Adaptive pricing, online learning and metric movement cost

Yishay Mansour

Many thanks to my co-authors:

Nicolo Cesa-Bianchi, Avrim Blum, Michal Feldman, Tomer Koren, Roi Livni, Gilles Stoltz, and Aviv Zohar

Talk outline

- Online learning:
  - Regret minimization
  - Full information
    - Best expert
  - Partial information
    - Multi-Arm Bandits

- Adaptive pricing
  - As Multi-Arm Bandits
  - Patient Buyers
- Metric Movement Cost
  - In Multi-Arm Bandits
  - New algorithms
    - Also, lower bounds

REGRET MINIMIZATION

[Blum & M] and [Cesa-Bianchi, M & Stoltz]

Regret Minimization: Setting

- ❑ Online decision making problem (single agent)
- ❑ At each time, the agent:
  - selects an action
  - observes the loss/gain
- ❑ Goal: minimize loss (or maximize gain)
- ❑ Environment model:
  - *stochastic* versus <u>*adversarial*</u>
- ❑ Performance measure:
  - *optimality* versus <u>*regret*</u>

Regret Minimization: Model

- Actions $A = \{1, \dots, N\}$
- Number time steps: $t \in \{1, \dots, T\}$
- At time step *t*:
  - The agent selects a distribution p_i^t over A
  - Environment returns costs $c_i^t \in [0,1]$
  - Online loss: $\ell^t = \sum_i c_i^t p_i^t$
  - Cumulative loss : $L_{online} = \sum_t \ell^t$
  - Regret: $L_{online} - L_{best} = L_{online} - \min_i \sum_t c_i^t$
- Information Models:
  - Full information: observes every action's cost
  - Partial information: observes only its own cost

Stochastic Costs

- Stochastic Costs:
  - for each action i,
  - c_i^t are *i.i.d.* r.v. (for diff. t)
- Full information
  - Observe $(c_1^t, \dots, c_N^t)$
- Greedy Algorithm:
  - selects the action with the lowest average cost.
  - $avg_i^t = \frac{1}{t} \sum_t c_i^t$
  - $a_t = \arg\max avg_i^t$
- Analysis sketch:
  - Two actions
  - Boolean cost (Bernoulli r.v.):
    $$\Pr[c_i = 1] = p_i$$
    $$p_2 - p_1 = \epsilon > 0$$
  - Concentration bound:
    $\Pr[avg_1^t > avg_2^t] < e^{-\epsilon^2 t}$
  - Expected regret:
  - $\epsilon E[n_2]$
    - $n_2 = \sum_t I(a_t = 2)$
  - $E[n_2] \approx \epsilon^{-2}$
  - Regret: ϵ^{-1}

Arbitrary costs

- Any hope to say anything?
- Surprising results:
  - Similar regert bounds to stochastic!
- Model:
  - Algorithm:
    - At each time selects distribution over actions
      - Mixed action
  - Adversary
    - Select loss per action
      - Can depend on the distribution!
    - Loss can be arbitrarily high!

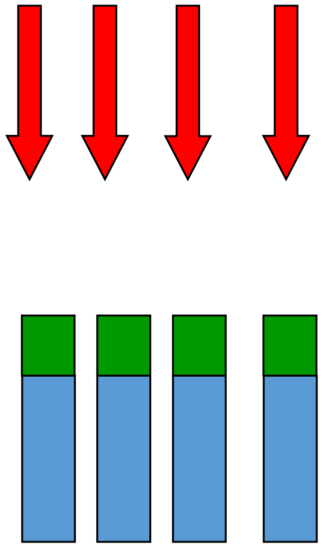
External regret

- Regret
  - $L_{online} - L_{best}$
    - If L_{best} is high,
    - L_{online} can be high
- Average regret: $(L_{online} - L_{best})/T$
  - Goal: Average external regret goes to zero
    - No regret
  - Hannan [1957]
- Explicit bounds
  - Littstone & Warmuth '94
  - CFHHSW '97
  - External regret = $O(\sqrt{T \log N})$
    - Similar to stochastic
      - $p_1 = \frac{1}{2} - \frac{1}{\sqrt{T}}$
      - $p_2 = \frac{1}{2} + \frac{1}{\sqrt{T}}$

How to do it ?

