Improved Multi-armed Bandit Mechanisms for Sponsored Search Auctions

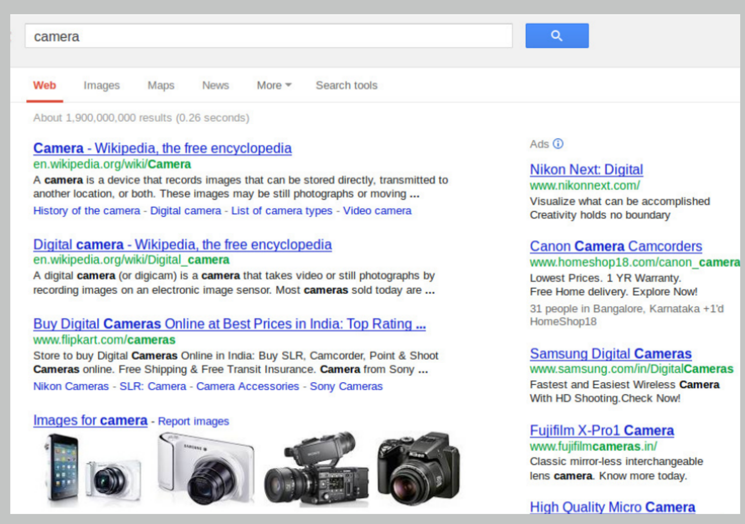
Divya Padmanabhan†, Satyanath Bhat†, Prabuchandran K.J.*, Shirish Shevade†, Y. Narahari†

†Dept. of CSA, IISc, Bangalore; * IBM IRL, Bangalore

Sponsored Search Auctions (SSA)

Consider a set of ads competing for a slot in a search engine's sponsored ads slots.

- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Expected Social Welfare for agent i, $W_i = \mu_i \theta_i$
- Planner wants to allocate a slot to the agent with maximum social welfare.

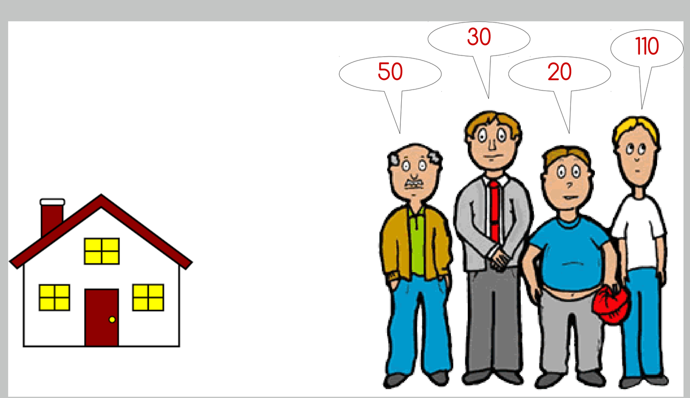


Multi-armed Bandit Algorithms

- Pull arms that have yielded high rewards? Or explore new arms?
- Explore-exploit dilemma
- Popular MAB algorithms:
  - Upper Confidence Bound (UCB)
  - Thompson Sampling
- *The samples we get are only of the clicks. We never get to observe samples of social welfare!*
- If θ_i known, use MAB algorithms

Mechanism Design

- Mechanism $< \mathcal{A}, \mathcal{P} >$ is a tuple with allocation rule and payment rule
- Need to design mechanisms satisfying some properties
  - Dominant Strategy Incentive Compatible (DSIC): Truthful reporting yields the best utility for every agent under all circumstances $u_i(\theta_i, b_{-i}, \rho, t; \theta_i) \geq u_i(b_i, b_{-i}, \rho, t; \theta_i)$
  - Individually Rational (IR): Truthful reporting yields a non-negative utility $u_i(\theta_i, b_{-i}, \rho, t; \theta_i) \geq 0$
- Example: House being auctioned, 4 participants



- Vickrey Clarke Grove (VCG) Mechanism [4]
- Externality based payment
- Highest bidder wins (110)
- Pays 2nd highest bid(50)
- Mechanism is DSIC, IR

