Pure Exploration Stochastic Multi-armed Bandits

Jian Li

Institute for Interdisciplinary Information Sciences

Tsinghua University

Outline

- Introduction
- Optimal PAC Algorithm (Best-Arm, Best-k-Arm):
  - Median/Quantile Elimination
- Combinatorial Pure Exploration
- Best-Arm – Instance optimality
- Conclusion

- Decision making with limited information

An "algorithm" that we use everyday

  - Initially, nothing/little is known
  - Explore (to gain a better understanding)
  - Exploit (make your decision)

- Balance between exploration and exploitation
  - We would like to explore widely so that we do not miss really good choices
  - We do not want to waste too much resource exploring bad choices (or try to identify good choices as quickly as possible)

The Stochastic Multi-armed Bandit

- Stochastic Multi-armed Bandit
  - Set of n arms
  - Each arm is associated with an unknown reward distribution supported on $[0,1]$ with mean θ_i
  - Each time, sample an arm and receive the reward independently drawn from the reward distribution

classic problems in stochastic control, stochastic optimization and online learning

The Stochastic Multi-armed Bandit

- Stochastic Multi-armed Bandit (MAB)

MAB has MANY variations!

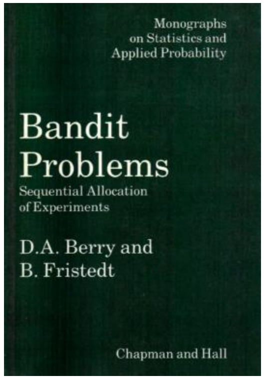
- Goal 1: Minimizing Cumulative Regret (Maximizing Cumulative Reward)
- Goal 2: (Pure Exploration) Identify the (approx) best K arms (arms with largest means) using as few samples as possible (Top-K Arm identification problem)
  - K=1 (best-arm identification)

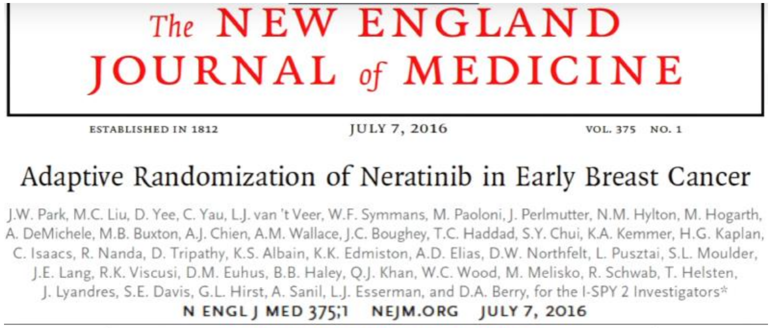
Stochastic Multi-armed Bandit

- Statistics, medical trials (Bechhofer, 54) ,Optimal control，Industrial engineering (Koenig & Law, 85), evolutionary computing (Schmidt, 06), Simulation optimization (Chen, Fu, Shi 08),Online learning (Bubeck Cesa-Bianchi,12)
- [Bechhofer, 58] [Farrell, 64] [Paulson, 64] [Bechhofer, Kiefer, and Sobel, 68],...., [Even-Dar, Mannor, Mansour, 02] [Mannor, Tsitsiklis, 04] [Even-Dar, Mannor, Mansour, 06] [Kalyanakrishnan, Stone 10] [Gabillon, Ghavamzadeh, Lazaric, Bubeck, 11] [Kalyanakrishnan, Tewari, Auer, Stone, 12] [Bubeck, Wang, Viswanatha, 12]....[Karnin, Koren, and Somekh, 13] [Chen, Lin, King, Lyu, Chen, 14]
- Books:
  - Multi-armed Bandit Allocation Indices, John Gittins, Kevin Glazebrook, Richard Weber, 2011
  - Regret analysis of stochastic and nonstochastic multi-armed bandit problems S. Bubeck and N. Cesa-Bianchi., 2012
  - ……

Applications

- **Clinical Trails**
  - **One arm – One treatment**
  - **One pull – One experiment**





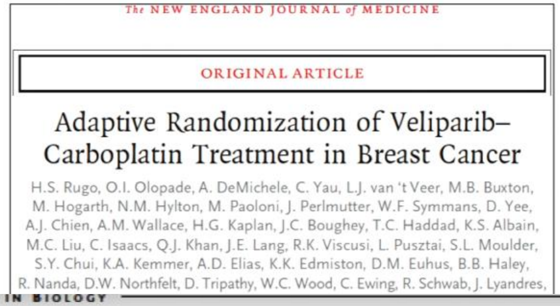
The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 — JULY 7, 2016 — VOL. 375 NO. 1

Adaptive Randomization of Neratinib in Early Breast Cancer

J.W. Park, M.C. Liu, D. Yee, C. Yau, L.J. van 't Veer, W.F. Symmans, M. Paoloni, J. Perlmutter, N.M. Hylton, M. Hogarth, A. DeMichele, M.B. Buxton, A.J. Chien, A.M. Wallace, J.C. Boughey, T.C. Haddad, S.Y. Chui, K.A. Kemmer, H.G. Kaplan, C. Isaacs, R. Nanda, D. Tripathy, K.S. Albain, K.K. Edmiston, A.D. Elias, D.W. Northfelt, L. Pusztai, S.L. Moulder, J.E. Lang, R.K. Viscusi, D.M. Euhus, B.B. Haley, Q.J. Khan, W.C. Wood, M. Melisko, R. Schwab, T. Helsten, J. Lyandres, S.E. Davis, G.L. Hirst, A. Sanil, L.J. Esserman, and D.A. Berry, for the I-SPY 2 Investigators*

