Supplement Human Interaction with Recommendation Systems

A APPENDIX

Proof of Proposition 3. We prove the result by showing that the best item cannot always be ranked at the top, because that would depress its score s_{it} sufficiently much that it cannot be at the top.

Fix a sample path $\omega \in \Omega$. Note that by assumption, each arm is optimal for a constant fraction of agents. Define

$$x_{it} = \frac{|\{\tau : a_\tau = i\}|}{t}. \tag{1}$$

Then, if $\liminf_t x_{it} < x_i^*$ for some sufficiently small $x_i^* > 0$, we incur linear regret almost surely. Instead, assume that each arm is sampled a constant fraction, $\liminf_t x_{it} > \delta_i$ for some δ_i for each arm i. We note that the expected reward for the item ranked highest is

$$Q_i + \frac{p}{p + (1-p)^K} = Q + \rho, \tag{2}$$

where we define $\rho = \frac{p}{p+(1-p)^K}$: With probability p this item is chosen because of a positive signal, and with probability $(1-p)^K$ it is chosen because none of the items have a positive signal. For the other items, the expected reward is $Q_i + 1$.

To understand limiting behavior of the item scores, it is thus important to understand how often an item is ranked first by the platform. Define c_t as the fraction (up to time t) that the first (best) item is *not* ranked at the top:

$$c_t = \frac{|\{\tau < t : \exists j > 1 : s_{1\tau} < s_{j\tau}\}|}{t}. \tag{3}$$

We note that if $\limsup_t c_t > c^*$ for some $c^* > 0$, then the regret is linear.

Informally, we proceed by bounding $\mathbb{P}(\text{item is ranked first} \mid \text{item is selected})$, and use that to understand the evolution of the averages of ratings the platform observes. To bound the above probability, we note that there are two extremes when the item is not ranked first; it is ranked second, or ranked last. If it is always ranked second if the item is not ranked first, it is less likely the item was ranked first given selection than when it is either ranked first or last. If, overall, the item is ranked first with fraction y, then we obtain

$$\lambda(y) \le \mathbb{P}(\text{item ranked first} \mid \text{item selected}) \le \lambda'(y) \tag{4}$$

where

$$\lambda(y) = \frac{(1-y)(p + (1-p)^K)}{(1-y)(p+(1-p)^K) + yp(1-p)}, \tag{5}$$

and

$$\lambda'(y) = \frac{(1-y)(p + (1-p)^K)}{(1-y)(p+(1-p)^K) + yp(1-p)^{K-1}} \tag{6}$$

correspond to the two extreme cases. Note that λ and λ' are both decreasing.[1]

Now suppose $\limsup c_t = c$. By the stong law of large numbers, the empirical average converges to its mean and thus

$$\limsup_t s_{1t} \le Q_1 + \lambda(c)\rho + (1 - \lambda(c)), \tag{7}$$

where the second term corresponds to the expected reward from being ranked first and the last term corresponds to the contribution from when the action is not ranked first. Similarly

$$\liminf_t s_{2t} \ge Q_2 + \lambda'(1-c)\rho + (1 - \lambda'(1-c)), \tag{8}$$

almost surely by the mean-converging condition.

We note for $c = 0$, this leads to

$$\limsup_t s_{1t} \le Q_1 + \rho \text{ and } \liminf_t s_{2t} \ge Q_2 + 1 \tag{9}$$

This is a contradiction if $\Delta < \frac{(1-p)^K}{p+(1-p)^K}$, as this would imply the score of the second arm is higher in the limit than that of the first arm, while the first item is always ranked before the second item ($c = 0$):

$$\begin{aligned} \limsup_t s_{1t} &= Q_2 + \Delta + \rho && (10) \\ &< Q_2 + \frac{(1-p)^K}{p+(1-p)^K} + \frac{p}{p+(1-p)^K} && (11) \\ &\le \liminf_t s_{2t}. && (12) \end{aligned}$$

Furthermore, since λ and λ' are continuous and monotone, there must exist some $c^* \in (0,1)$ such that

$$Q_1 + \lambda(c^*)\rho + 1 - \lambda(c^*) \le Q_2 + \lambda'(1-c^*) + 1 - \lambda'(1-c^*) \tag{13}$$

[1] Both have the form $\frac{(1-x)a}{(1-x)a+xb}$ for $a, b \in (0,1)^2$, which has a negative derivative for $x \in (0,1)$

almost surely. Thus, if the first item is the top ranked item fracion $1-c^*$ of the time, then its score is almost surely lower than the second item, which is a contradiction. This implies that $\limsup_t c_t > c^*$ almost surely, which proves that the regret is linear. □

Proof Proposition 4. To bound the regret, we look at individual arms and note that if at time t all scores s_{it} are reasonably accurate, i.e. $|s_{it} - Q_i| < \lambda$ for all i, at such time the regret is at most 2λ. Furthermore, if $\lambda < \frac{\Delta_{\min}}{2}$, then the regret is 0 as each agent is compelled to pick the best item for them. Finally, it is important to note that the regret at each period is at most 2.

We proceed as follows; we use concentration to bound the estimation error when we have observed enough sample values. Furthermore, we show that due to natural exploration, we have a high probability guarantee of observing samples for each item. When combined, they lead to a logarithmic regret bound.