MAB Mechanisms

- Allocation Rule
  - $\mathcal{A}$ is given by the MAB algorithm eg. UCB.
  - I_t arm pulled/allocated at time t,
  - W_*: Expected social welfare of the best agent
  - $Regret(\mathcal{A}) = \sum_{t=1}^{T} W_* - W_{I_t}$
  - MAB algorithm should not pull sub-optimal arms "very often" (logarithmic in T is alright) Eg. $T = 10^6$, $\log T = 6$
- Payment for selected agent: eg. VCG scheme

Existing MAB Mechanisms

- Characterisation of truthful MAB Mechanisms [1], [2]
  - DSIC truthful mechanism $\implies$ Exploration separated (No payment in the rounds where learning happens)
  - DSIC truthful, Deterministic mechanisms $\implies$ Regret $\Omega(T^{2/3})$ (Eg. $T = 10^6$, $\log T = 6$, $T^{2/3} = 10000$)

Some Key Observations:

- $\Omega(T^{2/3})$ exploration rounds needed for distinguishing arms that are arbitrarily close.
- Planner may not differentiate between arbitrarily close arms.
- Forced to suffer regret $\Omega(T^{2/3})$ even then!

Our Approach for Single Slot SSA: Δ-UCB [3]

- Allow the planner to specify a parameter Δ, the resolution to which arms must be distinguished
- $S_\Delta = \{i \in [K] : W_* - W_i < \Delta\}$
- Define Δ-regret $= \sum_{t=1}^{T} (W_* - W_{I_t}) \mathbb{1}[I_t \in [K] \setminus S_\Delta]$
- Design truthful mechanisms with logarithmic regret Δ-regret
- Δ-UCB Algorithm:

Exploration rounds

- Explore for $\gamma = 8K\theta_{max}^2 \log T / \Delta^2$ rounds, free for all agents
- Obtain LCB, UCB indices $\widehat{\mu}_{i,\gamma}^-, \widehat{\mu}_{i,\gamma}^+$ for every agent

Exploitation rounds $(\gamma + 1 \ldots T)$

- Choose agent $\hat{i}_* = \mathbf{arg\,max}_{i \in [K]} \widehat{\mu}_{i,\gamma}^+ b_i$ (Best agent according to estimated social welfare)
- $j = \mathbf{arg\,max}_{i \in [K] \setminus \{\hat{i}_*\}} \widehat{\mu}_{i,\gamma}^+ b_i$ (2nd best agent)
- $P_{\hat{i}_*} = \widehat{\mu}_{j,\gamma}^+ b_j / \widehat{\mu}_{\hat{i}_*,\gamma}^+$ (weighted VCG payment)

Properties of Δ-UCB

- Δ-UCB mechanism is dominant strategy incentive compatible (DSIC) and individually rational (IR).
- If the Δ-UCB mechanism is executed for a total time horizon of T rounds, it achieves an expected Δ-regret of $O(\log T)$.

Extension to Multi-slot SSA

- M slots to be allocated
- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Γ_m: Probability that an ad at slot m is observed
- Expected Social Welfare for agent i when allocated slot m, $W_{i,m} = \Gamma_m \mu_i \theta_i$, $\Gamma_m, \mu_i, \theta_i$ all are unknown!

Δ-UCB for Multi Slot SSA

A useful property of Γ_ms: $\Gamma_1 = 1 \geq \Gamma_2 \geq \ldots \Gamma_M$. Suppose Γ_ms are known.

Exploration Rounds:

- Explore for $\gamma = \lceil 8K\theta_{max}^2 \log T / \Delta^2 \rceil$ rounds, free for all agents
- Obtain LCB, UCB indices $\widehat{\mu}_{i,\gamma}^-, \widehat{\mu}_{i,\gamma}^+$ for every agent

Exploitation rounds $(\gamma + 1 \ldots T)$:

- Select the best M agents as per $\mathbf{arg\,max}_{i \in [K]} \widehat{\mu}_{i,\gamma}^+ b_i$
- Use weighted VCG payment to decide the amount agents pay to the planner $= \left(1/\Gamma_m \mu_{\widehat{K}(m)}^+\right) \sum_{l=m+1}^{M+1} (\Gamma_{l-1} - \Gamma_l)\, \widehat{\mu}_{K^{(l)},u}^+ b_{K^{(l)}} \rho_{\widehat{K}(m)}(t)$
- Payments are a function of Γ_m $\forall m$

If Γ_m unknown? Interestingly, can be shown that setting any value for Γ_m works! Need to ensure that $\Gamma_1 = 1 \geq \Gamma_2 \geq \ldots \Gamma_M$

Open Problems

- Find the payments so that planner gets maximum revenue.
- General scenario where Γ is unknown and depends on the ad as well.

