The K-armed dueling bandits problem

Yisong Yue, Josef Broder, Robert Kleinberg, Thorsten Joachims

Journal of Computer and System Sciences **78** (2012) 1538–1556.

Speaker: Joseph Chuang-Chieh Lin

Institute of Information Science
Academia Sinica
Taiwan

22 July 2016




Yisong Yue @Caltech

Josef Broder @Google

Robert Kleinberg @Cornell

Thorsten Joachims @Cornell

Outline

Motivations

- The conventional bandit problem :
  - Choose, in each of T iterations, one of the K possible bandits/arms/strategies $\mathcal{B} = \{b_1, \ldots, b_K\}$.
  - Receive the payoff in $[0, 1]$ (*initially unkown*) in each iteration.
  - **Goal:** Maximize the total payoff.
- It's difficult to elicit absolute-scale payoffs in some applications.
  - One can only rely on *relative* judgment of payoff.
- Given a collection of K bandits, we wish to find a sequence of noisy comparisons that has low regret.

Motivations

- The conventional bandit problem :
  - Choose, in each of T iterations, one of the K possible bandits/arms/strategies $\mathcal{B} = \{b_1, \ldots, b_K\}$.
  - Receive the payoff in $[0, 1]$ (*initially unkown*) in each iteration.
  - **Goal:** Maximize the total payoff.
- It's difficult to elicit absolute-scale payoffs in some applications.
  - One can only rely on *relative* judgment of payoff.
- Given a collection of K bandits, we wish to find a sequence of noisy comparisons that has low regret.

Motivations

- The conventional bandit problem :
  - Choose, in each of T iterations, one of the K possible bandits/arms/strategies $\mathcal{B} = \{b_1, \ldots, b_K\}$.
  - Receive the payoff in $[0, 1]$ (*initially unkown*) in each iteration.
  - **Goal:** Maximize the total payoff.
- It's difficult to elicit absolute-scale payoffs in some applications.
  - One can only rely on *relative* judgment of payoff.
- Given a collection of K bandits, we wish to find a sequence of noisy comparisons that has low regret.

Noisy comparisons

- $\Pr[b \succ b'] := \epsilon(b, b') + 1/2$.
  - $\epsilon(b, b') \in (-1/2, 1/2)$: a measure distinguishing b and b'.
    - $\epsilon(b, b') = -\epsilon(b', b)$
    - $\epsilon_{i,j} \equiv \epsilon(b_i, b_j)$.
  - $b \succ b' \Rightarrow \epsilon(b, b') > 0$.

⋆ The noisy comparisons are independent and $\Pr[b \succ b']$ is stationary over time.

Regrets

- $(b_1^{(t)}, b_2^{(t)})$: the bandits chosen at iteration t.
- b^*: the overall best bandit.
- T be time horizon.

Regrets

- The strong regret

$$R_T = \sum_{t=1}^{T} \max\{\epsilon(b^*, b_1^{(t)}), \epsilon(b^*, b_2^{(t)})\}.$$

- The weak regret

$$\tilde{R}_T = \sum_{t=1}^{T} \min\{\epsilon(b^*, b_1^{(t)}), \epsilon(b^*, b_2^{(t)})\}.$$

Modeling assumptions

Strong stochastic transitivity

For bandits $b_i \succ b_j \succ b_k$,

$$\epsilon_{i,k} \geq \max\{\epsilon_{i,j}, \epsilon_{j,k}\}.$$

Strong triangular inequality

For bandits $b_i \succ b_j \succ b_k$,

$$\epsilon_{i,k} \leq \epsilon_{i,j} + \epsilon_{j,k}.$$

The Algorithm

Explore then exploit

Algorithm 1 Explore then exploit

Input: T, $\mathcal{B} = \{b_1, \ldots, b_K\}$, *EXPLORE*
$(\hat{b}, \hat{T}) \leftarrow EXPLORE(T, \mathcal{B})$
for $t = \hat{T} + 1, \ldots, T$ **do**
  compare $\hat{b}$ and $\hat{b}$
end for

Algorithm 2 Interleaved Filter (IF).

