Supplemental Material

1 Proof of Algorithm 1

Proof. The proof follows standard techniques (such as [1]) once we rewrite our algorithm as FTRL in the space of $\mathbb{R}^{MK}$. First we extend the loss vector $\hat{l}_t \in \mathbb{R}^K$ to a vector $L_t \in \mathbb{R}^{MK}$ by padding zeros to irrelevant coordinates. Formally, $L_t = \hat{l}_t(i_t)e_{(j_t-1)K+i_t}$ where $e_1, \ldots, e_{KM}$ are standard basis vectors in $\mathbb{R}^{MK}$. Further let $G_t = -\sum_{s=1}^{t} L_s$ be the negative cumulative loss estimator up to time t. Define $\Psi^*(G) = \max_{P\in\Omega} \langle P, G\rangle - \frac{1}{\eta}\Psi(P)$, which is the convex conjugate of the function $\frac{1}{\eta}\Psi(P) + \mathbf{1}_\Omega(P)$ where $\mathbf{1}_\Omega(P)$ is 0 if $P \in \Omega$ and ∞ otherwise. With these notations we then have

$$\begin{aligned} P_t &= \arg\min_{P\in\Omega} \left\langle P, \sum_{s=1}^{t-1} L_s \right\rangle + \frac{1}{\eta}\Psi(P) \\ &= \arg\max_{P\in\Omega} \langle P, G_{t-1}\rangle - \frac{1}{\eta}\Psi(P) = \nabla\Psi^*(G_{t-1}). \end{aligned}$$

Next, note that the loss estimators are unbiased since $\mathbb{E}[\hat{l}_t(i)] = \mathbb{E}\left[p_t^{j_t}(i) \times \frac{l_t(i)}{p_t^{j_t}(i)}\right] = \mathbb{E}[l_t(i)]$ for all $i \in [K]$.

We can thus rewrite the regret as $\text{Reg} = \mathbb{E}\left[\langle P_*, G_T\rangle + \sum_{t=1}^{T} \langle \nabla\Psi^*(G_{t-1}), L_t\rangle\right]$ where $P_* = \arg\max_{P\in\Omega} \mathbb{E}\left[\langle P, G_T\rangle\right]$. Recalling the Bregman divergence associated with Ψ^* defined as

$$D_{\Psi^*}(G, G') = \Psi^*(G) - \Psi^*(G') - \langle \nabla\Psi^*(G'), G - G'\rangle .$$

we further rewrite the regret as

$$\begin{aligned} \text{Reg} &= \mathbb{E}\left[\langle P_*, G_T\rangle + \sum_{t=1}^{T} \left(\Psi^*(G_{t-1}) - \Psi^*(G_t) + D_{\Psi^*}(G_t, G_{t-1})\right)\right] \\ &= \mathbb{E}\left[\langle P_*, G_T\rangle + \Psi^*(G_0) - \Psi^*(G_T) + \sum_{t=1}^{T} D_{\Psi^*}(G_t, G_{t-1})\right]. \end{aligned}$$

The first three terms can be bounded as (note $G_0 = \mathbf{0}$)

$$\begin{aligned} \mathbb{E}\left[\langle P_*, G_T\rangle - \frac{1}{\eta}\min_{P\in\Omega}\Psi(P) - \langle P_*, G_T\rangle + \frac{1}{\eta}\Psi(P^*)\right] &\le \frac{\max_{P\in\Omega}\Psi(P) - \min_{P\in\Omega}\Psi(P)}{\eta} \\ &= \frac{D}{\eta}. \end{aligned}$$

