Multiarmed Bandits with Limited Expert Advice

Satyen Kale (Yahoo! Labs)
Motivation: A NIPS parable
Formalization

Set of \( N \) experts \( \{h_1, h_2, ..., h_N\} \)

Set of \( K \) arms \( \{1, 2, ..., K\} \)

Limit \( M \) on # of experts queried in each round

In each round \( t = 1, 2, ..., T \)

- Each expert \( h \) chooses arm \( h(t) \) [unknown to alg]
- Select a subset \( S_t \) of \( M \) experts, get their advice
- Select an arm \( a_t \) based on advice
- Suffer loss \( L_t(a_t) \)

\[ \text{Regret} = \sum_t L_t(a_t) - \min_h \sum_t L_t(h(t)) \]
Open Problem + Our Results

- Seldin, Crammer, Bartlett (COLT 2013) open problem: asked for upper and lower bounds
  - Conjecture: regret $\tilde{O}(\sqrt{KNT/M})$ possible.

- $M = 1$: bandit setting, EXP3 regret $= \tilde{O}(\sqrt{NT})$
- $M = N$: EXP4 regret $= \tilde{O}(\sqrt{KT})$

- Our result: regret bound is
  $$\tilde{O}\left(\sqrt{\min\{K, M\} NT / M}\right)$$
Algorithm

Group experts into $N/M$ bins of size $M$. Call them $B_1, B_2, \ldots, B_{N/M}$

$q_t(h_1) \ldots \ldots \ldots \ldots \ldots \ldots q_t(h_N)$

Run expert learning alg (e.g. MW or PolyInf)
In each round $t$:
• Sample expert $h_t \sim q_t$ and use $a_t = h_t(t)$
• Choose experts binned with $h_t$ to query
• Use following loss estimators in base expert learning alg

$$\hat{L}_t(h) := L_t(h(t)) \frac{1}{\Pr[h \in S_t \text{ and } h(t) = a_t]}$$
Analysis

0 Chosen expert sampled from dist of base alg
  0 Regret = regret of base alg with loss estimators

0 Regret of base alg \approx \sqrt{\sum_t \mathbb{E}[\hat{L}_t(h_t)^2] \log N}

0 We have

\mathbb{E}[\hat{L}_t(h_t)^2] \leq |\{(i, a) : \exists h \in B_i \text{ s.t. } h(t) = a\}| 

\leq \min\{K, M\} \frac{N}{M}

effective number of possible arms
Lower Bound

0 Information theoretic argument

0 Randomly select “best expert” $h^*$

0 In each round $t$, each expert chooses one of the $K$ arms u.a.r.

0 Arm $h^*(t)$ follows $\text{Bernoulli}(\frac{1}{2} - \varepsilon)$, and all other arms follow $\text{Bernoulli}(\frac{1}{2})$

0 Unless $h^*$ in $S_t$ and $A_t = h^*(t)$, regret $\geq \frac{\varepsilon}{2}$
Intuition via Balls-into-Bins

0 Need to bound $\Pr[h^* \text{ in } S_t \text{ and } A_t = h^*(t)]$

0 Regret $\approx \sqrt{T/\Pr[h^* \text{ in } S_t \text{ and } A_t = h^*(t)]}$

0 Bounded by $\mathbb{E}[\max_a \#\{h^* \text{ in } S_t \text{ s.t. } a = h^*(t)\}] / N$

0 Experts in $S_t =$ “balls”, $K$ arms = “bins”

0 Expected max load $= \max\{M/K, \log(K)\}$

0 So regret $\geq \sqrt{NT/\max\{M/K, \log(K)\}}$

$= \tilde{\Omega}\left(\sqrt{\min\{K, M\} / M} \cdot NT\right)$
Extensions

1. Global limit of MT queries over all rounds:

$$\tilde{\Omega} \left( \sqrt{\frac{\min\{K, M\}}{M}} NT \right)$$ lower bound still holds

1. Changing limits on queried experts $M_1, M_2, \ldots, M_T$:

- Re-bin experts on every round.
- Regret bound is

$$\tilde{\Theta} \left( \sqrt{\sum_{t=1}^{T} \frac{\min\{K, M_t\}}{M_t} N} \right)$$
Conclusions

- Near-optimal upper and lower bounds for the multiarmed bandits with limited expert advice.
- Extensions to global limits on queries and changing limits on queries.
- Open problem: close the logarithmic gap between upper and lower bounds.

Thank you!