Best Arm Identification: Almost Instance-Wise Optimality and the Gap Entropy Conjecture

Lijie Chen, Jian Li

Tsinghua University

June 25, 2016

Best Arm Identification: Pure Exploration

Fixed confidence setting.

- n stochastic arms, each with an associated Gaussian distribution $\mathcal{D}_i = \mathcal{N}(\mu_i, 1)$.
- Each time we can choose an arm and take a sample from that distribution.
- Want the arm with largest mean.
- **Goal:** Succeed w.p. $1 - \delta$ and minimize the samples we need.
- $\mu_{[i]}$: i^{th} largest mean, (Gap) $\Delta_{[i]} := \mu_{[1]} - \mu_{[i]}$.

Previous Result

Source	Sample Complexity
Even-Dar et al. [EDMM02]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln n + \ln \Delta_{[i]}^{-1} \right)$
Gabillon et al. [GGL12]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \sum_{i=2}^{n} \Delta_{[i]}^{-2} \right)$
Jamieson et al. [JMNB13]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \ln \left(\sum_{j=2}^{n} \Delta_{[j]}^{-2} \right) \right)$
Kalyanakrishnan et al. [KTAS12]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \sum_{i=2}^{n} \Delta_{[i]}^{-2} \right)$
Jamieson et al. [JMNB13]	$\ln \delta^{-1} \cdot \left(\ln \ln \delta^{-1} \cdot \sum_{i=2}^{n} \Delta_{[i]}^{-2} + \sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \Delta_{[i]}^{-1} \right)$
Karnin et al. [KKS13]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \ln \Delta_{[i]}^{-1} \right)$
Jamieson et al. [JMNB14]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \ln \Delta_{[i]}^{-1} \right)$
Chen et al. [CL15]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \ln \min(n, \Delta_{[i]}^{-1}) \right) + \Delta_{[2]}^{-2} \ln \ln \Delta_{[2]}^{-1}$

Table: Sample complexity upper bounds. We omit the big-O notations.

Source	Sample Complexity	Type
Mannor et al. [MT04]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \delta^{-1}$	instance-wise
Farrell [Far64]	$\Delta^{-2} \ln \ln \Delta^{-1}$	worst-case, two-arm
Chen et al. [CL15]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \ln n$	worst-case

Table: Sample complexity lower bounds. We omit the big-Ω notations.

Two Types of Optimality

- Instance-wise optimal: can't be improved on every instances up to a constant.
- All the algorithms listed above are worst case optimal.
  - Worst case optimal: can't be improved on some instances up to a constant.
- We want an instance-wise optimal algorithm.

Gap Entropy Conjecture

- Subtly: Due to the $\Delta^{-2} \ln \ln \Delta^{-1}$ worst-case lower bound for two-arm by Farrell [Far64], there is no instance-wise algorithm even for the two-arm case.
- We conjecture that the two-arm case is the only obstruction!
- Define

$$G_k = \{i \in [2, n] \mid 2^{-k} \le \Delta_{[i]} < 2^{-k+1}\}$$
$$H_k = \sum\nolimits_{i \in G_k} \Delta_{[i]}^{-2} \quad p_k = H_k / \sum\nolimits_j H_j.$$

- Our new quantity, Gap entropy

$$\mathrm{Ent}(I) = \sum\nolimits_{G_k \neq \emptyset} p_k \log p_k^{-1}.$$

- **Conjecture**: There is :
  - An upper bound: $O(\sum_{i=2}^{n} \Delta_{[i]}^{-2}(\mathrm{Ent}(I) + \ln \delta^{-1}) + \Delta_{[2]}^{-2} \ln \ln \Delta_{[2]}^{-1})$.
  - An instance-wise lower bound: $\Omega(\sum_{i=2}^{n} \Delta_{[i]}^{-2}(\mathrm{Ent}(I) + \ln \delta^{-1}))$.
- The best we can hope for!

Partial Progress

- In a recent work [CL15], we obtain an

$$O\left(\sum_{i=2}^{n} \Delta_{[i]}^{-2}\left(\ln \delta^{-1} + \ln\ln \min(n, \Delta_{[i]}^{-1})\right) + \Delta_{[2]}^{-2} \ln\ln \Delta_{[2]}^{-1}\right)$$

upper bound for BEST-1-ARM.

- Note that $\mathrm{Ent}(I) = O(\ln\ln n)$, our algorithm solves the case with maximum gap entropy.

- A worst case lower bound

$$\Omega\left(\sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln\ln n\right)$$

by constructing some instances with $\mathrm{Ent}(I) = \Theta(\ln\ln n)$.

Intuition : Upper bound

- In the framework of Karin, Koren and Somekh [KKS13].
- They assign r^{th} round a confidence level δ_r.
- Need to make sure $\sum_r \delta_r \le \delta$.
- The complexity is then $O(\sum_r H_r \ln \delta_r^{-1})$.
- They set $\delta_r = \Theta(\delta/r^2)$, so their complexity is

$$O\left(\sum_r H_r \cdot (\ln \delta^{-1} + \ln r)\right) = O\left(\sum_i \Delta_{[i]}^{-2} \ln \ln \Delta_{[i]}^{-1}\right).$$

- In [CL15], we use a better way to assign δ_r's.
- $\sum_r H_r \ln \delta_r^{-1}$ is minimized when we set $\delta_r = \delta \cdot \frac{H_r}{\sum_k H_k}$, and we will get the running time $H \cdot (\mathrm{Ent}(I) + \ln \delta^{-1})$.
- **Problem**: we don't know H_r's.

Ongoing Work

- We have some ideas on how to get an algorithm matching the upper bound.
- Despite that we are far from proving the conjectured lower bound, we have very strong evidence that it should be true.
- Joint work with Mingda Qiao (Tsinghua University).

Lijie Chen and Jian Li.
On the optimal sample complexity for best arm identification.
arXiv preprint arXiv:1511.03774, 2015.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour.
Pac bounds for multi-armed bandit and markov decision processes.
In *Computational Learning Theory*, pages 255–270. Springer, 2002.

RH Farrell.
Asymptotic behavior of expected sample size in certain one sided tests.
The Annals of Mathematical Statistics, pages 36–72, 1964.

Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric.
Best arm identification: A unified approach to fixed budget and fixed confidence.
In *Advances in Neural Information Processing Systems*, pages 3212–3220, 2012.

K. Jamieson, M. Malloy, R. Nowak, and S. Bubeck.
On Finding the Largest Mean Among Many.
ArXiv preprint arXiv:1306.3917, June 2013.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck.

lil'ucb: An optimal exploration algorithm for multi-armed bandits.
COLT, 2014.

Zohar Karnin, Tomer Koren, and Oren Somekh.
Almost optimal exploration in multi-armed bandits.
In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 1238–1246, 2013.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone.
Pac subset selection in stochastic multi-armed bandits.
In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 655–662, 2012.

Shie Mannor and John N Tsitsiklis.
The sample complexity of exploration in the multi-armed bandit problem.
The Journal of Machine Learning Research, 5:623–648, 2004.