Dynamic Pricing with Limited Supply

ACM EC 2012

Trans. on Economics and Computation 2015

Paradigmatic problem

- Seller with limited supply: k identical items to sell
- In each round $t = 1 \ldots n$, a new customer arrives
  - Seller offers 1 item at price $p_t \in [0,1]$
  - Customer i has private value $v_i \in [0,1]$
  - Customer accepts or rejects
- Until no more items or no more customers
- Dynamic pricing: update the price after each interaction
- Goal
  - Adjust price over time to maximize expected revenue
  - No bonus for leftover items

What's going on: Economics

- Interpretation: sales $\Leftrightarrow p_t \leq v_i$
- Where do the private value come from?
- Worst-case view: values chosen adversarially
  - Often leads to weak positive results
- Bayesian view: from a known distribution F
  - Strong assumption, sometimes unrealistic
- Prior-independent mechanisms are a compromise
  - Private values are sampled i.i.d. from unknown distribution F
  - Goal: be competitive against the optimal mechanism that knows the distribution F

Why Posted Price Mechanisms

- Demand curve
  - $S(p) = \Pr[sell\ at\ price\ p] = 1 - F(p)$
  - Fixed but unknown to seller
  - No parametric assumptions
- PPM 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 him in the future)
  - PPMs are truthful and group-strategy proof
  - Seller don’t need to estimate demand distribution in advance

Benchmarks with full information

- Revenue of the best fixed price

- Revenue of the optimal online mechanism
  - It is a posted price mechanism

- Revenue of the optimal offline mechanism
  - Reserve price [Myerson 1981]
  - Not constrained to posted prices

Weakest

Strongest

- Difference between the benchmarks is $O(\sqrt{k \log k})$ for regular distribution
  - $v - \frac{1-F(v)}{f(v)}$ is strictly increasing

Beyond best-fixed price

- Two prices better than one
  - Select from two prices randomly twice as good as the best fixed price
- Problem instance
  - Value $v_t = \begin{cases} 1 \; with \; prob \; \frac{\varepsilon k}{n} \\ \varepsilon \; otherwise \end{cases}$
  - Focus on prices $p = \{\varepsilon, 1\}$, $Rev(p) \leq \varepsilon k$ for both
  - Randomized policy $p = \begin{cases} \varepsilon \; with \; prob \; \frac{k}{n} \\ 1 \; otherwise \end{cases}$
  - $Rev(p) \geq \varepsilon k(2 - O(\frac{k}{n}))$

What is going on: Bandits

- First intuition: we want to sell at (unknown) best price
  - Offered price too low $\Rightarrow$ likely sale, wasted item
  - Offered price too high $\Rightarrow$ likely no sale, wasted customer
  - Learn something about the demand distribution
  - Explore-exploit tradeoff
  - Learn-and-earn
- With limited supply, the learning ability is handicapped
  - Can't afford to sell too many items while trying low prices
  - Without parametric assumptions, no long-range inference

Main results

Mechanism 1 Pricing strategy `CappedUCB` for n agents and k items

Parameter: $\delta \in (0, 1)$

1: $\mathcal{P} \leftarrow \{\delta(1+\delta)^i \in [0, 1] : i \in \mathbb{N}\}$ {"active prices"}

2: While there is at least one item left, in each round t
pick any price $p \in \text{argmax}_{p \in \mathcal{P}} I_t(p)$, where $I_t(p)$ is the "index" given by (4).

3: For all remaining agents, set price $p = \infty$.

- Optimal algorithm
  - For any supply k and any demand distribution
- Regret in expectation
  - $O(klogn)^{\frac{2}{3}}$
  - Compared with best fixed-price mechanism
- Arm$\Leftrightarrow$price p
- A random variable $Y_p = \begin{cases} p \; with \; prob \; S(p) \\ 0 \; with \; prob \; F(p) \end{cases}$

Fractional reward

- Total expected reward
- $Rew(p) = p \cdot E[\#sales\ at\ p\ ]$
- Difficult to work with directly, use approximation
- Easy upper bound: $E[\#sales\ at\ p] \leq \min(k, n \cdot S(p))$
- Use $v(p) = p \cdot \min(k, n \cdot S(p))$
  - fractional reward
- Claim: $v(p) - O(pk \log k) \leq Rew(p) \leq v(p)$
  - Prove by Chernoff bounds
  - $\Pr[|X - \mu| > \delta\mu] < 2e^{\frac{-\mu n \delta^2}{3}}$