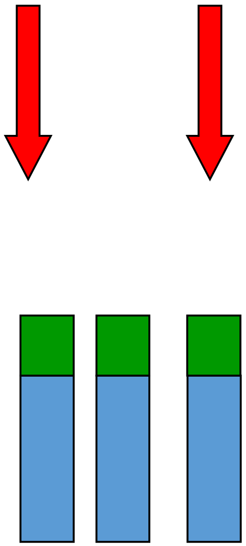
External Regret: Greedy

- Simple Greedy:
  - Go with best action so far.
- For simplicity loss is *{0,1}*
- Loss can be *N* times the best action
  - holds for any deterministic online algorithm
- Can not be worse:
  - $L_{online} < N\ L_{best}$



External Regret: Randomized Greedy

- Randomized Greedy:
  - *Random* best action.
- Loss is $\ln(N)$ times the best action
- Analysis:
  - At time time t
  - k_t best actions
  - Prob loss $\frac{1}{k_t}$
- Per increase in best loss:
  1/N + 1/(N-1) + … ≈ ln(N)

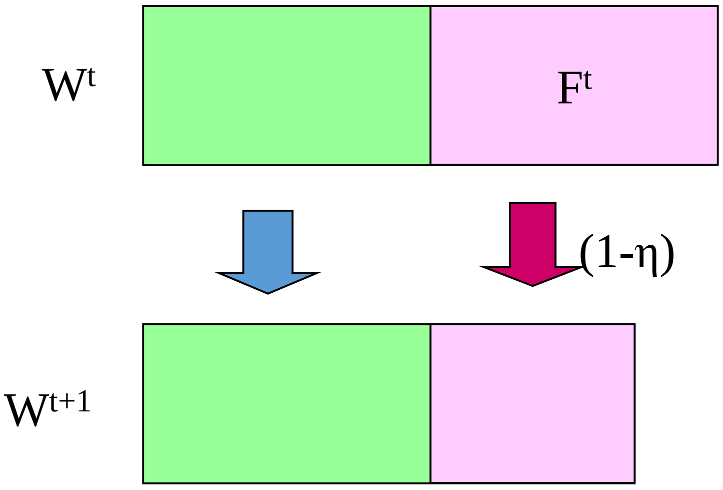


External Regret: PROD Algorithm

- Regret is $\sqrt{T log\ N}$
- PROD Algorithm:
  - plays sub-best actions
  - Uses exponential weights

    $w_i^t = (1-\eta)^{c_i^t}$

    - Normalize weights
- Analysis:
  - $W^t = \sum_i w_i^t$
  - $F^t = \sum_{i:c_i^t=1} w_i^t$
  - $W^{t+1} = W^t(1-\eta F^t)$
    - Also, expected loss: $L_{ON} = \sum F_t$



External Regret: Bounds Derivation

- **Bounding W^T**
- **Lower bound:**

  $W^T > (1-\eta)^{L_{min}}$
- **Upper bound:**

$$W^T = W^1 \Pi_t (1-\eta F^t)$$
$$\leq W^1 \Pi_t \exp\{-\eta F^t\}$$
$$= W^1 \exp\{-\eta L_{ON}\}$$

using $1-x \leq e^{-x}$

- **Combined bound:**

  $(1-\eta)^{L_{min}} \leq W^1 \exp\{-\eta L_{ON}\}$
- **Taking logarithms:**

$L_{min} \log(1-\eta) \leq \log(W^1) - \eta L_{ON}$

- **Final bound:**

$L_{ON} \leq L_{min} + \eta L_{min} + \log(N)/\eta$

- **Optimizing the bound:**

$$\eta = \sqrt{\log N / L_{min}}$$
$$L_{ON} \leq Lmin + 2\sqrt{L_{min} \log N}$$

External Regret: Summary

- How surprising are the results …
  - Near optimal result in online adversarial setting
    - very rear …
  - Lower bound: stochastic model
    - stochastic assumption does not help …
  - Models an "improved" greedy
    - Smoothed maximum
  - An "automatic" optimization methodology
    - Find the best fixed setting of parameters

