F(x)$ over k (or less) coordinates of x at a time. That is, $F(x)$ should have the following structure:

$$F(x) = \sum_{(i_1,i_2,\ldots,i_\ell)\in\ {}^mP_\ell} h_{i_1,i_2,\ldots,i_\ell}(x_{i_1}, x_{i_2}, \ldots, x_{i_\ell})$$

where $\ell \leq k$, ${}^mP_\ell$ is ℓ permutations of m and $h : \mathbb{R}^\ell \mapsto \mathbb{R}^a$. Moreover, when $F(x)$ can be written in form of Eq 2 , with $\ell = k$, an unbiased estimator of $F(x)$, based on $\{\sigma, x_{\sigma(1)}, \ldots, x_{\sigma(k)}\}$, is,

$$g(\sigma, x_{\sigma(1)}, \ldots, x_{\sigma(k)}) = \frac{\sum\limits_{(j_1,j_2,\ldots,j_k)\in S_k} h_{\sigma(j_1),\ldots,\sigma(j_k)}(x_{\sigma(j_1)}, \ldots, x_{\sigma(j_k)})}{\sum\limits_{(j_1,\ldots,j_k)\in S_k} p(\sigma(j_1), \ldots, \sigma(j_k))}$$

where S_k is the set of $k!$ permutations of [k] and $p(\sigma(1), \ldots, \sigma(k))$ is as in Eq 1 .

Proof. For a fixed $x \in \mathbb{R}^m$ and probability distribution $\mathbb{P}$, let the random permutation be $\sigma \sim \mathbb{P}(S_m)$ and the observed tuple be $\{\sigma, x_{\sigma(1)}, \ldots, x_{\sigma(k)}\}$. Let $\hat{G} = G(\sigma, x_{\sigma(1)}, \ldots, x_{\sigma(k)})$ be an unbiased estimator of $F(x)$ based on the random observed tuple. Taking expectation, we get:

$$\begin{aligned} F(x) = \mathbb{E}_{\sigma\sim\mathbb{P}}\left[\hat{G}\right] &= \sum_{\pi\in S_m} \mathbb{P}(\pi) G(\pi, x_{\pi(1)}, \ldots, x_{\pi(k)}) \\ &= \sum_{(i_1,i_2,\ldots,i_k)\in\ {}^mP_k} \sum_{\pi\in S_m} \mathbb{P}(\pi)\mathbb{1}(\pi(1) = i_1, \pi(2) = i_2, \ldots, \pi(k) = i_k) G(\pi, x_{i_1}, x_{i_2}, \ldots, x_{i_k}) \end{aligned}$$

We note that $\mathbb{P}(\pi) \in [0, 1]$ is independent of x for all $\pi \in S_m$. Then we can use the following construction of function $h(\cdot)$:

$$h_{i_1,i_2,\ldots,i_k}(x_{i_1}, \ldots, x_{i_k}) = \sum_{\pi\in S_m} \mathbb{P}(\pi)\mathbb{1}(\pi(1) = i_1, \pi(2) = i_2, \ldots, \pi(k) = i_k) G(\pi, x_{i_1}, x_{i_2}, \ldots, x_{i_k})$$

and thus,

$$F(x) = \sum_{(i_1,i_2,\ldots,i_k)\in\ {}^mP_k} h_{i_1,i_2,\ldots,i_k}(x_{i_1}, x_{i_2}, \ldots, x_i)$$

Hence, we conclude that for existence of an unbiased estimator based on the random observed tuple, it should be possible to decompose $F(x)$ over k (or less) coordinates of x at a time. The "less than k" coordinates arguement follows simply by noting that if $F(x)$ can be decomposed over ℓ coordinates at a time ($\ell < k$) and observation tuple is $\{\sigma, x_{\sigma(1)}, \ldots, x_{\sigma(k)})\}$, then any $k - \ell$ observations can be thrown away and the rest used for construction of the unbiased estimator.