Input: T, $\mathcal{B} = \{b_1, \ldots, b_K\}$
$\delta \leftarrow 1/(TK^2)$
Choose $\hat{b} \in \mathcal{B}$ randomly
$W \leftarrow \{b_1, \ldots, b_K\} \setminus \{\hat{b}\}$
$\forall b \in W$, maintain estimate $\hat{P}_{\hat{b},b}$ of $P(\hat{b} > b)$ according to (6)
$\forall b \in W$, maintain $1 - \delta$ confidence interval $\hat{C}_{\hat{b},b}$ of $\hat{P}_{\hat{b},b}$ according to (7), (8)
while $W \neq \emptyset$ **do**
  for $b \in W$ **do**
    compare $\hat{b}$ and b
    update $\hat{P}_{\hat{b},b}$, $\hat{C}_{\hat{b},b}$
  end for
  while $\exists b \in W$ s.t. $(\hat{P}_{\hat{b},b} > 1/2 \wedge 1/2 \notin \hat{C}_{\hat{b},b})$ **do**
    $W \leftarrow W \setminus \{b\}$ *//$\hat{b}$ declared winner against b*
  end while
  if $\exists b' \in W$ s.t. $(\hat{P}_{\hat{b},b'} < 1/2 \wedge 1/2 \notin \hat{C}_{\hat{b},b'})$ **then**
    while $\exists b \in W$ s.t. $\hat{P}_{\hat{b},b} > 1/2$ **do**
      $W \leftarrow W \setminus \{b\}$ *//pruning*
    end while
    $\hat{b} \leftarrow b'$, $W \leftarrow W \setminus \{b'\}$ *//b' declared winner against $\hat{b}$ (new round)*
    $\forall b \in W$, reset $\hat{P}_{\hat{b},b}$ and $\hat{C}_{\hat{b},b}$
  end if
end while
$\hat{T} \leftarrow$ Total Comparisons Made
return $(\hat{b}, \hat{T})$

The exploit algorithm

Algorithm 2 Interleaved Filter (IF).

```
Input: T, 𝓑 = {b_1, ..., b_K}
δ ← 1/(TK^2)
Choose b̂ ∈ 𝓑 randomly
W ← {b_1, ..., b_K} \ {b̂}
∀b ∈ W, maintain estimate P̂_{b̂,b} of P(b̂ > b) according to (6)
∀b ∈ W, maintain 1 − δ confidence interval Ĉ_{b̂,b} of P̂_{b̂,b} according to (7), (8)
while W ≠ ∅ do
    for b ∈ W do
        compare b̂ and b
        update P̂_{b̂,b}, Ĉ_{b̂,b}
    end for
    while ∃b ∈ W s.t. (P̂_{b̂,b} > 1/2 ∧ 1/2 ∉ Ĉ_{b̂,b}) do
        W ← W \ {b}    //b̂ declared winner against b
    end while
    if ∃b' ∈ W s.t. (P̂_{b̂,b'} < 1/2 ∧ 1/2 ∉ Ĉ_{b̂,b'}) then
        while ∃b ∈ W s.t. P̂_{b̂,b} > 1/2 do
            W ← W \ {b}    //pruning
        end while
        b̂ ← b',  W ← W \ {b'}    //b' declared winner against b̂ (new round)
        ∀b ∈ W, reset P̂_{b̂,b} and Ĉ_{b̂,b}
    end if
end while
T̂ ← Total Comparisons Made
return (b̂, T̂)
```

- $\hat{P}_{i,j} = \frac{\#\ b_i \text{ wins}}{\#\ \text{comparisons}}$.
  The empirical estimate of $\Pr[b_i \succ b_j]$ after t comparisons.
- Confidence interval:
  $\hat{C}_{i,j} := (\hat{P}_{i,j} - c_t, \hat{P}_{i,j} + c_t)$,
  where $c_t = \sqrt{4\log(1/\delta)/t}$.

