An Asymptotically Optimal Bandit Algorithm for Bounded Support Models

Junya Honda and Akimichi Takemura
The University of Tokyo

COLT 2010

Outline

- Introduction
- DMED policy
  - Proof of the optimality
  - Efficient computation
- Simulation results
- Conclusion

Outline

- Introduction
- DMED policy
  - Proof of the optimality
  - Efficient computation
- Simulation results
- Conclusion

Multiarmed bandit problem

- Model of a gambler playing a slot machine with multiple arms
- Example of a dilemma between exploration and exploitation

- K-armed stochastic bandit problem
  – Burnates-Katehakis derived an asymptotic bound of the regret

- Model of reward distributions with support in [0,1]
  – UCB policies by Auer et al. are widely used practically
  – Bound-achieving policies have not been known

  – We propose DMED policy, which achieves the bound

Notation

$\mathcal{A}$: family of distributions with support in [0,1]

$F_i \in \mathcal{A}$: probability distribution of arm $i = 1, \cdots, K$

$\mu_i = \mathrm{E}(F_i)$: expectation of arm i

($\mathrm{E}(F)$: expectation of distribution F)

$\mu^* = \max_i \mu_i$: maximum expectation of arms

$T_i(n)$: # of times that arm i has been pulled through the first n rounds

Goal: minimize the regret

$$\sum_{i:\mu_i<\mu^*} (\mu^* - \mu_i) T_i(n)$$

by reducing each $T_i(n)$ for suboptimal arm i

Asymptotic bound

Burnetas and Katehakis (1996)

- Under any policy satisfying a mild condition (consistency), for all $\boldsymbol{F} = (F_1, \cdots, F_K) \in \mathcal{A}^K$ and suboptimal i

$$\mathrm{E}_{\boldsymbol{F}}[T_i(n)] \geq \left(\frac{1}{D_{\min}(F_i, \mu^*)} - \mathrm{o}(1) \right) \log n$$

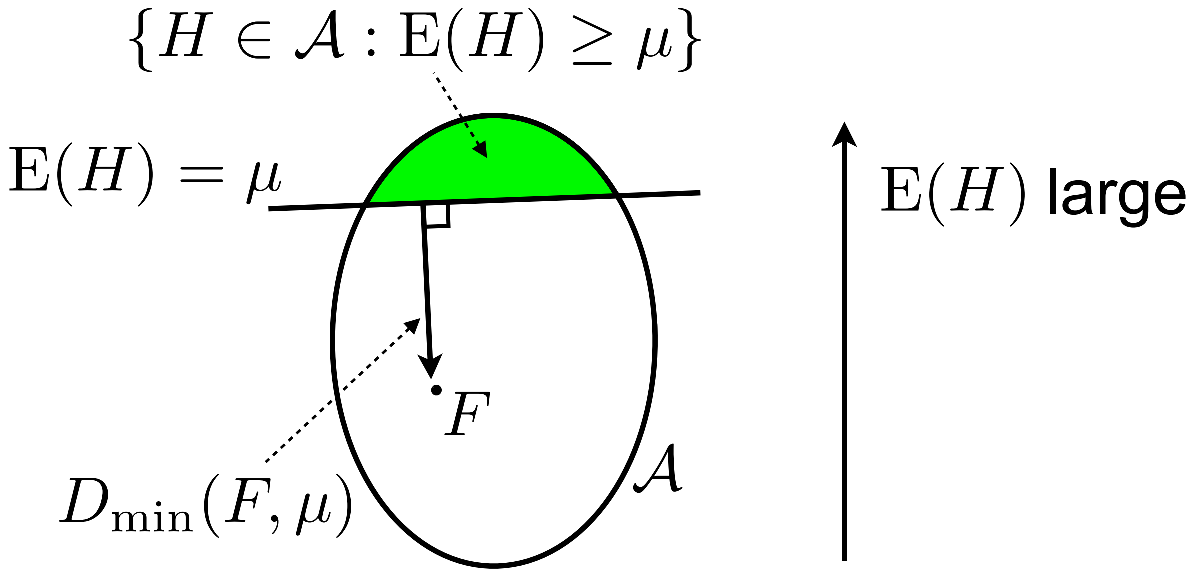
where

$$D_{\min}(F, \mu) = \min_{H \in \mathcal{A} : \mathrm{E}(H) \geq \mu} D(F||H)$$

$$D(F||H) = \mathrm{E}_F\left[\log \frac{\mathrm{d}F}{\mathrm{d}H} \right]$$: Kullback-Leibler divergence