References

[1] M. Babaioff, Y. Sharma, and A. Slivkins. Characterizing truthful multi-armed bandit mechanisms. *SIAM Journal on Computing*, 43(1):194–230, 2014.

[2] N. Gatti, A. Lazaric, M. Rocco, and F. Trovo. Truthful learning mechanisms for multi-slot sponsored search auctions with externalities. *Artificial Intelligence*, 227:93–139, 2015.

[3] D. Padmanabhan, S. Bhat, P. K. J., S. Shevade, and Y. Narahari. A dominant strategy truthful, deterministic multi-armed bandit mechanism with logarithmic regret. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*, 2017.

[4] W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961.

Divya Padmanabhan, PhD Student, Game Theory Lab, CSA, IISc   Email: divs1202@gmail.com

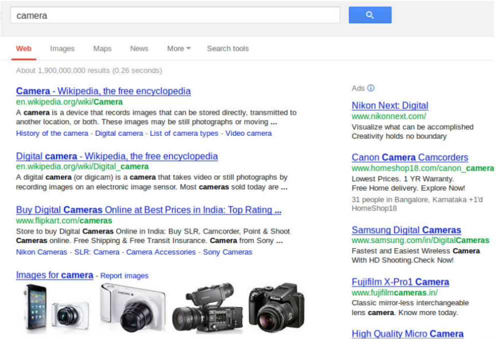
Improved Multi-armed Bandit Mechanisms for Sponsored Search Auctions

Divya Padmanabhan
PhD Student,
Dept. of CSA,
IISc, Bangalore

Advisors: Prof. Shirish Shevade,
Prof. Y. Narahari

April 7, 2017

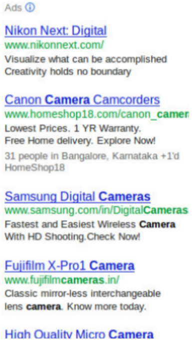
Sponsored Search Auctions (SSA)



- Ads shown against search results
- Advertiser pays search engine if his ad receives a click
- Search engine must decide:
  (1) how to allocate slots
  (2) payment for clicks

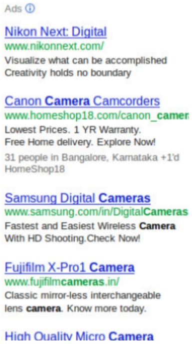
Sponsored Search Auctions

- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Expected Social Welfare for agent i, $W_i = \mu_i \theta_i$
- Planner wants to allocate a slot to the agent with maximum social welfare.



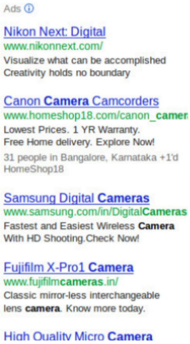
Sponsored Search Auctions

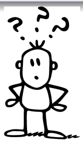
- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Expected Social Welfare for agent i, $W_i = \mu_i \theta_i$
- Planner wants to allocate a slot to the agent with maximum social welfare.
- Do we know μ_i? NO!



Sponsored Search Auctions

- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Expected Social Welfare for agent i, $W_i = \mu_i \theta_i$
- Planner wants to allocate a slot to the agent with maximum social welfare.
- Do we know μ_i? NO!
- Do we know θ_i? NO!





Multi-armed Bandits for Learning μ

Reward from Bernoulli with μ_1=0.9

More likely to get 1!

Reward from Bernoulli with μ_2=0.1

More likely to get 0!

At every time instant need to choose a machine/arm which will give highest reward $\in \{0, 1\}$. But means are unknown!

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

William R. Thompson. "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples". In: *Biometrika* 25.3/4 (1933), pp. 285–294.

Multi-armed Bandits for Learning μ

Reward from Bernoulli with μ_1=0.9

More likely to get 1!

Reward from Bernoulli with μ_2=0.1

More likely to get 0!

At every time instant need to choose a machine/arm which will give highest reward $\in \{0, 1\}$. But means are unknown!

Which arm to choose?

