Dynamic Pricing (part I)

Guest Lecturer: Alex Slivkins (MSR-NYC)

Columbia University, IEOR, Spring 2016

"Learning and Optimization for Sequential Decision Making"
Feb 29, 2016

Paradigmatic problem

- Seller with limited supply: k identical items to sell
- In each round $t = 1 \,..\, T$, a new customer arrives
  - seller offers 1 item @ price $p_t \in [0,1]$
  - customer accepts or rejects
- Until no more items or no more customers

Goal: adjust price over time, to maximize reward.
No bonus for leftover items!

- $S(p) = \Pr[\text{sale @ price } p]$ *demand curve*
  - fixed but unknown to seller

no parametric assumptions

- Compete with ***best fixed price***

What's going on: Economics

- interpretation: sale $\Leftrightarrow p_t \leq v_t$, where v_t is ***customer's value***.
- Where do values come from?
  - worst-case (CS) view: values chosen adversarially

    … often leads to weak positive results.
  - Bayesian (Econ) view: from a known distribution

    … strong assumption, sometimes unrealistic.
- Prior-independent mechanisms are a compromise
  - Private values are sampled IID from *unknown distribution*
  - Goal: be competitive against the optimal mechanism that knows the distribution (perhaps restricting it to a large, natural class).

What's going on: Economics

Why Posted Price Mechanisms? (PPM)

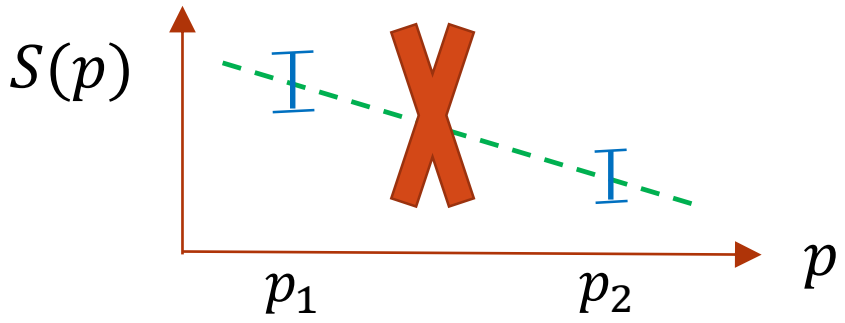
- PPMs are widely used in practice.
- Customers do not need to know their exact value, only need to evaluate a single price offer.
- Each customer reveals very little information to the seller (revealed info may hurt her in the future)

Moreover:

- PPMs are truthful and group-strategy proof.
- The optimal online mechanism for known distribution is a PPM

What is going on: Bandits

- First intuition: we want to sell at (unknown) "best price"
  - offered price too low ⇒ likely sale, wasted item
    offered price too high ⇒ likely no sale, wasted customer
  - … but we learn something about the demand distribution
  - *"explore-exploit tradeoff", "learn-and-earn"*
- With limited supply, our learning ability is handicapped:
  - can't afford to sell too many items while trying "low" prices
- Without parametric assumptions, no long-range inference



Outline

- ✓ Intro
- ❑ Unlimited supply

Limited supply:

- ❑ some observations
- ❑ explore-then-exploit
- ❑ lower bound
- ❑ a better algorithm (overview)

Unlimited supply ($k = T$)

- Multi-armed bandit problem
  - *arms* (prices) with fixed but unknown expected rewards
  - *bandit feedback*: only for chosen price
- Special feature: sale @price p => sale @any lower price
- *Reward function* $Rew(p) = p \cdot S(p)$ *(expected per-round reward)*
- Best fixed price p^*: maximizes $Rew(p)$
- algorithm's performance measured by
  $\text{Regret}(T) = T\ Rew(p^*) - E[\text{algorithm's total reward}]$

Reduction to bandits

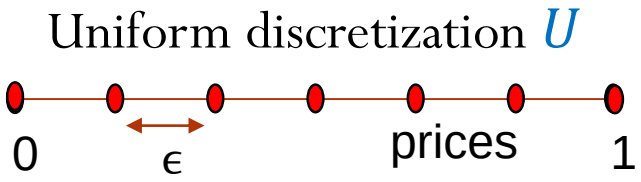
- Uniform discretization U, then run a bandit algorithm on U

