LECTURE 20

THEORETICAL ANALYSIS OF BANDIT ALGORITHMS

Setup: K coins

success probabilities: $p_1 \dots p_K$

T rounds. (e.g: $K=9$, $p_1=0.1, p_2=0.2, \dots, p_K=0.9$)

Each round: You pick a coin for tossing

1 \$ if heads

\$0 if tails

Given an "algorithm";

Reward (t) = Money made after t rounds.

t: number of rewards

$$\text{Regret}(t) = 0.9\,t - \text{Reward}(t)$$

$$= t \cdot \max_a p_a - \text{Reward}(t)$$

t: time

Averaged Regret (t)

$$= \mathbb{E}(\text{Regret}(t))$$

$$= t \cdot \max_a p_a - \mathbb{E}(\text{Reward}(t))$$

$$\boxed{\text{Averaged Regret increases with } t}$$

Averaged Regret $(t+1)$

$$= (t+1)\max_a p_a - \mathbb{E}[\text{Reward}(t+1)]$$

$$= t\max_a p_a + \mathbb{E}(\text{Reward}(t))$$

$$+ \max_a p_a - \mathbb{E}[\text{Reward}(t+1) - \text{Reward}(t)]$$

$$\max_a \underbrace{p_a - \mathbb{E}[\text{payoff in Round}(t+1)]}_{\geq 0}$$

So Averaged Regret $(t+1) \geq$ Averaged Regret (t)

Example ① $K = 9$

$p_1 = 0.1,\ p_2 = 0.2, \ldots, p_8 = 0.8, p_9 = 0.9$

Algorithm: At each round, picks one of coin 8 or coin 9 at random.

Qn: What is the Averaged Regret of this Algorithm?

Averaged Regret (t)

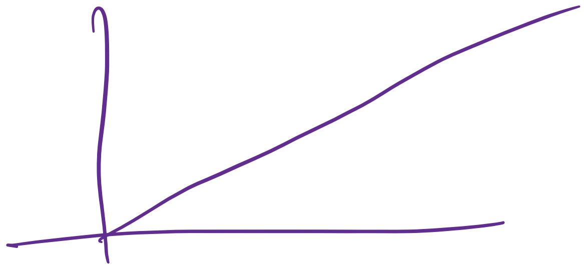
$$= t\max_a p_a - \mathbb{E}(\text{Reward}_t)$$

$$= t(0.9) - \mathbb{E}\sum_{s=1}^{t} I\{\text{success in } s^{th} \text{ round}\}$$

$$= t(0.9) - \sum_{s=1}^{t} \mathbb{P}\{\text{success in } s^{th} \text{ round}\}$$

$$= t(0.9) - \sum_{s=1}^{t}\left[\tfrac{1}{2}(0.8) + \tfrac{1}{2}(0.9)\right]$$

$$= t(0.9) - t(0.85) = t(0.9 - 0.85)$$
$$= t(0.05)$$



$\boxed{\text{Linear Regret Algorithm}}$

Example 2:

In round t:

0.8 coin $\rightarrow p_t$

$\boxed{0.9 \text{ coin} \rightarrow 1 - p_t}$

p_t : decreases with t.

(e.g $p_t = \frac{1}{t}$)

Averaged Regret for this algorithm.

$$= t(0.9) - \mathbb{E} \sum_{s=1}^{t} I\{\text{success in Round } s\}$$

$$= t(0.9) - \sum_{s=1}^{t} \mathbb{P}(\text{success in Round } s)$$

$$= t(0.9) - \sum_{s=1}^{t} \left[(0.8)(p_s) + (0.9)(1 - p_s)\right]$$

$$= t(0.9) - \sum_{s=1}^{t} \left[0.9 - 0.1\, p_s\right]$$

$$= 0.1 \sum_{s=1}^{t} p_s$$

$\boxed{\text{With } p_s = \frac{1}{s}}$

Averaged Regret

$$= 0.1\left(\underbrace{1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{t}}\right)$$

$\sim (0.1)(\log t)$

If $\rho_s = \frac{1}{s^2}$

$$\text{Averaged Regret} = 0.1 \sum_{s=1}^{t} \frac{1}{s^2}$$

$\leq$ Constant.