- Pull arms that have yielded high rewards?
- Explore new arms?
- Exploration-exploitation dilemma

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

William R. Thompson. "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples". In: *Biometrika* 25.3/4 (1933), pp. 285–294.

Multi-armed Bandits for Learning μ

Reward from
Bernoulli with μ_1=0.9

More likely to get 1!

Reward from
Bernoulli with μ_2=0.1

More likely to get 0!

At every time instant need to choose a machine/arm which will give highest reward $\in \{0, 1\}$. But means are unknown!

Which arm to choose?

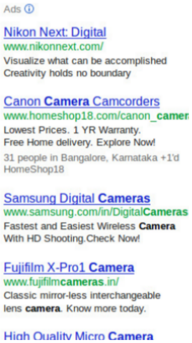
- Pull arms that have yielded high rewards?
- Explore new arms?
- Exploration-exploitation dilemma
- State of the art MAB algorithms:
  - Upper Confidence Bound (UCB)
  - Thompson Sampling

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

William R. Thompson. "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples". In: *Biometrika* 25.3/4 (1933), pp. 285–294.

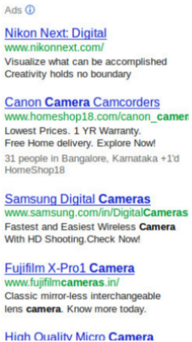
MAB Algorithms for Sponsored Search Auctions

- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Expected Social Welfare for agent i, $W_i = \mu_i \theta_i$
- Planner wants to allocate a slot to the agent with maximum social welfare.
- *The samples we get are only of the clicks. We never get to observe samples of social welfare!*



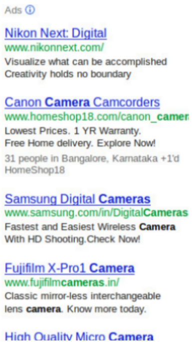
MAB Algorithms for Sponsored Search Auctions

- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Expected Social Welfare for agent i, $W_i = \mu_i \theta_i$
- Planner wants to allocate a slot to the agent with maximum social welfare.
- *The samples we get are only of the clicks. We never get to observe samples of social welfare!*
- If θ_i known, use MAB algorithms



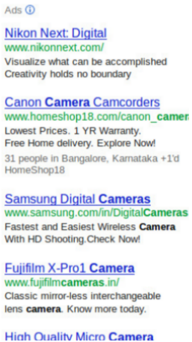
MAB Algorithms for Sponsored Search Auctions

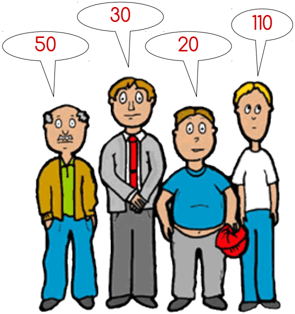
- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Expected Social Welfare for agent i, $W_i = \mu_i \theta_i$
- Planner wants to allocate a slot to the agent with maximum social welfare.
- *The samples we get are only of the clicks. We never get to observe samples of social welfare!*
- If θ_i known, use MAB algorithms
- Otherwise?



MAB Algorithms for Sponsored Search Auctions

- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Expected Social Welfare for agent i, $W_i = \mu_i \theta_i$
- Planner wants to allocate a slot to the agent with maximum social welfare.
- *The samples we get are only of the clicks. We never get to observe samples of social welfare!*
- If θ_i known, use MAB algorithms
- Otherwise? Mechanism Design (Auctions!)





Vickrey Clarke Grove (VCG) Mechanism

- Allocation rule: Highest bidder wins ($b_4 = 110$)
- Payment rule: Winner pays second highest bid ($p_4 = 50$)
- VCG Mechanism is dominant strategy incentive compatible (DSIC) truthful, maximizes social welfare.

William Vickrey. "Counterspeculation, auctions, and competitive sealed tenders". In: *The Journal of Finance* 16.1 (1961), pp. 8–37.

MAB Mechanisms

Learn mean of observable samples using multi-armed bandits and elicit private valuations truthfully with mechanism design.

Mechanism Design with Multi-armed Bandit Algorithms

- Allocation Rule
  - $\mathcal{A}$ is given by the MAB algorithm eg. UCB.
  - I_t arm pulled/allocated at time t,
  - W_*: Expected social welfare of the best agent

MAB Mechanisms

Learn mean of observable samples using multi-armed bandits and elicit private valuations truthfully with mechanism design.