External Regret and classification

- Connections to Machine Learning:
- *H* – the hypothesis class
- *cost* – an abstract loss function
  - no need to specify in advance
- Learning setting – online
  - learner: observes point, predicts, observes loss
- Regret guarantee:
  - compares to the best classifier in *H*.
  - Given the sequence of inputs

Partial Information

Multi-Arm Bandits

Partial Information

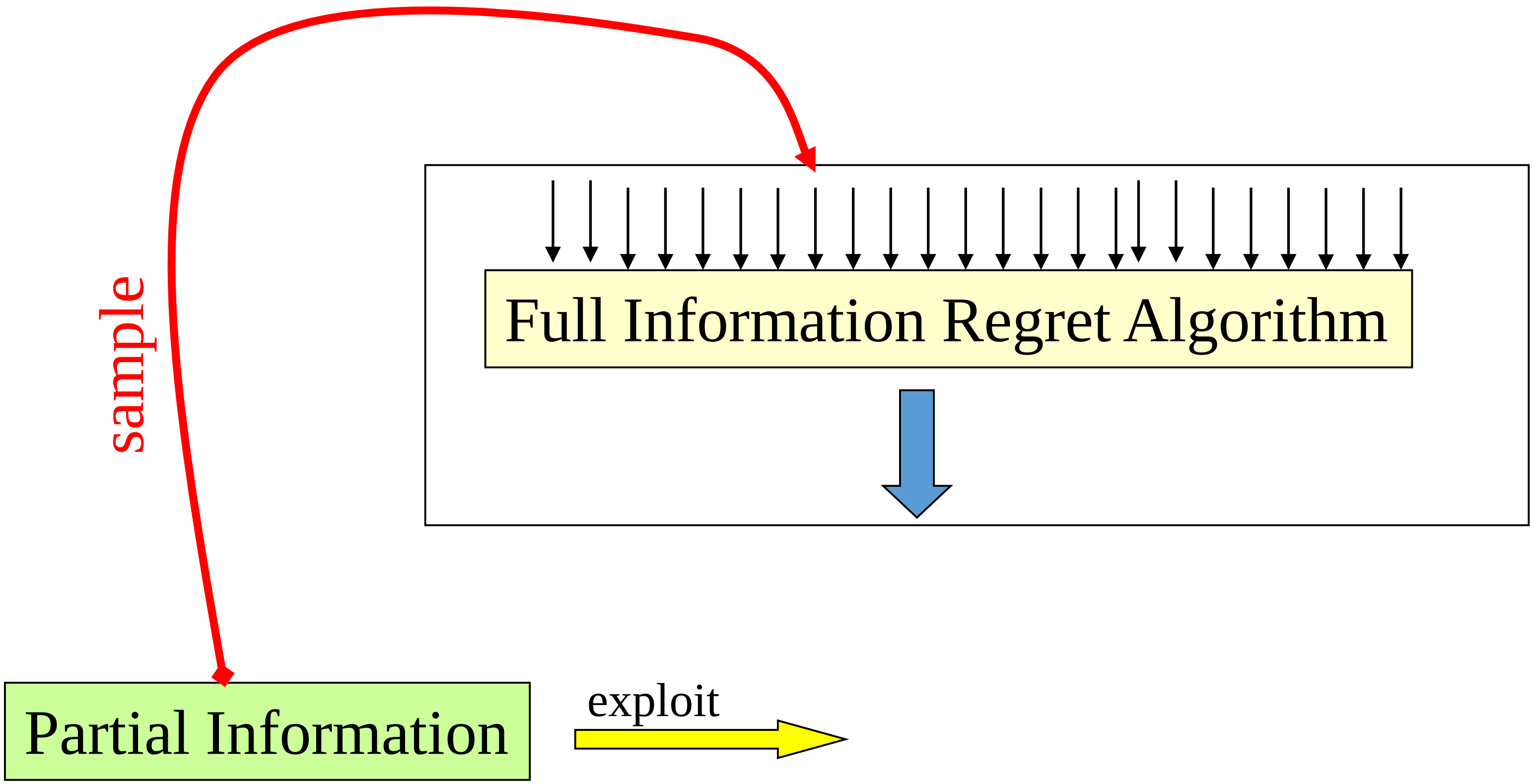
- Partial information (Multi-Arm Bandits):
  - Agent selects action i
  - Observes the loss of action i
  - No information regarding the loss of other actions