The construction of the unbiased estimator proceeds as follows:

Let $F(x) = \sum_{i=1}^m h_i(x_i)$ and feedback is for top-1 item ($k = 1$). The unbiased estimator according to Lemma. 1 is:

$$g(\sigma, x_{\sigma(1)}) = \frac{h_{\sigma(1)}(x_{\sigma(1)})}{p(\sigma(1))} = \frac{h_{\sigma(1)}(x_{\sigma(1)})}{\sum_\pi \mathbb{P}(\pi)\mathbb{1}(\pi(1) = \sigma(1))}$$

Taking expectation w.r.t. σ, we get:

$$\mathbb{E}_\sigma[g(\sigma, x_{\sigma(1)})] = \sum_{i=1}^m \frac{h_i(x_i)(\sum_\pi \mathbb{P}(\pi)\mathbb{1}(\pi(1) = i))}{\sum_\pi \mathbb{P}(\pi)\mathbb{1}(\pi(1) = i)} = \sum_{i=1}^m h_i(x_i) = F(x)$$

Now, let $F(x) = \sum_{i \neq j = 1}^{m} h_{i,j}(x_i, x_j)$ and the feedback is for top-2 item ($k = 2$). The unbiased estimator according to Lemma. 1 is:

$$g(\sigma, x_{\sigma(1)}, x_{\sigma(2)}) = \frac{h_{\sigma(1),\sigma(2)}(x_{\sigma(1)}, x_{\sigma(2)}) + h_{\sigma(2),\sigma(1)}(x_{\sigma(2)}, x_{\sigma(1)})}{p(\sigma(1), \sigma(2)) + p(\sigma(2), \sigma(1))}$$

We will use the fact that for any 2 permutations σ_1, σ_2, which places the same 2 objects in top-2 positions but in opposite order, estimators based on σ_1 (i.e, $g(\sigma_1, x_{\sigma_1(1)}, x_{\sigma_1(2)})$) and σ_2 (i.e, $g(\sigma_2, x_{\sigma_2(1)}, x_{\sigma_2(2)})$) have same numerator and denominator. For eg., let $\sigma_1(1) = i, \sigma_1(2) = j$. Numerator and denominator for $g(\sigma_1, x_{\sigma_1(1)}, x_{\sigma_1(2)})$ are $h_{i,j}(x_i, x_j) + h_{j,i}(x_j, x_i)$ and $p(i,j) + p(j,i)$ respectively. Now let $\sigma_2(1) = j, \sigma_2(2) = i$. Then numerator and denominator for $g(\sigma_2, x_{\sigma_2(1)}, x_{\sigma_2(2)})$ are $h_{j,i}(x_j, x_i) + h_{i,j}(x_i, x_j)$ and $p(j,i) + p(i,j)$ respectively.

Then, taking expectation w.r.t. σ, we get:

$$\begin{aligned}
\mathbb{E}_{\sigma} g(\sigma, x_{\sigma(1)}, x_{\sigma(2)}) &= \sum_{i \neq j = 1}^{m} \frac{(h_{i,j}(x_i, x_j) + h_{j,i}(x_j, x_i))p(i,j)}{p(i,j) + p(j,i)} \\
&= \sum_{i > j = 1}^{m} \frac{(h_{i,j}(x_i, x_j) + h_{j,i}(x_j, x_i))(p(i,j) + p(j,i))}{p(i,j) + p(j,i)} \\
&= \sum_{i > j = 1}^{m} (h_{i,j}(x_i, x_j) + h_{j,i}(x_j, x_i)) = \sum_{i \neq j = 1}^{m} h_{i,j}(x_i, x_j) = F(x)
\end{aligned}$$

This chain of logic can be extended for any $k \geq 3$. Explicitly, for general $k \leq m$, let $\mathbb{S}(i_1, i_2, \ldots, i_k)$ denote all permutations of the set $\{i_1, \ldots, i_k\}$. Then, taking expectation of the unbiased estimator will give:

$$\begin{aligned}
&\mathbb{E}_{\sigma} g(\sigma, x_{\sigma(1)}, \ldots, x_{\sigma(k)}) \\
&= \sum_{(i_1, i_2, \ldots, i_k) \in {}^{m}P_k} \frac{\left(\sum_{(j_1, \ldots, j_k) \in \mathbb{S}(i_1, \ldots, i_k)} h_{j_1, \ldots, j_k}(x_{j_1}, \ldots, x_{j_k}) \right) p(i_1, \ldots, i_k)}{\sum_{(j_1, \ldots, j_k) \in \mathbb{S}(i_1, \ldots, i_k)} p(j_1, \ldots, j_k)} \\
&= \sum_{i_1 > i_2 > \ldots > i_k = 1}^{m} \frac{\left(\sum_{(j_1, \ldots, j_k) \in \mathbb{S}(i_1, \ldots, i_k)} h_{j_1, \ldots, j_k}(x_{j_1}, \ldots, x_{j_k}) \right) \left(\sum_{(j_1, \ldots, j_k) \in \mathbb{S}(i_1, \ldots, i_k)} p(j_1, \ldots, j_k) \right)}{\sum_{(j_1, \ldots, j_k) \in \mathbb{S}(i_1, \ldots, i_k)} p(j_1, \ldots, j_k)} \\
&= \sum_{i_1 > i_2 > \ldots > i_k = 1}^{m} \left(\sum_{(j_1, \ldots, j_k) \in \mathbb{S}(i_1, \ldots, i_k)} h_{j_1, \ldots, j_k}(x_{j_1}, \ldots, x_{j_k}) \right) = \sum_{(i_1, i_2, \ldots, i_k) \in {}^{m}P_k} h_{i_1, i_2, \ldots, i_k}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = F(x)
\end{aligned}$$

Note: For $k = m$, i.e., when the full feedback is received, the unbiased estimator is:

$$\begin{aligned}
g(\sigma, x_{\sigma(1)}, \ldots, x_{\sigma(m)}) &= \frac{\sum_{(j_1, j_2, \ldots, j_m) \in S_m} h_{\sigma(j_1), \ldots, \sigma(j_m)}(x_{\sigma(j_1)}, \ldots, x_{\sigma(j_m)})}{\sum_{(j_1, \ldots, j_m) \in S_m} p(\sigma(j_1), \ldots, \sigma(j_m))} \\
&= \frac{\sum_{(i_1, i_2, \ldots, i_m) \in {}^{m}P_m} h_{i_1, \ldots, i_m}(x_{i_1}, \ldots, x_{i_m})}{1} = F(x)
\end{aligned}$$

Hence, with full information, the unbiased estimator of $F(x)$ is actually $F(x)$ itself, which is consistent with the theory of unbiased estimator.

□

Proof of Lemma 4:

Proof. The first equality is true because

$$\begin{aligned}\|X^\top\|_{1\to p} &= \sup_{v\neq 0}\frac{\|X^\top v\|_p}{\|v\|_1} = \sup_{v\neq 0}\sup_{u\neq 0}\frac{\langle X^\top v, u\rangle}{\|v\|_1\|u\|_q}\\ &= \sup_{u\neq 0}\sup_{v\neq 0}\frac{\langle v, Xu\rangle}{\|v\|_1\|u\|_q} = \sup_{u\neq 0}\frac{\|Xu\|_\infty}{\|u\|_q} = \|X\|_{q\to\infty}.\end{aligned}$$

The second is true because

$$\begin{aligned}\|X\|_{q\to\infty} &= \sup_{u\neq 0}\frac{\|Xu\|_\infty}{\|u\|_q} = \sup_{u\neq 0}\max_{j=1}^{m}\frac{|\langle X_{j:}, u\rangle|}{\|u\|_q}\\ &= \max_{j=1}^{m}\sup_{u\neq 0}\frac{|\langle X_{j:}, u\rangle|}{\|u\|_q} = \max_{j=1}^{m}\|X_{j:}\|_p.\end{aligned}$$

□

Proof of Lemma 5 : We restate the lemma before giving the proof:

Lemma 5: For parameter γ in Algorithm 1 , R_D being the bound on ℓ_2 norm of the feature vectors (rows of document matrix X), m being the upper bound on number of documents per query, U being the radius of the Euclidean ball denoting the space of ranking parameters and $R_{\max}$ being the maximum possible relevance value (in practice always ≤ 5), let $C^\phi \in \{C^{sq}, C^{svm}, C^{KL}\}$ be polynomial functions of R_D, m, U, R_{max}, where the degrees of the polynomials depend on the surrogate $(\phi_{sq}, \phi_{svm}, \phi_{KL})$. Then we have,