Mechanism Design with Multi-armed Bandit Algorithms

- Allocation Rule
  - $\mathcal{A}$ is given by the MAB algorithm eg. UCB.
  - I_t arm pulled/allocated at time t,
  - W_*: Expected social welfare of the best agent
  - $Regret(\mathcal{A}) = \sum_{t=1}^{T} W_* - W_{I_t}$

MAB Mechanisms

Learn mean of observable samples using multi-armed bandits and elicit private valuations truthfully with mechanism design.

Mechanism Design with Multi-armed Bandit Algorithms

- Allocation Rule
  - $\mathcal{A}$ is given by the MAB algorithm eg. UCB.
  - I_t arm pulled/allocated at time t,
  - W_*: Expected social welfare of the best agent
  - $Regret(\mathcal{A}) = \sum_{t=1}^{T} W_* - W_{I_t}$
  - MAB algorithm should not pull sub-optimal arms "very often"

MAB Mechanisms

Learn mean of observable samples using multi-armed bandits and elicit private valuations truthfully with mechanism design.

Mechanism Design with Multi-armed Bandit Algorithms

- Allocation Rule
  - $\mathcal{A}$ is given by the MAB algorithm eg. UCB.
  - I_t arm pulled/allocated at time t,
  - W_*: Expected social welfare of the best agent
  - $Regret(\mathcal{A}) = \sum_{t=1}^{T} W_* - W_{I_t}$
  - MAB algorithm should not pull sub-optimal arms "very often" (logarithmic in T is alright)

Learn mean of observable samples using multi-armed bandits and elicit private valuations truthfully with mechanism design.

Mechanism Design with Multi-armed Bandit Algorithms

- Allocation Rule
  - $\mathcal{A}$ is given by the MAB algorithm eg. UCB.
  - I_t arm pulled/allocated at time t,
  - W_*: Expected social welfare of the best agent
  - $Regret(\mathcal{A}) = \sum_{t=1}^{T} W_* - W_{I_t}$
  - MAB algorithm should not pull sub-optimal arms "very often" (logarithmic in T is alright)
- Payment for selected agent: eg. VCG scheme

MAB Mechanisms: Our Observation

- Characterisation of truthful MAB Mechanisms[4]

  - DSIC truthful mechanism $\implies$ Exploration separated
    (No payment in the rounds where learning happens)

[4] Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. "Characterizing truthful multi-armed bandit mechanisms". In: *SIAM Journal on Computing* 43.1 (2014), pp. 194–230.

[5] Nicola Gatti et al. "Truthful learning mechanisms for multi-slot sponsored search auctions with externalities". In: *Artificial Intelligence* 227 (2015), pp. 93–139.

MAB Mechanisms: Our Observation

- Characterisation of truthful MAB Mechanisms[4]

  - DSIC truthful mechanism $\implies$ Exploration separated (No payment in the rounds where learning happens)
  - DSIC truthful, Deterministic mechanisms $\implies$ Regret $\Omega(T^{2/3})$ (Eg. $T = 10^6$, $\log T = 6$, $T^{2/3} = 10000$)

[4] Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. "Characterizing truthful multi-armed bandit mechanisms". In: *SIAM Journal on Computing* 43.1 (2014), pp. 194–230.

[5] Nicola Gatti et al. "Truthful learning mechanisms for multi-slot sponsored search auctions with externalities". In: *Artificial Intelligence* 227 (2015), pp. 93–139.

MAB Mechanisms: Our Observation

- Characterisation of truthful MAB Mechanisms[4]

- DSIC truthful mechanism $\implies$ Exploration separated
  (No payment in the rounds where learning happens)
- DSIC truthful, Deterministic mechanisms $\implies$ Regret $\Omega(T^{2/3})$
  (Eg. $T = 10^6$, $\log T = 6$, $T^{2/3} = 10000$)

- Subsequent research influenced by this characterisation, similar results in multi-slot SSA[5]

- Settle for weaker notions of truthfulness to bring regret down from $\Omega(T^{2/3})$

[4] Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. "Characterizing truthful multi-armed bandit mechanisms". In: *SIAM Journal on Computing* 43.1 (2014), pp. 194–230.