- How can we handle this case?

Partial Information

- Simple reduction to Full Info
  - Work in blocks of size *B*
  - explore each action once in each block
    - Random positions
  - Otherwise uses Full Info action distribution
  - At the end of a block:
    - Feeds the explored actions to Full Info
- Regret:
  - Regret of Full Info on *T/B* time steps
    - each of magnitude in [0,*B*]
  - Exploration regret N*T/B*

Information model: Full versus Partial



Information: Full versus Partial

- Analysis:
  - Regret of FI on *T/B* time steps (each of size B)
    - Exploitation Regret ~ $\sqrt{B\ T}$
    - Exploration Regret N in block
    - Number of blocks T/B
- Optimizing: Set $B = N^{2/3}T^{1/3}$
- Regret guarantee: $N^{1/3}T^{2/3}$
- Benefit:
  - Vanishing regret
  - Non-optimal regret bound

Information: Full versus Partial

- Importance Sampling:
  - maintain weights as before.
  - update the selected action *k* by loss c_k^t/p_k^t
  - Expectation is maintain
  - Need to argue directly on the algorithm.

- Used in: [ACFS] and others
  - Regret Bound about $\sqrt{TN}$

More elaborate regret notions

- Time selection functions [Blum & M]
  - determines the relevance of the next time step
  - identical for all actions
  - multiple time-selection functions
- Wide range regret [Lehrer, Blum & M]
  - Any set of modification functions
    - mapping histories to actions
- Many more information models:
  - Graph Observability
  - Delayed feedback

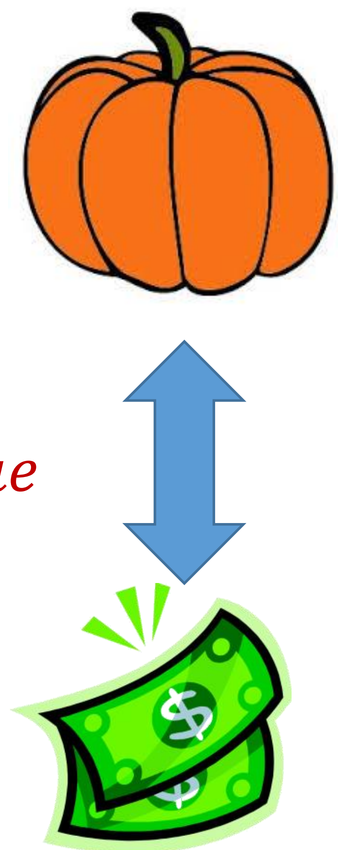
Adaptive Pricing

Pricing a single item: Classical Model

- Single seller
  - Single item
  - Unlimited supply
- Stream of T buyers
  - Buyer t has value v_t
- At time t:
  - Seller offers price p_t
  - Buyer buys if $v_t \geq p_t$
    - if buys, then revenue p_t

Pricing a single item: Classic Model

- **Revenue**
  - $OnlineRevnue = \sum_t I(v_t \geq p_t)p_t$
- **Regret**
  - Compare to the best fixed price
  - $Revenue(p) = \sum_t I(v_t \geq p)p$
  - $Regert = \max_p Revenue(p) \ - OnlineRevenue$
- **Seller Objective**
  - Maximize Revenue
  - Minimize Regret



