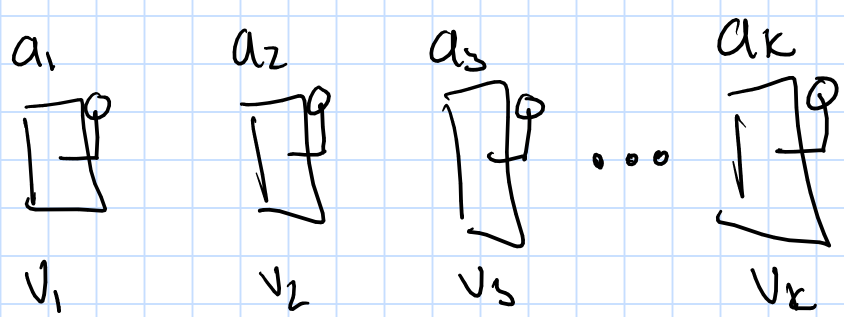
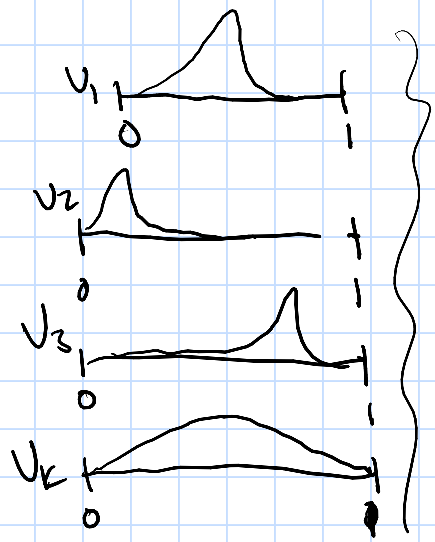
Admin

- Homework #5 due 04/30/2021 (Semi Supervised learning)
- Final Project offically due 05/05/2021
  - Both partners must submit the project
  - IEEE format
- Exam 04/09 – 04/12

Multi arm bandits

The multi arm bandit (MAB) is a sequential allocation problem defined by a set of actions. Our goal is to maximize our reward.





$\mu_1 = 0.9$, $\sigma^2 = 0.01$

$\mu_2 = 0.92$, $\sigma^2 = 0.1$

Approachs

- Randomized
- Round Robin

The Stochastic Bandit Problem

Known Parameters: number of arms K, number of rounds $n \geq K$

Unknown Parameters: K probability distributions $\nu_1, \dots, \nu_K$ on $[0,1]$

for $t = 1, \dots, n$

1. The forecaster 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 and reveals it to the forecaster

<u>Defs:</u>

I_t: arm sampled at time t from $\{1, \dots, K\}$

t: round of play

ν: distributions

$X_{I_t, t}$: the reward sampled from ν_{I_t} at time t

ν_i has a mean μ_i

$$\mu^* = \max_{i \in [K]} \mu_i, \quad i^* = \arg\max_{i \in [K]} \mu_i$$

Regret

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

Pseudo Regret

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

In a stochastic setting, it is easy to show that

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

UCB1

$$\bar{X}_i + \sqrt{\frac{2\log(t)}{n_i}}$$

$\bar{X}_i \rightarrow$ average samples from machine i

$n_i \rightarrow$ # of times we sampled machine i