Stochastic Bandits

Known: number of arms k, number of rounds $n > k$

Unknown: k probability distributions $\nu_1, \dots, \nu_k$ on $[0,1]$

for $t = 1, \dots, n$

1. Forcaster chooses $I_t \in \{1, \dots, k\}$
2. Given I_t, the environment draws reward $X_{I_t,t} \sim \nu_{I_t}$ independently from the past

Regret

$$\max_{i \in [k]} \sum_{t=1}^{n} X_{i,t} - \sum_{t=1}^{n} X_{I_t,t} = R_n$$

$$\mu^* = \max_{i \in [k]} \mu_i$$

$$i^* = \arg\max_{i \in [k]} \mu_i$$

Pseudo Regret

$$\max_{i \in [k]} \mathbb{E}\left[\sum_{t=1}^{n} X_{i,t} - \sum_{t=1}^{n} X_{I_t,t}\right] = \bar{R}_n \leq \{\}$$

$$\boxed{\Rightarrow \bar{R}_n = \mu^* n - \sum_{t=1}^{n} \mathbb{E}[\mu_{I_t}]}$$

UCB1 Strategy

Choose the arm w/ the largest

$$a_i = \bar{x}_i + \sqrt{\frac{2 \log(t)}{n_i}}$$

$\bar{x}_i \to$ sample mean of i^{th} arm

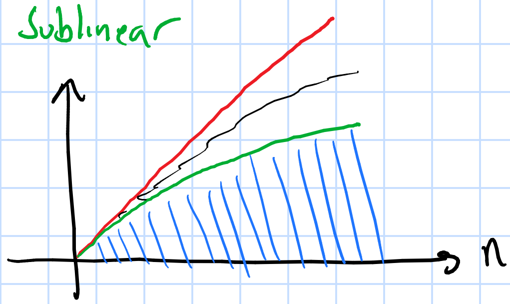
$n_i \to$ # of times i^{th} arm was sampled

Thm: Let $\Delta_i = \mu^* - \mu_i$. For all $k > 1$, if UCB1 is run on k machines having arbitrary distributions $V_1, \dots, V_k$ then its expected regret after n plays is at most

$$\left[8 \sum_{i : \mu_i \neq \mu^*} \frac{\log(n)}{\Delta_i} \right] + \left(1 + \frac{\pi^2}{3}\right) \sum_{j=1}^{k} \Delta_j$$

	Per/round	cumulative
t_1	r_1	r_1
t_2	r_2	$r_1 + r_2$
t_3	r_3	$r_1 + r_2 + r_3$
t_4	r_4	$r_1 + r_2 + r_3 + r_4$

✱ Sublinear



ε-Greedy $\quad \varepsilon \to (0,1)$

① $S_i \sim \mathrm{Uni}(0,1)$

② if $S_i \geq \varepsilon$

✱ Choose I_t to be the largest mean

else

✱ Choose I_t at random

The Adversarial Bandit

Known parameters: number of arms $k \geq 2$, # of rounds

for $t = 1, \dots, n$

① The forecaster chooses $I_t \in \{1, \dots, k\}$

② The adversary chooses a gain $g_t = (g_{1t}, g_{2t}, \dots g_{kt})$

③ The forecaster gets reward $g_{I_t, t} \in [0,1]^k$
(none of the other gains are revealed)

Exp3

Inputs: $\gamma \in [0,1]$, $W_i(1) = 1 \quad i \in [k]$, n

for $t = 1, \dots n$

① $$p_i(t) = (1-\gamma)\frac{W_i(t)}{\sum_j W_j(t)} + \frac{\gamma}{k}$$

② Draw I_t from $p_1(t), \dots, p_k(t)$

③ $X_{I_t, t} \in [0,1]$

④ for $j = 1, \dots, k$

⑤ $$\hat{X}_j(t) = \begin{cases} X_j(t)/p_j(t) & j = I_t \\ 0 & \text{else} \end{cases}$$

⑥ $$W_j(t+1) = W_j(t) \exp(\gamma \hat{X}_j(t)/k)$$