N ENGL J MED 375;1 NEJM.ORG JULY 7, 2016

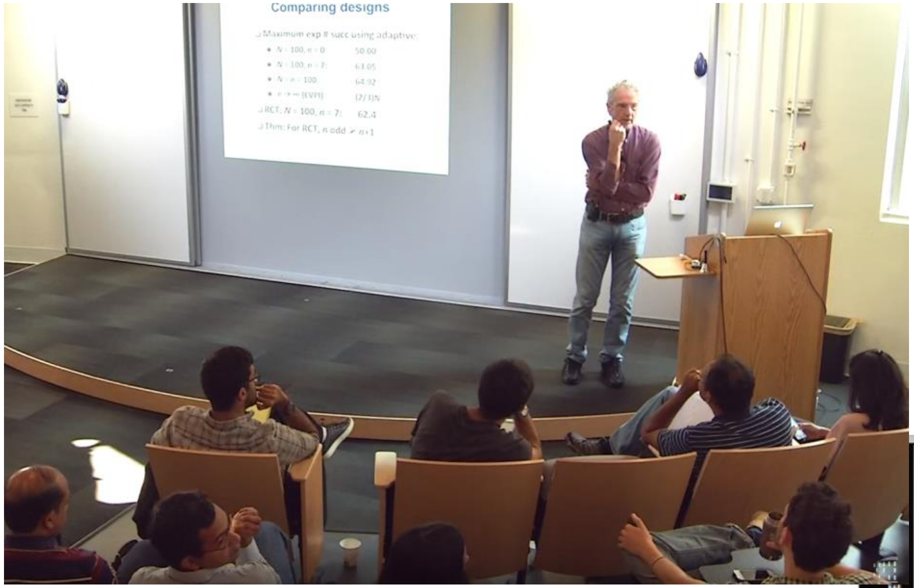


The NEW ENGLAND JOURNAL of MEDICINE

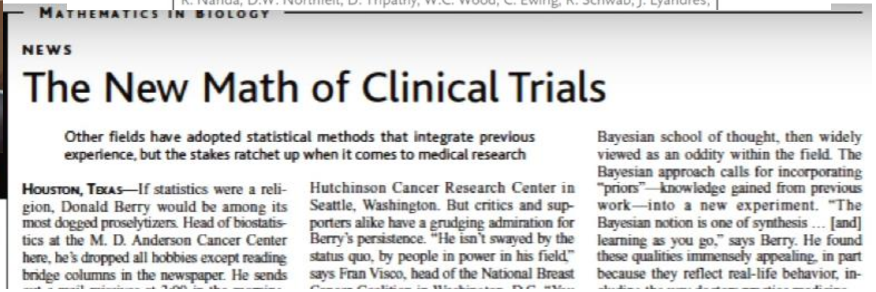
ORIGINAL ARTICLE

Adaptive Randomization of Veliparib–Carboplatin Treatment in Breast Cancer

H.S. Rugo, O.I. Olopade, A. DeMichele, C. Yau, L.J. van 't Veer, M.B. Buxton, M. Hogarth, N.M. Hylton, M. Paoloni, J. Perlmutter, W.F. Symmans, D. Yee, A.J. Chien, A.M. Wallace, H.G. Kaplan, J.C. Boughey, T.C. Haddad, K.S. Albain, M.C. Liu, C. Isaacs, Q.J. Khan, J.E. Lang, R.K. Viscusi, L. Pusztai, S.L. Moulder, S.Y. Chui, K.A. Kemmer, A.D. Elias, K.K. Edmiston, D.M. Euhus, B.B. Haley, R. Nanda, D.W. Northfelt, D. Tripathy, W.C. Wood, C. Ewing, R. Schwab, J. Lyandres,



Don Berry, University of Texas
MD Anderson Cancer Center



MATHEMATICS IN BIOLOGY

NEWS

The New Math of Clinical Trials

Other fields have adopted statistical methods that integrate previous experience, but the stakes ratchet up when it comes to medical research

HOUSTON, TEXAS—If statistics were a religion, Donald Berry would be among its most dogged proselytizers. Head of biostatistics at the M. D. Anderson Cancer Center here, he's dropped all hobbies except reading bridge columns in the newspaper. He sends

Hutchinson Cancer Research Center in Seattle, Washington. But critics and supporters alike have a grudging admiration for Berry's persistence. "He isn't swayed by the status quo, by people in power in his field," says Fran Visco, head of the National Breast

Bayesian school of thought, then widely viewed as an oddity within the field. The Bayesian approach calls for incorporating "priors"—knowledge gained from previous work—into a new experiment. "The Bayesian notion is one of synthesis ... [and] learning as you go," says Berry. He found these qualities immensely appealing, in part because they reflect real-life behavior, in-

Applications

- Crowdsourcing:
- Workers are noisy

0.95

0.99

0.5

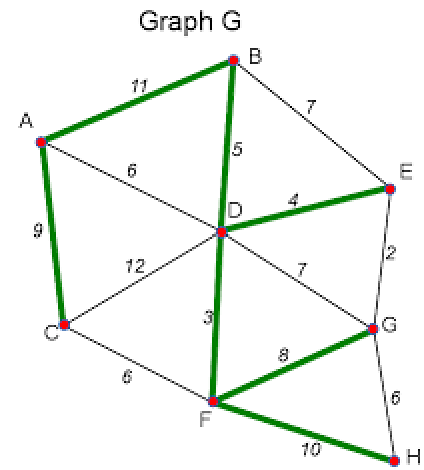
- How to identify reliable workers and exclude unreliable workers ?
- Test workers by golden tasks (i.e., tasks with known answers)

❖ Each test costs money. How to identify the best K workers with minimum amount of money?

Top-K Arm Identification	
Worker	Bernoulli arm with mean θ_i (θ_i: i-th worker's reliability)
Test with golden task	Obtain a binary-valued sample (correct/wrong)

Applications

We want to build a MST.
But we don't know the true cost of each edge.
Each time we can get a sample from an edge,
which is a noisy estimate of its true cost.



Combinatorial Pure Exploration

- A general combinatorial constraint on the feasible set of arms
  - Best-k-arm: the uniform matroid constraint
  - First studied by [Chen et al. NIPS14]

Outline

- Introduction
- Optimal PAC Algorithm (Best-Arm, Best-k-Arm):
  - Median/Quantile Elimination
- Combinatorial Pure Exploration
- Best-Arm – Instance optimality
- Conclusion

PAC

- PAC learning: find an ϵ-optimal solution with probability $1-\delta$
- ϵ-optimal solution for best-arm
  - (additive/multiplicative) ϵ-optimality
    - The arm in our solution is ϵ away from the best arm
- ϵ-optimal solution for best-k-arm
  - (additive/multiplicative) Elementwise ϵ-optimality (this talk)
    - The ith arm in our solution is ϵ away from the ith arm in OPT
  - (additive/multiplicative) Average ϵ-optimality
    - The average mean of our solution is ϵ away from the average of OPT

Chernoff-Hoeffding Inequality

Proposition *Let* $X_i (1 \leq i \leq n)$ *be independent random variables with values in* $[0,1]$. *Let* $X = \frac{1}{n}\sum_{i=1}^{n} X_i$. *The following statements hold:*

For every $t > 0$, *we have that*

$$\Pr\left[|X - \mathbf{E}[X]| > t\right] < 2\exp(-2t^2 n).$$

For every $\epsilon > 0$, *we have that*

$$\Pr\left[|X < (1-\epsilon)\,\mathbf{E}[X]\right] < \exp(-\epsilon^2 n\,\mathbf{E}[X]/2), \text{ and}$$

$$\Pr\left[|X > (1+\epsilon)\,\mathbf{E}[X]\right] < \exp(-\epsilon^2 n\,\mathbf{E}[X]/3).$$

Naïve Solution (Best-Arm)

- Uniform Sampling

Sample each coin M times

Pick the coins with the largest empirical mean

empirical mean: #heads/M

How large M needs to be (in order to achieve ϵ-optimality)??