It remains to bound $\mathbb{E}\left[D_{\Psi^*}(G_t, G_{t-1})\right]$. By Taylor's theorem, there exists $\tilde{G}_t$ on the segment connecting G_{t-1} and G_t such that $D_{\Psi^*}(G_t, G_{t-1}) = \frac{1}{2} L_t^\top \nabla^2 \Psi^*(\tilde{G}_t) L_t$. Moreover, using properties of convex conjugates (see for example [1]) we have $\nabla^2 \Psi^*(\tilde{G}_t) \preceq \eta \nabla^{-2} \Psi(\nabla \Psi^*(\tilde{G}_t))$. Realizing that for any $P \in \Omega$, $\eta \nabla^{-2} \Psi(P)$ is a diagonal matrix with ηP on the diagonal, we further bound the Bregman divergence by

$$D_{\Psi^*}(G_t, G_{t-1}) \leq \frac{\eta}{2} \nabla \Psi^*(\tilde{G}_t)_{(j_t-1)K+i_t} \hat{l}_t^2(i_t).$$

Note that $\tilde{G}_t$ is the same as G_{t-1} for all coordinates except the $((j_t-1)K+i_t)$-th one, where the value could only be smaller (if not equal) by the non-negativity of losses. By the convexity of Ψ^* (and thus monotonicity of $\nabla \Psi^*$), we then have

$$\begin{aligned} D_{\Psi^*}(G_t, G_{t-1}) &\leq \frac{\eta}{2} \nabla \Psi^*(G_{t-1})_{(j_t-1)K+i_t} \hat{l}_t^2(i_t) \\ &= \frac{\eta}{2} p_t^{j_t}(i_t) \hat{l}_t^2(i_t) = \frac{\eta l_t^2(i_t)}{2 p_t^{j_t}(i_t)} \leq \frac{\eta}{2 p_t^{j_t}(i_t)}. \end{aligned}$$

Taking expectation on both sides gives $\mathbb{E}\left[D_{\Psi^*}(G_t, G_{t-1})\right] \leq \frac{\eta K}{2}$.

Combing everything above we arrive at

$$\text{Reg} \leq \frac{D}{\eta} + \frac{\eta T K}{2},$$

which is of order $O(\sqrt{TKD})$ with the optimal choice of learning rate $\eta = \sqrt{\frac{D}{TK}}$. Finally, note that

$$D \leq -\min_{P \in \Omega} \Psi(P) = \sum_{j=1}^{M} \sum_{i=1}^{K} p_1^j(i) \ln \frac{1}{p_1^j(i)} \leq M \ln K$$

by the fact the entropy of a distribution over K items is nonnegative and is at most $\ln K$. This finishes the proof. □

2 Unknown Context Distributions

The Fair CB algorithm described in the paper assumes that the context distribution q is known to the learner. We provide an extension of our algorithm to the case where the context distribution q is unknown. We include regret guarantee of the algorithm, while we leave empirical results for future work.

A natural idea is to maintain an empirical context distribution based on the observations and to use it as a proxy for q. Specifically, to avoid changing the feasible set too often, we divide the entire horizon into $O(\log_2 T)$ epochs, where epoch k contains rounds $\tau_k, \ldots, \tau_{k+1} - 1$ with $\tau_k = 2^{k-1}$. Within epoch $k > 1$, we let q_k be the empirical context distribution using observations from the last $k-1$ epochs:

$$q_k(j) = \frac{1}{\tau_k - 1} \sum_{t=1}^{\tau_k - 1} \mathbf{1}\{j_t = j\}, \ \forall j \in [M]. \tag{1}$$

Note that by standard concentration argument (specifically Bernstein inequalities and union bound), we have with probability at least $1 - 1/T$,

$$\|q - q_k\|_1 \leq \epsilon_k \triangleq 4\sqrt{\frac{M \ln(TM)}{\tau_k - 1}} + \frac{2M \ln(TM)}{\tau_k - 1}. \tag{2}$$

