Lecture 2

- Update on enrollment
- More relaxed environment
- First half in person, the second half remotely

1. Multi-armed bandits

1) - Indroduced by William R. Thompson in 1933.

- Name comes from 1950s by Mosteller & Bush studying animal learning

2) Become popular in different applications.

- adaptive experimental design (recognized by US Food & Drug Administration)
- new recommendation
- dynamic pricing.
- ad placement → e.g.

3). A simple example & a naive approach.

$r(a_1) \sim N(\mu_1, 1)$

$r(a_2) \sim N(\mu_2, 1)$

It costs me \$1 to pull the slot machine once

I have \$100, what is the best strategy

to maximize the total reward without knowing

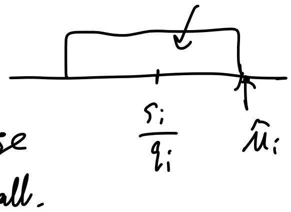
μ_1 & μ_2 ?

- Naive approach:
  - Try a_1 25 times
  - Try a_2 25 times.
  - Pull the one with higher empirical means for the
  - rest of 50 times.
- why it is not optimal?
  - when $\mu_1 = 10$, $\mu_2 = -10$.
  - when $\mu_1 = 1$, $\mu_2 = -0.9$

trade off btw exploration & exploitation.

- A better Algorithm:
  - UCB: $UCB_i = \frac{s_i}{q_i} + \sqrt{\frac{2\log(\frac{1}{\delta})}{q_i}}$

μ_i is in this interval with $1-\delta$ prob



  - why it is better? when $|\mu_1 - \mu_2|$ is large / when $|\mu_1 - \mu_2|$ is small.

- What theoretical guarantee do we have?

  - $\mathbb{A} = \{a_1, \cdots, a_k\}$.
  - reward of a_i : $X_{a_i} \sim p(x \mid \theta_i)$
  - Horizon : total # of pulls.
  - policy : a mapping from the history data → distribution is action space.

$$h_{i-1} \to A_i$$

  - Goal : maximize $\mathbb{E}\left[\sum_{i=1}^{n} X_{A_i}\right]$

e.g. UCB: history data At the beginning of each round i, is summarized by $(s_{t-1}^i, q_{t-1}^i)_{i=1}^k$

$$A_t = \max_i \left\{ \frac{s_{t-1}^i}{q_{t-1}^i} + \sqrt{\frac{2\log(1/\delta)}{q_{t-1}^i}} \right\}.$$

  - regret : $R_n = n \max_a \mu_a - \mathbb{E}\left[\sum_{i=1}^{n} X_{A_i}\right]$

$\Leftrightarrow$ min regret

→ Thm : k-armed 1-subgaussian bandit prob.

for $\forall$ horizon n, if $\delta = \frac{1}{n^2}$, then

$$R_n = n \max_{a \in A} \mu_a - \mathbb{E}\left[\sum_{t=1}^{n} X_t\right] \leq 3 \sum_{i=1}^{k} \Delta_i + \sum_{i: \Delta_i > 0} \frac{16 \log(n)}{\Delta_i}$$

(where $\Delta_i = \mu_i - \max_{a \in A} \mu_a$)

High-level idea: WLOG, μ_1 is the optimal arm

- When the regret > 0? when suboptimal arm $i > 1$ will be selected

selected at least one of them happens.

1) $UCB_i(t-1) \geq \mu_*$ → when this happens sufficiently large, $UCB_i \to \mu_i < \mu_*$ and then it won't happen

2) $UCB_{i^*}(t-1) < \mu_*$ → this will unlikely happen b/c UCB is the ucb of i^*-th arm.

It happens when $UCB_i > UCB_1$

① $UCB_1 > \mu_1$ ← this happens w.p. $1-\delta$

$UCB_i > UCB_1 > \mu_1$

↳ why the # of times this will happen has an upper bd?

↑ subgaussian could give it an explicit upper bd.

$UCB_i \to \mu_i < \mu_1 \to X$

② $UCB_1 < \mu_1$ ← this happens w.p. $\frac{\delta}{2}$

$\downarrow$ The UCB value of the optimal arm is always $> \mu$.