Pricing and Multi-Arm Bandits

- Finite Set of action A
- At time t:
  - Select action $a_t \, \varepsilon \, A$
  - Observe gain $g_t[a_t] \, \varepsilon \, [0,1]$
- Cumulative Gain:
- $OnlineGain = \sum_t g_t[a_t]$
- $Gain(a) = \sum_t g_t[a]$
- Regret:

$$Regret = \max_a Gain(a) - OnlineGain$$

A = Discrete Prices

Gain = $I(v_t \geq p_t) \, p_t$

Online Revenue

Multi-Arm Bandit: Recall

- Choose the best action until now
  - With a “soft-max”
- Maintain a distribution p_t over actions
  - Change distribution slowly
  - Concentrate on the high gains
- Full information
  - Exponential weights
  - Regret $\sqrt{T \log K}$
- Partial information
- Estimating the gain
  - Importance sampling:
    - $g_t[a_t]/p_t[a_t]$
  - Unbiased estimator
  - Bound second moment
    - Lower bound probabilities
- Regret: $\sqrt{T\ K}$

Multi-Arm Bandit and Pricing

- Has a history:
- A Two-Armed Bandit Theory of Market Pricing
  - ROTHSCHILD, 1974
- Kleinberg and Leighton:
  - Use discrete prices
  - Regret $= T^{2/3}$
  - Upper and lower bound

- Why not Regret $= T^{1/2}$?
- Discretization
  - number of prices $T^{1/3}$
    - Prices = actions
- Additional loss
  - Discretization size $T^{-1/3}$
- Regret $\sqrt{KT} + \epsilon T$
  - $K = T^{1/3}$; $\epsilon = T^{-1/3}$

Patient buyers

- Procrastination is the hallmark of human nature
  - And it even has good effects
- Modeling:
  - Buyers are not: “buy-it or leave-it”
  - Allow buyers laxity over time
  - Trying to buy at the best price
- Strategic issues:
  - Seller:
    - need to plan for strategic buyers
  - Buyers: Need to anticipate seller
    - Indirectly other buyers

Patient buyers

- ❑ Our Model:
- ❑ Each buyer:
  - has a (small) time window
  - Buys at the best price in window
- ❑ Seller
  - Publishes prices in advance
    - ➢ For the maximum window size
- ❑ Buyer strategy:
  - Buy at the lowest price in its window
    - ➢ If below its value.
- ❑ Seller Strategy ?!

Challenges for the seller

- Changing prices:
  - Increasing price: No problem
  - Decreasing price: might lose revenue
- MAB with switching cost:
  - Pay 1 each time you change an action
  - Benchmark (by definition) does not change action

- Lower bound:
  - MAB with switching cost
    - [Dekel et al]
  - Regret $= \Theta(k^{1/3}\, T^{2/3})$
    - k actions, T time steps

Lower bound on the regret

- Reduction to switching cost:
- Simple case:
  - Three valuations {0, ½ ,1}
  - Window size 2
- Merge with random buyers:
  - value ½ and window=1
  - value 1 and window=2
- Each price reduction
  - With prob. ¼ Loses ½
  - Otherwise identical

- Still need to take care of the feedback.
  - Prices feedback is richer

Theorem (lower bound):

For patient buyers the seller has regret at least

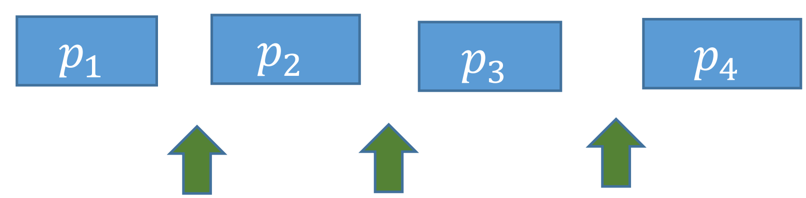
$$\Omega(T^{2/3})$$

Simple Block MAB Algorithm

- ❑ Partition time to blocks of size B
  - ○ T/B blocks, k prices
- ❑ Fix the price in each block
  - ○ Switching only between blocks
- ❑ Standard regret bound
  - ○ Inside: $\sqrt{BkT}$
  - ○ Between: T/B
- ▪ Optimize block size
- ▪ $B = \left({}^{T}/_{k}\right)^{1/3}$
  - ▪ Regret $k^{1/3}T^{2/3}$