Accordingly, for epoch $k > 1$ we define the feasible set Ω_k as

$$\Omega_k = \left\{ P = (p^1, ..., p^M) \,\middle|\, \begin{array}{c} p^1, ..., p^M \in \Delta_K \text{ and} \\ \sum_{j=1}^M q_k(j) p^j(i) \geq v - \epsilon_k, \forall i \in [K] \end{array} \right\}, \tag{3}$$

where we introduce a small slack ϵ_k to the fairness constraint v. The reason of relaxing the constraint is to make sure that Ω_k always contains Ω with high probability. Indeed, conditioning on the event Eq. (2), for any $P \in \Omega$ we have

$$\sum_{j=1}^M q_k(j) p^j(i) \geq \sum_{j=1}^M q(j) p^j(i) - \|q - q_k\|_1 \geq v - \epsilon_k$$

and thus $P \in \Omega_k$. On the other hand, relaxing the constraint means that the algorithm no longer always strictly satisfies the fairness requirement. Instead, we measure the fairness of the algorithm by the average amount of violation of the fairness constraint, defined as

$$\text{Vio} = \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^T \max\left\{0, v - \min_{i \in [K]} \sum_{j=1}^M q(j) p_t^j(i)\right\}\right]$$

where p_t^j is again the distribution of arm i_t given the history and $j_t = j$.

Our final algorithm simply runs a new instance of Algorithm 1 with feasible set Ω_k on epoch k. See Algorithm 1 for the pseudocode. In the following theorem, we show that the algorithm ensures the same regret bound while keeping the per-round fairness violation to be arbitrarily small as long as T is large enough.

Theorem 1. *Algorithm 1 ensures*

$$\textit{Reg} = O\left(\sqrt{TMK \ln K}\right) \textit{ and } \textit{Vio} = O\left(\sqrt{\frac{M \ln(TM)}{T}} + \frac{M \ln(TM) \ln T}{T}\right).$$

Proof. Clearly we only need to condition on the event Eq. (2) since it happens with probability at least $1 - 1/T$. With the fact $P_* \in \Omega_k$ for all k, the regret guarantee is a simple application of Theorem 1. Indeed, let $K = O(\log_2 T)$ be the total number of epochs, we have

$$\begin{aligned} \text{Reg} &= \sum_{k=1}^K \sum_{t=\tau_k}^{\min\{\tau_{k+1}-1, T\}} \mathbb{E}\left[\left\langle p_t^{j_t} - p_*^{j_t}, l_t \right\rangle\right] \\ &= \sum_{k=1}^K O\left(\sqrt{\tau_k MK \ln K}\right) = O\left(\sqrt{TMK \ln K}\right). \end{aligned}$$

Algorithm 1 Fairness CB with Unknown Context Distribution

Input: fairness constraint parameter v
Define: $\tau_k = 2^{k-1}$, $\Psi_k(P) = \frac{1}{\eta_k} \sum_{j=1}^{M} \sum_{i=1}^{K} \psi(p^j(i))$ where $\psi(p) = p \ln p$ and $\eta_k = \sqrt{M \ln K / (\tau_k K)}$
For $t = 1$, sample an arm uniformly at random
for $k = 2, 3, \ldots$ **do**
    Update q_k and Ω_k according to Eq. (1) and Eq. (3)
    for $t = \tau_k, \ldots, \tau_{k+1} - 1$ **do**
        Compute $P_t = \arg\min_{P \in \Omega_k} \sum_{s=\tau_k}^{t-1} \left\langle p^{j_s}, \hat{l}_s \right\rangle + \Psi_k(P)$
        Observe j_t and play $i_t \sim p_t^{j_t}$
        Construct loss estimator $\hat{l}_t(i) = \frac{l_t(i)}{p_t^{j_t}(i)} \mathbf{1}\{i_t = i\},\ \forall i$
    end for
end for

The amount of violation is also clear due to the construction of Ω_k:

$$\text{Vio} \leq \frac{1}{T} \sum_{k=1}^{K} \tau_k \epsilon_k = O\left(\sqrt{\frac{M \ln(TM)}{T}} + \frac{M \ln(TM) \ln T}{T} \right).$$

This finishes the proof. □

References

[1] J. D. Abernethy, C. Lee, and A. Tewari. Fighting bandits with a new kind of smoothness. In *Advances in Neural Information Processing Systems*, pages 2197–2205, 2015.