- $\text{Regret} = \text{Regret}_U + OPT - OPT_U$

  bandit discretization error

- Round down p^* to the nearest price U, call it q^*
  Selling $@q^*$ loses at most ϵ per each sale.
  So, discretization error $\leq \epsilon n$
- Pick ϵ in advance to optimize regret

Reduction to bandits



- simple algorithm:
  explore-then-exploit
  - pick arm u.a.r. from U for T_0 rounds, then pick est. "best arm" and stick with it
  - pick ϵ, T_0 in advance to optimize regret

Regret $\tilde{O}(T^{3/4})$

- better approach: *adaptive exploration*
  - ***adapt*** to observations to zoom in on better arms
    e.g., UCB1 or Thompson Sampling
  - pick ϵ in advance to optimize regret

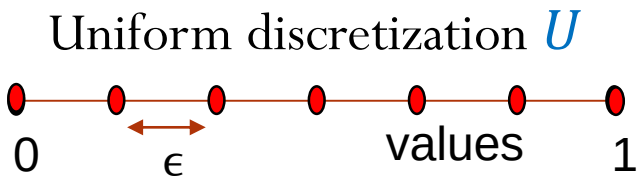
Optimal regret $\tilde{O}(T^{2/3})$

Lower bound on regret

- Theorem: Regret $\geq \Omega(T^{2/3})$
- Family of problem instances:
  - values v_t are only multiples of ϵ

    (note: suffices to consider prices p_t of the same form)
  - *needle-in-a-haystack:* choose sale probabilities $S(p)$ so that

$$Rew(p) \equiv p \cdot S(p) = \begin{cases} 1/4 + \epsilon/2, & p = p_0 \\ 1/4, & p \geq 1/4 \\ 0 & p < 1/4 \end{cases}$$



- Bandit problem with IID rewards, $n = \frac{3}{4\epsilon}$ arms, one "needle"

  $\Longrightarrow$ Any algorithm has regret $\Omega(n/\epsilon)$ for some p_0.
- Plug in $\epsilon = T^{-1/3}$

Outline

- ✓ Intro
- ✓ Unlimited supply

Limited supply:

- ❑ some observations
- ❑ explore-then-exploit
- ❑ lower bound
- ❑ a better algorithm (overview)

Limited supply ($k < T$)

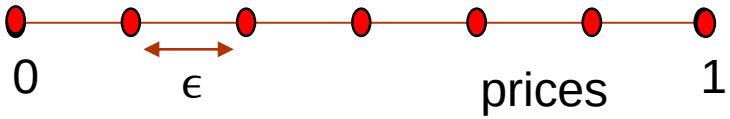
- [even with explore-then-exploit]
  exploitation is limited by #items left after exploration
- maximizing expected *per-round* reward is not the right goal.
  need to think about expected ***total*** reward
- Best fixed price $p^* = \operatorname{argmax}_p \text{Rew}(p)$

  $\text{Rew}(p)$ expected total reward for fixed price p
- $\text{Regret}(k, T) = \text{Rew}(p^*) - \text{Rew}(\text{algorithm})$

  now want regret sublinear in k = #items

Explore-then-exploit fails for $k < \sqrt{T}$

- Uniform discretization,
  T_0 rounds of exploration (u.a.r.)
  then pick "est. best arm" & stick with it.