Naïve Solution (Best-Arm)

- Uniform Sampling

Sample each coin M times

Pick the coins with the largest empirical mean

empirical mean: #heads/M

How large M needs to be (in order to achieve ϵ-optimality)??

$$M = O(\frac{1}{\epsilon^2}\left(\log n + \log\frac{1}{\delta}\right)) = O(\log n)$$

Then, by Chernoff Bound, we can have

$$\Pr[|\mu_i - \widehat{\mu}_i| \le \epsilon] = \delta/n$$

True mean of arm i

Emp mean of arm i

So the total number of samples is $O(n\log n)$

Is this necessary?

Naïve Solution

- Uniform Sampling
- What if we use M=O(1) (let us say M=10)
  - E.g., consider the following example (K=1):
    - 0.9, 0.5, 0.5, ……………………, 0.5 (a million coins with mean 0.5)
    - Consider a coin with mean 0.5,

      Pr[All samples from this coin are head]=(1/2)^10
    - With const prob, there are more than 500 coins whose samples are all heads

Can we do better??

- Consider the following example:
  - 0.9, 0.5, 0.5, ..., 0.5 (a million coins with mean 0.5)
  - Uniform sampling spends too many samples on bad coins.
  - Should spend more samples on good coins
    - However, we do not know which one is good and which is bad……
  - Sample each coin M=O(1) times.
    - If the empirical mean of a coin is large, we DO NOT know whether it is good or bad
    - But if the empirical mean of a coin is very small, we DO know it is bad (with high probability)

Median/Quantile-Elimination

PAC algorithm for best-k arm

For i=1,2,….

Sample each arm M_i times M_i : *increasing expoentially*

Eliminate one quarter arms

Until less 4k arms

When $\mathrm{n} \leq 4k$, use uniform sampling

We can find a solution with additive error ϵ

Our algorithm

Algorithm 1: ME-AS

input: B, ϵ, δ, k
for $\mu = 1/2, 1/4, \dots$ **do**
  S = ME$(B, \epsilon, \delta, \mu, k)$;
  $\{(a_i, \hat{\theta}^{US}(a_i)) \mid 1 \le i \le k\}$ = US$(S, \epsilon, \delta, (1 - \epsilon/2)\mu, k)$;
  if $\hat{\theta}^{US}(a_k) \ge 2\mu$ **then**
    return $\{a_1, \dots, a_k\}$;

Algorithm 2: Median Elimination (ME)

input: B, ϵ, δ, μ, k
$S_1 = B$, $\epsilon_1 = \epsilon/16$, $\delta_1 = \delta/8$, $\mu_1 = \mu$, and $\ell = 1$;
while $|S_\ell| > 4k$ **do**
  sample every arm $a \in S_\ell$ for $Q_\ell = (12/\epsilon_\ell^2)(1/\mu_\ell)\log(6k/\delta_\ell)$ times;
  for *each arm* $a \in S_\ell$ **do**
    its empirical value $\hat{\theta}(a)$ = the average of the Q_ℓ samples from a;
  $a_1, \dots, a_{|S_\ell|}$ = the arms sorted in non-increasing order of their empirical values;
  $S_{\ell+1} = \{a_1, \dots, a_{|S_\ell|/2}\}$;
  $\epsilon_{\ell+1} = 3\epsilon_\ell/4, \delta_{\ell+1} = \delta_\ell/2, \mu_{\ell+1} = (1 - \epsilon_\ell)\mu_\ell$, and $\ell = \ell + 1$;
return S_ℓ;

Algorithm 3: Uniform Sampling (US)

input: $S, \epsilon, \delta, \mu_s, k$
sample every arm $a \in S$ for $Q = (96/\epsilon^2)(1/\mu_s)\log(4|S|/\delta)$ times;
for *each arm* $a \in S$ **do**
  its US-empirical value $\hat{\theta}^{US}(a)$ = the average of the Q samples from a;
$a_1, \dots, a_{|S|}$ = the arms sorted in non-increasing order of their US-empirical values;
return $\{(a_1, \hat{\theta}^{US}(a_1)), \dots, (a_k, \hat{\theta}^{US}(a_k))\}$

(worst case) Optimal bounds

Table 1: Comparison of our and previous results (all bounds are in expectation)

problem		sample complexity	source
k -AS	upper bound	$O\left(\frac{n}{\epsilon^2}\log\frac{n}{\delta}\right)$	[14]
		$O\left(\frac{n}{\epsilon^2}\log\frac{k}{\delta}\right)$	NIPS15
	lower bound	$\Omega\left(\frac{n}{\epsilon^2}\log\frac{k}{\delta}\right)$	[11]
		$\Omega\left(\frac{n}{\epsilon^2}\log\frac{k}{\delta}\right)$	NIPS15
k_{avg} -AS	upper bound	$O(\frac{n}{\epsilon^2}\log\frac{n}{\delta})$	[14]
		$O\left(\frac{n}{\epsilon^2}\cdot\left(1+\frac{\log(1/\delta)}{k}\right)\right)$	ICML14
	lower bound	$\Omega\left(\frac{n}{\epsilon^2}\left(1+\frac{\log(1/\delta)}{k}\right)\right)$	ICML14

Additive version

Original Idea for best-arm [Even-Dar COLT02]
We solve the average (additive) version in [Zhou, Chen, L ICML'14]
We extend the result to both (multiplicative) elementwise and average in [Cao, L, Tao, Li, NIPS'15]

(worst case) Optimal bounds

Table 1: Comparison of our and previous results (all bounds are in expectation)

problem		sample complexity	source
k-AS	upper bound	$O\left(\frac{n}{\epsilon^2}\frac{1}{\theta_k(B)}\log\frac{n}{\delta}\right)$	[14]
		$O\left(\frac{n}{\epsilon^2}\frac{1}{\theta_k(B)}\log\frac{k}{\delta}\right)$	**new**
	lower bound	$\Omega\left(\frac{n}{\epsilon^2}\log\frac{k}{\delta}\right)$	[11]
		$\Omega\left(\frac{n}{\epsilon^2}\frac{1}{\theta_k(B)}\log\frac{k}{\delta}\right)$	**new**
k_{avg}-AS	upper bound	$O(\frac{n}{\epsilon^2}\frac{1}{\theta_k(B)}\log\frac{n}{\delta})$	[14]
		$O\left(\frac{n}{\epsilon^2}\frac{1}{(\theta_{\mathsf{avg}}(B))^2}\left(1+\frac{\log(1/\delta)}{k}\right)\right)$	[16]
		$O\left(\frac{n}{\epsilon^2}\frac{1}{\theta_{\mathsf{avg}}(B)}\left(1+\frac{\log(1/\delta)}{k}\right)\right)$	**new**
	lower bound	$\Omega\left(\frac{n}{\epsilon^2}\left(1+\frac{\log(1/\delta)}{k}\right)\right)$	[16]
		$\Omega\left(\frac{n}{\epsilon^2}\frac{1}{\theta_{\mathsf{avg}}(B)}\left(1+\frac{\log(1/\delta)}{k}\right)\right)$	**new**