$$\mathbb{E}_t\left[\|\tilde{z}_t\|^2\right] \leq \frac{C^\phi}{\gamma}.$$

Proof. All our unbiased estimators are of the form $X^\top f(s, R, \sigma)$. We will actually get a bound on $f(s, R, \sigma)$ by using Lemma 4 and $p \to q$ norm relation, to equate out X:

$$\begin{aligned}\|\tilde{z}\|_2 = \|X^\top f(s, R, \sigma)\|_2 &\leq \|X^\top\|_{1\to 2}\|f(s, R, \sigma)\|_1\\ &\leq R_D\|f(s, R, \sigma)\|_1\end{aligned}$$

since $R_D \geq \max_{j=1}^m \|X_{j:}\|_2$.

Squared Loss: The unbiased estimator of gradient of squared loss, as given in the main text, is:

$$\tilde{z} = X^\top(2(s - \frac{R_{\sigma(1)}e_{\sigma(1)}}{p(\sigma(1))}))$$

where $p(\sigma(1)) = \sum_{\pi\in S_m}\mathbb{P}(\pi)\mathbb{1}(\pi(1) = \sigma(1))$ ($\mathbb{P} = \mathbb{P}_t$ is the distribution at round t as in Alg. 1)

Now we have:

$$\|s - \frac{R_{\sigma(1)}e_{\sigma(1)}}{p(\sigma(1))}\|_1 \leq mR_DU + \frac{R_{max}}{p(\sigma(1))} \leq \frac{mR_DUR_{max}}{p(\sigma(1)}$$

Thus, taking expectation w.r.t σ, we get:

$$\mathbb{E}_\sigma\|\tilde{z}\|_2^2 \leq m^2R_D^4U^2R_{max}^2\mathbb{E}_\sigma\frac{1}{p(\sigma(1))^2} = m^2R_D^4U^2R_{max}^2\sum_{i=1}^m\frac{p(i)}{p^2(i)}$$

Now, since $p(i) \geq \frac{\gamma}{m}$, $\forall\ i$, we get: $\mathbb{E}_\sigma\|\tilde{z}\|_2^2 \leq \frac{C^{sq}}{\gamma}$, where $C^{sq} = m^4R_D^4U^2R_{max}^2$.

RankSVM Surrogate: The unbiased estimator of gradient of the RankSVM surrogate, as given in the main text, is:

$$\tilde{z} = X^\top\left(\frac{h_{s,\sigma(1),\sigma(2)}(R_{\sigma(1)}, R_{\sigma(2)}) + h_{s,\sigma(2),\sigma(1)}(R_{\sigma(2)}, R_{\sigma(1)})}{p(\sigma(1), \sigma(2)) + p(\sigma(2), \sigma(1))}\right)$$

where $h_{s,i,j}(R_i,R_j) = \mathbb{1}(R_i > R_j)\mathbb{1}(1+s_j > s_i)(e_j - e_i)$ and $p(\sigma(1),\sigma(2)) = \sum_{\pi \in S_m} \mathbb{P}(\pi)\mathbb{1}(\pi(1) = \sigma(1), \pi(2) = \sigma(2))$ ($\mathbb{P} = \mathbb{P}_t$ as in Alg. 1)).

Now we have:

$$\|\frac{h_{s,\sigma(1),\sigma(2)}(R_{\sigma(1)},R_{\sigma(2)}) + h_{s,\sigma(2),\sigma(1)}(R_{\sigma(2)},R_{\sigma(1)})}{p(\sigma(1),\sigma(2)) + p(\sigma(2),\sigma(1))}\|_1 \leq \frac{2}{p(\sigma(1),\sigma(2)) + p(\sigma(2),\sigma(1))}$$

Thus, taking expectation w.r.t σ, we get:

$$\mathbb{E}_\sigma \|\tilde{z}\|_2^2 \leq 4R_D^2 \mathbb{E}_\sigma \frac{1}{(p(\sigma(1),\sigma(2)) + p(\sigma(2),\sigma(1)))^2} \leq 4R_D^2 \sum_{i>j}^{m} \frac{p(i,j)+p(j,i)}{(p(i,j)+p(j,i))^2}$$

Now, since $p(i,j) \geq \frac{\gamma}{m^2}$, $\forall\ i,j$, we get: $\mathbb{E}_\sigma \|\tilde{z}\|_2^2 \leq \frac{C^{svm}}{\gamma}$, where $C^{svm} = O(m^4 R_D^2)$.