- If $T_0 \leq \frac{T}{k}$. Two problem instances

$$v_t = \begin{cases} 1 & \text{w. prob } \frac{k}{T} \\ 0 & \text{othw} \end{cases} \qquad \text{Best exp. reward} = k$$

$$v_t = \begin{cases} \frac{1}{2} & \text{w. prob } \frac{k}{T} \\ 0 & \text{othw} \end{cases} \qquad \text{Best exp. reward} = k/2$$

with const prob., $\{v_t = 0 \text{ for all } t \leq T_0\}$

$\Longrightarrow$ with const prob., cannot tell apart the two instances $\Longrightarrow$ regret $\frac{k}{2}$

- If $T_0 \geq \frac{T}{k}$ then for problem instance with value $v_t \equiv 1$,
  all items are sold in exploration, at average price $\frac{1}{2}$ $\Longrightarrow$ regret $\frac{k}{2}$

Tool: fractional reward

- Total expected reward $Rew(p) = \mathrm{p} \cdot \mathrm{E}[\#\text{sales } @p]$

  difficult to work with directly, use approximation
- Easy upper bound: $\mathrm{E}[\#\text{sales } @p] \leq \min(k, T \cdot S(p))$
- Use $f(p) = p \cdot \min(k, T \cdot S(p))$ "fractional reward"
- Claim: $f(p) - O(p\sqrt{k \log k}) \leq Rew(p) \leq f(p)$
- Proof ($\leq$): sell at price p, let $X_t = \mathbf{1}(\text{sale @round t}), X = \sum_t X_t$.
  - $\mu := E[X] = T \cdot S(p)$
  - by Chernoff Bounds, $|X - \mu| \leq \mathrm{O}(\sqrt{\mu \log k})$ WHP
  - $\#\text{sales} = p \cdot \min(k, X) \geq \min\left(k, \mu - \mathrm{O}(\sqrt{\mu \log k})\right)$
    $\geq \min(k, \mu - \mathrm{O}(\sqrt{k \log k}))$
  - $Rew(p) = p \cdot \#sales \geq p \cdot \min(k, \mu) - \mathrm{O}(p)\sqrt{\mu \log k}$

Outline

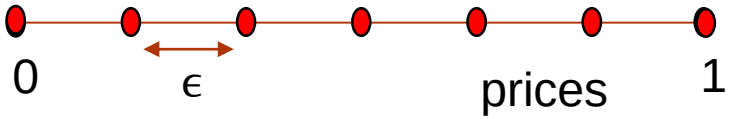
- ✓ Intro
- ✓ Unlimited supply

Limited supply:

- ✓ some observations
- ❑ explore-then-exploit
- ❑ lower bound
- ❑ a better algorithm (overview)

Explore-then-exploit for $k \sim T$

- Uniform discretization U,
  $T_0 \leq k$ rounds of exploration (u.a.r.)
  then pick the "exploit arm" p_0 & stick with it.



- Here's how we pick p_0 after exploration:
  - Est. sales prob $\hat{S}(p) = \frac{\#\text{sales } @p}{\#\text{rounds } @p}$
  - Est. fractional reward $\hat{f}(p) = p \cdot \min\left(k, T \cdot \hat{S}(p)\right)$.
  - Pick exploit arm: $p_0 = argmax_{p \in U}\, \hat{f}(p)$
- Theorem: Regret $\leq O(T^{3/4} \log T)$

Theorem: Regret $\leq O(T^{3/4} \log T)$

Assume: $\left|\hat{S}(p) - S(p)\right| \leq \delta \quad \forall p \in U$, where $\delta = O(\log T)/\sqrt{\epsilon T_0}$

- $\left|\hat{f}(p) - f(p)\right| \leq 2pT\delta$

$\hat{f}(p) = p \cdot \min\left(k, T \cdot \hat{S}(p)\right)$

- p^*: best arm; pick closest arm $q^* \leq p^*$ in discretization
- By defn of exploit arm: $\hat{f}(p_0) \geq \hat{f}(q^*)$
- $f(p_0) \geq f(q^*) - 4T\delta$ ($f(\cdot)$ vs. $\hat{f}(\cdot)$)

  $\geq Rew(q^*) - 4T\delta$ ($f(\cdot)$ vs. $Rew(\cdot)$)

  $\geq Rew(p^*) - 4T\delta - \epsilon T$ (discretization)
- $\{\, Rew(p_0)$ in exploitation $\} \geq Rew(p_0) - T_0$

  $\geq f(p_0) - T_0 - \gamma\,, \quad \gamma = O(\sqrt{k \log k})$ ($f(\cdot)$ vs. $Rew(\cdot)$)