Multiplicative version: θ_k: true mean of the k-th arm

We solve the average (additive) version in [Zhou, Chen, L ICML'14]
We extend the result to both (multiplicative) elementwise and average in [Cao, L, Tao, Li, NIPS'15]

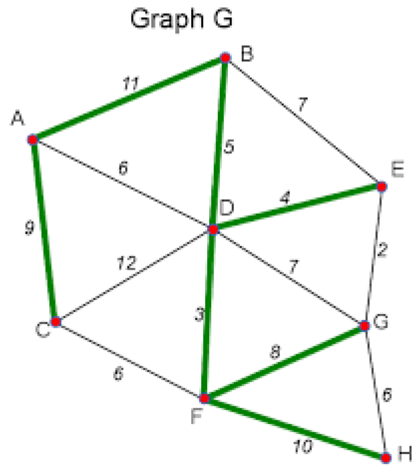
Outline

- Introduction
- Optimal PAC Algorithm (Best-Arm, Best-k-Arm):
  - Median/Quantile Elimination
- Combinatorial Pure Exploration
- Best-Arm – Instance optimality
- Conclusion

A More General Problem

Combinatorial Pure Exploration

- A general combinatorial constraint on the feasible set of arms
  - Best-k-arm: the uniform matroid constraint
  - First studied by [Chen et al. NIPS14]
  - E.g., we want to build a MST. But each time get a noisy estimate of the true cost of each edge



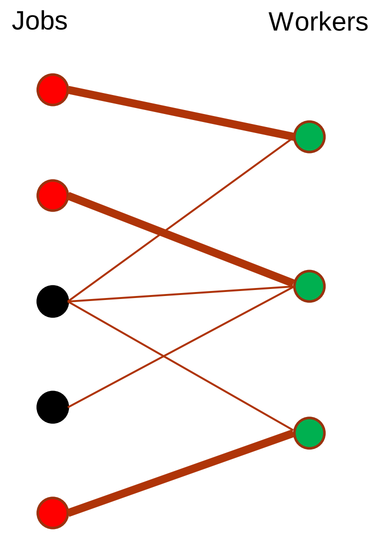
- We obtain improved bounds for general matroid constaints
  - Our bounds even improve previous results on Best-k-arm

[Chen, Gupta, **L.** COLT'16]

Application

- A set of jobs
- A set of workers
- Each worker can only do one job
- Each job has a reward distribution

- Goal: choose the set of jobs with the largest total expected reward



Feasible sets of jobs that can be completed form a transversal matroid

Our Results

- PAC: Strong eps-optimality (stronger than elementwise opt)
  - Ours: $O(n\varepsilon^{-2} \cdot (\ln k + \ln \delta^{-1}))$
  - Generalizes [Cao et al.][Kalyanakrishnan et al.]
  - Optimal: Matching the LB in [Kalyanakrishnan et al.]

- PAC: Average eps-optimality
  - Ours: $O(n\varepsilon^{-2}(1 + \ln \delta^{-1}/k))$. (under mild condition)
  - Generalizes [Zhou et al.]
  - Optimal (under mild condition): matching the lower bound in [Zhou et al.]

Our Results

- A generalized definition of gap

$$\Delta_e^{\mathcal{M},\mu} := \begin{cases} \mathsf{OPT}(\mathcal{M}) - \mathsf{OPT}(\mathcal{M}_{S\setminus\{e\}}) & e \in \mathsf{OPT}(\mathcal{M}) \\ \mathsf{OPT}(\mathcal{M}) - (\mathsf{OPT}(\mathcal{M}_{/\{e\}}) + \mu(e)) & e \notin \mathsf{OPT}(\mathcal{M}) \end{cases}$$

- Exact identification
  - [Chen et al.] $\left(\sum_{e\in S} \Delta_e^{-2}(\ln \delta^{-1} + \ln n + \ln \sum_{e\in S} \Delta_e^{-2})\right)$
  - Previous best-k-arm [Kalyanakrishnan]:

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

  - Ours: $O\left(\sum_{e\in S} \Delta_e^{-2}(\ln \delta^{-1} + \ln k + \ln \ln \Delta_e^{-1})\right)$
  - Our result is even better than previous best-k-arm result
  - Our result matches Karnin'et al. result for best-1-arm

Our technique

- Attempt: try to adapt the median/quantile elimination technique
- Key difficulty:
  - We cannot just eliminate half of elements, due to the matroid constraint!

Our technique

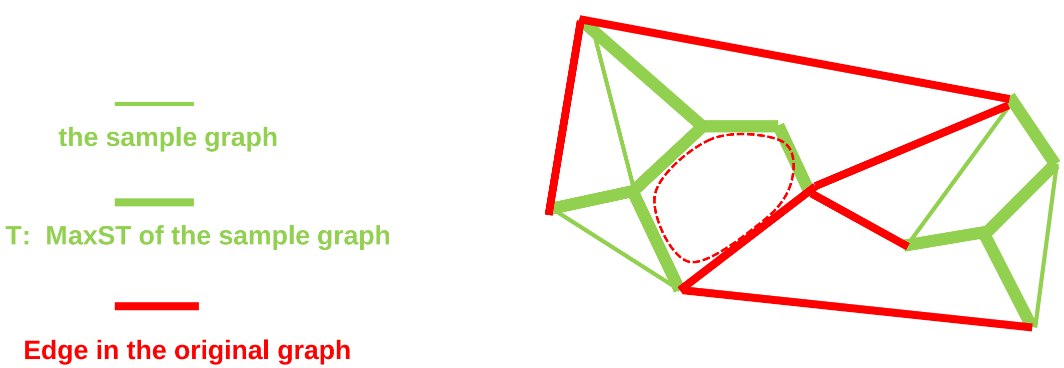
- Attempt: try to adapt the median/quantile elimination technique
- Key difficulty:
  - We cannot just eliminate half of elements, due to the matroid constraint!

- Sampling-and-Pruning technique
  - Originally developed by Karger, and used by Karger, Klein, Tarjan for the expected linear time MST
  - First time used in Bandit literature
  - IDEA: Instead of using a single threshold to prune elements, we use the solution for a sampled set to prune.