Visualization of $D_{\min}$

$$D_{\min}(F,\mu) = \min_{H \in \mathcal{A}: \mathrm{E}(H) \geq \mu} D(F||H)$$



Outline

- Introduction
- DMED policy
  - Proof of the optimality
  - Efficient computation
- Simulation results
- Conclusion

DMED policy

- Deterministic Minimum Empirical Divergence policy

For each loop, DMED chooses arms to pull in this way:

1. For each arm i, check the condition

empirical distribution of arm i at the n-th round

$$T_i(n) D_{\min}(\hat{F}_i(n), \hat{\mu}^*(n)) \leq \log n$$

maximum sample mean at the n-th round

(The condition is always true for the currently best arm)

2. Pull all of arms such that the condition is true

Main theorem

Under DMED policy, for all suboptimal arm i ,

$$\mathrm{E}_{\boldsymbol{F}}[T_i(n)] \leq \left(\frac{1}{D_{\min}(F_i, \mu^*)} + \mathrm{o}(1) \right) \log n$$

Asymptotic bound：

$$\mathrm{E}_{\boldsymbol{F}}[T_i(n)] \geq \left(\frac{1}{D_{\min}(F_i, \mu^*)} - \mathrm{o}(1) \right) \log n$$

DMED is asymptotically optimal

Intuitive interpretation (1)

- Assume $K = 2$ and consider the event
  - $\hat{\mu}_1(n) < \hat{\mu}_2(n) = \hat{\mu}^*(n)$
  - $T_1(n) \ll T_2(n)$
- How likely is arm 1 actually the best?

  - $\mu_2 \approx \hat{\mu}_2$ is far more likely than $\mu_1 \approx \hat{\mu}_1$

- How likely is the hypothesis $\mu_1 \geq \hat{\mu}_2$?

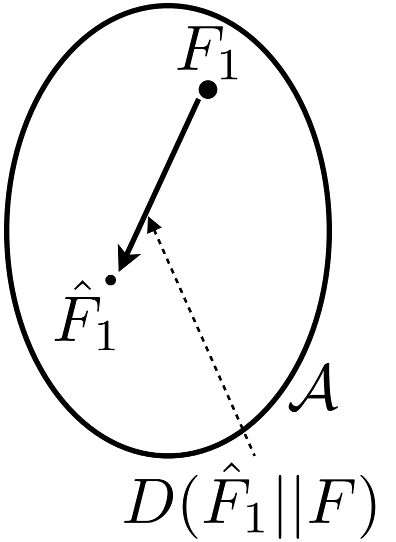
Intuitive interpretation (2)

- By Sanov's theorem in the large deviation theory,

$$P[\text{empirical distribution from } F_1 \text{ come close to } \hat{F}_1]$$

$$\approx \exp(-\boxed{T_1(n)} D(\hat{F}_1 || F_1))$$

number of samples



Intuitive interpretation (2)

