Dueling bandits

immediate

September 26, 2014

We define the set $T_p = \{2^{p-1}, \ldots, 2^p - 1\} = \{s \in \mathbb{N} : \lfloor \log_2 s \rfloor = p - 1\}$.

Algorithm 1: Improved Doubler

```
initialization x_1 fixed in X, L = {x_1}, f̂_0 = 0;
t ← 1;
p ← 1;
while true do
    for j = 1 to 2^{p-1} do
        choose x_t uniformly from L;
        y_t ← advance(S);
        play (x_t; y_t), observe choice b_t;
        feedback(S; b_t + f̂_{p-1});
        t ← t + 1;
    L the multi-set of arms played as y_t in the last for-loop;
    f̂_p ← f̂_p + Σ_{s∈T_p} b_s/2^{p-1} − 1/2;
    p ← p + 1;
```

Observe that if $t \in T_p$

$$\mathbb{E}\left[b_t \middle| \left\{y_s, s \in T_{p-1}\right\}, y_t\right] \quad = \quad \sum_{s \in T_{p-1}} \frac{\mu(y_t) - \mu(y_s) + 1}{2^{p-1}} = \frac{\mu(y_t) + 1}{2} - \sum_{s \in T_{p-1}} \frac{\mu(y_s)}{2^{p-1}} \,,$$

and that

$$\mathbb{E}\left[\sum_{s \in T_{p-1}} b_s / 2^{p-2} - 1/2 \middle| \bigcup_{r=p-2}^{p-1} \{y_s, s \in T_r\}\right] \quad = \quad \sum_{s \in T_{p-1}} \frac{\mu(y_s)}{2^{p-1}} - \sum_{s \in T_{p-2}} \frac{\mu(y_s)}{2^{p-2}} \,.$$

Let us denote $f_t = b_t + \hat{f}_{p-1} = b_t + \sum_{r=1}^{\lfloor \log_2 t \rfloor} \sum_{s \in T_r} b_s / 2^{r-1} - \lfloor \log_2 t \rfloor / 2$ the feedback that we introduce in S. Using the recurrence defining $\hat{f}_p$ we obtain

$$\mathbb{E}\left[f_t \middle| x_1, \bigcup_{r=1}^{p-1} \{y_s, s \in T_r\}, y_t \right] = \frac{\mu(y_t) - \mu(x_1) + 1}{2} .$$

Since the above right term is $\sigma(x_1, y_t)$- measurable we conclude that

$$\mathbb{E}\left[f_t | x_1, y_t \right] = \frac{\mu(y_t) - \mu(x_1) + 1}{2} .$$

Let $y_{t_1} = \ldots = y_{t_k}$ and let $f = \sum_{j=1}^{k} f_{t_j}/k$. Observe that $f = \sum_{s=1}^{t_k} a_s b_s / t_k - \sum_{j=1}^{k} \log_2 t_j / 2k$ with $a_s \in [0, A_s]$. We will later specify this bound. Since the $a_s b_s$ are independent and $\mathbb{P}[a_s b_s \in [0, A_s]] = 1$ we can apply the Hoeffding's inequality:

$$\mathbb{P}\left[f - \mathbb{E}[f] \geq \epsilon \middle| x_1, y_{t_1} = \ldots = y_{t_k} \right] \leq \exp\left(-\frac{2t_k^2 \epsilon^2}{\sum_{s=1}^{t_k} A_s^2} \right)$$

Assume the convention $t_0 = 1$ and set $S_j = \{s \in \mathbb{N} : 2^{\lfloor \log_2 t_{j-1} \rfloor} \leq s < 2^{\lfloor \log_2 t_j \rfloor}\}$ for $j = 1, \ldots, k$ and $S_{k+1} = \{s \in \mathbb{N} : 2^{\lfloor \log_2 t_k \rfloor} \leq s < t_k\}$. For each $1 \leq j \leq k+1$, $s \in S_j$, $A_s = t_k((k-j+1)/2^{\lfloor \log_2 s \rfloor} + 1)/k$ if $s = t_i$ for some $1 \leq i < j$ and $A_s = t_k(k-j+1)/(2^{\lfloor \log_2 s \rfloor} k)$ otherwise. We say that $t_j \in S_{t(j)}$. The function t is obviously non decreasing, $t(j) \geq j+1$ and $t(k) = k+1$.

$$\begin{aligned}
\frac{k^2}{t_k^2} \sum_{s=1}^{t_k} A_s^2 = \sum_{j=1}^{k+1} \sum_{s \in S_j} \frac{(k-j+1)^2}{2^{2\lfloor \log_2 s \rfloor}} + \sum_{j=1}^{k} \frac{k - t(j) + 1}{2^{\lfloor \log_2 t_j \rfloor - 1}} + k = \\
\sum_{j=1}^{k} (k-j+1)^2 \left(\frac{1}{2^{\lfloor \log_2 t_{j-1} \rfloor + 1}} - \frac{1}{2^{\lfloor \log_2 t_j \rfloor + 1}} \right) + \sum_{j=1}^{k} \frac{4(k - t(j)) + 4}{2^{\lfloor \log_2 t_j \rfloor + 1}} + k \\
= \frac{k^2}{2} - \sum_{j=1}^{k} \frac{2(k-j)+1}{2^{\lfloor \log_2 t_j \rfloor + 1}} + \sum_{j=1}^{k} \frac{4(k-t(j))+4}{2^{\lfloor \log_2 t_j \rfloor + 1}} + k .
\end{aligned}$$

This implies that

$$k^2 \geq \frac{k^2}{2} + \sum_{j=1}^{k} \frac{2(k-j)-1}{2^{\lfloor \log_2 t_j \rfloor + 1}} + k \geq \frac{k^2}{t_k^2} \sum_{s=1}^{t_k} A_s^2 \geq \frac{k^2}{2} - \sum_{j=1}^{k} \frac{2(k-j)+1}{2^{\lfloor \log_2 t_j \rfloor + 1}} \geq \frac{k^2}{2} \left(1 - \frac{1}{t_k} \right) .$$

To obtain the convergence of Improved Doubler using UCB at the same rate as UCB we need that $\sum_{s=1}^{t_k} A_s^2 = O(t_k^2/k)$, but we just showed that it is not possible.

$$\sum_{s=1}^{t} t^{-\beta/s} \tag{0.1}$$