An Optimal Bidimensional Multi Armed Bandit Auction for Multi unit Procurement

Satyanath Bhat
Joint work with: Shweta Jain, Sujit Gujar, Y. Narahari

Department of Computer Science and Automation,
Indian Institute of Science
Bangalore

August 3, 2015

Outline

- Preliminaries:
  - Optimal auction: An Economics perspective
  - The stochastic Multi-armed bandit problem
- Motivation
- Problem statement
- Main results
- Empirical evaluations
- Future work

Optimal Auction

Simple Economics of Optimal Auctions

Optimal Auctions $\equiv$ Third Degree Price Discrimination in Monopolistic Market[a]

[a] Jeremy Bulow and John Roberts. "The simple economics of optimal auctions". In: *The Journal of Political Economy* (1989), pp. 1060–1090.

Monopoly

- Single Seller
- Price set by Seller
- Profit maximization :price vs demand

Price Discrimination

- First Degree:- Maximum price to each customer.
- Second Degree:- Charge prices based on quantity.
- Third Degree:- Separate market for different sets of customers, different prices set for different groups.

Third Degree Price Discrimination

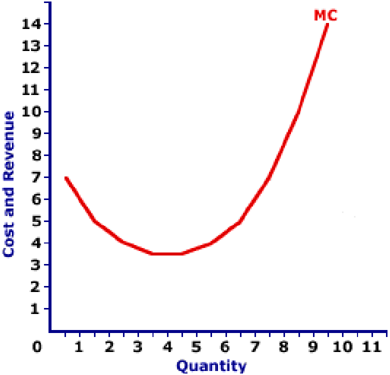


Image Courtesy: http://www.amosweb.com

Third Degree Price Discrimination

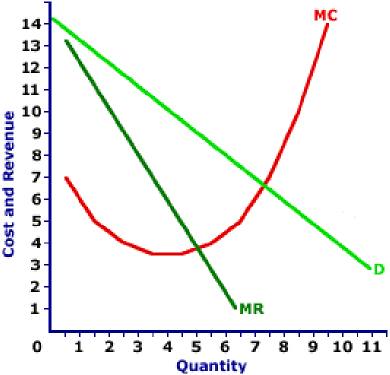


Image Courtesy: http://www.amosweb.com

Third Degree Price Discrimination

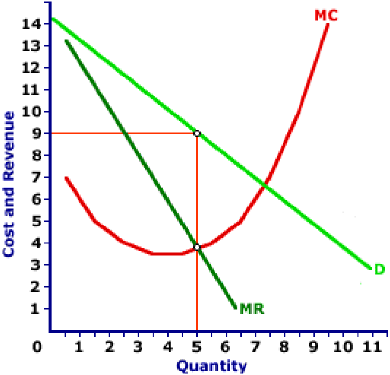


Image Courtesy: http://www.amosweb.com

Third Degree Price Discrimination

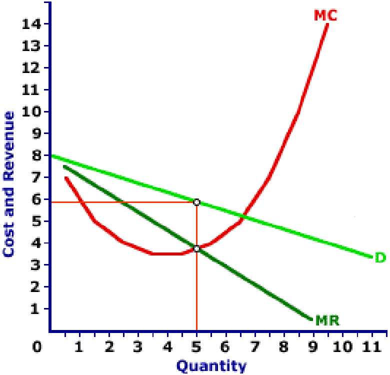


Image Courtesy: http://www.amosweb.com

Optimal Auction

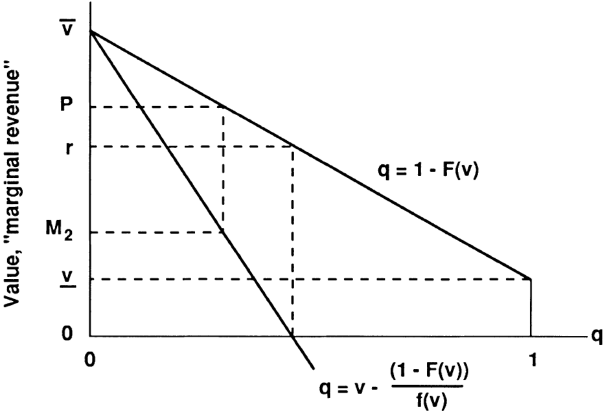


Fig. 1.—Construction of an optimal auction

Multi-armed bandits

Estimating the bias of a coin confidently

- Toss a coin with bias p, t times
- Random variables $X_1, \cdots, X_t \in \{0, 1\}$
- Estimator $\hat{p} = \sum_i X_i / t$
- Hoeffding's Inequality: $P((\hat{p} - p) > \epsilon) \leq e^{-2t\epsilon^2}$