Contribution of this paper

Theorem 1

Running Algorithm 1 with $\mathcal{B} = \{b_1, \dots, b_K\}$, time horizon T ($T \geq K$), then **IF** incurs expected regret (weak & strong) bounded by

$$\mathbf{E}[R_T] = O(\mathbf{E}[R_T^{IF}]) = O\left(\frac{K}{\epsilon_{1,2}} \log T\right).$$

Theorem 2

For any fixed $\epsilon > 0$ and any algorithm ϕ for the K-armed dueling bandit problem, there exists a problem instance such that

$$R_T^{\phi} = \Omega\left(\frac{K}{\epsilon} \log T\right),$$

where $\epsilon = \min_{b \neq b^*} \Pr[b^* \succ b]$.

Crucial lemmas

Lemma 1

The probability that **IF** makes a mistake resulting in the elimination of the best bandit b_1 is $\leq 1/T$.

- $\mathbf{E}[R_T] \leq (1 - 1/T)\mathbf{E}[R_T^{IF}] + (1/T) \cdot O(T) = O(\mathbf{E}[R_T^{IF}])$.
  - R_T^{IF}: the regret incurred from **IF**.

Crucial lemmas (contd.)

Lemma 2

Assuming **IF** is mistake-free, then with high probability,

$$R_T^{IF} = O\left(\frac{K \log K}{\epsilon_{1,2}} \log T\right)$$

for both weak and strong regret.

Lemma 3

Assuming **IF** is mistake-free, then

$$\mathbf{E}[R_T^{IF}] = O\left(\frac{K}{\epsilon_{1,2}} \log T\right)$$

for both weak and strong regret.

Some more terminologies

- **IF** makes a "mistake": it draws a false conclusion regarding a bandit pair.
- A "match": all the comparisons **IF** makes between two bandits.
- A "round": all the matches played by the *incumbent* bandit $\hat{b}$.

The Main Analysis

Justification of the confidence intervals

Lemma 4

- For $\delta = 1/(TK^2)$, the number of comparisons in a match b/w b_i, b_j is

$$O\left(\frac{1}{\epsilon_{i,j}^2}\log(TK)\right).$$

- Pr[an inferior bandit is declared the winner at some time $t \leq T$] $\leq \delta$.

- **IF** makes a mistake at time $t \Rightarrow 1/2 + \epsilon_{i,j} \notin \hat{C}_{i,j}$.
- Note: $\mathbf{E}[\hat{P}_{i,j}] = 1/2 + \epsilon_{i,j}$.
- $\Pr[1/2 + \epsilon_{i,j} \notin \hat{C}_{i,j}] = \Pr[|\hat{P}_{i,j} - \mathbf{E}[\hat{P}_{i,j}]| \geq c_t] \leq 2 \cdot e^{-2t \cdot c_t^2} = 2/(T^8K^{16})$.
- $$\Pr\left[\bigcup_{t=1}^{T}\{1/2 + \epsilon_{i,j} \notin \hat{C}_{i,j}\}\right] \leq \frac{2T}{T^8K^{16}} \leq \frac{1}{TK^2} = \delta.$$

Justification of the confidence intervals

Lemma 4

- For $\delta = 1/(TK^2)$, the number of comparisons in a match b/w b_i, b_j is

$$O\left(\frac{1}{\epsilon_{i,j}^2}\log(TK)\right).$$

- Pr[an inferior bandit is declared the winner at some time $t \le T$] $\le \delta$.

- **IF** makes a mistake at time $t \quad \Rightarrow \quad 1/2 + \epsilon_{i,j} \notin \hat{C}_{i,j}$.
- Note: $\mathbf{E}[\hat{P}_{i,j}] = 1/2 + \epsilon_{i,j}$.
- $\Pr[1/2 + \epsilon_{i,j} \notin \hat{C}_{i,j}] = \Pr[|\hat{P}_{i,j} - \mathbf{E}[\hat{P}_{i,j}]| \ge c_t] \le 2 \cdot e^{-2t \cdot c_t^2} = 2/(T^8K^{16})$.
-

$$\Pr\left[\bigcup_{t=1}^{T}\{1/2 + \epsilon_{i,j} \notin \hat{C}_{i,j}\}\right] \le \frac{2T}{T^8K^{16}} \le \frac{1}{TK^2} = \delta.$$

Proof of Lemma 4 (contd.)