High level idea (for MaxST)

Sample-Prune

- Sample a subset of edges (uniformly and random, w.p. 1/100)
- Find the MaxST *T* over the sampled edges
- Use *T* to prune a lot of edges (w.h.p. we can prune a constant fraction of edges)
- Iterate over the remaining edges

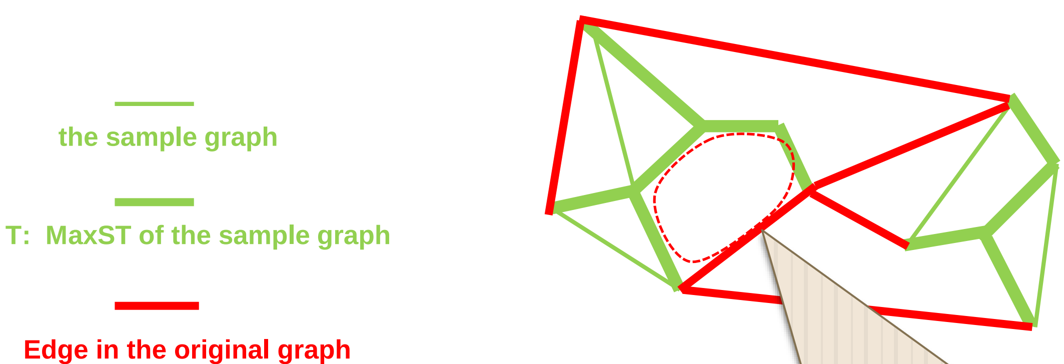


High level idea (for MaxST)

Sample-Prune

- Sample a subset of edges (uniformly and random, w.p. 1/100)
- Find the MaxST T over the sampled edges
- Use T to prune a lot of edges (w.h.p. we can prune a constant fraction of edges)
- Iterate over the remaining edges

OB: If e is the lightest edge in a cycle, e can not appear in the MaxST.
There is a generalization of this statement in the more general matroid context.



Consider an edge in the original graph. If it is the lightest edge in the cycle, it can be pruned.

Our technique

- Sampling-and-Pruning technique
  - Originally developed by Karger, and used by Karger, Klein, Tarjan for the expected linear time MST

Algorithm 3: PAC-SamplePrune $(\mathcal{S}, \varepsilon, \delta)$

Data: A PAC-BASIS instance $\mathcal{S} = (S, \mathcal{M})$, with $\mathrm{rank}(\mathcal{M}) = k$, approximation error ε, confidence level δ.

Result: A basis I in $\mathcal{M}$.

if $|S| \leq 2p^{-2} \cdot \max(4 \cdot \ln 8\delta^{-1}, k)$ **then**
  Return Naïve-I $(\mathcal{S}, \varepsilon, \delta)$

Sample a subset $F \subseteq S$ by choosing each element with probability p independently.
$\alpha \leftarrow \varepsilon/3$, $\lambda \leftarrow \varepsilon/12$
$I \leftarrow$ PAC-SamplePrune$(\mathcal{S}_F = (F, \mathcal{M}_F), \alpha, \delta/8)$
$\hat{\mu} \leftarrow$ UniformSample $(S, \lambda, \delta \cdot p/8k)$
$S' \leftarrow I \cup \{e \in S \setminus I \mid I_{\hat{\mu}}^{\geq \hat{\mu}_e - \alpha - 2\lambda}$ does not block $e\}$
Return PAC-SamplePrune $(\mathcal{S}_{S'} = (S', \mathcal{M}_{S'}), \alpha, \delta/4)$

See our paper for the details!

Outline

- Introduction
- Optimal PAC Algorithm (Best-Arm, Best-k-Arm):
  - Median/Quantile Elimination
- Combinatorial Pure Exploration
- Best-Arm – Instance optimality?
- Conclusion

2 Arms (A/B test)

- Distinguish two coins（w.p. 0.999）

0.5/0.5 0.499999/0.500001

Needs approx. 10^10 samples

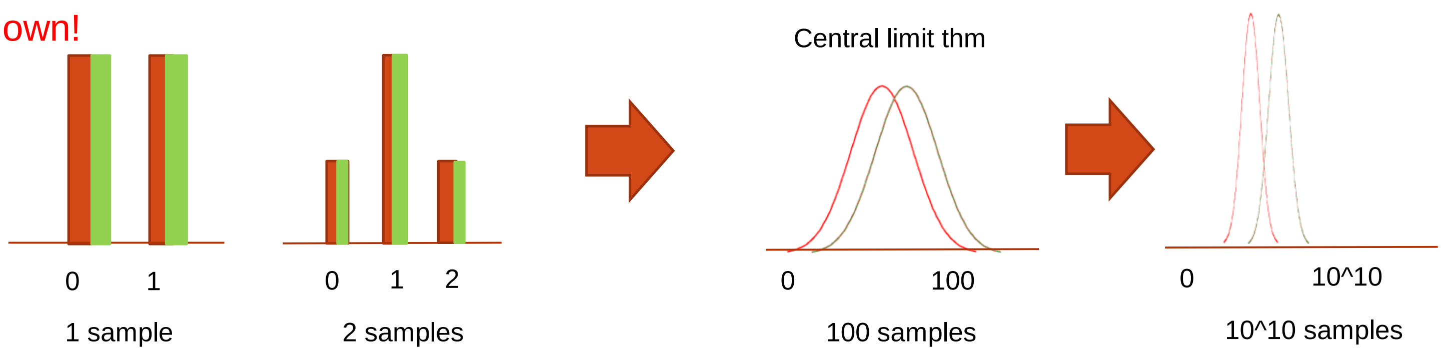
$$(\theta_1 - \theta_2)^{-2} = \Delta^{-2}$$

Sufficient：Chernoff-Hoeffding inequality

Necessary：Total variational distance/Hellinger distance

Assuming Δ is known!

1960s



2 Arms (A/B test)

- Distinguish two coins（w.p. 0.999）

0.5/0.5

0.499999/0.500001

Needs 10^10 samples

What if Δ is unknown? $\Delta^{-2}\log\log\Delta^{-1}$

Sufficient：Guess+Verify (loglog term due to union bound)
Necessary：Farrell’s lower bound in 1964
(based on Law of Iterative Logarithm)