K armed bandit

- Samples from each arm distributed as a Bernoulli r.v
- Arm i is equivalent to a coin producing a mean reward μ_i
- $N_i(t)$: No. of times arm i is pulled in total t trials
- $S_i(t)$: sum of the rewards produced by arm i upto t trials
- $\hat{\mu}_i = S_i(t)/N_i(t)$
- Define UCB for $\hat{\mu}_i{}^+ =: \hat{\mu}_i + \sqrt{\frac{2\log t}{N_i(t)}}$
- Arm with highest UCB is pulled(Algorithm UCB1[a])

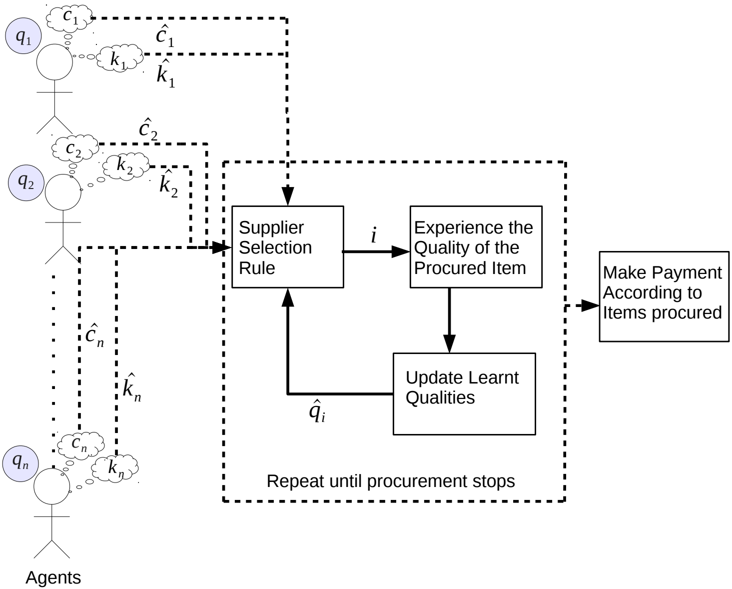
[a] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. "Finite-time Analysis of the Multiarmed Bandit Problem". In: *Journal of Machine Learning* 47.2-3 (2002), pp. 235–256.

Motivation

A hospital wishes to procure a large number of units of a generic drug from a pool of suppliers.

- The efficiency of a drug is stochastic and the mean varies across suppliers
- For each supplier: the cost of production per unit and the capacity is private to him
- Hospital aims to learn the mean efficiency of the drug from each supplier and minimize its cost

Model



q_1
c_1
k_1
$\hat{c}_1$
$\hat{k}_1$
$\hat{c}_2$
c_2
k_2
q_2
$\hat{k}_2$
Supplier Selection Rule
i
Experience the Quality of the Procured Item
Make Payment According to Items procured
$\hat{c}_n$
$\hat{k}_n$
Update Learnt Qualities
$\hat{q}_i$
c_n
q_n
k_n
Repeat until procurement stops
Agents

Problem

Optimization Problem:

$$\max \sum_{i=1}^{n} \Big(x_i R q_i - t_i \Big)$$
$$\text{s.t. } x_i \in \{0, 1, \ldots, \hat{k}_i\} \ \ , \ \sum_i x_i \leq L$$

Challenges: Unknown qualities $q_i^\prime s$, strategic costs $c_i^\prime s$ and strategic capacities $k_i^\prime s$

Goal: Design an incentive compatible MAB mechanism, to procure L units of the item while learning the qualities of the suppliers

Auctions with known qualities

Theorem (BIC and IR characterization)

A mechanism is BIC and IR iff $\forall i \in N$,

1. $X_i(\hat{c}_i, \hat{k}_i; q)$ is non-increasing in $\hat{c}_i$, $\forall q$ and $\forall \hat{k}_i \in [\underline{k}_i, k_i]$
2. $\rho_i(\hat{c}_i, \hat{k}_i; q)$ is non-negative, and non-decreasing in $\hat{k}_i$
3. $\rho_i(\hat{c}_i, \hat{k}_i; q) = \rho_i(\bar{c}_i, \hat{k}_i; q) + \int_{\hat{c}_i}^{\bar{c}_i} X_i(z, \hat{k}_i; q)dz$

Auctions with known qualities

Theorem (Optimal payment structure)

Suppose the allocation rule maximizes

$$\sum_{i=1}^{n} \int_{\underline{c}_1}^{\bar{c}_1} \cdots \int_{\underline{c}_n}^{\bar{c}_n} \int_{\underline{k}_1}^{\bar{k}_1} \cdots \int_{\underline{k}_n}^{\bar{k}_n} \left(Rq_i - \left(c_i + \frac{F_i(c_i|k_i)}{f_i(c_i|k_i)} \right) \right)$$
$$x_i(c_i, k_i, c_{-i}, k_{-i}) f_1(c_1, k_1) \ldots f_n(c_n, k_n)\, dc_1 \ldots dc_n\, dk_1 \ldots dk_n$$

subject to monotonicity condition (1). Also suppose that the payment is given by

$$T_i(c_i, k_i; q) = c_i X_i(c_i, k_i; q) + \int_{c_i}^{\bar{c}_i} X_i(z, k_i; q) dz$$

then such a payment scheme and allocation scheme constitute an optimal auction satisfying BIC and IR.