UCB algorithm for total rewards

- Discretize prices
  - Only look at prices $p = \delta(1+\delta)^i \in [0,1]$
  - A finite set of prices U
- In each round
  - each price p is assigned a numerical score $Index(p)$
  - pick price $argmax_{p \in U} \in Index(p)$
  - With high probability $Index(p) \geq p \cdot E[\# \ sales \ at \ p]$
  - High-prob upper confidence bound on expected revenue in total rounds

Total revenue index

Selling rate

- Expected total revenue at price p
  - Use approximation: $p \cdot \min(k, nS(p))$
- Replace $S(p)$ with a high-prob upper bound
  - $Index(p) = p \cdot \min\{k, n\left(\hat{S}_t(p) + r_t(p)\right)\}$
- $\hat{S}_t(p) = \frac{\#\, of\ sells\ at\ p\ before\ round\ t}{N_t(p)}$
  - the average selling rate at price p so far
  - $N_t(p) = \#times\ price\ p\ was\ chosen\ before\ round\ t$
- Confidence radius $|\hat{S}_t(p) - S(p)| \leq r_t(p)$ WHP
- $r_t(p) = \sqrt{\frac{\alpha \hat{S}_t(p)}{N_t(p)+1}} + \frac{\alpha}{N_t(p)+1}$ reflecting uncertainty
- Index implicitly combines explore & exploit

High probability events

For each discretized price p

- Event 1: confidence radius
  - $|\hat{S}_t(p) - S(p)| \leq r_t(p) \leq 3(\sqrt{\frac{\alpha S(p)}{N_t(p)+1}} + \frac{\alpha}{N_t(p)+1})$
- Event 2: sales
  - $|X - S| \leq \beta(S) \triangleq O(\sqrt{S \log n}) + \log n$
  - $X_t = \mathbf{1}\{sale\ in\ round\ t\}, X = \sum_t X_t, S = \sum_t S(p_t)$
- Event 3: total reward
  - $\sum_t p_t\,(X_t - S(p_t)) \leq \beta(S)$
- $E[X_t | X_1, \ldots, X_{t-1}] = S(p)$
- p_t is determined by $(X_1, \ldots, X_{t-1})$

Badness of a price p

- Best discretized price q^* maximizes $v(\cdot)$ on U
- Compare per-round expected reward from p and $\frac{v(q^*)}{n}$
    - $\Delta(p) = \max\{0, \frac{v(q^*)}{n} - p \cdot S(p)\}$
- Analysis of a single price
    - Upper-bound $N(p) \cdot \Delta(p) \leq O(\log n)(1 + \frac{k}{n}\frac{1}{\Delta(p)})$
    - $N(p)$ is total #times price p is chosen
- Global analysis
    - Upper-bound regret in terms of $\sum_{p \in U} N(p) \cdot \Delta(p)$
    - Divide prices into two sets $\Delta(p) \geq \varepsilon$

Bound the total reward

- Best price p^*: maximizes $v(p)$ on $[0,1]$
- Best discretized price q^*: maximizes $v(p)$ on U
- Discretization Error $\triangleq v(p^*) - v(q^*) \leq \delta k$
- $Regret\ (n) \leq v(p^*) - Reward \leq O(klogn)^{\frac{2}{3}}$

$$\text{Regret} = \underbrace{\text{Regret}_U}_{\text{bandit}} + \underbrace{OPT - OPT_U}_{\text{discretization error}}$$

Better regret for regular demands

- Reward function $R(p) = p \cdot S(p)$
- Regular demands
  - if $R(\cdot)$ is concave: $R''(\cdot) \leq 0$
- Analysis uses an upper bound on
  - $H_{\delta,U} = | p \in U : R(p^*) - R(p) \leq \delta|$
- By concavity, $C \triangleq R'(s_F) > 0$, $R'\left(\frac{k}{n}\right) > C, \forall \frac{k}{n} \leq s_F$
  - A better upper bound on $H_{\delta,U}$
- Same algorithm but a different discretization step
- Regret: $c_F \sqrt{k} \log n$
  - Constant c_F depends on the demand curve, but not on T

Thank you ^^