[5] Nicola Gatti et al. "Truthful learning mechanisms for multi-slot sponsored search auctions with externalities". In: *Artificial Intelligence* 227 (2015), pp. 93–139.

MAB Mechanisms: Our Observation

- Characterisation of truthful MAB Mechanisms[4]

- DSIC truthful mechanism $\implies$ Exploration separated (No payment in the rounds where learning happens)
- DSIC truthful, Deterministic mechanisms $\implies$ Regret $\Omega(T^{2/3})$ (Eg. $T = 10^6$, $\log T = 6$, $T^{2/3} = 10000$)

- Subsequent research influenced by this characterisation, similar results in multi-slot SSA[5]

- Settle for weaker notions of truthfulness to bring regret down from $\Omega(T^{2/3})$

Our Observation

- $\Omega(T^{2/3})$ exploration rounds needed for distinguishing arms that are arbitrarily close.

- Planner may not differentiate between arbitrarily close arms.

- Forced to suffer regret $\Omega(T^{2/3})$ even then!

[4] Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. "Characterizing truthful multi-armed bandit mechanisms". In: *SIAM Journal on Computing* 43.1 (2014), pp. 194–230.

[5] Nicola Gatti et al. "Truthful learning mechanisms for multi-slot sponsored search auctions with externalities". In: *Artificial Intelligence* 227 (2015), pp. 93–139.

Our Approach for Single Slot SSA: Δ-UCB

- Allow the planner to specify a parameter Δ, the resolution to which arms must be distinguished

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

Our Approach for Single Slot SSA: Δ-UCB

- Allow the planner to specify a parameter Δ, the resolution to which arms must be distinguished
- $S_\Delta = \{i \in [K] : W_* - W_i < \Delta\}$

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

Our Approach for Single Slot SSA: Δ-UCB

- Allow the planner to specify a parameter Δ, the resolution to which arms must be distinguished
- $S_\Delta = \{i \in [K] : W_* - W_i < \Delta\}$
- Define Δ-regret $= \sum_{t=1}^{T} (W_* - W_{I_t}) \mathbb{1}\left[I_t \in [K] \setminus S_\Delta\right]$

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

Our Approach for Single Slot SSA: Δ-UCB

- Allow the planner to specify a parameter Δ, the resolution to which arms must be distinguished
- $S_\Delta = \{i \in [K] : W_* - W_i < \Delta\}$
- Define Δ-regret $= \sum_{t=1}^{T} (W_* - W_{I_t}) \mathbb{1}\left[I_t \in [K] \setminus S_\Delta\right]$
- Design truthful mechanisms with logarithmic ~~regret~~ Δ-regret

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

Our Approach for Single Slot SSA: Δ-UCB

- Allow the planner to specify a parameter Δ, the resolution to which arms must be distinguished
- $S_\Delta = \{i \in [K] : W_* - W_i < \Delta\}$
- Define Δ-regret $= \sum_{t=1}^{T} (W_* - W_{I_t}) \mathbb{1}\left[I_t \in [K] \setminus S_\Delta\right]$
- Design truthful mechanisms with logarithmic ~~regret~~ Δ-regret

- Exploration rounds
  - Explore for $\gamma = 8K\theta^2_{max} \log T/\Delta^2$ rounds, free for all agents
  - Obtain LCB, UCB indices $\widehat{\mu}^-_{i,\gamma}, \widehat{\mu}^+_{i,\gamma}$ for every agent

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear).* 2017.

Our Approach for Single Slot SSA: Δ-UCB

- Allow the planner to specify a parameter Δ, the resolution to which arms must be distinguished
- $S_\Delta = \{i \in [K] : W_* - W_i < \Delta\}$
- Define Δ-regret $= \sum_{t=1}^{T} (W_* - W_{I_t}) \mathbb{1}\left[I_t \in [K] \setminus S_\Delta\right]$
- Design truthful mechanisms with logarithmic ~~regret~~ Δ-regret

