Recap Stochastic Bandits

Finite arms
- Non-adaptive - uniform exploration, eps-greedy
- Adaptive - successive elimination, UCB (upper confidence bnd) sampling

Infinite arms/actions

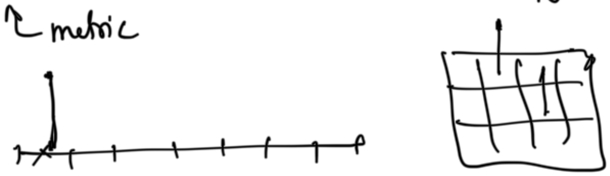
Structured bandits - Lipschitz, Linear, GP, ...

Lipschitz bandits

$$\to |\mu(x) - \mu(x')| \le L\,\underbrace{|x - x'|}_{D(x,x')} \quad \forall x, x' \in \mathcal{X}$$

↑ metric

① Fixed discretization - N bins



2-d

$$E[R(T)] = \underbrace{T\mu^*(x) - T\mu^*_N}_{\text{discretization error}} + \underbrace{T\mu^*_N - \sum_{t=1}^{T} \mu^*(x_t)}_{\text{cumulative regret from N-armed bandit}}$$

a_t

1-dim
$$\le T\frac{L}{N} + \sqrt{NT\log T}$$
$$= O((L\log T)^{1/3}\, T^{2/3})$$

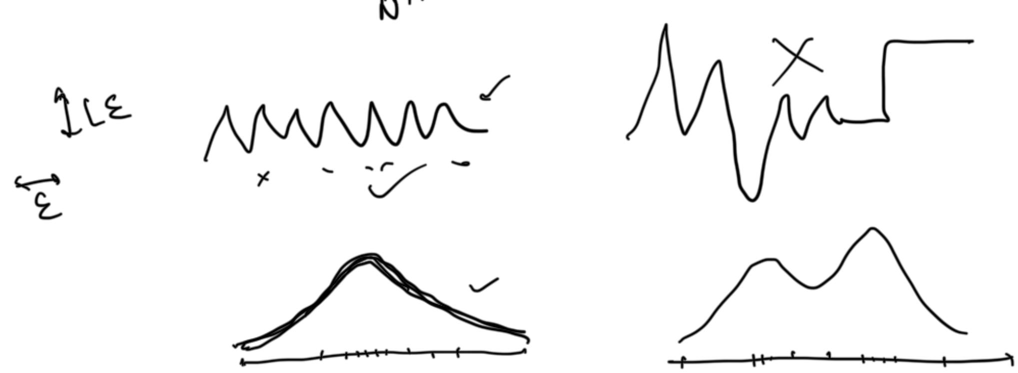
Optimal in worst case for L-Lipschitz rewards

Lower bound $\Omega(L^{1/3} T^{2/3})$

d-dim
$$\le T\frac{L}{N^{1/d}} + \sqrt{NT\log T} \simeq (L\log T)^{\frac{1}{d+2}}\, T^{\frac{d+1}{d+2}}$$

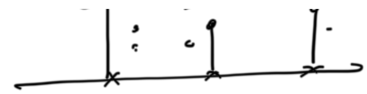
UB
LB. (upto log)

$\exists\, \varepsilon$



② Adaptive Discretization - Zooming Algorithm

more points/arms in promising regions



active arms $S \leftarrow \phi$

For $t = 1, 2, \ldots$

if some arm is not covered by confidence ball of active arms $\quad x : |x - a| > \varepsilon_t(a)$, $a \in S$

then pick any such arm & add to S

Play active arm with largest $\hat{\mu}_t(x) + 2\varepsilon_t(x)$

For any x $\quad |\hat{\mu}_t(x) - \mu(x)| \le \underbrace{\sqrt{\frac{2\log 1/\delta}{n_t(x)}}}_{\varepsilon_t(x)}$ $\quad$ wp $\ge 1 - \delta$ ☆ $\qquad \underline{\underline{L=1}}$

Need to hold $\forall x$ active & all $t = 1, \ldots T$ ← easy

hard ∵ infinitely many arms

Let a_t be arm activated at time t.