- By the stopping condition of **IF**, the match terminates at any time t if $\hat{P}_{i,j} - c_t > 1/2$.
  - If $n > t$, then $\hat{P}_{i,j} - c_t \leq 1/2$.
- $\Pr[n > t] \leq \Pr[\hat{P}_{i,j} - c_t \leq 1/2] = \Pr[\hat{P}_{i,j} - 1/2 - \epsilon_{i,j} \leq c_t - \epsilon_{i,j}] = \Pr[\mathbf{E}[\hat{P}_{i,j}] - \hat{P}_{i,j} \geq \epsilon_{i,j} - c_t]$.
- Set $m \geq 8$ and $t \geq \lceil 2m \log(TK^2)/\epsilon_{i,J}^2 \rceil$ (then $c_t \leq \epsilon_{i,j}/2$), we will have

$$\Pr\left(n \geq \frac{m}{\epsilon_{i,j}^2} \log(TK)\right) \leq \frac{1}{(TK)^m}.$$

Regret per match

Lemma 5

Assume that b_1 has not been removed and $T \geq K$, then w.h.p. the accumulated weak/strong regret **from any match** is

$$O\left(\frac{1}{\epsilon_{1,2}} \log T\right).$$

- Suppose $\hat{b} = b_j$ is playing a match against b_i.
- By Lemma 4, any match played by b_j contains at most

$$O\left(\frac{1}{\epsilon_{1,j}^2} \log(TK)\right) = O\left(\frac{1}{\epsilon_{1,2}^2} \log(TK)\right) \text{ comparisons.}$$

  - **Note:** All matches within a round are played *simultaneously*.

Proof of Lemma 5 (contd.)

- The accumulated weak regret is bounded by

$$\epsilon_{1,j} \cdot O\left(\frac{1}{\epsilon_{1,j}^2}\log(TK)\right) = O\left(\frac{1}{\epsilon_{1,j}}\log(TK)\right)$$
$$= O\left(\frac{1}{\epsilon_{1,2}}\log(T)\right).$$

Mistake bound

- **IF** eliminates the best bandit b_1 if
  - an inferior bandit defeats b_1, or
  - b_1 is removed during the pruning step (lines 16–18).
- Consider the second case.

Lemma 6

For all triples of bandits $b, b', \hat{b}$ such that $b \succ b'$, the probability that **IF** eliminates b in a pruning step, where

- b' wins a match against $\hat{b}$ while
- b is empirically inferior to $\hat{b}$,

is $\leq \delta$.

Proof of Lemma 6

- $X_1, X_2, \ldots$: an infinite sequence of i.i.d. Bernoulli random variables with $\mathbf{E}[X_i] = \Pr[\hat{b} \succ b']$.
- $Y_1, Y_2, \ldots$: an infinite sequence of i.i.d. Bernoulli random variables with $\mathbf{E}[Y_i] = \Pr[\hat{b} \succ b]$.
  - X_i (resp. Y_i) represents the outcome of the ith comparison b/w $\hat{b}$ & b' (resp. $\hat{b}$ & b).
- If b is eliminated in a pruning step at the end of a match consisting of n comparisons b/w b' and $\hat{b}$, then

$$X_1 + \ldots + X_n < n/2 - \sqrt{4n\log(1/\delta)},$$

$$Y_1 + \ldots + Y_n > n/2.$$

Proof of Lemma 6 (contd.)

- Define $Z_i = Y_i - X_i$, we have

$$Z_1 + \ldots + Z_n > \sqrt{4n \log(1/\delta)}.$$

  - $(Z_i)_{i=1}^{\infty}$ are i.i.d., and $|Z_i| \leq 1$, $\forall i$.

⋆ $\mathbf{E}[Z_i] = \Pr[\hat{b} \succ b] - \Pr[\hat{b} \succ b'] \leq 0$.

- Taking Hoeffding's inequality & union bound...

Proof of Lemma 1

Lemma 1

The probability that **IF** makes a mistake resulting in the elimination of the best bandit b_1 is $\leq 1/T$.

- For every i, the probability that b_1 is eliminated in a match against b_i is $\leq \delta$ (Lemma 4).
- For all i, j, the probability that b_1 is eliminated in a pruning step resulting from a match where b_i defeats b_j is $\leq \delta$ (Lemma 6).

