UCBoost: A Boosting Approach to Tame Complexity and Optimality for Stochastic Bandits

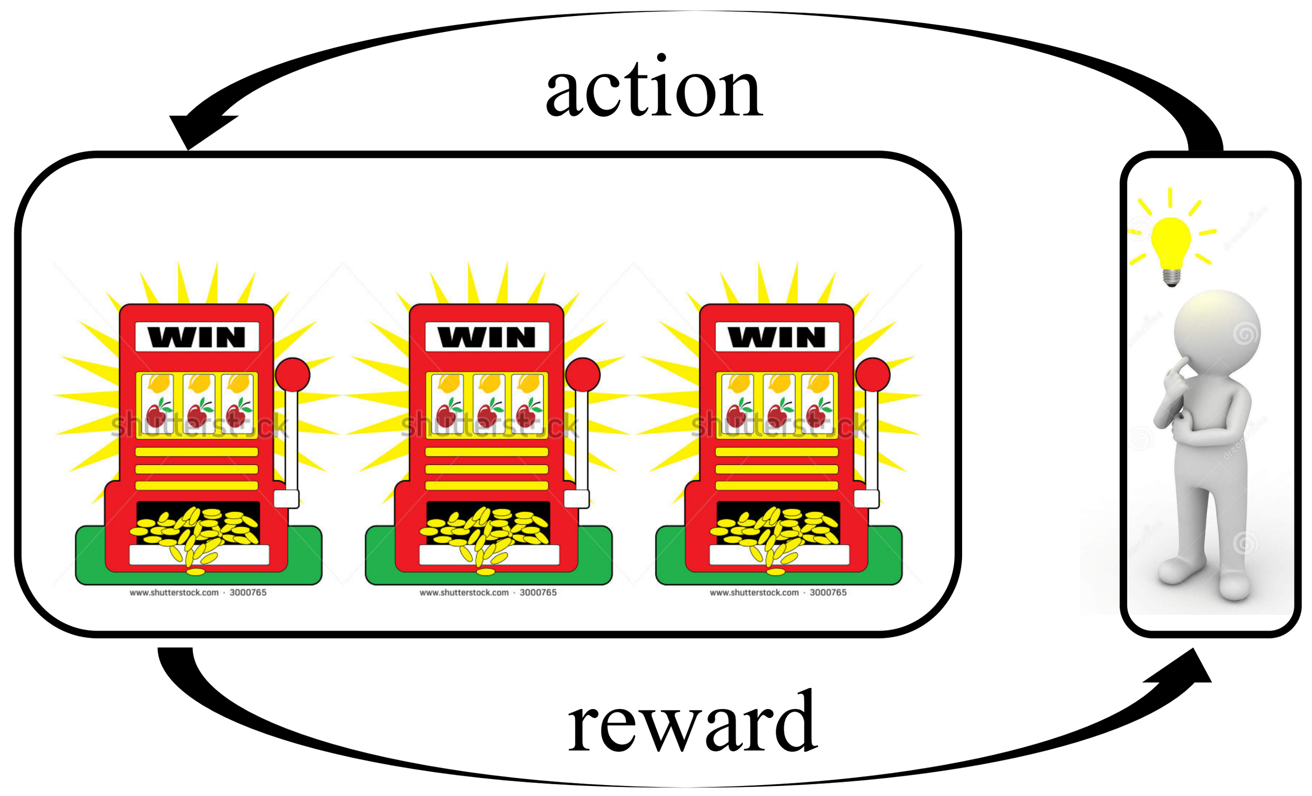
Fang Liu[1], Sinong Wang[1], Swapna Buccapatnam[2] and Ness Shroff[1]

[1]The Ohio State University

[2]AT&T Labs Research

What is Stochastic Bandit?