events are independent.

$$\Pr(\text{☆ holds for } a_t) = \sum_x \Pr(a_t = x) \cdot \Pr(\text{☆ holds for } x)$$

$$\ge 1 - \delta$$

Apply union bond $\forall a_t$ & all t $\quad \ge 1 - \delta T^2 \cong 1 - \frac{1}{T^2} \qquad \delta \simeq \frac{1}{T^4}$

Assume this high prob event from now on.

$$R(T) = T\mu^* - \sum_{t=1}^{T} \mu(x_t)$$

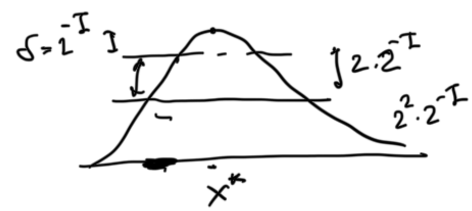
$$= T\delta + \sum_{i=1}^{I = \log_2 1/\delta} R_i(T)$$

Consider active arms $\Delta(x) \le \delta$ ✓ $\quad = \mu^* - \mu(x)$

& active arms gap $\Delta(x) > \delta = 2^{-I}$

$2^{-i} \le \Delta(x) \le 2 \cdot 2^{-i}$

$$R_i(T) = \sum_{x : 2^{-i} \le \Delta(x) \le 2 \cdot 2^{-i}} n(x) \cdot \Delta(x)$$



We first prove the following lemma:

Lemma: $\underline{\underline{\Delta(x)}} \le \underline{\underline{3\varepsilon_t(x)}} = O\left(\sqrt{\frac{\log T}{n_x}}\right)$ for each x, t. $\quad$ whp

Let x^* be covered by active arm y

Let x be chosen at time t. (x - active arm)

$$\mu(x)+3\varepsilon_t(x) \geq \hat{\mu}_t(x)+2\varepsilon_t(x) \geq \hat{\mu}_t(y)+2\varepsilon_t(y) \geq \mu(y)+\varepsilon_t(y) \geq \mu(x^*)$$

since y covers x^*, $|\mu(y)-\mu(x^*)| \leq |y-x| \leq \varepsilon_t(y)$

$$\Delta(x) = \mu(x^*)-\mu(x) \leq 3\varepsilon_t(x)$$

∎

$$\Rightarrow n(x) = O\left(\frac{\log T}{\Delta^2(x)}\right)$$

$$R_i(T) = \sum_{x:\, 2^{-i} \leq \Delta(x) \leq 2 \cdot 2^{-i}} n(x) \cdot \Delta(x)$$

$$= O\Big(\sum_{x:\, 2^{-i} \leq \Delta(x) \leq 2\cdot 2^{-i}} \frac{\log T}{\Delta(x)}\Big) = O\Big(\sum_{x:\, 2^{-i} \leq \Delta(x) \leq 2\cdot 2^{-i}} \frac{\log T}{2^{-i}}\Big)$$

How many arms are activated in $\{x:\ 2^{-i} \leq \Delta(x) \leq 2 \cdot 2^{-i}\} = \Delta_i$?

To bound how many arms activated in Δ_i, we will use the lemma above to argue that two active arms can't be too close.

Let x, y are activated in Δ_i

x activated before y.

When y is activated, it is not covered by x.

$$\Rightarrow D(x,y) > \varepsilon_t(x) \geq \frac{\Delta(x)}{3} \quad \text{from lemma.}$$

$$D(x,y) \geq \frac{1}{3} \min \Delta(x), \Delta(y) \geq \frac{2^{-i}}{3}$$

$$\Rightarrow R_i(T) = O\Big(\frac{\log T}{2^{-i}} \underbrace{N_{2^{-i}/3}(x:\ 2^{-i} \leq \Delta(x) \leq 2\cdot 2^{-i})}_{\text{Zooming dimension.}}\Big)$$

$$\inf_{d \geq 0} \{ N_{r/3}(\Delta_r) \leq c \cdot r^{-d} \} \; \forall r > 0$$

$$R(T) = \delta T + O\left(\frac{\log T}{\delta} \delta^{-d} \right)$$

$$\simeq O\left(T^{\frac{d+1}{d+2}} (\log T)^{\frac{1}{d+2}} \right)$$

whp

d - zooming dim
NOT necessarily ambient dim.

