Week 7 - Lecture 34

UCB Algorithm

combines exploration & exploitation

Index in UCB for arm i at time t is :

$$UCB_i(t) = \hat{\mu}_i(t-1) + \sqrt{\frac{\alpha \log t}{T_i(t-1)}}$$

where $T_i(t-1) = \sum_{s=1}^{t-1} \mathbb{1}_{\{I_s = i\}}$

Thus :

$$\underbrace{\hat{\mu}_i(t-1)}_{\text{random quantity}} = \frac{\sum_{s=1}^{t-1} \mathbb{1}\{I_s = i\} X_{is}}{\underbrace{T_i(t-1)}_{\text{random quantity}}}$$

Let :

$$\hat{\mu}_{iu}(t-1) = \frac{\sum_{s=1}^{t-1} X_{is}}{u}$$

fixing no. of samples observed till (t-1) to be u

In UCB we have finally:

$$I_t = \arg\max_i UCB_i(t-1)$$

Let $\mu^* = \max_i \mu_i$ $\therefore$ $\underset{\substack{\uparrow \\ \text{optimal arm}}}{i^*} = \arg\max_i \mu_i$

$\forall i, \; i \neq i^*$ ← sub-optimal arms

$\Delta_i = \mu^* - \mu_i$ ← sub-optimality gap of arm i

$\Delta = \min\limits_{i \neq i^*} \Delta_i$ ← sub-optimality gap

Assuming that optimal arm is unique

$\therefore \quad \Delta > 0$

We know:

$$R(\pi, n) = \sum_{i=1}^{K} \mathbb{E}\left[T_i(n)\right] \Delta_i$$

<u>Proof of regret bound for UCB</u>

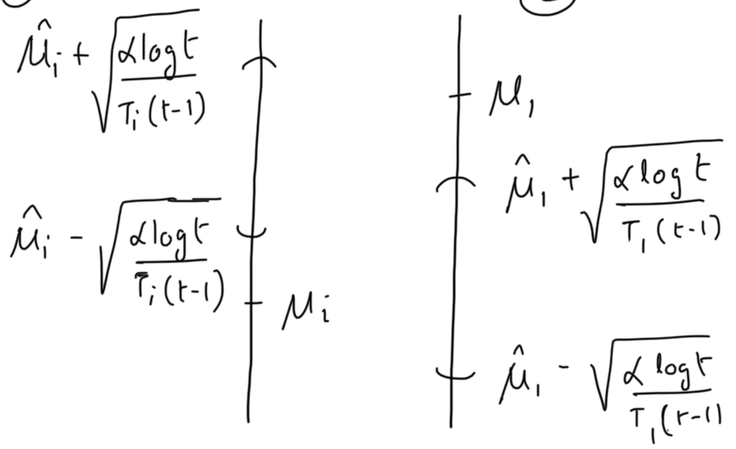
Assume that sub-optimal arm is played in round t.

$I_t = i \quad , i \neq i^*$

[Assuming $i^* = 1$, arm 1 is optimal, without loss of generality.]

How might this have come about:

1 2



3 arm is not sampled enough

1 $\hat{\mu}_i - \sqrt{\dfrac{\alpha \log t}{T_i(t-1)}} > \mu_i$

2 $\hat{\mu}_1 + \sqrt{\dfrac{\alpha \log t}{T_i(t-1)}} \leq \mu_1$

3 $T_i(t-1) \leq \dfrac{\alpha \log n}{\Delta_i^2/4} \Rightarrow \Delta_i \leq 2\sqrt{\dfrac{\alpha \log n}{T_i(t-1)}}$

<u>Claim</u>: If $I_t = i$ in round t at least one of

①, ② or ③ must hold

To prove this claim :

Suppose none of this holds then $I_t \neq i$

Assume any of ①, ②, ③ holds

Then :

Actually should be

$$\underbrace{\hat{\mu}_1 + \sqrt{\frac{\alpha \log t}{T_1(t-1)}}}_{= \, UCB_1(t)} \geq \mu_1 = \mu_i + \Delta_i \geq \left(\mu_i + 2\sqrt{\frac{\alpha \log n}{T_i(t-1)}}\right)$$

$$\geq \left(\mu_i + 2\sqrt{\frac{\alpha \log t}{T_i(t-1)}}\right)$$

← playing till round t, where $t \leq n$ ∴ this holds

$$> \hat{\mu}_i + \sqrt{\frac{\alpha \log t}{T_i(t-1)}} = UCB_i(t)$$

$\Rightarrow I_t \neq i$ a contradiction !

All the conditions have been shown to have been violated, and if this happens then $UCB_1(t)$ dominates $UCB_i(t)$, thus the optimal arm should have been played.

Thus if sub-optimal arm is played then one of the three conditions should hold.

Let, $u = \left\lceil \dfrac{\alpha \log n}{\Delta_i^2} \right\rceil$ (u is an integer thus taking ceal, but may not write ceal all the time)

$$T_i(n) = \sum_{t=1}^{n} 1\{I_t = i\}$$ [t can be any round even last round]

$$\mathbb{E}[T_i(n)] = \mathbb{E}\left[T_i(n)\, 1\{T_i(n) \le u\}\right] + \mathbb{E}\left[T_i(n)\, 1\{T_i(n) > u\}\right]$$

$$\le u + \mathbb{E}\left[\sum_{t=1}^{n} 1\{I_t = i\}\, 1\{T_i(n) > u\}\right]$$

$$= u + \mathbb{E}\left[\sum_{t=1}^{n} 1\{I_t = i, T_i(n) > u\}\right]$$

joint condition

Here the 3 condition is violated

$$= u + \mathbb{E}\left[\sum_{t=1}^{n} 1\{I_t = i, \text{1 or 2 holds}\}\right]$$

$$= u + \mathbb{E}\left[\sum_{t=1}^{n} 1\{I_t = i, \text{1 holds}\}\right] + \mathbb{E}\left[\sum_{t=1}^{n} 1\{I_t = i, \text{2 holds}\}\right]$$

$$\Rightarrow \mathbb{E}\left[T_i(n)\right] \leqslant u + \sum_{t=1} \Pr\left\{ I_t = i \,,\, 1 \text{ holds} \right\}$$

$$+ \sum_{t=1}^{n} \Pr\left\{ I_t = i,\ 2 \text{ holds} \right\}$$

$$\leqslant u + \sum_{t=1}^{n} \Pr\left\{ 1 \text{ holds} \right\} + \sum_{t=1}^{n} \Pr\left\{ 2 \text{ holds} \right\}$$

(upper bounding)

$$\Pr\left\{ 1 \text{ holds} \right\} = \Pr\left\{ \hat{\mu}_i - \sqrt{\frac{\alpha \log t}{T_i(t-1)}} > \mu_i \right\} \quad \text{—— } i$$

$$\Pr\left\{ 2 \text{ holds} \right\} = \Pr\left\{ \hat{\mu}_1 + \sqrt{\frac{\alpha \log t}{T_i(t-1)}} < \mu_1 \right\} \quad \text{—— } ii$$

We know how to bound i & ii,

will continue with that in the next lecture . . .