2D-OPT

Algorithm 1: 2D-OPT Mechanism

Input: $\forall i$, Bids $b_i = (\hat{c}_i \, \hat{k}_i)$, reward parameter R

Output: An optimal, DSIC, IR Mechanism $\mathcal{M} = (x, t)$

Allocation is given by $x = \text{ALLOC}(N, \hat{c}, \hat{k}, q, L)$

for $i \in N$ && $x_i \neq 0$ **do**

  $G_i := Rq_i - H_i(b_i)$

  $y = \text{ALLOC}(N \setminus \{i\}, \hat{c}_{-i}, (\hat{k}_{-i} - x_{-i}), q_{-i}, x_i)$

  Payment to i, $t_i =$

$$\sum_{k \in N \setminus \{i\}} y_k \max(G_i^{-1}(Rq_k - H_k(b_k)), \bar{c}_i) + \Big(x_i - \sum_k y_k\Big)\bar{c}_i$$

end

Auctions with unknown qualities

Well-behaved allocation Rule An allocation rule x is called a *Well Behaved Allocation Rule* if:

- x_i^j depends only on the agent's bids and the reward realization till j units and is non decreasing in terms of costs
- For any three distinct agents $\{\alpha, \beta, \gamma\}$ such that j^{th} round unit is allocated to β. A change of bid by agent α should not transfer allocation of j^{th} round unit from β to γ if other quantities are fixed till j units
- For all reward realizations s, $x_i(c_i, k_i; s)$ is non-decreasing with increase in capacity k_i

Theorem

For a well behaved allocation rule, there exists a transformation that produces the transformed allocation $(\tilde{x})$ and payment $(\tilde{t})$ such that the resulting mechanism $\mathcal{M} = (\tilde{x}, \tilde{t})$ is stochastic BIC and IR.

2D-UCB

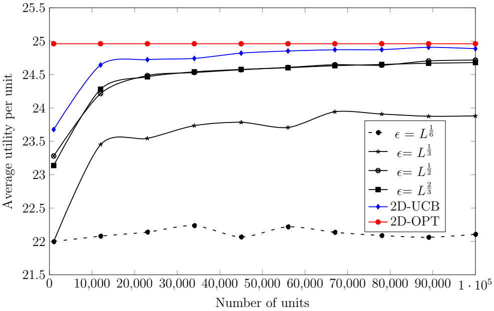
Algorithm 2: 2D-UCB Mechanism

Input: $\forall i \in N$, bids $\hat{c}_i \in [\underline{c}_i, \overline{c}_i]$, $\hat{k}_i \in [\underline{k}_i, k_i]$, parameter $\mu \in (0,1)$, Reward parameter R
Output: A mechanism $\mathcal{M} = (x, t)$

$\forall i \in N$, $\hat{q}_i^+ = 1$, $\hat{q}_i^- = 0$, $n_i = 1$
Obtain modified bids as (α, β)
$= ((\alpha_1(\hat{c}_1), \beta_1(\hat{c}_1)), \ldots, (\alpha_n(\hat{c}_n), \beta_n(\hat{c}_n))$ using resampling
Allocate one unit to all agents and estimate empirical quality $\hat{q}$
$\hat{q}_i = \tilde{q}_i(i)/n_i$, $\hat{q}_i^+ = \hat{q}_i + \sqrt{\frac{1}{2n_i} ln(t)}$
for $t = n$ *to* L **do**
  Compute $H_i = \alpha_i + \frac{F_i(\alpha_i|\hat{k}_i)}{f_i(\alpha_i|\hat{k}_i)}$
  Let $i = \arg\max_{\{j s.t. k_j > n_j\}} R\hat{q}_j^+ - H_j$ and $\hat{G}_i = R\hat{q}_i^+ - H_i$
  if $\hat{G}_j > 0$ **then**
    Procure the unit from agent i and update $\hat{q}_i$
    $\hat{q}_i^+ = \hat{q}_i + \sqrt{\frac{2}{n_i} ln(t)}$
  else
    break \\ Don't allocate future units to anyone
Make payment to each agent i, $\tilde{T}_i = \hat{c}_i n_i + P_i$, where,

$$P_i = \begin{cases} \frac{1}{\mu} n_i(\overline{c} - \hat{c}_i), & \text{if} \beta_i > \hat{c}_i \\ 0, & \text{otherwise.} \end{cases}$$

Empirical Evaluations



Observations:

- All the mechanisms approach 2D-OPT
- The performance of 2D-UCB is superior as it approaches 2D-OPT faster

Future Work

- An extension where allocation happens to subset of agents at any round would be interesting
- The complete characterization of a learning algorithm in this space is still open
- A theoretical analysis of regret is also a possible direction of future work

THANK YOU