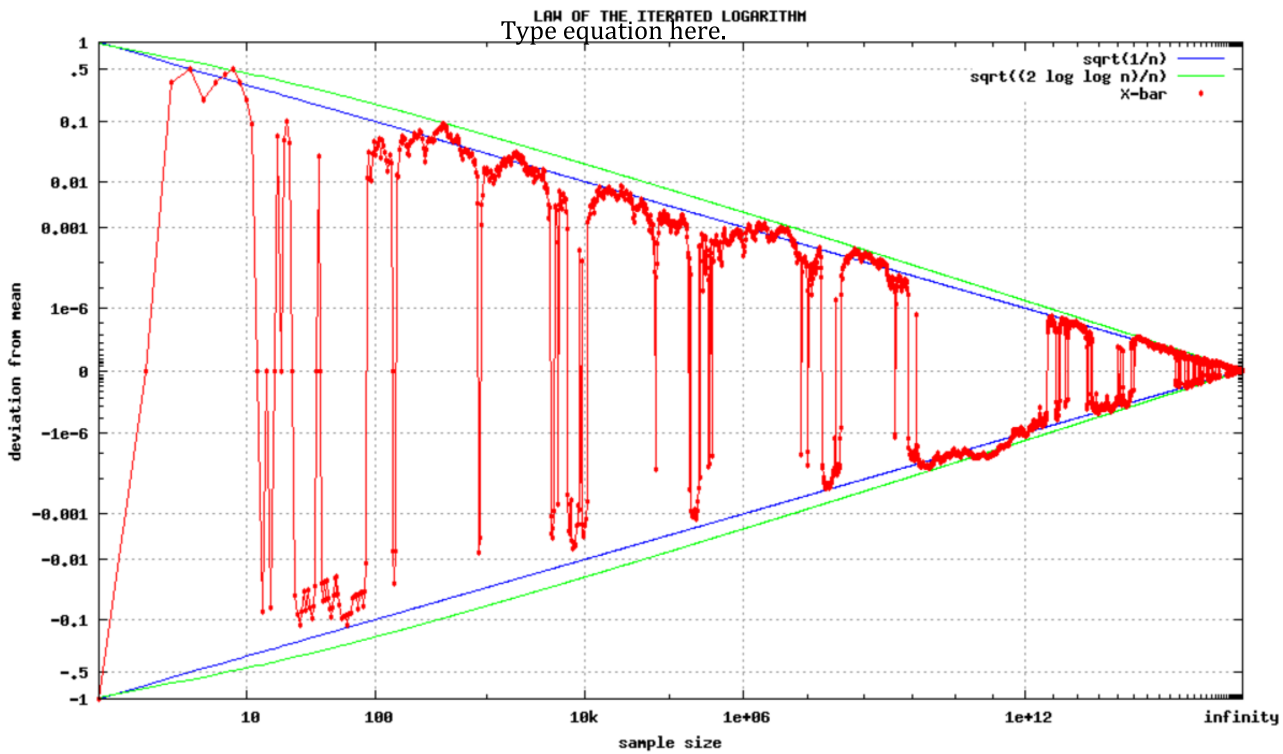
$$\limsup_{\Delta \to 0} \frac{T_{\mathbb{A}}[\Delta]}{\Delta^{-2} \ln \ln \Delta^{-1}} > 0$$

Law of Iterative Logarithm

LIL: $\limsup_t \left|\sum_{i=1}^{t} X_i\right| / \sqrt{2t \log\log t} = 1$ almost surely where $X_i \sim \mathcal{N}(0,1)$ for all i.

Type equation here.

$\frac{1}{t}\sum_{i=1}^{t} X_i$



Both axes are non-linearly transformed

2 Arms

A subtle issue:

- If $\limsup_{\Delta \to +0} T(\Delta)\Delta^2 = +\infty$

  then we can design an algorithm A such that

$$\liminf_{\Delta \to +0} \frac{T_{\mathbb{A}}(\Delta)}{T(\Delta)} = 0.$$

Hence, we cannot get a $\Delta^{-2}\log\log\Delta^{-1}$ lower bound for every instance

- No instance optimal algorithm possible
- So the story is not over! (lower bound – density result, shortly)

Best Arm Identification

- Find the best arm out of n arms, with means $\mu_{[1]}, \mu_{[n]}, .., \mu_{[n]}$
- Formulated by Bechhofer in 1954

- Again, if we want to get the exact best arm, the bound has to depend on the gaps $\Delta_{[i]} = \mu_{[1]} - \mu_{[i]}$

- Some classical results:
  - Mannor-T $\Omega\left(\sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \delta^{-1}\right)$ $\Delta_{[i]} = \mu_{[1]} - \mu_{[i]}$

It is an instance-wise lower bound

Are we done? – a misclaim

Source	Sample Complexity
Even-Dar et al. [12]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2}\left(\ln \delta^{-1} + \ln n + \ln \Delta_{[i]}^{-1}\right)$
Gabillon et al. [16]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2}\left(\ln \delta^{-1} + \ln \sum_{i=2}^{n} \Delta_{[i]}^{-2}\right)$
Jamieson et al. [19]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2}\left(\ln \delta^{-1} + \ln\ln \left(\sum_{j=2}^{n} \Delta_{[j]}^{-2}\right)\right)$
kalyanakrishnan et al. [23]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2}\left(\ln \delta^{-1} + \ln \sum_{i=2}^{n} \Delta_{[i]}^{-2}\right)$
Jamieson et al. [19]	$\ln \delta^{-1} \cdot \left(\ln\ln \delta^{-1} \cdot \sum_{i=2}^{n} \Delta_{[i]}^{-2} + \sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \Delta_{[i]}^{-1}\right)$
Karnin et al.[24], Jamieson et al.[20]	$\boxed{\sum_{i=2}^{n} \Delta_{[i]}^{-2}\left(\ln \delta^{-1} + \ln\ln \Delta_{[i]}^{-1}\right)}$
This paper (Thm 2.5)	$\sum_{i=2}^{n} \Delta_{[i]}^{-2}\left(\ln \delta^{-1} + \ln\ln \min(n, \Delta_{[i]}^{-1})\right) + \Delta_{[2]}^{-2} \ln\ln \Delta_{[2]}^{-1}$
This paper (clustered instances) Thm B.22	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \delta^{-1} + \Delta_{[2]}^{-2} \ln\ln \Delta_{[2]}^{-1}$

Mannor-Tsitsiklis lower bound: $\Omega\left(\sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \delta^{-1}\right)$

Farrell's lower bound (2 arms): $\Delta_{[2]}^{-2} \ln\ln \Delta_{[2]}^{-1}$

Attempting to believe : Karnin's upper bound is tight

Jamieson et al.: "The procedure cannot be improved in the sense that the number of samples required to identify the best arm is within a constant factor of a lower bound based on the law of the iterated logarithm (LIL)".

Are we done? – a misclaim

Source	Sample Complexity
Even-Dar et al. [12]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln n + \ln \Delta_{[i]}^{-1}\right)$
Gabillon et al. [16]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \sum_{i=2}^{n} \Delta_{[i]}^{-2}\right)$
Jamieson et al. [19]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \ln \left(\sum_{j=2}^{n} \Delta_{[j]}^{-2}\right)\right)$
kalyanakrishnan et al. [23]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \sum_{i=2}^{n} \Delta_{[i]}^{-2}\right)$
Jamieson et al. [19]	$\ln \delta^{-1} \cdot \left(\ln \ln \delta^{-1} \cdot \sum_{i=2}^{n} \Delta_{[i]}^{-2} + \sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \Delta_{[i]}^{-1}\right)$
Karnin et al.[24], Jamieson et al.[20]	$\boxed{\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \ln \Delta_{[i]}^{-1}\right)}$
This paper (Thm 2.5)	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \ln \min(n, \Delta_{[i]}^{-1})\right) + \Delta_{[2]}^{-2} \ln \ln \Delta_{[2]}^{-1}$
This paper (clustered instances) Thm B.22	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \delta^{-1} + \Delta_{[2]}^{-2} \ln \ln \Delta_{[2]}^{-1}$

Mannor-Tsitsiklis lower bound: $\Omega\left(\sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \delta^{-1}\right)$