KL based Surrogate: The unbiased estimator of gradient of the KL based surrogate, as given in the main text, is:

$$\tilde{z} = X^\top \left(\frac{(\exp(s_{\sigma(1)}) - \exp(R_{\sigma(1)}))e_{\sigma(1)}}{p(\sigma(1))} \right)$$

where $p(\sigma(1)) = \sum_{\pi \in S_m} \mathbb{P}(\pi)\mathbb{1}(\pi(1) = \sigma(1))$ ($\mathbb{P} = \mathbb{P}_t$ as in Alg. 1)).

Now we have:

$$\|\frac{(\exp(s_{\sigma(1)}) - \exp(R_{\sigma(1)}))e_{\sigma(1)}}{p(\sigma(1))}\|_1 \leq \frac{\exp(R_D U)}{p(\sigma(1))}$$

Thus, taking expectation w.r.t σ, we get:

$$\mathbb{E}_\sigma \|\tilde{z}\|_2^2 \leq R_D^2 \exp(2R_D U) \mathbb{E}_\sigma \frac{1}{p(\sigma(1))^2}$$

Following the same arguement as in squared loss, we get: $\mathbb{E}_\sigma \|\tilde{z}\|_2^2 \leq \frac{C^{KL}}{\gamma}$, where $C^{KL} = m^2 R_D^2 \exp(2R_D U)$.

□

Proof of Lemma 2 :

Proof. Let $m = 3$. Collection of all terms which are functions of 1st coordinate of R, i.e, R_1, in the gradient of RankSVM is: $\mathbb{1}(R_1 > R_2)\mathbb{1}(1+s_2 > s_1)(e_2 - e_1) + \mathbb{1}(R_2 > R_1)\mathbb{1}(1+s_1 > s_2)(e_1 - e_2) + \mathbb{1}(R_1 > R_3)\mathbb{1}(1+s_3 > s_1)(e_3 - e_1) + \mathbb{1}(R_3 > R_1)\mathbb{1}(1+s_1 > s_3)(e_1 - e_3)$. Now let $s_1 = 1, s_2 = 0, s_3 = 0$. Then the collection becomes: $\mathbb{1}(R_2 > R_1)(e_1 - e_2) + \mathbb{1}(R_3 > R_1)(e_1 - e_3) = (\mathbb{1}(R_2 > R_1) + \mathbb{1}(R_3 > R_1))e_1 - \mathbb{1}(R_2 > R_1)e_2 - \mathbb{1}(R_3 > R_1)e_3$. Now, if the gradient can be decomposed over each coordinate of R, then the collection of terms associated with R_1 should *only and only be a function* of R_1. Specifically, $(\mathbb{1}(R_2 > R_1) + \mathbb{1}(R_3 > R_1))$ (the non-zero coefficient of e_1) should be a function of only R_1 (similarly for e_2 and e_3).

Now assume that the $(\mathbb{1}(R_2 > R_1) + \mathbb{1}(R_3 > R_1))$ can be expressed as a function of R_1 only. Then the difference between the coefficient's values, for the following two cases: $R_1 = 0, R_2 = 0, R_3 = 0$ and $R_1 = 1, R_2 = 0, R_3 = 0$, would be same as the difference between the coefficient's values, for the following two cases: $R_1 = 0, R_2 = 1, R_3 = 1$ and $R_1 = 1, R_2 = 1, R_3 = 1$ (Since the difference would be affected only by change in R_1 value). It can be clearly seen that the change in value between the first two cases is: $0 - 0 = 0$, while the change in value between the second two cases is: $2 - 0 = 2$. Thus, we reach a contradiction. □

Proof of Lemma 3 :

Proof. The term associated with the 1st coordinate of R, i.e, R_1, in the gradient of ListNet is $=$ $\sum_{i=1}^m \left(\frac{-\exp(R_i)}{\sum_{j=1}^m \exp(R_j)} + \frac{\exp(s_i)}{\sum_{j=1}^m \exp(s_j)} \right) e_i$ (in fact, the same term is associated with every coordinate of R).

