Dynamic Pricing (Part II)

Guest Lecturer: Alex Slivkins (MSR-NYC)

Columbia University, IEOR, Spring 2016
“Learning and Optimization for Sequential Decision Making”
Mar 2, 2016
Paradigmatic problem

- **Seller** with limited supply: $k$ identical items to sell
- In each round $t = 1 \ldots T$, 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
  - Compete with **best fixed price**

No parametric assumptions
Limited supply \((k < T)\)

- maximizing expected \textit{per-round} reward is not the right goal. need to think about expected \textit{total} reward
- Best fixed price \(p^* = \arg\max_p \text{Rew}(p)\)
  \[
  \text{Rew}(p) \text{ expected total reward for fixed price } p
  \]
- \(\text{Regret}(k, T) = \text{Rew}(p^*) - \text{Rew}(\text{algorithm})\)
  want regret sublinear in \(k = \#\text{items}\)
- Lower bound on regret: \(\Omega(k^{2/3})\)
- Tool: \(f(p) = p \cdot \min(k, T \cdot S(p))\) “fractional reward”
  \[
  \text{Claim}: f(p) - O(p\sqrt{k \log k}) \leq \text{Rew}(p) \leq f(p)
  \]
UCB algorithm for total rewards

- Uniform discretization $U$

- Want: in each round, pick price
  $$\arg\max_{p \in U} UCB(\text{REW}(p))$$

- Approximate with frac. reward

- Algorithm: in each round $t$, pick price $p \in U$ with maximal “index”
  $$I_t(p) = p \cdot \min(k, T \cdot UCB(S(p)))$$

Recap

- $f(p) = p \cdot \min(k, T \cdot S(p))$

- Ave. sales rate + conf. term
Two thoughts from last lecture

- Adaptive vs non-adaptive exploration
- Clean event
Outline

Preparation for the algorithm
- better bound on discretization error
- sharper UCB
- clean events

Main argument (assuming clean event)
- “badness” of a price
- analysis of a single round
- argue about total reward

Beyond the basic model (time permitting)
Discretization error, revisited

- bounding discretization error by $\epsilon T$ is not good enough!
- Use frac. reward \( f(p) = p \cdot \min(k, T \cdot S(p)) \)
- Best price $p^*$: maximizes $f(p)$ on $[0,1]$
  
  Best discretized price $q^*$: maximizes $f(p)$ on $U$
  
  Discretization Error \( := f(p^*) - f(q^*) \leq \epsilon k \)
- Round down $p^*$ to the nearest price in $U$, call it $q$
  \[
  f(q^*) \geq (p^* - \epsilon) \cdot \min(k, p \cdot S(p^*)) \geq p^* \cdot \min(k, p \cdot S(p^*)) - \epsilon \cdot \min(k, p \cdot S(p^*)) \geq f(p^*) - \epsilon k
  \]
Better UCB

- \( \hat{S}_t(p) = \frac{\text{#sales \at \p \ before \ round \ t}}{N_t(p)} \), 
  where \( N_t(p) = \text{#times \ p \ was \ chosen \ before \ round \ t} \)
- **Confidence radius:**  \( |\hat{S}_t(p) - S(p)| \leq r_t(p) \)  WHP
  Then \( S^{UCB}(p) = \hat{S}_t(p) + r_t(p) \).
- Standard:  \( r_t(p) = \sqrt{\alpha / N_t(p)} \), where \( \alpha = \text{const} \times \log T \).
- Better:  \( r_t(p) = \sqrt{\alpha \hat{S}(p) / N_t(p)} + \alpha / N_t(p) \).
  - matches the standard conf. radius in the worst case.
  - much better for very small \( S(p) \):  \( \alpha / N_t(p) \).
“Clean” events (WHP)

- Event #1: confidence radius. 
\[ |\hat{S}_t(p) - S(p)| \leq r_t(p) \leq 3 \left( \frac{\sqrt{\alpha S(p)}}{N_t(p)} + \frac{\alpha}{N_t(p)} \right) \]

- Notation: consider our algorithm on unlimited supply instance.
\[ X_t = 1 \text{(sale in round } t); \ X = \sum_t X_t; \ S = \sum_t S(p_t). \]

