Dueling

immediate

June 27, 2014

We define the set $T_p = \{2^{p-1}, \ldots, 2^p - 1\}$.

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 + ∑_{s∈T_p} b_s/2^{p-1} − 1/2;
    p ← p + 1;
```

Observe that if $t \in T_p$

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

and that

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

Using the recurrence defining $\hat{f}_p$ we obtain

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

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

$$\mathbb{E}\left[b_t + \hat{f}_{p-1} \,|x_1, y_t\right] \quad = \quad \frac{\mu\,(y_t) - \mu\,(x_1) + 1}{2} \; .$$