DATE
PAGE

CS 747: Weekly Quiz 3

VIBHAV AGGARWAL
190050128

a) We have,

$$ucb_a^t = \hat{p}_a^t + \sqrt{\frac{1}{2u_a^t} \ln\left(\frac{1}{\delta(t)}\right)}$$

$$\geq \hat{p}_a^t + \sqrt{\frac{1}{2u_a^t} \ln\left(\frac{1}{\delta'(t)}\right)} \quad [\because \delta'(t) \geq \delta(t) > 0]$$

$$= \hat{p}_a^t + \sqrt{\frac{1}{2u_a^t} \cdot \{u_a^t KL(\hat{p}_a^t, ucb\text{-}kl_a^t)\}}$$

[$\because$ Due to the def^n of $ucb\text{-}kl_a^t$]

$$= \hat{p}_a^t + \sqrt{\frac{1}{2} \cdot KL(\hat{p}_a^t, ucb\text{-}kl_a^t)}$$

$$\geq \hat{p}_a^t + \sqrt{\frac{1}{2} \cdot 2(\hat{p}_a^t - ucb\text{-}kl_a^t)^2}$$

[Pinsker's Inequality]

$$= \hat{p}_a^t + |\hat{p}_a^t - ucb\text{-}kl_a^t|$$

$$= \hat{p}_a^t + ucb\text{-}kl_a^t - \hat{p}_a^t \quad [\because ucb\text{-}kl_a^t \geq \hat{p}_a^t]$$

$$= ucb\text{-}kl_a^t$$

Hence, $ucb_a^t \geq ucb\text{-}kl_a^t$ as desired.

DATE
PAGE

b) We should not expect a lower regret from the ucb-proposed algorithm because a tighter upper bound may not always result in lower regret.

Take, for example, the extreme case when the upper confidence bound is simply $\hat{p}_a$. This is the greedy strategy and we know this incurs linear regret.

In fact, we cannot achieve a lower regret than the one given by KL-UCB because it matches the constant given by Lai and Robbins.