- By Sanov's theorem in the large deviation theory,

$$P[\text{empirical distribution from } F_1 \text{ come close to } \hat{F}_1]$$

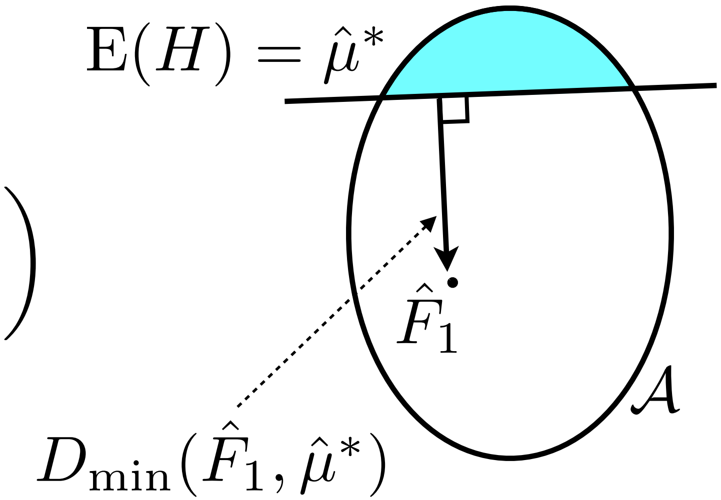
$$\approx \exp(-T_1(n) D(\hat{F}_1 || F_1))$$

- Maximum likelihood of $\mu_1 \geq \hat{\mu}^*$ is

$$\max_{H \in \mathcal{A}:\mathrm{E}(H) \geq \hat{\mu}^*} \exp(-T_1(n) D(\hat{F}_1 || H))$$

$$= \exp\left(-T_1(n) \min_{H \in \mathcal{A}:\mathrm{E}(H) \geq \hat{\mu}^*} D(\hat{F}_1 || H)\right)$$

$$= \exp(-T_1(n) D_{\min}(\hat{F}_1, \hat{\mu}^*))$$



Intuitive interpretation (3)

- Maximum likelihood that arm i is actually the best:

$$\exp(-T_i(n)D_{\min}(\hat{F}_i, \hat{\mu}^*))$$

- In DMED policy, arm i is pulled when

$$T_i(n)D_{\min}(\hat{F}_i, \hat{\mu}^*) \leq \log n$$

  - Arm i is pulled if
    - the maximum likelihood is large
    - round number n is large

Outline

- Introduction
- DMED policy
  - Proof of the optimality
  - Efficient computation
- Simulation results
- Conclusion

Proof of the optimality

- Assume $K = 2$ and $\mu_2 < \mu_1 = \mu^*$ (arm 1 is the best)
- Two events are essential for the proof:

  A_n : Estimators $\hat{F}_i(n), \hat{\mu}_i(n)$ are already close to F_i, μ_i

  B_n : $\hat{\mu}_2(n) \approx \mu_2$, but $\hat{\mu}_1(n) < \mu_2 \, (< \mu_1)$ (arm 1 seems inferior)

"Arm 2 is pulled at the n-th round"

$$T_2(N) = \sum_{n=1}^{N} \Big(\mathrm{I}[\{J_n = 2\} \cap A_n] + \mathrm{I}[\{J_n = 2\} \cap B_n] + \mathrm{I}[\{J_n = 2\} \cap A_n^c \cap B^c] \Big)$$

arm pulled at the n-th round

Proof of the optimality

- Assume $K = 2$ and $\mu_2 < \mu_1 = \mu^*$ (arm 1 is the best)
- Two events are essential for the proof:

A_n : Estimators $\hat{F}_i(n), \hat{\mu}_i(n)$ are already close to F_i, μ_i

B_n : $\hat{\mu}_2(n) \approx \mu_2$, but $\hat{\mu}_1(n) < \mu_2 \,(< \mu_1)$ (arm 1 seems inferior)

$$T_2(N) = \sum_{n=1}^{N} \Bigg(\overset{\displaystyle \frac{\log n}{D_{\min}(F_2, \mu_1)}}{\overset{\wr\wr}{\mathrm{I}[\{J_n = 2\} \cap A_n]}} + \overset{\mathrm{O}(1)}{\overset{\|}{\mathrm{I}[\{J_n = 2\} \cap B_n]}} + \underset{\mathrm{O}(1)}{\underset{\|}{\mathrm{I}[\{J_n = 2\} \cap A_n^c \cap B^c]}} \Bigg)$$

After the convergence

- Arm 2 is pulled when $T_2(n) D_{\min}(\hat{F}_2(n), \hat{\mu}^*(n)) \leq \log n$

- On the event A_n, $D_{\min}(\hat{F}_2(n), \hat{\mu}^*(n)) \approx D_{\min}(F_2, \mu^*)$ holds because $D_{\min}(F, \mu)$ is continuous

⇨ If A_n is true, arm 2 is pulled only while

$$T_2(n) \lesssim \frac{\log n}{D_{\min}(F_2, \mu^*)}$$

is true.