  $\geq Rew(p^*) - (T_0 + \gamma + 4T\delta + \epsilon T).$ (plug in prev. eq.)
- Take $\epsilon = T^{-1/4}$ and $T_0 = T^{3/4} \implies$ done

Explore-then-exploit: analysis

Claim: $\left\{\left|\hat{S}(p) - S(p)\right| \leq \delta\right\}$ WHP

$\delta = O(\log T)/\sqrt{\epsilon T_0}$

w. prob. $\geq 1 - T^{-c}$

- $T_0 \leq k \Longrightarrow$ cannot run out of supply during exploration
- Fix price $p \in U$. Let $N = \#\text{rounds } @p$ during exploration.
- $E[N] = \frac{T_0}{\#\text{prices}} = \epsilon T_0$. Chernoff Bounds $\Longrightarrow N > \epsilon T_0/2$ WHP.
- $X_j = \mathbf{1}(\text{ sale at the j−th time price p is chosen })$
  $Y_m = X_1 + \cdots + X_m$ sum of IID random variables in [0,1]
  Chernoff Bounds $\Longrightarrow \left|\frac{Y_m}{m} - S(p)\right| \leq \delta(m)$ WHP
- Take union bound over all m.
  Then: $\left|\frac{Y_N}{N} - S(p)\right| \leq \delta\left(\frac{\epsilon T_0}{2}\right)$
  done because $\hat{S}(p) = Y_N/N$.

$\delta(m) = O(\log T)/\sqrt{m}$

Lower bound on regret

(old) cannot have regret $o(T^{2/3})$ for unlimited supply.
(new) cannot have regret $o(k^{2/3})$ for arbitrarily large k, T.

Proof idea: reduce to the LB for unlimited supply.

- Suppose we have an algorithm $\mathcal{A}$ which breaks (new).
  Construct algorithm which breaks (old).
- Suffices to break (old) for arb. large time horizon T'.
  Take any k, T for which $\mathcal{A}$ achieves regret $o(k^{2/3})$.
  We solve any unlimited supply instance I_0 with $T' = k/4$.
- Algorithm: use $\mathcal{A}$ on a simulated problem instance I_{sim}:
  - in each round, with prob. $k/2T$, ask next customer from I_0; else, just return "no sale".
- I_{sim} can't run out of supply, so its restriction to I_0 "works"

Outline

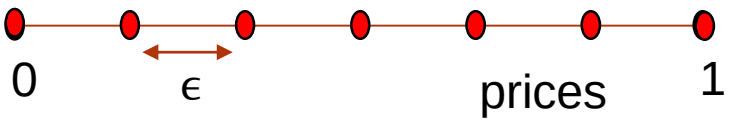
- ✓ Intro
- ✓ Unlimited supply

Limited supply:

- ✓ some observations
- ✓ explore-then-exploit
- ✓ lower bound
- ❑ a better algorithm (overview)

UCB algorithm for total rewards

- Uniform discretization U



- Want: in each round, pick price

$$\operatorname*{argmax}_{p \in U} \mathrm{UCB}(\mathrm{REW}(p))$$

- Approximate

approx. #sales @p

$$\mathrm{REW}(p) \approx p \cdot \min\left(k, T \cdot S(p)\right) \quad := f(p)$$

- Algorithm: in each round, pick price $p \in U$ with maximal

$$\mathrm{Index}(p) = p \cdot \min(k, n \cdot \mathrm{UCB}(S(p)))$$

ave. sales rate + conf. term

References

- (unlimited supply) Robert Kleinberg and Frank Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. FOCS 2003.
- (limited supply: explore-then-exploit) Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. Operations Research, 2009.
- (limited supply: optimal algorithm & lower bounds) Moshe Babaioff, Shaddin Dughmi, Robert Kleinberg and Alex Slivkins. Dynamic Pricing with Limited Supply. ACM EC 2012. Trans. on Economics and Computation, 2015.
- (limited supply: treatment of explore-then-exploit in this lecture) Alex Slivkins, unpublished.