Bandit algorithms in revenue optimization

('The value of knowing a demand curve' - Kleinberg & Leighton)

- We now see how bandit algorithms (and in particular, UCB) can be used to perform revenue optimization without knowing prices
- Model -
  - n buyers arrive sequentially
  - Seller makes a 'posted price' take-it-or-leave-it offer to each buyer.
  - Each buyer has i.i.d value $V_t \sim F$, $\in [0,1]$ a.s.
    If posted price $= P_t$, then buyer purchases iff $P_t \leq V_t$
  - F is unknown
- If F is known, can use 'monopolist price'

$$p^* = \arg\max_p \Big[\underbrace{p\,[1-F(p)]}_{R(p)} \Big]$$

- Assume F is regular $\Rightarrow$ unique p^*, $R(p)$ quasi concave
- If instead we know $V_1, V_2, \ldots, V_n$, can choose P_{OPT} to maximize revenue

- Thm - Assuming $R(p) = p\bar{F}(p)$ has unique global maximum p^*, and $R''(p^*)$ exists and is strictly negative, then there is an online pricing strategy which achieves an expected regret of $O(\sqrt{n \log n})$

Notes

- Let $\bar{R}^*_n = \mathbb{E}\left[\sum_{t=1}^{n} p^* \mathbb{1}\{V_t \geq p^*\}\right] = nR(p^*)$

$$\bar{R}^{OPT}_n = \max_p \left[\sum_{t=1}^{n} p\, \mathbb{1}\{V_t \geq p\}\right] \quad \left(\text{oracle bound!}\right)$$

$$\bar{R}^{\pi}_n = \mathbb{E}\left[\sum_{t=1}^{n} P_t\, \mathbb{1}\{V_t \geq P_t\}\right], \text{ where } P_t \equiv \text{Policy } \pi$$

We will actually get that $\bar{R}^*_n - \bar{R}^{\pi}_n = O(\sqrt{n \log n})$

<u>and</u> $\bar{R}^{OPT}_n - \bar{R}^*_n = O(\sqrt{n \log n})$

Thus we have small regret w.r.t an <u>oracle bound</u> - this is stronger than competing with p^*

- Why regret? It was known that there are randomized pricing algos s.t $\dfrac{\bar{R}^{\pi}_n}{\bar{R}^*_n} \geq \dfrac{1}{1+\epsilon}$ for any $\epsilon > 0$. Regret captures the lower order dependence on n
- This was the first regret bound for an 'infinite' arm setting.

Pf - Main Idea - Choose appropriate discretization of $[0,1]$

- In particular, given K, we consider the prices $\{1/K, 2/K, \ldots, K/K\}$ as 'arms' (we later choose $K = (n/\log n)^{1/4}$)
  - Now we can use UCB.
- For $p_i = i/K$, the payoff for t^{th} person is $X_i = \begin{cases} i/K & \text{if } V_t > i/K \\ 0 & \text{ow} \end{cases}$

$$\Rightarrow \mu_i = \mathbb{E}[X_i] = \frac{i}{K} \bar{F}\left(\frac{i}{K}\right)$$

- ~~Let~~ Given K, let $(i^*, \mu^*) \equiv$ best arm i^*/K

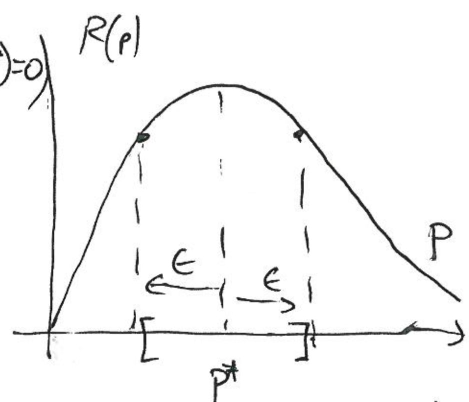
$$\Delta_i \equiv (\mu^* - \mu_i)$$

  - We want to show we get close to $n\mu^*$, and also that $n\mu^*$ is close to $np^* \bar{F}(p^*)$.

Lemma: $\exists$ constants C_1, C_2 s.t. $C_1(p^*-p)^2 \le R(p^*) - R(p) \le C_2(p^*-p)^2$ for all $p \in [0,1]$

Pf - For $p \in [p^*-\epsilon, p^*+\epsilon]$, use Taylor ($\because R'(p^*) = 0$)

$$R(p) \approx R(p^*) + C(p-p^*)^2, \quad C = \frac{R''(p^*)}{2}$$



For $p \in [0, p^*-\epsilon] \cup [p^*+\epsilon, 1]$, since the set is compact and $R(p) < R(p^*)$, we can find B_1, B_2 s.t. $B_1(p^*-p)^2 \le R(p^*) - R(p) \le B_2(p^*-p)^2$. Now take $C_1 = \min(C, B_1)$

Lemma - $\mu^* > p^* \bar{F}(p^*) - c_2/K^2$

Pf - $\exists\ i$ s.t. $|p^* - i/K| \leq 1/K$

$$\Rightarrow \left| R(p^*) - R(i^*/K) \right| \leq c_2/K^2$$

Lemma - Suppose we sort Δ_i as $\tilde{\Delta}_0 \leq \tilde{\Delta}_1 \leq \dots \leq \tilde{\Delta}_K$.

Then $\tilde{\Delta}_j \geq c_1 (j/2K)^2$

Pf - At most j elements of $\{1/K, 2/K, \dots, K/K\}$ lie within distance $j/2K$ of i^*/K.

- Now consider $\bar{R}_n^\pi - n\mu^*$

From UCB - $-(\bar{R}_n^\pi - n\mu^*) \leq \sum_{i : \mu_i < \mu^*} \left(\frac{8 \log n}{\Delta_i} + 2 \right)$

$$\leq \frac{32K^2}{c_1^2} \cdot \frac{\pi^2}{6} \log n + 2K$$

$$= O(K^2 \log n)$$

- On the other hand, $nR(p^*) - n\mu^* \leq \frac{n c_2}{K^2} = O\left(\frac{n}{K^2}\right)$

- Choosing $K = \left(\frac{n}{\log n}\right)^{1/4} \Rightarrow nR(p^*) - \bar{R}_n^\pi \leq O(\sqrt{n \log n})$

- However we do not know n

  - Doubling trick

- Use $n_0 = 1,\ n_1 = 2,\ n_2 = 4 \ldots,\ n_\ell = 2^\ell$

- this continues till $2^{\ell} \leq n \Rightarrow \ell^* = O(\log_2 n)$

- However regret $= \sum_{\ell=0}^{\ell^*} (n_\ell \log n_\ell)^{1/2} = \sum_{\ell=0}^{\log n} (\ell\, 2^\ell)^{1/2}$

$$= O\left(\sqrt{n \log n}\right)$$

- Finally, we can also show $\overline{R}_n^{OPT} - \overline{R}_n^* = O(\sqrt{n \log n})$

  - this follows from Chernoff bounds - See KL'03

- This result is also near-optimal

Thm (KL'03) - No policy π can achieve $\overline{R}_n^{*} - \overline{R}_n^{*\pi} = o(\sqrt{n})$

- Intuition - Consider 2 coins of prob $1/2$, $1/2 + \epsilon$

  We need $\Omega(1/\epsilon^2)$ trials to accurately identify the better coin

  Now we can use this to construct a worst-case F s.t regret $= \Omega(\sqrt{n})$