⇨

$$\sum_{n=1}^{N} \mathrm{I}[\{J_n = 2\} \cap A_n] \lesssim \frac{\log N}{D_{\min}(F_2, \mu^*)}$$

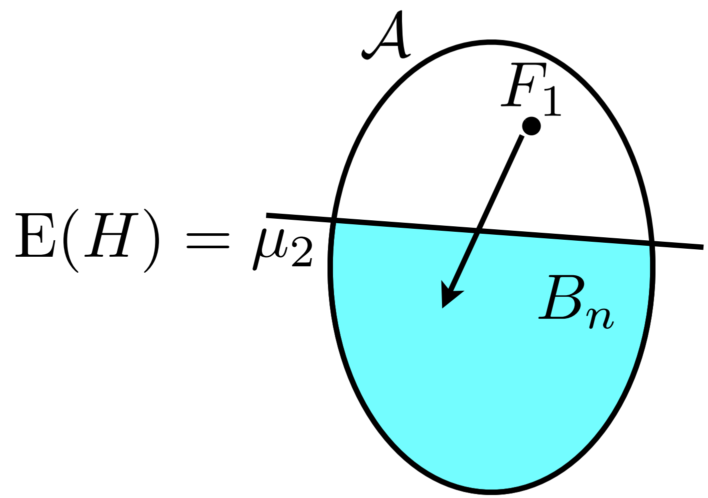
Before the convergence (1)

- B_n: $\hat{\mu}_2 \approx \mu_2$ and $\hat{\mu}_1 < \mu_2 \, (< \mu_1)$
- We will show

$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[\{J_n = 2\} \cap B_n]\right] = \mathrm{O}(1)$$

$$\leq$$

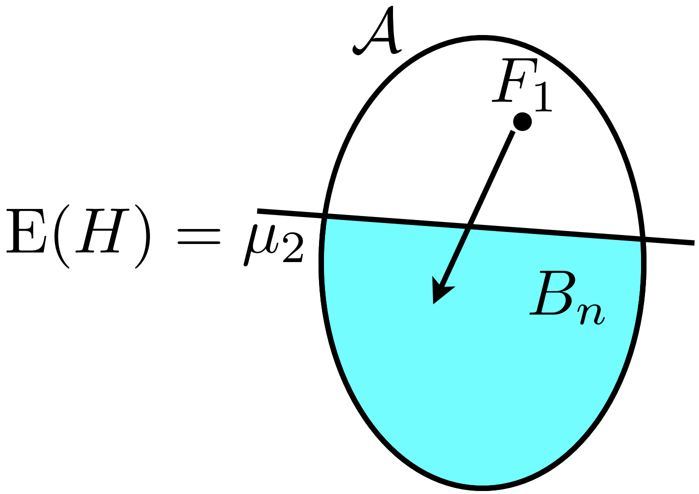
$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n]\right]$$



Before the convergence (1)

- B_n: $\hat{\mu}_2 \approx \mu_2$ and $\hat{\mu}_1 < \mu_2\,(< \mu_1)$
- We will show

$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n]\right] = \mathrm{O}(1)$$