Specifically, $f(R) = \left(\frac{-\exp(R_1)}{\sum_{j=1}^m \exp(R_j)} + \frac{\exp(s_1)}{\sum_{j=1}^m \exp(s_j)} \right)$ is the non-zero coefficient of e_1, associated with R_1. Now, if $f(R)$ would have only been a function of R_1, then $\frac{\partial^2 f(R)}{\partial R_1 \partial R_j}$, $\forall\ j \neq 1$ would have been zero. It can be clearly seen this is not the case.

Now, the term associated jointly with R_1 and R_2, in the gradient of ListNet is same as before, i.e, $\sum_{i=1}^m \left(\frac{-\exp(R_i)}{\sum_{j=1}^m \exp(R_j)} + \frac{\exp(s_i)}{\sum_{j=1}^m \exp(s_j)} \right) e_i$ (since R_1 and R_2 are present in all the summation terms of the gradient).

Specifically, $f(R) = \left(\frac{-\exp(R_i)}{\sum_{j=1}^m \exp(R_j)} + \frac{\exp(s_i)}{\sum_{j=1}^m \exp(s_j)} \right)$ is the non-zero coefficient of e_1. Now, if $f(R)$ would have only been a function of R_1 and R_2, then $\frac{\partial^3 f(R)}{\partial R_1 \partial R_2 \partial R_j}$, $\forall j \neq 1, j \neq 2$ would have been zero. It can be clearly seen this is not the case.

The same arguement can be extended for any $k < m$.

□

Proof of Theorem. 5.2:

Proof. We will first fix the setting of the online game. We consider $m = 3$ and fixed the document matrix $X \in \mathbb{R}^{3\times 3}$ to be the identity. At each round of the game, the adversary generates the fixed X and the learner chooses a score vector $s \in \mathbb{R}^3$. Making the matrix X identity makes the distinction between weight vectors w and scores s irrelevant since $s = Xw = w$. We note that allowing the adversary to vary X over the rounds only makes him more powerful, which can only increase the regret. We also restrict the adversary to choose binary relevance vectors. Once again, allowing adversary to choose multi-graded relevance vectors only makes it more powerful. Thus, in this setting, the adversary can now choose among $2^3 = 8$ possible relevance vectors. The learner's action set is infinite, i.e., the learner can choose any score vector $s = Xw = \mathbb{R}^m$. The loss function $\phi(s, R)$ is any NDCG calibrated surrogate and feedback is the relevance of top-ranked item at each round, where ranking is induced by sorted order (descending) of score vector. We will use p to denote randomized adversary one-short strategies, i.e. distributions over the 8 possible relevance score vectors. Let $s_p^* = \text{argmin}_s \mathbb{E}_{R\sim p}\phi(s, R)$. We note that in the definition of NDCG calibrated surrogates, Ravikumar et al. [2011] assume that the optimal score vector for each distribution over relevance vectors is unique and we subscribe to that assumption. The assumption was taken to avoid some boundary conditions.

It remains to specify the choice of U, a bound on the Euclidean norm of the weight vectors (same as score vectors for us right now) that is used to define the best loss in hindsight. It never makes sense for the learner to play anything outside the set $\cup_p s_p^*$ so that we can set $U = \max\{\|s\|_2 \ : \ s \in \cup_p s_p^*\}$.

The paragraph following Lemma 6 of Thm. 3 in Piccolboni and Schindelhauer [2001] gives the main intuition behind the argument the authors developed to prove hopelessness of finite action partial monitoring games. To make our proof self contained, we will explain the intuition in a rigorous way.

Key insight: Two adversary strategies $p, \tilde{p}$ are said to be indistinguishable from the learner's feedback perspective, if for every action of the learner, the probability distribution over the feedbacks received by learner is the same for p and $\tilde{p}$. Now assume that adversary always selects actions according to one of the two such indistinguishable strategies. Thus, the learner will always play one of s_p^* and $s_{\tilde{p}}^*$. By uniqueness, $s_p^* \neq s_{\tilde{p}}^*$. Then, the learner incurs a constant (non-zero) regret on any round where adversary plays according to p and learner plays $s_{\tilde{p}}^*$, or if the adversary plays according to $\tilde{p}$ and learner plays s_p^*. We show that in such a setting, adversary can simply play according to $(p + \tilde{p})/2$ and the learner suffers an expected regret of $\Omega(T)$.