- ❑ Optimizing over continuous prices
  - ○ Discretization regret T/k
  - ○ Number of prices
    - ➢ $k = T^{1/4}$
  - ○ Total Regret $T^{3/4}$

$B\sqrt{k(T/B)}$ regret inside blocks



T/B regret when switching blocks

Improved MAB: metric space

- Where are we losing:
  - Discrete prices
  - switching cost
    - Each has regret $T^{2/3}$
- Together the regret is higher
  - $T^{3/4}$
- Can we do a better?
- Observation:
  - Price change from p_1 to p_2
  - Loss is at most $|p_1 - p_2|$

- Metric over actions (prices):
  - Each action i has a price p_i
  - Switching from p_i to p_j has cost

    $|p_i - p_j|$
  - A simple line metric over the actions.
- Benchmark:
  - Best static price has no movement cost!

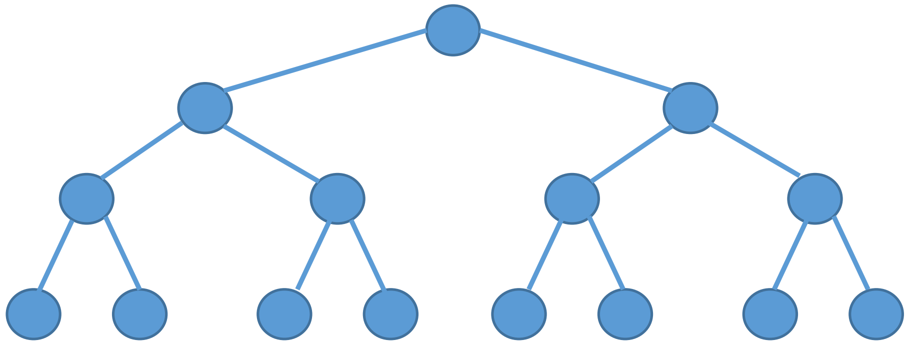
Bounding the switching effect

- What happens if we ran a "standard" MAB
  - Many switches
- Goal:
  - Switch often to similar prices
  - Switch rarely to far prices
- Has also an intuitive appeal

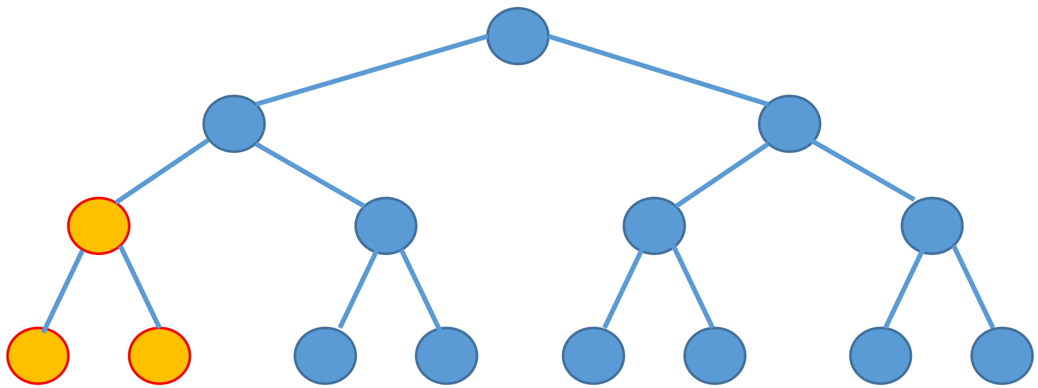
- Basic idea:
  - Change $\geq 2^d$ with prob $\leq 2^{-d}$
  - Fix the prob of the change
- Look at the expectation
  - Compensate for the slow changes
  - Allow big changes in distribution