- Event #2: sales.
\[ |X - S| \leq \beta(S) := O(\sqrt{S \log T} + \log T). \]

- Event #3: total reward.
\[ \sum_t p_t (X_t - S(p_t)) \leq \beta(S) \]

For each discretized price \( p \)

- For events #2 and #3, it is essential that:
  - \( E[X_t \mid X_1, \ldots, X_{t-1}] = S(p) \)
  - \( p_t \) is determined by \( (X_t, \ldots, X_{t-1}) \).
Outline

Preparation for the algorithm
✓ better bound on discretization error
✓ sharper UCB
✓ clean events

Main argument (assuming clean event)
☐ “badness” of a price
☐ analysis of a single round
☐ argue about total reward

Beyond the basic model (time permitting)
"Badness" of a price $p$

- Recall: Best (fractional) discretized price $q^*$: maximizes $f(\cdot)$ on $U$

\[
f(p) = p \cdot \min(k, T \cdot S(p))
\]

- Compare per-round exp. reward from $p$ and $f(q^*)/T$:

\[
\Delta(p) = \max \left( 0, \frac{f(q^*)}{T} - p \cdot S(p) \right)
\]

- Analysis of a single round: upper-bound $N(p) \cdot \Delta(p)$, where $N(p)$ is total #times price $p$ is chosen.

- "Global" analysis:
  upper-bound regret in terms of $\sum_{p \in U} N(p) \cdot \Delta(p)$. 
Analysis of a single round

**Lemma**: \( N(p) \cdot \Delta(p) \leq O(\log T) \left( 1 + \frac{k}{T \Delta(p)} \right) \)

\[
\begin{align*}
  f(p) &= p \cdot \min(k, T \cdot S(p)) \\
  I_t(p) &:= p \cdot \min(k, T \cdot (\hat{S}(p) + r_t(p)))
\end{align*}
\]

- By defn. of conf. radius:
  \[
  f(p) \leq I_t(p) \leq p \cdot \min(k, T \cdot (S(p) + 2r_t(p)))
  \]
- The “UCB trick”:
  \[
  (1) \quad f(q^*) \leq I_t(q^*) \leq I_t(p_t) \leq p_t \cdot \min(k, T \cdot (S(p_t) + 2r_t(p_t)))
  \]
  Then \( \Delta(p_t) \leq 2 \cdot p_t \cdot r_t(p_t) \); plugging in the “clean event”:
  \[
  (2) \quad \Delta(p_t) \leq O(p_t) \cdot \left( \sqrt{\alpha S(p) / N_t(p)} + \alpha / N_t(p) \right)
  \]
- Also from (1), if \( \Delta(p_t) > 0 \) then \( S(p_t) \leq \frac{k}{T} \) (omitting the details).
- Plug this into (2) and rearrange the terms; for each price \( p \), consider the last round \( t \) when \( p \) is chosen.
Bound \( \mathcal{W} := \sum_{p \in U} \Delta(p) \cdot N(p) \)

- A trick from analysis of UCB1: fix some \( \delta > 0 \), prices with \( \Delta(p) \leq \delta \) contribute \( \leq \delta \) per round. So:
  \[
  \mathcal{W} = \delta T + \sum_{p \in U: \Delta(p) \geq \delta} \Delta(p) \cdot N(p)
  \]

- Now plug in the previous lemma:
  \[
  \mathcal{W} \leq \delta T + O(\log T) \sum_{p \in U: \Delta(p) \geq \delta} \left( 1 + \frac{k}{T \Delta(p)} \right)
  \leq \delta T + O(\log T) \left( \frac{1}{\epsilon} + \frac{k}{T \epsilon \delta} \right).
  \]
Analysis of the total reward

- \( Rew_0 \): total exp. reward of our algorithm on problem instance with same demand curve & unlimited supply.
- **Lemma**: \( Rew \geq \min(f(q^*), Rew_0) - \beta \).
  - Short but subtle proof (omitted), uses “clean events”.
- By definition of “badness”
  \[
  Rew_0 = \sum_t p_t \cdot S(p) \geq \sum_t f(q^*)/T - \Delta(p_t) = f(q^*) - \sum_{p \in U} \Delta(p) \cdot N(p)
  \]
- Final computation:
  \[
  Rew \geq f(q^*) - \beta - \sum_{p \in U} \Delta(p) \cdot N(p)
  \]
  \[
  \geq f(p^*) - \epsilon k - \beta - \delta T + O(\log T) \left(\frac{1}{\epsilon} + \frac{k}{T \epsilon \delta}\right).
  \]
- Adjust parameters: \( \delta = \epsilon \frac{k}{T} \) and \( \epsilon = k^{-1/3} (\log T)^{2/3} \).