Farrell's lower bound (2 arms): $\Delta_{[2]}^{-2} \ln \ln \Delta_{[2]}^{-1}$

Attempting to believe : Karnin's upper bound is tight

- Of course, to completely close the problem, we need to show the remaining generalization from Farrell's LB to n arms: $\sum \Delta_{[i]}^{-2} \log\log \Delta_{[i]}^{-1}$

Are we done? – a misclaim

Source	Sample Complexity
Even-Dar et al. [12]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln n + \ln \Delta_{[i]}^{-1} \right)$
Gabillon et al. [16]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \sum_{i=2}^{n} \Delta_{[i]}^{-2} \right)$
Jamieson et al. [19]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \ln \left(\sum_{j=2}^{n} \Delta_{[j]}^{-2} \right) \right)$
kalyanakrishnan et al. [23]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \sum_{i=2}^{n} \Delta_{[i]}^{-2} \right)$
Jamieson et al. [19]	$\ln \delta^{-1} \cdot \left(\ln \ln \delta^{-1} \cdot \sum_{i=2}^{n} \Delta_{[i]}^{-2} + \sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \Delta_{[i]}^{-1} \right)$
Karnin et al.[24], Jamieson et al.[20]	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \ln \Delta_{[i]}^{-1} \right)$
This paper (Thm 2.5)	$\boxed{\sum_{i=2}^{n} \Delta_{[i]}^{-2} \left(\ln \delta^{-1} + \ln \ln \min(n, \Delta_{[i]}^{-1}) \right) + \Delta_{[2]}^{-2} \ln \ln \Delta_{[2]}^{-1}}$
This paper (clustered instances) Thm B.22	$\sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \delta^{-1} + \Delta_{[2]}^{-2} \ln \ln \Delta_{[2]}^{-1}$

Mannor-Tsitsiklis lower bound: $\Omega \left(\sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \delta^{-1} \right)$

Farrell's lower bound (2 arms): $\Delta_{[2]}^{-2} \ln \ln \Delta_{[2]}^{-1}$

Attempting to believe : Karnin's upper bound is tight

- Of course, to completely close the problem, we need to show the remaining generalization from Farrell's LB to arbitrary $\sum \Delta_{[i]}^{-2} \log\log \Delta_{[i]}^{-1}$

Impossible!

New Upper and Lower Bounds

- Our new upper bound (strictly better than Karnin's)

$$O\Big(\underbrace{\Delta_{[2]}^{-2}\ln\ln\Delta_{[2]}^{-1}}_{\text{Farrell's LB}} + \underbrace{\sum\nolimits_{i=2}^{n}\Delta_{[i]}^{-2}\ln\delta^{-1}}_{\text{M-T LB}} + \underbrace{\sum\nolimits_{i=2}^{n}\Delta_{[i]}^{-2}\ln\ln\min(n,\Delta_{[i]}^{-1})}_{\ln\ln n \text{ term seems strange........}}\Big)$$

New Upper and Lower Bounds

- Our new upper bound (strictly better than Karnin's)

$$O\Big(\underbrace{\Delta_{[2]}^{-2}\ln\ln\Delta_{[2]}^{-1}}_{\text{Farrell's LB}} + \underbrace{\sum_{i=2}^{n}\Delta_{[i]}^{-2}\ln\delta^{-1}}_{\text{M-T LB}} + \underbrace{\sum_{i=2}^{n}\Delta_{[i]}^{-2}\ln\ln\min(n,\Delta_{[i]}^{-1})}_{\text{lnln}n\text{ term seems strange........}}\Big)$$

- It turns out the lnln*n* term is fundamental.
- Our new lower bound (not instance-wise)

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

High Level Idea of Our Algorithm

- Sketch of ExpGap-Halving [Karnin et al.]

ExpGap-Halving
$\boldsymbol{r = 1}$
Repeat
  $\epsilon_r = O(2^{-r})$
  Find an ϵ_r-optimal arm a_r using Median-Elimination
  Estimate $u_{[a_r]}$
  Uniformly sample all remaining arms
  Eliminate arms with empirical means $\leq \hat{u}_{[a_r]}$
  $r = r + 1$
Until S is a singleton

High Level Idea of Our Algorithm

- Sketch of ExpGap-Halving [Karnin et al.]

Several previous elimination algorithms, e.g., eliminate ½ arms, eliminate arms below a threshold.
This is the most aggressive one.

ExpGap-Halving
$\boldsymbol{r = 1}$
Repeat
  $\epsilon_r = O(2^{-r})$
  Find an ϵ_r-optimal arm a_r using Median-Elimination
  Estimate $u_{[a_r]}$
  Uniformly sample all remaining arms
  Eliminate arms with empirical means $\leq \hat{u}_{[a_r]}$
  $r = r + 1$
Until S is a singleton

High Level Idea of Our Algorithm

- Our idea

ExpGap-Halving
$\boldsymbol{r = 1}$
Repeat
  $\epsilon_r = O(2^{-r})$
  Find an ϵ_r-optimal arm a_r using Median-Elimination
  Estimate $u_{[a_r]}$
  Uniformly sample all remaining arms
  Eliminate arms with empirical means $\leq \hat{u}_{[a_r]}$
  $r = r + 1$
Until S is a singleton

Can be wasteful if we can't eliminate a lot of arms.

Don't be too aggressive. Do elimination only when we have a lot of arms to eliminate.

High Level Idea of Our Algorithm

Do elimination only when we have a lot of arms to eliminate.

Do this test by **Sampling arms**

DistrBasedElimination
$\boldsymbol{r = 1}$
Repeat
    $\epsilon_r = O(2^{-r})$
    Find an ϵ_r-optimal arm a_r using Median-Elimination
    Estimate $u_{[a_r]}$
    If (we can eliminate a lot of arms)
        Uniformly sample all remaining arms
        Eliminate arms with empirical means $\leq \hat{u}_{[a_r]}$
    else
        Don't do anything
    $r = r + 1$
Until S is a singleton

Our Algorithm

- A lot of details
- The analysis is intricate – need a potential function to amortize the cost

Algorithm 3: FractionTest$(S, c_l, c_r, \delta, t, \varepsilon)$

Data: Arm set S, range parameters c_l, c_r, confidence level δ, threshold t, approximate parameter ε.

cnt $\leftarrow 0$
tot $\leftarrow \ln(2 \cdot \delta^{-1})(\varepsilon/3)^{-2}/2$
for $i = 1$ *to* tot **do**
  Pick a random arm $a_i \in S$ uniformly.
  $\hat{\mu}_{[a_i]} \leftarrow$ UniformSample$(\{a_i\}, (c_r - c_l)/2, \varepsilon/3)$
  if $\hat{\mu}_{[a_i]} < (c_l + c_r)/2$ **then** cnt $\leftarrow$ cnt $+ 1$
if cnt/tot $> t$ **then**
  Return True
else
  Return False

Algorithm 4: Elimination(S, c_l, c_r, δ)

Data: Arm set S, range parameters c_l, c_r, confidence level δ.
Result: A set of arms after elimination.