Uniform Confidence Bound.

① Basic Exploration $(m=1)$

② In Round (t),

$$\begin{cases} T_a(t) = \#\text{times coin } a \text{ is picked in the first } t \text{ rounds} \\ X_a(t) = \#\text{ successes from coin } a \text{ in the first } t \text{ rounds.} \end{cases}$$

$$UCB_a(t, \delta) = \frac{X_a(t)}{T_a(t)} + \underbrace{\sqrt{\frac{\log(1/\delta)}{2\, T_a(t)}}}$$

$UCB_a(t-1, \delta)$

pick the coin a which maximizes this.

Idea $\to$ $\mathbb{P}\left[\, p_a \leq \underbrace{UCB_a(t, \delta)}\,\right] \geq 1-\delta$

Hoeffding for Binomial.

$X \sim \text{Bin}(n, p)$, $t \geq np$

$$\mathbb{P}(X \geq t) \leq \underbrace{\exp\left[-\frac{2(t-np)^2}{n}\right]}_{=\delta}$$

Solve for t:

$$t = np + \sqrt{\frac{n}{2}\log\frac{1}{\delta}}$$

$$\frac{t}{n} = p + \sqrt{\frac{\log 1/\delta}{2n}}$$

$$\mathbb{P}\left[X \geq np + \sqrt{\frac{n}{2}\log\frac{1}{\delta}}\right] \leq \delta$$

$$\mathbb{P}\left[p \leq \frac{X}{n} - \sqrt{\frac{\log 1/\delta}{2n}}\right] \leq \delta$$

$$\mathbb{P}\left[p > \underbrace{\frac{X}{n} - \sqrt{\frac{\log 1/\delta}{2n}}}_{\text{LCB (Lower Confidence Bound)}}\right] \geq 1-\delta$$

UCB from Hoeffding:

$$\mathbb{P}\left[p < \frac{X}{n} + \sqrt{\frac{\log 1/\delta}{2n}}\right] \geq 1-\delta$$

UCB Bandit Algorithm

After $t-1$ rounds, consider the maximum possible value for each p_a subject to a given prob δ.

$\boxed{UCB_a(t-1, \delta)} \rightarrow \frac{X_a(t-1)}{T_a(t+1)} + \sqrt{\frac{\log 1/\delta}{2T_a(t-1)}}$

$p_a \rightarrow \boxed{\begin{matrix} T_a(t-1) \\ X_a(t-1) \end{matrix}}$

Qn: What is the Averaged Regret of UCB Bandit Algorithm?

Ans: Logarithmic. $\left(\delta_t := \frac{1}{t^3}\right)$

Sketch of proof:

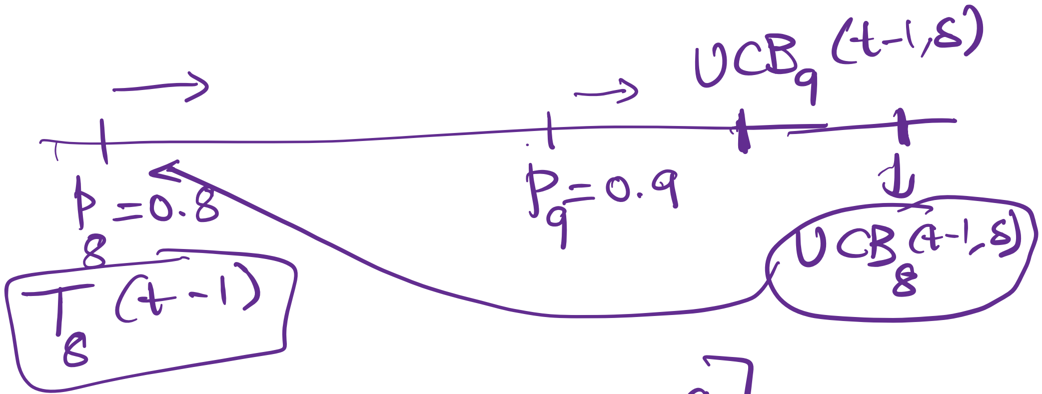
$p_1 = 0.1,\ p_2 = 0.2, \ldots, p_8 = 0.8,\ p_9 = 0.9$

In round t, what are the chances of picking coin 8 over coin 9.