Assume that the adversary selects $\{R_1, \ldots, R_T\}$ from product distribution $\otimes p$. Let the number of times the learner plays s_p^* and $s_{\tilde{p}}^*$ be denoted by random variables N_1^p and N_2^p respectively, where N^p shows the exclusive dependence on p. It is always true that $N_1^p + N_2^p = T$. Moreover, let the expected per round regret be ϵ_p when learner plays $s_{\tilde{p}}^*$, where the expectation is taken over the randomization of adversary. Now, assume that

Table 1: Relevance and probability vectors.

p	0.0	0.1	0.15	0.05	0.2	0.3	0.2	0.0
$\tilde{p}$	0.0	0.3	0.0	0.0	0.15	0.15	0.4	0.0
Rel.	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
	0	1	1	0	1	0	0	1
	0	1	0	1	0	1	0	1
	0	0	1	1	0	0	1	1

adversary selects $\{R_1, \ldots, R_T\}$ from product distribution $\otimes\tilde{p}$. The corresponding notations become $N_1^{\tilde{p}}$ and $N_2^{\tilde{p}}$ and $\epsilon_{\tilde{p}}$. Then,

$$\mathbb{E}_{(R_1,\ldots,R_T)\sim\otimes p}\mathbb{E}_{(s_1,\ldots,s_T)}[\text{Regret}((s_1,\ldots,s_T),(R_1,\ldots,R_T))] = 0\cdot\mathbb{E}[N_1^p] + \epsilon_p\cdot\mathbb{E}[N_2^p]$$

and

$$\mathbb{E}_{(R_1,\ldots,R_T)\sim\otimes\tilde{p}}\mathbb{E}_{(s_1,\ldots,s_T)}[\text{Regret}((s_1,\ldots,s_T),(R_1,\ldots,R_T))] = \epsilon_{\tilde{p}}\cdot\mathbb{E}[N_1^{\tilde{p}}] + 0\cdot\mathbb{E}[N_2^{\tilde{p}}]$$

Since p and $\tilde{p}$ are indistinguishable from perspective of learner, $\mathbb{E}[N_1^p] = \mathbb{E}[N_1^{\tilde{p}}] = \mathbb{E}[N_1]$ and $\mathbb{E}[N_2^p] = \mathbb{E}[N_2^{\tilde{p}}] = \mathbb{E}[N_2]$. That is, the random variable denoting number of times s_p^* is played by learner does not depend on adversary distribution (same for $s_{\tilde{p}}^*$.). Using this fact and averaging the two expectations, we get:

$$\mathbb{E}_{(R_1,\ldots,R_T)\sim\frac{\otimes p+\otimes\tilde{p}}{2}}\mathbb{E}_{(s_1,\ldots,s_T)}[\text{Regret}((s_1,\ldots,s_T),(R_1,\ldots,R_T))] = \frac{\epsilon_{\tilde{p}}}{2}\cdot\mathbb{E}[N_1] + \frac{\epsilon_p}{2}\cdot\mathbb{E}[N_2] \geq \min(\frac{\epsilon_p}{2},\frac{\epsilon_{\tilde{p}}}{2})\cdot\mathbb{E}[N_1+N_2] = \epsilon\cdot T$$

Since $\sup_{R_1,\ldots,R_T}\mathbb{E}[\text{Regret}((s_1,\ldots,s_T),(R_1,\ldots,R_T))] \geq \mathbb{E}_{(R_1,\ldots,R_T)\sim\frac{\otimes p+\otimes\tilde{p}}{2}}\mathbb{E}_{(s_1,\ldots,s_T)}[\text{Regret}((s_1,\ldots,s_T),(R_1,\ldots,R_T))]$, we conclude that for every learner algorithm, adversary has a strategy, s.t. learner suffers an expected regret of $\Omega(T)$.

Now, the thing left to be shown is the existence of two indistinguishable distributions p and $\tilde{p}$, s.t. $s_p^* \neq s_{\tilde{p}}^*$.