Linear Bandits, Gaussian bandits. NN bandits.

Concentration Bounds Dependent data.

Independent data

Hoeffding's inequality $X_1 \ldots X_n$ iid mean μ, $a_i \leq X_i \leq b_i$ a.s.

then $P\left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| \geq \varepsilon \right) \leq e^{-2n^2\varepsilon^2 / \sum_{i=1}^n (b_i - a_i)^2}$

$e^{-2n\varepsilon^2/\ldots}$

Union bound

$$P(A \cup B) \leq P(A) + P(B).$$

Bernstein inequality $X_1 \ldots X_n$ iid mean μ with $E[e^{t(X_i-\mu)}] \leq e^{\frac{\mathrm{var}(X_i) t^2/2}{1 - b|t|}}$ for any $t \in \left(-\frac{1}{b}, \frac{1}{b}\right)$, $b > 0$

then $P\left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| \geq \varepsilon \right) \leq 2 e^{\frac{-n\varepsilon^2/2}{\mathrm{var}(X) + b\varepsilon}}$

eg. if $|X_i| \leq c$ $b = c/3$

if var(x) is small.

$$\approx e^{-n\varepsilon}$$

Martingale - seq^n of random variables s.t. $Z_1 Z_2 \ldots Z_n \ldots$

$$\forall n \quad E[Z_{n+1} | Z_1 \ldots Z_n] = Z_n$$

For our purposes, martingales behave like ind. r.v..

→ Azuma-Hoeffding inequality $\{Z_i\}_{i=1}^n$, $Z_1 = 0$ be a martingale with a.s. bounded increments $\underbrace{|Z_i - Z_{i-1}|}$ $\leq b_i$ then

$$P(Z_n \geq \varepsilon) \leq \exp\left\{-\frac{\varepsilon^2}{2\sum_{i=1}^{n} b_i^2}\right\} \qquad \forall \varepsilon > 0, n$$

Eq. $\{X_i\}_{i=1}^{n}$ be a martingale difference seq^n with $|X_i| \leq b_i$ a.s.

$$P\left(\sum_{i=1}^{n} X_i \geq \varepsilon\right) \leq \exp\left\{-\frac{\varepsilon^2}{2\sum_{i=1}^{n} b_i^2}\right\}$$

$$\underbrace{Z_n = \sum_{i=1}^{n} X_i} \qquad E[X_{n+1}] = 0. \qquad E\left[\sum_{i=1}^{n+1} X_i \mid X_1 \ldots X_n\right] = \sum_{i=1}^{n} X_i$$

$$\overset{\shortparallel}{} \underbrace{E[X_{n+1}]}_{=0} + \sum_{i=1}^{n} X_i$$

$\Rightarrow$ Bernstein inequality for Martingales (Freedman's inequality)

$\{X_i\}_{i=1}^{n}$ be a martingale difference seq^n with $|X_i| \leq b_i$ a.s.

$$P\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp\left\{\frac{-\varepsilon^2}{\underbrace{\sum_{i=1}^{n} E[X_i^2 \mid X_{1:i-1}]}_{\text{var}} + \frac{b\varepsilon}{3}}\right\} \simeq e^{-\varepsilon}$$

if var is small

$\underline{\mu^*(x) = \theta^{*T} x}$ Linear reward. $\qquad \mu(x) = \theta^T x$

whp $\rightarrow \hat{\mu}_t(x) - \mu(x) \approx \theta^* - \hat{\theta}_t$ $\qquad \theta, \theta^* -$ d-dim

$$\Rightarrow \|\theta^* - \hat{\theta}_t\|^2_{V_t} \leq \beta_t \simeq \sigma^2 d$$

$$\rightarrow \hat{\theta}_t = \Big(\underbrace{\sum_{s=1}^{t} x_s x_s^T}\Big)^{-1} x_s^T y_s + \lambda I \qquad V_t = \sum_{s=1}^{t} x_s x_s^T$$