Intuition: This should be small when t is large.

$$\mathbb{P}\left[UCB_8(t-1, \delta) > UCB_9(t-1, \delta)\right]$$

$$\mathbb{P}\left[\frac{X_8(t-1)}{T_8(t-1)} + \sqrt{\frac{\log 1/\delta}{2T_8(t-1)}} > \frac{X_9(t-1)}{T_9(t-1)} + \sqrt{\frac{\log 1/\delta}{2T_9(t-1)}}\right]$$



$$\leq \mathbb{P}\left[\underbrace{UCB_8(t-1,\delta)} \geq 0.9\right]$$

$$= \mathbb{P}\left[\frac{X_8(t-1)}{T_8(t-1)} + \underbrace{\sqrt{\frac{\log 1/\delta}{2T_8(t-1)}}} \geq 0.9\right]$$

$$= \mathbb{P}\left[\frac{X_8(t-1)}{T_8(t-1)} - 0.8 \geq 0.9 - 0.8 - \sqrt{\frac{\log 1/\delta}{2T_8(t-1)}}\right]$$

$$= \mathbb{P}\left[\frac{X_8(t-1)}{T_8(t-1)} - 0.8 \geq \sqrt{\frac{\log 1/\delta}{2T_8(t-1)}}\right]$$

$$0.1 \geq 2\sqrt{\frac{\log 1/\delta}{2T_8(t-1)}}$$

$$\iff T_8(t-1) \geq \frac{2\log 1/\delta}{(0.1)^2}$$

$$\leq \delta$$

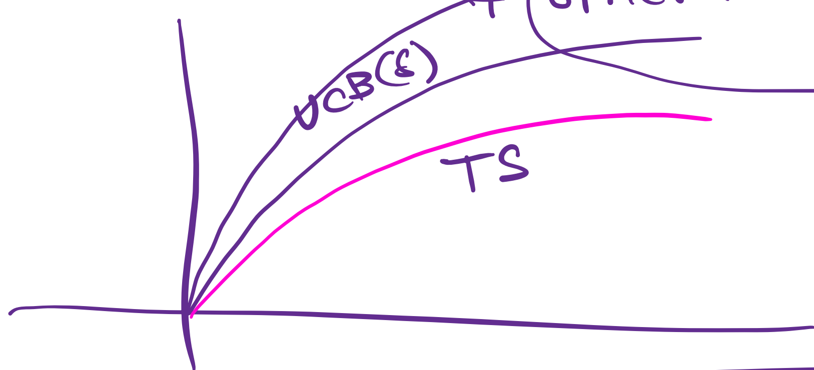
δ : confidence level :

$\mathbb{P}$(picking 0.8 coin instead of 0.9 when $T_8(t-1) \geq \frac{2 \log 1/\delta}{(0.1)^2}$] $\leq \delta$

Suppose $\delta = \frac{1}{t}$

Averaged.

$\Rightarrow$ Regret $\leq$ $(\log T) \times$ constants + other smaller order terms.



General Bandit Set up

K arms

arms = coins

pull = pick

Means : $\boxed{\mu_1}, \ldots, \mu_K$ (you want larger μ)

Reward : For arm a :

Reward distribution with mean μ_a

(e.g. coin case : Reward distribution $= Ber(p_a)$)

More generally : Reward distn : $N(\mu_a, 1)$

$$\text{Reward}(t) = \sum_{s=1}^{t} \text{Reward in round } s.$$

$$\text{Regret}(t) = t \cdot \max_a \mu_a - \text{Reward}(t)$$

$$\text{Averaged Regret}(t) = \mathbb{E}(\text{Regret}(t))$$

① <u>UCB Algorithm</u>:

Assume Reward Distribution has bounded support: $[\alpha, \beta]$

Hoeffding (General) to get UCB.

② <u>Thompson Sampling</u>

$\boxed{\mu}$ Might be difficult to choose prior. (Natural ~~to~~ try Uniform)

Does not use specific formula for the Reward Distribution.