- Repeated game between agent and environment with random rewards

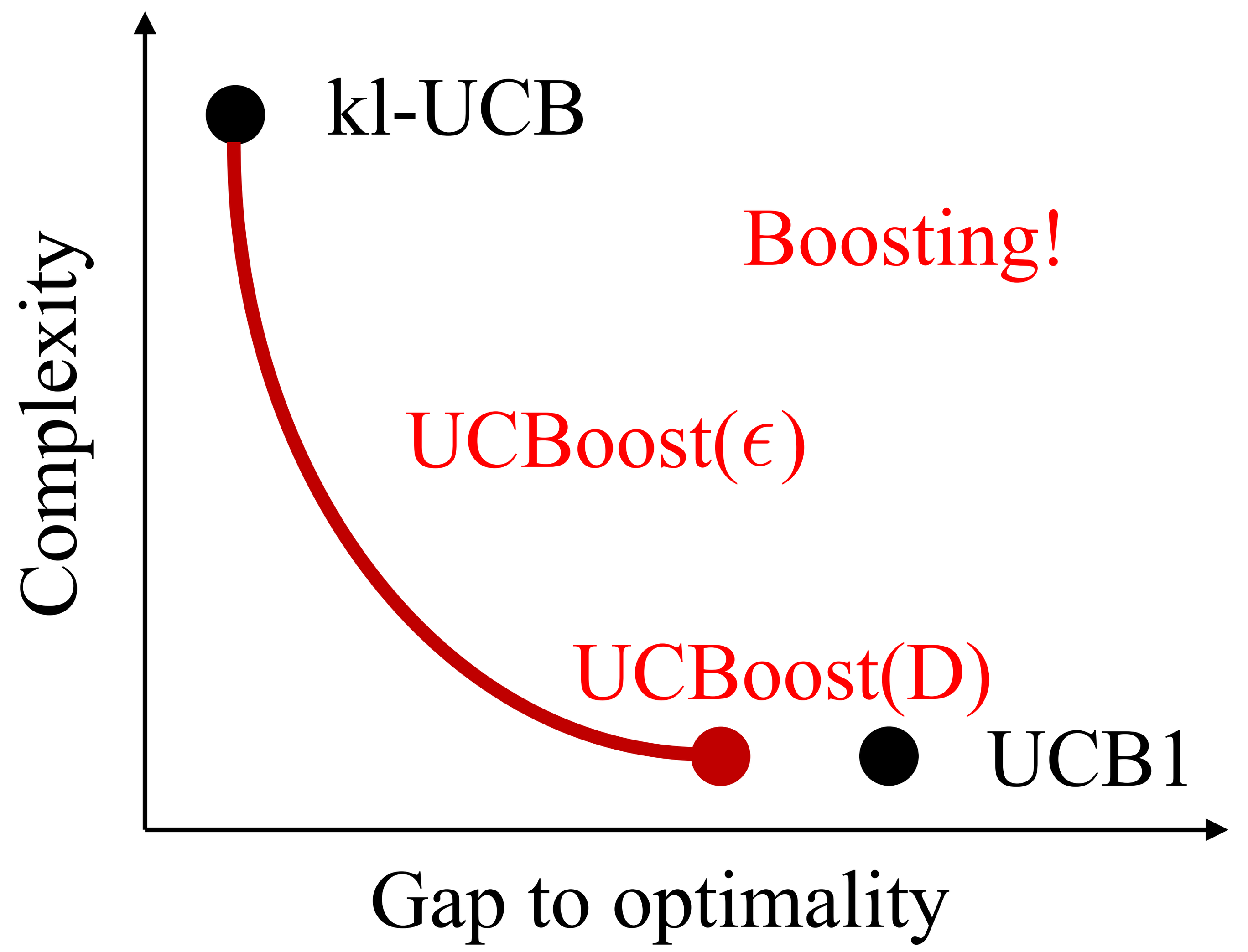


Complexity vs Optimality

- Theoretical bounds

	kl-UCB	UCBoost(ϵ)	UCBoost(D)	UCB1
Regret/log(T)	$O\left(\sum_a \frac{\mu^*-\mu_a}{d_{kl}(\mu_a,\mu^*)}\right)$	$O\left(\sum_a \frac{\mu^*-\mu_a}{d_{kl}(\mu_a,\mu^*)-\epsilon}\right)$	$O\left(\sum_a \frac{\mu^*-\mu_a}{d_{kl}(\mu_a,\mu^*)-1/e}\right)$	$O\left(\sum_a \frac{\mu^*-\mu_a}{2(\mu^*-\mu_a)^2}\right)$
Complexity	unbounded	$O(\log(1/\epsilon))$	$O(1)$	O(1)

- UCBoost connect the dots smoothly



UCBoost

- UCB kernel is a distance function d

$$P(d):\max_{q\in\Theta}\quad q \qquad s.t.\quad d(p,q)\le\delta$$

- UCBoost ensemble a set D of distance functions (i.e. UCBs) by taking the minimum.
- For each d in D, P(d) closed-form

Why taking the minimum?