- Exploration rounds
  - Explore for $\gamma = 8K\theta_{max}^2 \log T/\Delta^2$ rounds, free for all agents
  - Obtain LCB, UCB indices $\widehat{\mu}_{i,\gamma}^-, \widehat{\mu}_{i,\gamma}^+$ for every agent
- Exploitation rounds $(\gamma + 1 \ldots T)$
  - Choose agent $\hat{i}_* = \arg\max_{i \in [K]} \widehat{\mu}_{i,\gamma}^+ b_i$
  - $j = \arg\max_{i \in [K] \setminus \{\hat{i}_*\}} \widehat{\mu}_{i,\gamma}^+ b_i$ (2nd best agent)
  - $P_{\hat{i}_*} = \widehat{\mu}_{j,\gamma}^+ b_j / \widehat{\mu}_{\hat{i}_*,\gamma}^+$ (weighted VCG payment)

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

Theorem

Δ-UCB mechanism is dominant strategy incentive compatible (DSIC) truthful and individually rational (IR).

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

Properties of Δ-UCB

Theorem

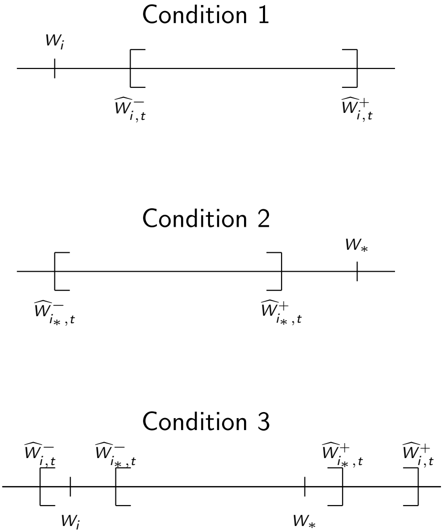
Δ-UCB mechanism is dominant strategy incentive compatible (DSIC) truthful and individually rational (IR).

Theorem

If the Δ-UCB mechanism is executed for a total time horizon of T rounds, it achieves an expected Δ-regret of $O(\log T)$.

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear).* 2017.

Proof Idea of Logarithmic Δ-regret



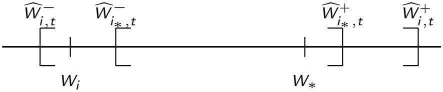
If sub-optimal arm is pulled, one of these 3 cases must occur, Bound the number of times that any of these cases occur

Proof Idea of Logarithmic Δ-regret

Condition 1

Condition 2

Condition 3



If sub-optimal arm is pulled, one of these 3 cases must occur, Bound the number of times that any of these cases occur

How?

- Occurrence of conditions 1 and 2 bounded using concentration inequalities
- Condition 3 eliminated by fixing number of exploration rounds appropriately

Extensions and Future Work

Multi-slot SSA

- Γ_m: Probability that an ad at slot m is observed
- Expected Social Welfare for agent i when allocated slot m, $W_{i,m} = \Gamma_m \mu_i \theta_i$
- $\Gamma_m, \mu_i, \theta_i$ all are unknown!

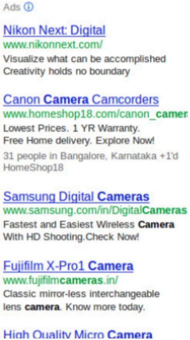
Open Problems

- Find the payments so that planner gets maximum revenue.
- General scenario where Γ is unknown and depends on the ad as well.

THANK YOU!

Extension to Multi-slot SSA

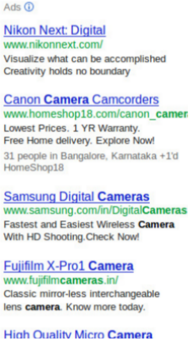
- M slots to be allocated
- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Γ_m: Probability that an ad at slot m is observed
- Expected Social Welfare for agent i when allocated slot m, $W_{i,m} = \Gamma_m \mu_i \theta_i$



Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

Extension to Multi-slot SSA

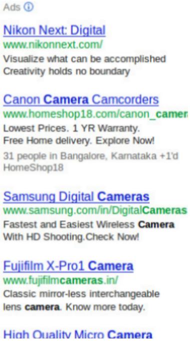
- M slots to be allocated
- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Γ_m: Probability that an ad at slot m is observed
- Expected Social Welfare for agent i when allocated slot m, $W_{i,m} = \Gamma_m \mu_i \theta_i$
- $\Gamma_m, \mu_i, \theta_i$ all are unknown!



Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

Extension to Multi-slot SSA

- M slots to be allocated
- Ad i has click through rate μ_i
- Agent i has valuation per click θ_i
- Γ_m: Probability that an ad at slot m is observed
- Expected Social Welfare for agent i when allocated slot m, $W_{i,m} = \Gamma_m \mu_i \theta_i$
- $\Gamma_m, \mu_i, \theta_i$ all are unknown!



Suppose Γ_ms were known?

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

A useful property of Γ_ms: $\Gamma_1 = 1 \geq \Gamma_2 \geq \ldots \Gamma_M$

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

A useful property of Γ_ms: $\Gamma_1 = 1 \geq \Gamma_2 \geq \ldots \Gamma_M$

Δ-UCB for multi Slot SSA

- Exploration rounds
  - Explore for $\gamma = \lceil 8K\theta^2_{max} \log T/\Delta^2 \rceil$ rounds, free for all agents
  - Obtain LCB, UCB indices $\widehat{\mu}^-_{i,\gamma}, \widehat{\mu}^+_{i,\gamma}$ for every agent

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear).* 2017.

Our Approach for Multi-slot SSA

A useful property of Γ_ms: $\Gamma_1 = 1 \geq \Gamma_2 \geq \ldots \Gamma_M$

Δ-UCB for multi Slot SSA

- Exploration rounds
  - Explore for $\gamma = \lceil 8K\theta_{max}^2 \log T/\Delta^2 \rceil$ rounds, free for all agents
  - Obtain LCB, UCB indices $\widehat{\mu}_{i,\gamma}^-, \widehat{\mu}_{i,\gamma}^+$ for every agent
- Exploitation rounds $(\gamma + 1 \ldots T)$
  - Select the best M agents as per $\arg\max_{i \in [K]} \widehat{\mu}_{i,\gamma}^+ b_i$
  - Use weighted VCG payment to decide the amount agents pay to the planner $= \left(1/\Gamma_m \mu_{\widehat{K}(m)}^+\right) \sum_{l=m+1}^{M+1} (\Gamma_{l-1} - \Gamma_l)\, \widehat{\mu}_{K^{(l)},u}^+ b_{K^{(l)}} \rho_{\widehat{K}^{(m)}}(t)$
  - Payments are a function of $\Gamma_m \ \forall m$

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear).* 2017.

Analysis of Δ-UCB (multi-slot)

$$S_{\Delta,m} = \{i \in [K] : W_{*,m} - W_{i,m} < \Delta\}\,. \tag{1}$$

$$\Delta\text{-regret} = \sum_{t=1}^{T}\sum_{m=1}^{M}(W_{*,m} - W_{I_{t,m},m})\mathbb{1}\left[I_{I_t,m} \in [K] \setminus S_{\Delta,m}\right]$$

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

Analysis of Δ-UCB (multi-slot)

$$S_{\Delta,m} = \{ i \in [K] : W_{*,m} - W_{i,m} < \Delta \} \,. \tag{1}$$

$$\Delta\text{-regret} = \sum_{t=1}^{T} \sum_{m=1}^{M} (W_{*,m} - W_{I_{t,m},m}) \mathbb{1} \left[I_{I_t,m} \in [K] \setminus S_{\Delta,m} \right]$$

Theorem

In the multi-slot setting, Δ-UCB is Dominant Strategy Incentive Compatible (DSIC) and Individually Rational (IR).

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.

Analysis of Δ-UCB (multi-slot)

$$S_{\Delta,m} = \{i \in [K] : W_{*,m} - W_{i,m} < \Delta\}\,. \tag{1}$$

$$\Delta\text{-regret} = \sum_{t=1}^{T} \sum_{m=1}^{M} (W_{*,m} - W_{I_{t,m},m}) \mathbb{1}\left[I_{I_t,m} \in [K] \setminus S_{\Delta,m}\right]$$

Theorem

In the multi-slot setting, Δ-UCB is Dominant Strategy Incentive Compatible (DSIC) and Individually Rational (IR).

Theorem

If the Δ-UCB mechanism is executed in the multiple slot scenario for a total time horizon of T rounds, it achieves an expected Δ-regret of $O(\log T)$.

Divya Padmanabhan et al. "A Dominant Strategy Truthful, Deterministic Multi-Armed Bandit Mechanism with Logarithmic Regret". In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17) (To appear)*. 2017.