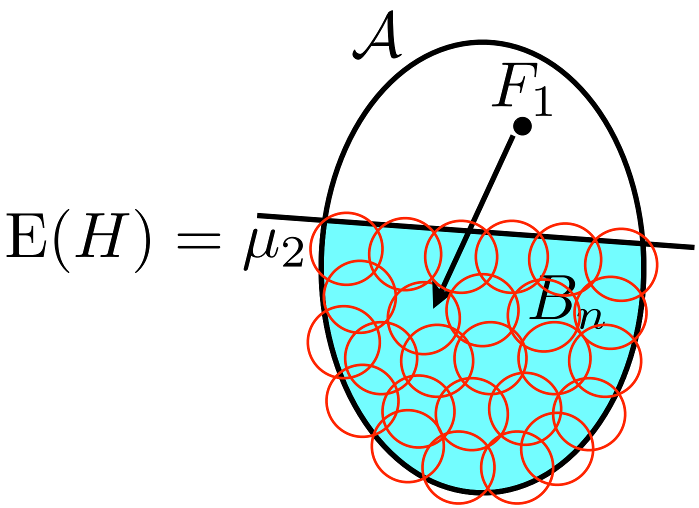


- Focus on $\hat{F}_1(n)$ of the event B_n
- $\mathcal{A}$ is compact (w.r.t. Lévy distance)

Before the convergence (1)

- B_n: $\hat{\mu}_2 \approx \mu_2$ and $\hat{\mu}_1 < \mu_2 \, (< \mu_1)$
- We will show

$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n]\right] = \mathrm{O}(1)$$

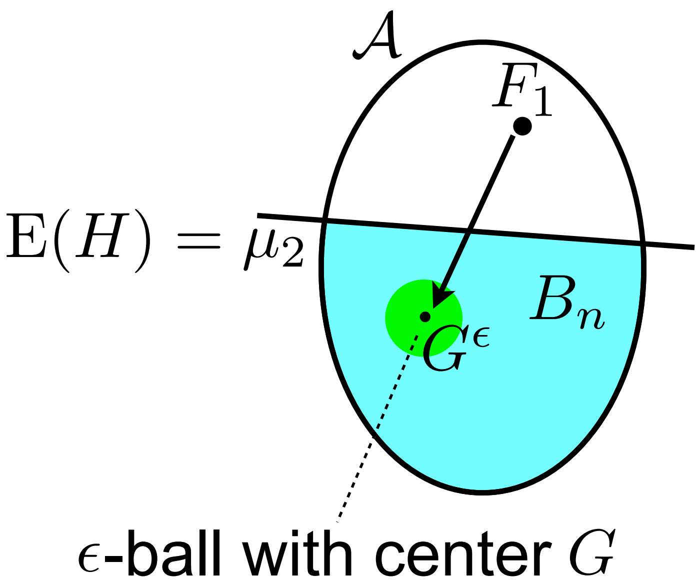


- Focus on $\hat{F}_1(n)$ of the event B_n
- $\mathcal{A}$ is compact (w.r.t. Lévy distance)

Before the convergence (1)

- B_n: $\hat{\mu}_2 \approx \mu_2$ and $\hat{\mu}_1 < \mu_2 \, (< \mu_1)$
- We will show

$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n]\right] = \mathrm{O}(1)$$



- Focus on $\hat{F}_1(n)$ of the event B_n
- $\mathcal{A}$ is compact (w.r.t. Lévy distance)

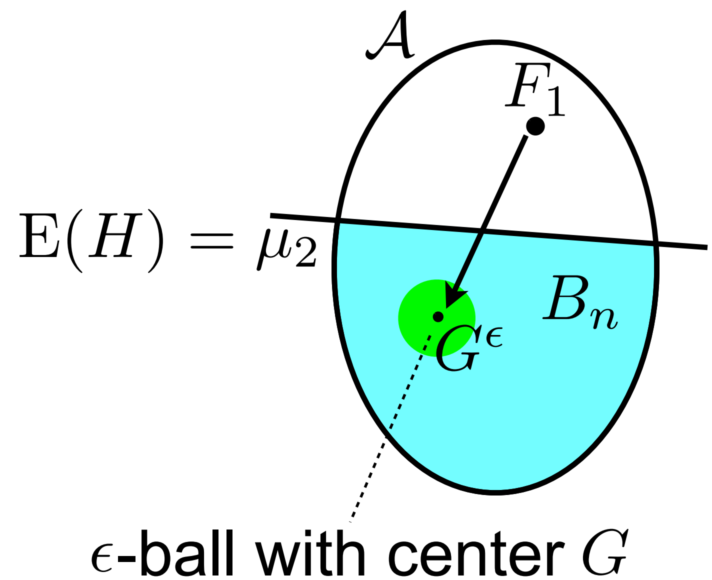
⇨ It is sufficient to show for arbitrary $G \in \mathcal{A}$ s.t. $\mathrm{E}(G) \leq \mu_2$