⋆ The probability that **IF** makes a mistake resulting in eliminating b_1 is $\leq \delta(K-1) + \delta(K-1)^2 < \delta K^2 = 1/T$.

The regret upper bound

Lemma 2

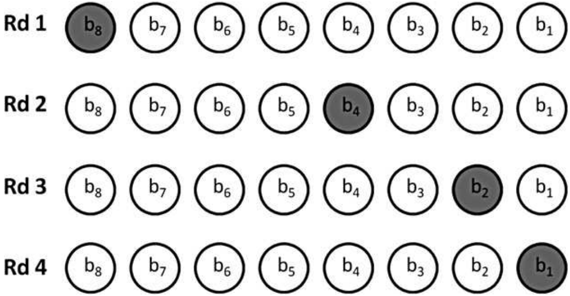
Assuming **IF** is mistake-free, then with high probability,

$$R_T^{IF} = O\left(\frac{K \log K}{\epsilon_{1,2}} \log T\right)$$

for both weak and strong regret.

- We wish to prove that the number of candidate bandits (i.e., # rounds) is $O(\log K)$ w.h.p.
- Model the sequence of candidate bandits as a random walk.

Random walk model



- p_i: the prob. b_i will be the incumbent in the following round.
  - $p_{j-1} \leq \ldots \leq p_1$ ($\because$ strong stochastic transitivity).
- The "worst case": $p_{j-1} = \ldots = p_1 = 1/(j-1)$ (assuming no mistakes are made).

Random Walk Model (contd.)

Random Walk Model

Define a random walk graph with K nodes labeled $b_1, \ldots, b_K$. Each node b_j ($j > 1$) transitions to b_i for $j > i \geq 1$ with prob. $1/(j-1)$ (uniform). The final node b_1 is an absorbing node.

Proposition 1

If S and $\tilde{S}$ are random variables corresponding to the number of rounds in **IF** and the Random Walk Model, resp., then

$$\forall x: \quad \Pr[S \geq x] \leq \Pr[\tilde{S} \geq x].$$

Analysis of the Random Walk Model

Lemma 7

Let X_i $(1 \le i \le K)$ be an indicator random variable corresponding to whether a random walk starting at b_K visits b_i in the Random Walk Model. Then

$$\Pr[X_i = 1] = \frac{1}{i},$$

and for all $W \subseteq \{X_1, \dots, X_{K-1}\}$,

$$\Pr[\wedge_{i \in W} X_i] = \prod_{X_i \in W} \Pr[X_i].$$

- We can express the number of steps taken by a random walk from b_K to b_1 as $S_k = 1 + \sum_{i=1}^{K-1} X_i$. Then,

$$\mathbf{E}[S_k] = 1 + \sum_{i=1}^{K-1} \mathbf{E}[X_i] = 1 + H_{K-1} \approx \log K.$$

Analysis of the Random Walk Model (contd.)

Lemma 8

Assuming **IF** is mistake-free, then it runs for $O(\log K)$ rounds w.h.p..

Corollary 1

Assuming **IF** is mistake-free, then it plays $O(K \log K)$ matches w.h.p.

- $O(\log T/\epsilon_{1,2})$ accumulated regret per match (Lemma 5).

▷ Lemma 2 (i.e., $R_T^{IF} = O((K \log K) \log T/\epsilon_{1,2})$) follows.

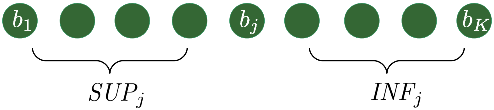
Expected Regret Upper Bound

Expected regret upper bound

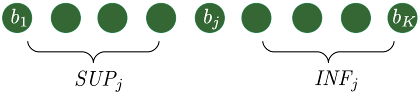
Lemma 9

Assuming **IF** is mistake-free, then it plays $O(K)$ matches *in expectation*.