$S_1 \leftarrow S$
$c_m \leftarrow (c_l + c_r)/2$
for $r = 1$ *to* $+\infty$ **do**
  $\delta_r = \delta/(10 \cdot 2^r)$
  if FractionTest$(S_r, c_l, c_m, \delta_r, 0.075, 0.025)$ **then**
    UniformSample$(S_r, (c_r - c_m)/2, \delta_r)$
    $S_{r+1} \leftarrow \left\{a \in S_r \mid \hat{\mu}_{[a]} > (c_m + c_r)/2\right\}$
  else
    Return S_r

Algorithm 1: DistrBasedElim(S, δ)

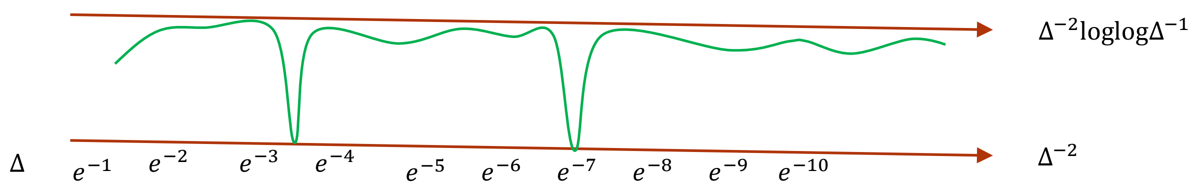
$h \leftarrow 1$
$S_1 \leftarrow S$
for $r = 1$ *to* $+\infty$ **do**
  if $|S_r| = 1$ **then**
    Return the only arm in S_r
  $\varepsilon_r \leftarrow 2^{-r}$
  $\delta_r \leftarrow \delta/50r^2$
  $a_r \leftarrow$ MedianElim$(S_r, \varepsilon_r/4, 0.01)$.
  $\hat{\mu}_{[a_r]} \leftarrow$ UniformSample$(\{a_r\}, \varepsilon_r/4, \delta_r)$
  if FractionTest$(S_r, \hat{\mu}_{[a_r]} - 1.5\varepsilon_r, \hat{\mu}_{[a_r]} - 1.25\varepsilon_r, \delta_r, 0.4, 0.1)$ **then**
    $\delta_h \leftarrow \delta/50h^2$
    $b_r \leftarrow$ MedianElim$(S_r, \varepsilon_r/4, \delta_h)$
    $\hat{\mu}_{[b_r]} \leftarrow$ UniformSample$(\{b_r\}, \varepsilon_r/4, \delta_h)$
    $S_{r+1} \leftarrow$ Elimination$(S_r, \hat{\mu}_{[b_r]} - 0.5\varepsilon_r, \hat{\mu}_{[b_r]} - 0.25\varepsilon_r, \delta_h)$
    $h \leftarrow h + 1$
  else
    $S_{r+1} \leftarrow S_r$

Our Lower Bound

- (almost) all previous lower bound for bestarm (even best-k-arm) can be seen as a directed sum result:
  - Solving the bestarm is as hard as solving n copies of 2 arm problems
    - E.g., Mannor-Tsitsiklis lower bound: $\Omega\left(\sum_{i=2}^{n} \Delta_{[i]}^{-2} \ln \delta^{-1}\right)$
    - We can (randomly) embed a 2-arm instance in an n-arm instance
    - By the lower bound of 2-arm, we can show an lower bound for n-arm

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

- However, our new lower bound is NOT a directed sum result!
  - Solving the bestarm is HARDER than solving n copies of 2 arm problems!
  - One subtlety: an 2-arm instance does NOT have a $\Delta^{-2}\mathrm{loglog}\Delta^{-1}$ lower bound!
  - We need a "density" $\Delta^{-2}\mathrm{loglog}\Delta^{-1}$ lower bound for 2 arms as the basis



Any algorithm must be slow for most Δ

  - We also need a more involved embedding argument to take advantage of the above density result

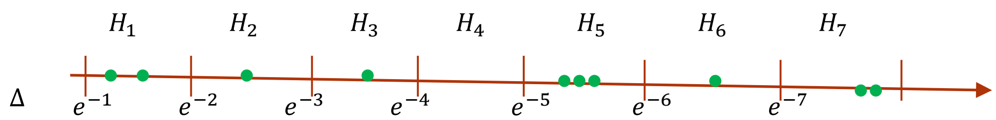
Outline

- Introduction
- Optimal PAC Algorithm (Best-Arm, Best-k-Arm):
  - Median/Quantile Elimination
- Combinatorial Pure Exploration
- Best-Arm – Instance optimality
- Conclusion

Open Question

- (almost) Instance optimal algorithm for best arm

$$H_i = \sum_{u \in G_i} \Delta_{[u]}^{-2}$$



- Gap Entropy: $\mathsf{Ent}(I) = \sum_{G_i \neq \emptyset} p_i \log p_i^{-1}.$ $\quad p_i = H_i / \sum_j H_j.$

- Gap Entropy Conjecture:
  - An instance-wise lower bound $\mathcal{L}(I, \delta) = \Theta\left(H(I)(\ln \delta^{-1} + \mathsf{Ent}(I))\right).$

    $H(I) = \sum_{i=2}^{n} \Delta_{[i]}^{-2}.$

  - An algorithm with sample complexity:

$$O\left(\mathcal{L}(I, \delta) + \Delta_{[2]}^{-2} \ln \ln \Delta_{[2]}^{-1}\right).$$

Future Direction

- Learning + Stochastic Optimization
  - Online/Bandit convex optimization
  - Bayesian mechanism design without full distr. infor.
  - A LOT of problems in this domain

Thanks.

lapordge@gmail.com

Ref

- Farrell. Asymptotic behavior of expected sample size in certain one sided tests. The Annals of Mathematical Statistics 1964
- E. Even-Dar, S. Mannor, and Y. Mansour. Pac bounds for multi-armed bandit and markov decision processes. In COLT 2002
- S. Mannor and J. N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. JMLR, 2004
- Z. Karnin, T. Koren, and O. Somekh. Almost optimal exploration in multi-armed bandits. In ICML, 2013
- K. Jamieson, M. Malloy, R. Nowak, and S. Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. COLT, 2014
- S. Chen, T. Lin, I. King, M. R. Lyu, and W. Chen. Combinatorial pure exploration of multi-armed bandits.In NIPS, 2014
- Y. Zhou, X. Chen, and J. Li. Optimal pac multiple arm identification with applications to crowdsourcing. In ICML2014
- W. Cao, J. Li, Y. Tao, and Z. Li. On top-k selection in multi-armed bandits and hidden bipartite graphs. In NIPS 2015
- L. Chen, J. Li. On the Optimal Sample Complexity for Best Arm Identification, ArXiv, 2016
- L. Chen, A. Gupta, and J. Li. Pure exploration of multi-armed bandit under matroid constraints. In COLT2016.