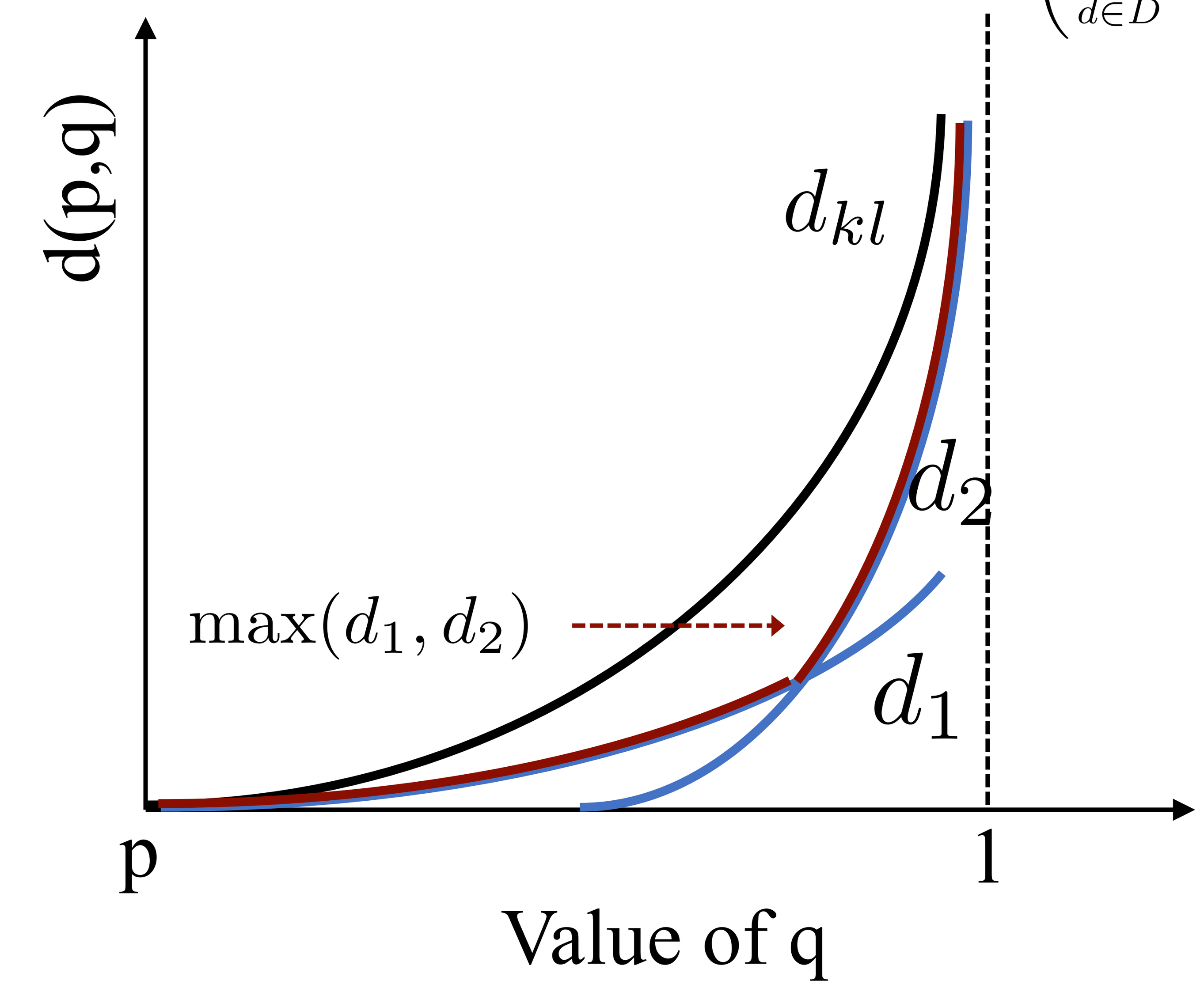
Philosophy of voting:

- If the ordering is known, follow the leader. No majority vote.
- UCBoost takes the minimum, thus the tightest UCB.