Tree Metric

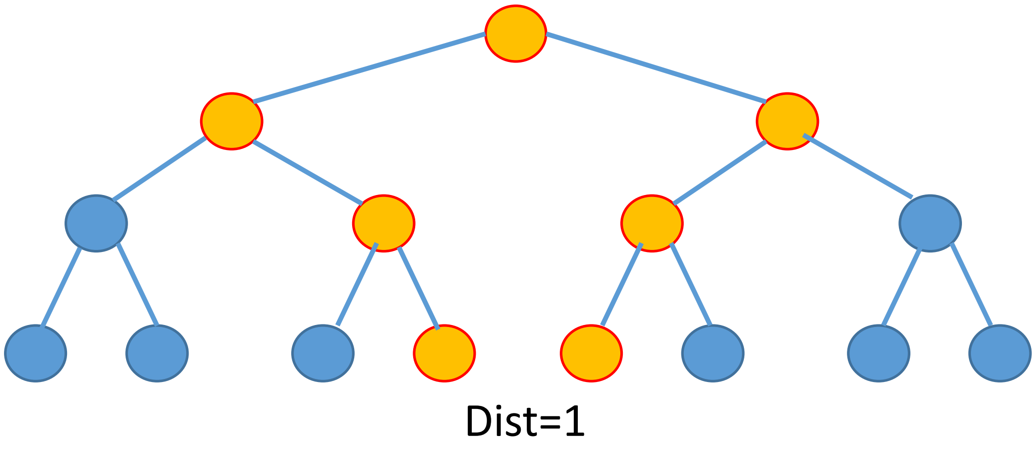
- The leafs are labeled by numbers in [0,1]
  - Equally spaced
- Distance between leaves:
  - (Size of subtree of LCA)/K
  - Upper bounds the real distance
    - Very loose upper bound
- Note: Benchmark does not move
  - Just need an upper bound



Tree Metric

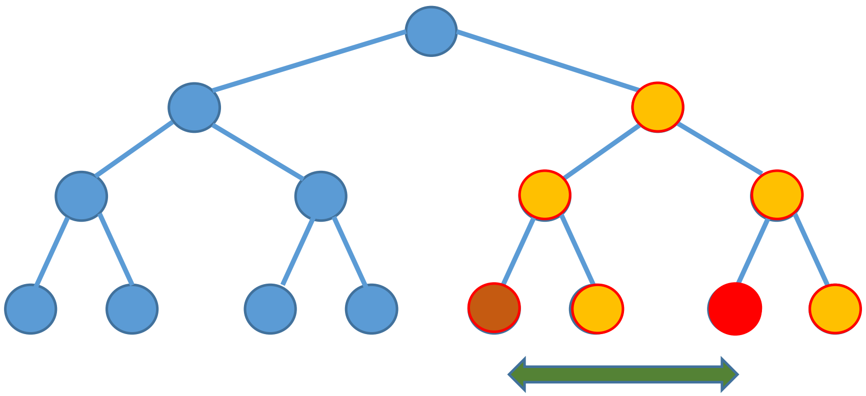


Tree Metric



Lazy sampling

- Lazy sampling
  - Given previous action (price)
  - Select a random subtree
    - That includes it
    - Geometric dist
  - Sample only actions in that subtree
- Movement
  - Move $2^i/K$ with prob 2^{-i}
  - Expected movement $(\log K)/K$
- Need to take care of quality!



Analyzing the sampling

- For a static distribution
  - OK, in expectation
- Our case:
  - Dynamic changing distribution
- Basic idea:
  - Rebalance the subtree
  - Maintain ratios across subtrees

- Analysis:
  - Biased estimator

  - Show that for any subtree:

$$E\left[\frac{I\{i_t \in A_s\}}{p_t(A_s)}\right] = 1$$

  - Enough for the regret analysis to go through!