"good event"

$G_i = \{ \mu_1 < \min_{t\in[n]} UCB_1(t,\delta) \} \cap \{ \hat{\mu}_{i u_i} + \sqrt{\frac{2}{u_i}\log(\frac{1}{\delta})} < \mu_1 \}$

$\uparrow$ The empirical mean of i-th arm with u_i pulls.

$\uparrow$ After U_i pulls of i-th arm, its UCB value $< \mu_1$

u_i to be determined later.

(*) Key pt: exploring i-th arm u_i times, its UCB is smaller than the smallest UCB for optimal arm.

We will show two things:

1). If G_i occurs, then arm i will be played at most u_i times : $T_i(n) \le u_i$

2). ~~The~~ G_i^c occurs with low prob. $\mathbb{P}(G_i^c) \le n\delta + e^{-\frac{u_i c^2 \Delta_i^2}{2}}$

holds for $\forall u_i, c\in(0,1)$

$\Rightarrow \mathbb{E}[T_i(n)] = \mathbb{E}[T_i(n) \mid G_i(n)] + \mathbb{E}[T_i(n) \mid G_i^c(n)]$

$\le u_i + \mathbb{P}(G_i^c(n))\, n$

$\le u_i + (n\delta + e^{-\frac{u_i c^2 \Delta_i^2}{2}})$

plug in $u_i = \left\lceil \frac{2\log(1/\delta)}{(1-c)^2\Delta_i^2} \right\rceil$ (if $u_i \ge n$, (*) always holds). & $\delta = \frac{1}{n^2}$

$\Rightarrow \mathbb{E}[T_i(n)] = \left\lceil \frac{2\log(n^2)}{(1-c)^2\Delta_i^2} \right\rceil + 1 + n^{1-2c^2/(1-c)^2}$

plug in $c = \frac{1}{2}$

$\Rightarrow \mathbb{E}[T_i(n)] \le 3 + \frac{16\log(n)}{\Delta_i^2}$

$R_n = \sum_i \Delta_i \mathbb{E}[T_i(n)] = \sum_i 3\Delta_i + \frac{16\log(n)}{\Delta_i}$

- proof of 1)
- proof of 2).
- Intro of σ- subguassian
- generalization

pf of 1) : Contradiction: If i-th arm is pulled u_i times, then $UCB_i < \mu_1$, it won't be pulled any more

If $T_i(n) > u_i$ $\Rightarrow \exists t$ $T_i(t-1) = u_i$, $A_t = i$

$$\Rightarrow UCB_i(t-1, \delta) = \hat{u}_i(t-1) + \sqrt{\frac{2\log(\frac{1}{\delta})}{T_i(t-1)}}$$

$$= \hat{\mu}_{i u_i} + \sqrt{\frac{2\log(1/\delta)}{u_i}}$$

$$< \mu_1 < UCB_1(t-1, \delta)$$

$\Rightarrow A_t = \arg\max_j UCB_j(t-1, \delta) \neq i$ ⊗

pf of 2)

$$G_i^c = \underbrace{\{\mu_1 \geq \min_{t \in [n]} UCB_1(t, \delta)\}}_{A} \cup \underbrace{\{\hat{\mu}_{1 u_i} + \sqrt{\quad} \geq \mu_1\}}_{B}.$$

at least one of UCB_1 value is less than μ_1

$$A \leq \mathbb{P}\left[\bigcup_{s=[n]} \left\{\mu_1 \geq \hat{\mu}_{1s} + \sqrt{\frac{2\log(1/\delta)}{s}}\right\}\right]$$

$$\leq \sum_{s=1}^{n} \mathbb{P}\left(\mu_1 \geq \hat{\mu}_{1s} + \sqrt{\frac{2\log(1/\delta)}{s}}\right)$$

$$\leq n\delta$$

when first arm follows 1-sub guassian, This prob $\leq \delta$

$$B = \mathbb{P}\left(\hat{\mu}_{i,u_i} + \sqrt{} \geq \mu_1 \right)$$

want the empirical mean close to true mean w.h.p

$$= \mathbb{P}\left(\underline{\hat{\mu}_{i,u_i} - \mu_i} \geq \mu_1 - \mu_i - \underbrace{\sqrt{\frac{2\log(1/\delta)}{u_i}}}_{\substack{u_i \text{ large enough s.t. } < (1-c)(\mu_1 - \mu_i) \\ \text{for } c \text{ determined later}}} \right)$$

$$\leq \mathbb{P}\left(\hat{\mu}_{i,u_i} - \mu_i \geq c\Delta_i \right)$$

$$\leq e^{-\frac{u_i c^2 \Delta_i^2}{2}}$$

when the i-th arm follows 1-subgaussian. The prob has this upper bd.