To use a concentration bound on the estimation error, we define event

$$A_m(i,\lambda) = \left\{ \exists s \in \{m,\dots,T\} : \frac{1}{s}\left|\sum_{j=1}^{s} \varepsilon_{ij}\right| > \lambda \right\}. \tag{14}$$

That is, $A_m(i,\lambda)$ is the bad event that after m pulls, there is some time t that the score s_{it} is off by more than λ.

Furthermore, we define events

$$B_m(i,M) = \{|S| < m : \tau \in S \iff a_\tau = i \text{ and } \tau < M\} \tag{15}$$

that indicate whether within M time steps, at least m users reported values for item i.

Using these two events, we can bound the expected regret by

$$\begin{aligned}\mathbb{E}[\text{regret}(T)] \leq \sum_{i=1}^{K} 2(\mathbb{P}(A_m(i,\lambda)) + \mathbb{P}(B_m(i,M)))T \\ + 2M + \lambda T \mathbb{I}_{\lambda > \frac{\Delta_{\min}}{2}}\end{aligned} \tag{16}$$

Bounding A_m Using the standard σ-sub-Gaussian concentration bound (see, for example, Wainwright [2015, Chapter 2]), we have

$$\mathbb{P}(A_m(i,\lambda)) \leq \mathbb{P}\left(\exists s \in \{m,\dots,t\} : \frac{1}{s}\left|\sum_{i=1}^{s}\varepsilon_i\right| > \lambda\right) \tag{17}$$

$$\leq \sum_{s=m}^{t} \mathbb{P}\left(\frac{1}{s}\left|\sum_{i=1}^{s}\varepsilon_i\right| > \lambda\right) \tag{18}$$

$$\leq 2\sum_{s=m}^{t} \exp\left(-\frac{s\lambda^2}{2\sigma^2}\right) \tag{19}$$

$$\leq 2\int_{m}^{t+1} \exp\left(-\frac{s\lambda^2}{2\sigma^2}\right) ds \tag{20}$$

$$\leq \frac{4\sigma^2}{\lambda^2}\exp\left(-\frac{m\lambda^2}{2\sigma^2}\right) \tag{21}$$

Now set

$$m = \frac{2\sigma^2(\log(T) - \log(\lambda))}{\lambda^2}, \tag{22}$$

and obtain

$$\mathbb{P}(A_m(i,\lambda)) \leq \frac{4\sigma^2}{\lambda^2}\exp\left(-\frac{m\lambda^2}{2\sigma^2}\right) = \frac{4\sigma^2}{\lambda T} \tag{23}$$

Bounding B_m From the above, we know that the estimation error concentrates well after observing m selections. Now we show that with high probability, it does not take too long to wait for m selections.

First note that the probability of selection of any item at any time t is at least $2^{1-K}\gamma$. This follows from the conditions imposed on F_i. For $M > m$, we note that the probability that we have not observed m selections is lower bounded by a Binomial random variable $Z \sim B(M, 2^{1-K}\gamma)$ since preferences are independent between agents. Consider

$$M = \frac{2\alpha m}{2^{1-K}\gamma} = \frac{4\alpha\sigma^2(\log(T) - \log(\lambda))}{\lambda^2 2^{1-K}\gamma} \tag{24}$$

where $\alpha = \max\left(1, 2\lambda^2/\sigma^2\right)$.

First we note that in this case, $\mathbb{E}(Z) = 2\alpha m \geq 2m$ and thus

$$\mathbb{P}(B_m) \leq \mathbb{P}\left(Z \leq \frac{1}{2}\mathbb{E}(Z)\right) \tag{25}$$

$$\leq \exp\left(-\frac{\mathbb{E}(Z)}{8}\right) \tag{26}$$

$$\leq \exp\left(-\frac{\alpha\sigma^2(\log(T) - \log(\lambda))}{2\lambda^2}\right) \tag{27}$$

$$\leq \frac{\lambda}{T} \tag{28}$$

where third inequality is a standard Chernoff bound and the second to last step follows from the condition on α.

Plugging these bounds on $A_m(i, \lambda)$ and $B_m(i, M)$ in to our bound for regret (16), we obtain

$$\begin{aligned}\mathbb{E}[\text{regret}(T)] \leq 2\left(\frac{4\sigma^2}{\lambda} + \lambda\right)K \\ + \frac{8\alpha\sigma^2 K(\log(T) - \log(\lambda))}{\lambda^2 2^{1-K}\gamma} + \lambda K T \mathbb{I}_{\lambda > \frac{\Delta_{\min}}{2}}\end{aligned} \tag{29}$$

and thus if we set $\lambda = \frac{\Delta_{\min}}{2}$, we find

$$\begin{aligned}\mathbb{E}[\text{regret}(T)] \leq \left(\frac{16\sigma^2}{\Delta_{\min}} + \Delta_{\min}\right)K \\ + \frac{32\alpha\sigma^2 K(\log(T) - \log(\Delta_{\min}) + \log(2))}{\Delta_{\min}^2 2^{1-K}\gamma}\end{aligned} \tag{30}$$

as desired. □

References

Martin Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint.* Forthcoming, 2015.