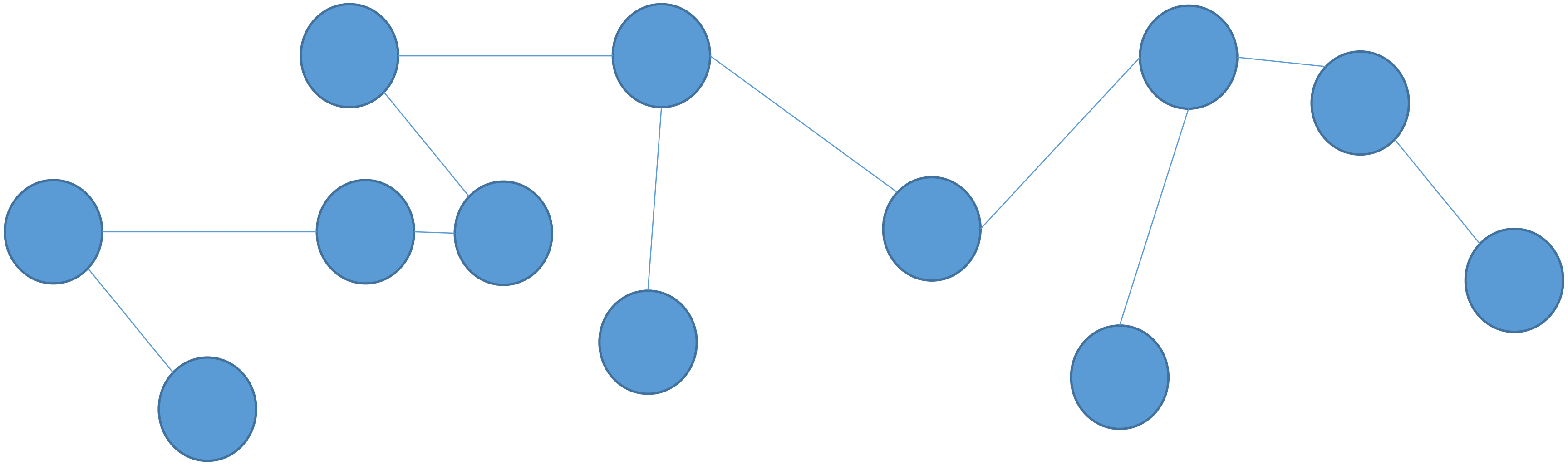
Results for patient buyers

- Upper bound: $\Theta\left(\max(T^{2/3}, \sqrt{kT})\right)$
- Discretizing prices
  - Additional regret T/k
- Optimizing for discretization
  - $k = T^{1/3}$

- Lower Bound:
- 2 prices + patient buyers
  - $\tilde{\Omega}\left(T^{2/3}\right)$
- Regular buyers, continuous price
  - Kleinberg & Leighton $\Omega\left(T^{2/3}\right)$

What about general Metric ???

$$MRegret = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(x_t) - \min_{x^* \in A} \sum_{t=1}^{T} \ell_t(x^*) + dist(x_t, x_{t-1})\right]$$



Moving to general metric spaces

Lower bound

- Packing number $N_p(\epsilon)$
- Lower bound: $\Omega(\epsilon^{\frac{1}{3}} N_p(\epsilon)^{\frac{1}{3}} T^{\frac{2}{3}})$
- Let $P = \sup_{\epsilon>0} \epsilon \cdot N_p(\epsilon)$
- Lower bound: $\Omega(P^{1/3} T^{2/3})$

Upper bound

- Covering number $N_c(\epsilon)$
  - Bound using HST
  - Let $C = \sup_{\epsilon>0} \epsilon \cdot N_c(\epsilon)$
  - Run Slowly-Moving-Bandit
- Upper bound: $\tilde{O}(\max(\sqrt{KT}, C^{1/3} T^{2/3}))$

Non discrete metric spaces: Minkowski dimension $O(T^{\frac{d}{d+1}})$

Concluding remarks: Patient Buyers and MAB

- Patient buyers
  - More realistic buyer model
- Fixed window
  - Discounted utility?
- Metric MAB
  - Competitive analysis and regret minimization
- Other Applications
  - other online problems?
  - Losses correlated over time