- B_j: # matches played by b_j when it is NOT the incumbent.
- $B_j = INF_j + SUP_j$, where
  - INF_j: # matches played by b_j against b_i for $i > j$.
  - SUP_j: # matches played by b_j against b_i for $i < j$.
- Then $\sum_{j=1}^{K} \mathbf{E}[B_j] = \sum_{j=1}^{K} (\mathbf{E}[INF_j] + \mathbf{E}[SUP_j])$.

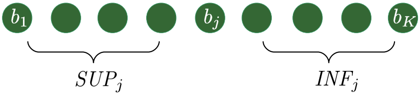


Proof of Lemma 9 (contd.)



- $$\mathbf{E}[INF_j] \leq 1 + \sum_{i=j+1}^{K-1} \frac{1}{i} = 1 + H_{K-1} - H_i.$$

Proof of Lemma 9 (contd.)



- Assume that b_j does NOT lose a match (not to be eliminated) to any superior incumbent b_i before b_i is defeated unless $b_i = b_1$.
- $\mathcal{E}_{j,t}$: b_j is pruned after the t-th round where the incumbent bandit is superior to b_j, conditioned on NOT being pruned in the first $t-1$ such rounds.
- $G_{j,t}$: # matches beyond the first $t-1$ played by b_j against a superior incumbent, conditioned on playing $\geq t-1$ such matches.
- $\mathbf{E}[G_{j,t}] = 1 + \Pr[\mathcal{E}^c_{j,t}] \cdot \mathbf{E}[G_{j,t+1}]$.

⋆ $\mathbf{E}[SUP_j] \leq \mathbf{E}[G_{j,1}] = 1 + \Pr[\mathcal{E}^c_{j,1}] \cdot \mathbf{E}[G_{j,2}] \leq 1 + 1/2 + 1/4 + \ldots = 2$.
($\because \Pr[\mathcal{E}_{j,t}] \leq 1/2,\ \forall j \neq 1, t$)

Proof of Lemma 9 (contd.)

- Thus,

$$\begin{aligned}
\sum_{j=1}^{K}(\mathbf{E}[INF_j] + \mathbf{E}[SUP_j]) &\leq \sum_{j=1}^{K}(1 + H_{K-1} - H_j) + 2K \\
&= \sum_{j=1}^{K}\left(1 + \sum_{i=j+1}^{K-1}\frac{1}{i}\right) + 2K \\
&= \sum_{j=1}^{K}(j-1)\frac{1}{j} + 3K \\
&= O(K).
\end{aligned}$$

The Lower Bound

The lower bound

Theorem 2

For any fixed $\epsilon > 0$ and any algorithm ϕ for the K-armed dueling bandit problem, there exists a problem instance such that

$$R_T^{\phi} = \Omega\left(\frac{K}{\epsilon}\log T\right),$$

where $\epsilon = \min_{b \neq b^*} \Pr[b^* \succ b]$.

Construction of the problem instances

A family of K problem instances

- In instance j, let b_j be the best bandit, order the remaining ones by their indices.
    - In instance j, we have $b_j \succ b_k$ for all $k \neq j$ and we have $b_i \succ b_k$ whenever $i < k$.
- $\Pr[b_i \succ b_k] := 1/2 + \epsilon$ whenever $b_i \succ b_k$.
- q_j: the *distribution* on T-step histories induced by ϕ under instance j.

- $n_{j,T}$: # comparisons involving b_j scheduled by ϕ up to time T.

Proving the lower bound

Lemma 10

Let ϕ be an algorithm for the K-armed dueling bandits problem, such that $R_T^{\phi} = o(T^a)$ for all $a > 0$. Then for all j,

$$\mathbf{E}_{q_1}[n_{j,T}] = \Omega\left(\frac{\log T}{\epsilon^2}\right).$$

- If $R_T^{\phi} \neq o(T^a)$, then Theorem 2 holds trivially.
- On instance j, ϕ incurs regret $\geq \epsilon$ every time when it plays a match involving $b_j \neq b_1$.

$$R_T^{\phi} \geq \sum_{j \neq 1} \epsilon \cdot \mathbf{E}_{q_1}[n_{j,T}] = \Omega\left(\frac{K}{\epsilon} \log T\right).$$

Proof of Lemma 10

- $\mathcal{E}_j$: the event that $n_{j,T} < \log(T)/\epsilon^2$.
- $J := \{j \mid q_1(\mathcal{E}_j) < 1/3\}$.
- For each $j \in J$:

$$\mathbf{E}_{q_1}[n_{j,T}] \geq q_1(\mathcal{E}_j^c)(\log(T)/\epsilon^2) = \Omega\left(\frac{\log(T)}{\epsilon^2}\right).$$

- Hence, it remains to show that $\mathbf{E}_{q_1}[n_{j,T}] = \Omega(\log(T)/\epsilon^2)$ for each $j \notin J$.