**Theorem**: \( \text{Regret}(T) \leq f(p^*) - Rew \leq (k \log T)^{2/3} \).
Outline

Preparation for the algorithm
✓ better bound on discretization error
✓ sharper UCB
✓ clean events

Main argument (assuming clean event)
✓ “badness” of a price
✓ analysis of a single round
✓ argue about total reward

Beyond the basic model (time permitting)
Better regret for regular demands

- Better regret if \( R(\cdot) \) is concave: \( R''(\cdot) \leq 0 \) (regular demands)

- How does it help:
  - analysis uses an upper bound on \( H_{\delta,U} = |\{p \in U : R(p^*) - R(p) \leq \delta \}| \)
  - by concavity, \( R(\cdot) \) is essentially quadratic near \( p^* \), \( \Rightarrow \) a better upper bound on \( H_{\delta,U} \).

- Same algorithm (UCB for total rewards), but a different discretization step \( \varepsilon \)

- regret \( C \times (k \log T)^{1/2} \), where constant \( C \) depends on the demand curve, but not on \( T \).
Beyond best fixed price

*All-knowing* benchmarks (known demand curve)

- best fixed price $p^*$
- optimal pricing policy
- optimal offline mechanism (Myerson 1981).

All benchmarks are within $O(\sqrt{k \log k})$ for regular demands (*)

In general, optimal pricing policy can be much better than $p^*$

(*) I.e., if the reward function $R(p) = p \cdot S(p)$ is concave.

Two prices better than one!

**Example:** Distribution $D$ over two prices twice as good as $p^*$

- Problem instance: value $v_t = \begin{cases} 1 & \text{w/ prob } \epsilon k/T \\ \epsilon & \text{otherwise} \end{cases}$
- WLOG focus on prices $p \in \{\epsilon, 1\}$. For both, $\text{REW}(p) \leq \epsilon k$.
- Distribution $D$: $\begin{cases} p = \epsilon & \text{w/ prob } k/T \\ p = 1 & \text{otherwise} \end{cases}$

Then $\text{REW}(D) \geq \epsilon k \left(2 - O\left(k/T\right)\right)$. 

**Beyond best fixed price**
Generalizations

• selling multiple products, limited supply of each
  • action = price vector (price for each product)

• each product consumes some \textit{primitive resources}
  • action = price vector (price for each product)

• bundling & volume pricing
  • given: collection of allowed bundles
  • action = price vector (price for each allowed bundle)
Contextual dynamic pricing

- Seller with \( k \) identical items
- In each round \( t \),
  - new customer arrives, with known profile \( x_t \)
  - seller offers 1 item @ price \( p_t \in [0,1] \)
  - customer accepts with (unknown) probability \( S(p_t|x_t) \)
- Until no more items or no more customers
- Goal: adjust price over time, to maximize revenue

**Contextual bandits**: in each round, observable “context”. All probabilities depend on both action and context.
Closely related: dynamic procurement

“Dynamic pricing for buying” (vs. selling)

- **Employer** with many tasks, limited budget
- In each round $t$, a new **worker** arrives
  - employer offers price $p_t \in [0,1]$
  - worker accepts or rejects
- Until **out of workers** or **out of money**
- $\text{Pr}[\text{accept @ price } p]$ fixed but unknown

**Goal:** adjust price over time, to maximize #tasks

**Extensions:** e.g. multiple types of tasks with per-type budgets
References for the two lectures (*)


- *(limited supply: treatment of explore-then-exploit in this lecture)* Alex Slivkins, unpublished.

- *(Beyond the basic model)* Ashwin Badanidiyuru, Robert Kleinberg and Alex Slivkins. Bandits with Knapsacks. FOCS 2013.

(*) These references are only for the material presented in the lectures. For more background, see the “bandits with knapsacks” paper (full version).