$\left(\hat{\mu}_{i,u_i} - \mu_i \text{ is } \frac{1}{\sqrt{u_i}}\text{-subgaussian} \right)$

$$\mathbb{P}(G_i^c) \leq n\delta + e^{-\frac{u_i c^2 \Delta_i^2}{2}}$$

Formal Def

σ-subgaussian $:\Longleftrightarrow \mathbb{E}[e^{\lambda X}] \leq e^{\frac{\lambda^2 \sigma^2}{2}}$ for $\forall \lambda \in \mathbb{R}$

$\mathbb{P}(X \geq \varepsilon) \leq e^{-\frac{\varepsilon^2}{2\sigma^2}}$, $\mathbb{P}(X \leq -\varepsilon) \leq e^{-\frac{\varepsilon^2}{2\sigma^2}}$

useful ineq.

Property of σ-subgaussian :

1. $\mathbb{V}[X] \leq \sigma^2$,
2. cX is $|c|\sigma$-subgaussian
3. If X_1, X_2 are σ_1, σ_2-subgaussian

   then $X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$-subgaussian

Lemma: Define $\{X_i\}_{i=1}^n$ are i.i.d with $\mathbb{E}[X_i] = \mu$,

If $X_i - \mu$ is σ-subgaussian, then

$$\hat{\mu} - \mu = \frac{1}{n}\sum_{i=1}^n X_i - \mu \text{ is } \hat{\sigma} = \frac{\sigma}{\sqrt{n}}\text{-subgaussian}$$

$$\& \ \mathbb{P}(X \geq \varepsilon) \leq e^{-\frac{\varepsilon^2 n}{2\sigma^2}} \quad \& \ \mathbb{P}(X \leq -\varepsilon) \leq e^{-\frac{\varepsilon^2 n}{2\sigma^2}}$$

Thm 2 K-armed 1-subgaussian bandit prob.

for $\forall$ horizon n, if $\delta = \frac{1}{n^2}$, then

$$R_n = 8\sqrt{kn\log(n)} + 3\sum_i \Delta_i$$

pf: $R_n = \sum_i \Delta_i \mathbb{E}[T_i(n)]$

$$= \sum_{i:\Delta_i<\Delta} \Delta_i \mathbb{E}[T_i(n)] + \sum_{i:\Delta_i\geq\Delta} 3\Delta_i + \frac{16\log(n)}{\Delta_i}$$

$$= n\Delta + \frac{k\,16\log(n)}{\Delta} + \sum_{\Delta_i>\Delta} 3\Delta_i$$

$$\leq 8\sqrt{kn\log(n)} + 3\sum_i \Delta_i$$

How to obtain the UCB for σ-subguassian.

$$\mathbb{P}(\mu > \hat{\mu}_n + c) \geq \delta$$

$$\mathbb{P}(\hat{\mu}_n - \mu < -c) \leq e^{-\frac{nc^2}{2\sigma^2}} = \delta \Rightarrow \frac{nc^2}{2\sigma^2} = \log\left(\frac{1}{\delta}\right)$$

$$\Rightarrow c = \sqrt{\frac{2\sigma^2\log\left(\frac{1}{\delta}\right)}{n}}$$

δ smaller $\rightarrow$ more exploration

δ larger $\rightarrow$ more exploration

UCB Algo

$$\text{UCB}_i(t-1,\delta) = \begin{cases} \infty \\ \underbrace{\hat{\mu}_i(t-1)}_{\text{Empirical mean}} + \underbrace{\sqrt{\frac{2\log\left(\frac{1}{\delta}\right)}{T_i(t-1)}}}_{1-\delta \text{ upper bound for 1-subguassian}} \end{cases}$$

HW: Try UCB on $N(\mu_1, \sigma)$, $N(\mu_2, \sigma)$ for different $\Delta = \mu_1 - \mu_2$

$\text{Ber}(p_1)$, $\text{Ber}(p_2)$. for different $\Delta = p_1 - p_2$.