Proof of Lemma 10

- $\mathcal{E}_j$: the event that $n_{j,T} < \log(T)/\epsilon^2$.
- $J := \{j \mid q_1(\mathcal{E}_j) < 1/3\}$.
- For each $j \in J$:

$$\mathbf{E}_{q_1}[n_{j,T}] \geq q_1(\mathcal{E}_j^c)(\log(T)/\epsilon^2) = \Omega\left(\frac{\log(T)}{\epsilon^2}\right).$$

- Hence, it remains to show that $\mathbf{E}_{q_1}[n_{j,T}] = \Omega(\log(T)/\epsilon^2)$ for each $j \notin J$.

Proof of Lemma 10 (contd.)

- $\mathbf{E}_{q_j}[T - n_{j,T}] = o((T^a)/\epsilon)$.
  - Regret ϵ is incurred for every comparison not involving b_j.
- By Markov's inequality,

$$q_j(\mathcal{E}_j) = q_j(\{T - n_{j,T} > T - \log(T)/\epsilon^2\}) \leq \frac{\mathbf{E}_{q_j}[T - n_{j,T}]}{T - \log(T)/\epsilon^2} = o(T^{a-1}).$$

  - Choose a sufficiently large T so that $q_j(\mathcal{E}_j) < 1/3$ for each j.

Karp & Kleinberg @SODA 2007

For any event $\mathcal{E}$ and distributions p, q with $p(\mathcal{E}) \geq 1/3$ and $q(\mathcal{E}) < 1/3$,

$$KL(p||q) \geq \frac{1}{3} \ln \left(\frac{1}{3q(\mathcal{E})} - \frac{1}{e} \right).$$

Proof of Lemma 10 (contd.)

Karp & Kleinberg @SODA 2007

For any event $\mathcal{E}$ and distributions p, q with $p(\mathcal{E}) \geq 1/3$ and $q(\mathcal{E}) < 1/3$,

$$KL(p||q) \geq \frac{1}{3} \ln \left(\frac{1}{3q(\mathcal{E})} - \frac{1}{e} \right).$$

- We have

$$KL(q_1||q_j) \geq \frac{1}{3} \ln \left(\frac{1}{o(T^{a-1})} \right) - \frac{1}{e} = \Omega(\log T).$$

Proof of Lemma 10 (contd.)

- On the other hand, by the chain rule for KL-divergence,

$$KL(q_1||q_j) \leq \mathbf{E}_{q_1}[n_{j,T}] \cdot KL(1/2+\epsilon||1/2-\epsilon) \leq 16\epsilon^2 \cdot \mathbf{E}_{q_1}[n_{j,T}].$$

  - If a comparison does not involve b_j, then the distribution on the comparison outcome will be the same under q_1 and q_j.
  - $KL(1/2+\epsilon||1/2-\epsilon)$: the KL-divergence b/w two Bernoulli distributions $\text{Ber}(1/2+\epsilon), \text{Ber}(1/2-\epsilon)$.

KL-divergence

For two probability mass functions $p(x_1, \ldots, x_r)$ and $q(x_1, \ldots, x_r)$,

$$KL(p(x_1, \ldots, x_r||q(x_1, \ldots, x_r)) = \sum_{x_1} \ldots \sum_{x_r} p(x_1, \ldots, x_r) \log \frac{p(x_1, \ldots, x_r)}{q(x_1, \ldots, x_r)}.$$

- Hence, $\mathbf{E}_{q_1}[n_{j,T}] = \Omega(\log(T)/\epsilon^2)$ for $j \notin J$.


No climbing
Thank you