	UCB1	UCB2	UCB3	UCBoost
	0.9	0.8	0.6	0.6
	0.8	0.75	0.7	0.7
	0.2	0.2	0.3	0.2
decision	1	1	2	2

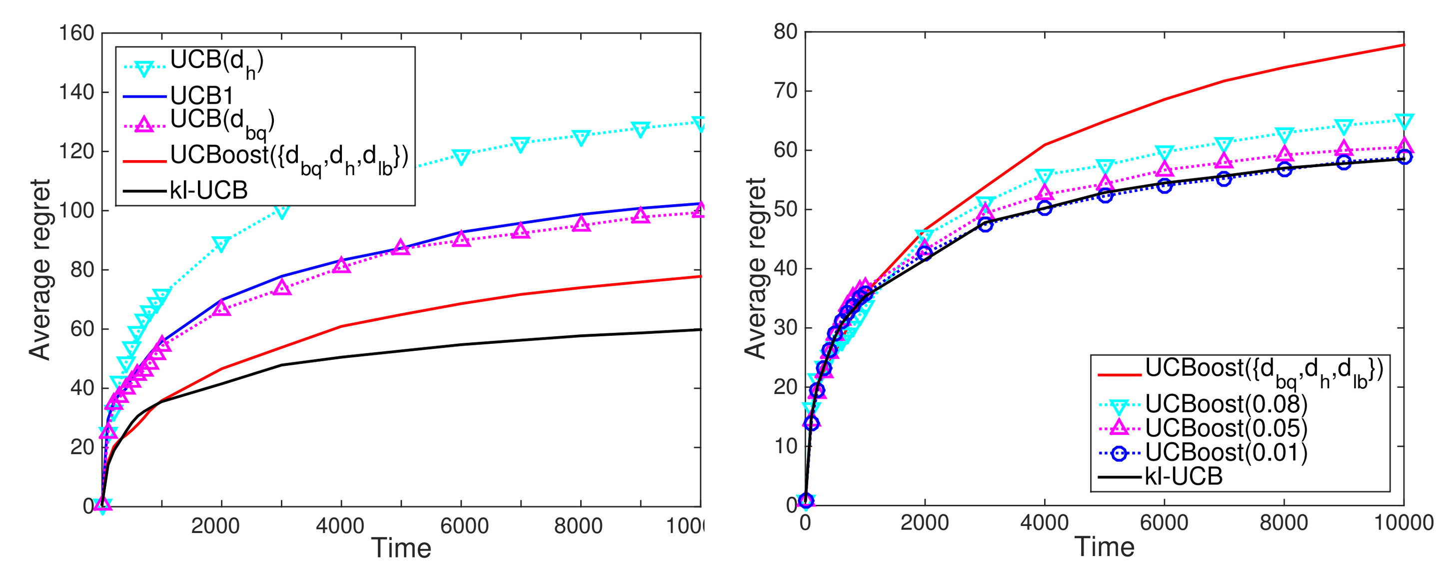
Geometric view of UCBoost:

- The kernel of UCBoost is $\max_{d\in D} d$
- Take the minimum = solve $P\left(\max_{d\in D} d\right)$



Numerical Results

- Bernoulli case



- Computation time

Scenario	kl-UCB	UCBoost(ϵ) $\epsilon=0.01(0.001)$	UCBoost(ϵ) $\epsilon=0.05(0.005)$	UCBoost(ϵ) $\epsilon=0.08$	UCBoost($\{d_{bq},d_h,d_{lb}\}$)	UCB1
Bernoulli 1	$933\mu s$	$7.67\mu s$	$6.67\mu s$	$5.78\mu s$	$1.67\mu s$	$0.31\mu s$
Bernoulli 2	$986\mu s$	$8.76\mu s$	$7.96\mu s$	$6.27\mu s$	$1.60\mu s$	$0.30\mu s$
Beta	$907\mu s$	$8.33\mu s$	$6.89\mu s$	$5.89\mu s$	$2.01\mu s$	$0.33\mu s$