Characterization of indistinguishable strategies in our problem setting: Two adversary's strategies p and $\tilde{p}$ will be indistinguishable, in our problem setting, if for every score vector s, the relevances of the top-ranked item, according to s, are same for relevance vector drawn from p and $\tilde{p}$. Since relevance vectors are restricted to be binary, mathematically, it means that $\forall s$, $\mathbb{P}_{R\sim p}(R_{\pi_s(1)} = 1) = \mathbb{P}_{R\sim\tilde{p}}(R_{\pi_s(1)} = 1)$ (actually, we also need $\forall s$, $\mathbb{P}_{R\sim p}(R_{\pi_s(1)} = 0) = \mathbb{P}_{R\sim\tilde{p}}(R_{\pi_s(1)} = 0)$, but due to the binary nature, $\mathbb{P}_{R\sim p}(R_{\pi_s(1)} = 1) = \mathbb{P}_{R\sim\tilde{p}}(R_{\pi_s(1)} = 1) \implies \mathbb{P}_{R\sim p}(R_{\pi_s(1)} = 0) = \mathbb{P}_{R\sim\tilde{p}}(R_{\pi_s(1)} = 0)$). Since the equality has to hold $\forall s$, this implies $\forall j \in [m]$, $\mathbb{P}_{R\sim p}(R_j = 1) = \mathbb{P}_{R\sim\tilde{p}}(R_j = 1)$ (as every item will be ranked at top by some score vector). Hence, $\forall j \in [m]$, $\mathbb{E}_{R\sim p}[R_j] = \mathbb{E}_{R\sim\tilde{p}}[R_j] \implies \mathbb{E}_{R\sim p}[R] = \mathbb{E}_{R\sim\tilde{p}}[R]$. It can be seen clearly that the chain of implications can be reversed. Hence, $\forall s$, $\mathbb{P}_{R\sim p}(R_{\pi_s(1)} = 1) = \mathbb{P}_{R\sim\tilde{p}}(R_{\pi_s(1)} = 1) \iff \mathbb{E}_{R\sim p}[R] = \mathbb{E}_{R\sim\tilde{p}}[R]$.

Explicit adversary strategies: Following from the discussion so far and Theorem 5.1, if we can show existence of two strategies p and $\tilde{p}$ s.t. $\mathbb{E}_{R\sim p}[R] = \mathbb{E}_{R\sim\tilde{p}}[R]$, but $\text{argsort}\left(\mathbb{E}_{R\sim p}\left[\frac{G(\mathbf{R})}{Z_m(R)}\right]\right) \neq \text{argsort}\left(\mathbb{E}_{R\sim\tilde{p}}\left[\frac{G(\mathbf{R})}{Z_m(R)}\right]\right)$, we are done.

The 8 possible relevance vectors (adversary's actions) are $(R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8) = (000, 110, 101, 011, 100, 010, 001, 111)$. Let the two probability vectors be: $p = (0.0, 0.1, 0.15, 0.05, 0.2, 0.3, 0.2, 0.0)$ and $\tilde{p} = (0.0, 0.3, 0.0, 0.0, 0.15, 0.15, 0.4, 0.0)$. The data is provided in table format in Table. 1.

Under the two distributions, it can be checked that $\mathbb{E}_{R\sim p}[R] = \mathbb{E}_{R\sim\tilde{p}}[R] = (0.45, 0.45, 0.4)^\top$.

However, $\mathbb{E}_{R\sim p}\left[\frac{G(\mathbf{R})}{Z_m(R)}\right] = (0.3533, 0.3920, 0.3226)^\top$, but $\mathbb{E}_{R\sim\tilde{p}}\left[\frac{G(\mathbf{R})}{Z_m(R)}\right] = (0.3339, 0.3339, 0.4000)^\top$. Hence, $\text{argsort}\left(\mathbb{E}_{R\sim p}\left[\frac{G(\mathbf{R})}{Z_m(R)}\right]\right) = [2, 1, 3]^\top$ but $\text{argsort}\left(\mathbb{E}_{R\sim\tilde{p}}\left[\frac{G(\mathbf{R})}{Z_m(R)}\right]\right) \in \{[3, 1, 2]^\top, [3, 2, 1]^\top\}$.

□