$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n \cap \{\hat{F}_1(n) \in G^{\epsilon}\}]\right] = \mathrm{O}(1)$$

Before the convergence (1)

- B_n: $\hat{\mu}_2 \approx \mu_2$ and $\hat{\mu}_1 < \mu_2\,(< \mu_1)$
- We will show

$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n]\right] = \mathrm{O}(1)$$



- Focus on $\hat{F}_1(n)$ of the event B_n
- $\mathcal{A}$ is compa[...] (w.r.t. Lévy distance)

➡ It is suffici[...]t to show for arbitrary $G \in \mathcal{A}$ s.t. $\mathrm{E}(G) \le \mu_2$

Take the summation over finite balls

$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n \cap \{\hat{F}_1(n) \in G^\epsilon\}]\right] = \mathrm{O}(1)$$

Before the convergence (2)

- B_n: $\hat{\mu}_2 \approx \mu_2$ and $\hat{\mu}_1 < \mu_2 \, (< \mu_1)$
- We will show

$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n \cap \{\hat{F}_1(n) \in G^{\epsilon}\}]\right] = \mathrm{O}(1)$$

$$\leq$$

$$\sum_{t=1}^{\infty} \mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n \cap \{\hat{F}_1(n) \in G^{\epsilon}\} \cap \{T_1(n) = t\}]\right]$$

Before the convergence (3)

- We will show

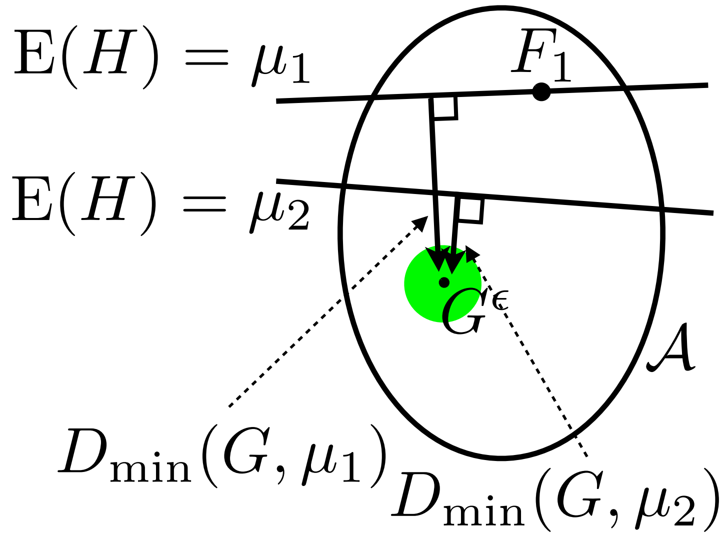
$$\sum_{t=1}^{\infty} \mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n \cap \{\hat{F}_1(n) \in G^{\epsilon}\} \cap \{T_1(n) = t\}]\right] = \mathrm{O}(1)$$

$$\begin{aligned}
&\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n \cap \{\hat{F}_1(n) \in G^{\epsilon}\} \cap \{T_1(n) = t\}]\right] \\
&\quad \le P_{F_1}[\{\hat{F}_1(n) \in G^{\epsilon}\} \cap \{T_1(n) = t\}] \\
&\qquad \times \max\left\{\sum_{n=1}^{N} \mathrm{I}[B_n \cap \{\hat{F}_1(n) \in G^{\epsilon}\} \cap \{T_1(n) = t\}]\right\} \\
&\quad \le \exp\Big(- t\big(D_{\min}(G, \mu_1) - D_{\min}(G, \mu_2)\big)\Big)
\end{aligned}$$

Before the convergence (4)

$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n \cap \{\hat{F}_1(n) \in G^\epsilon\} \cap \{T_1(n) = t\}]\right]$$

$$\leq \exp\Big(-t\big(D_{\min}(G,\mu_1) - D_{\min}(G,\mu_2)\big)\Big)$$

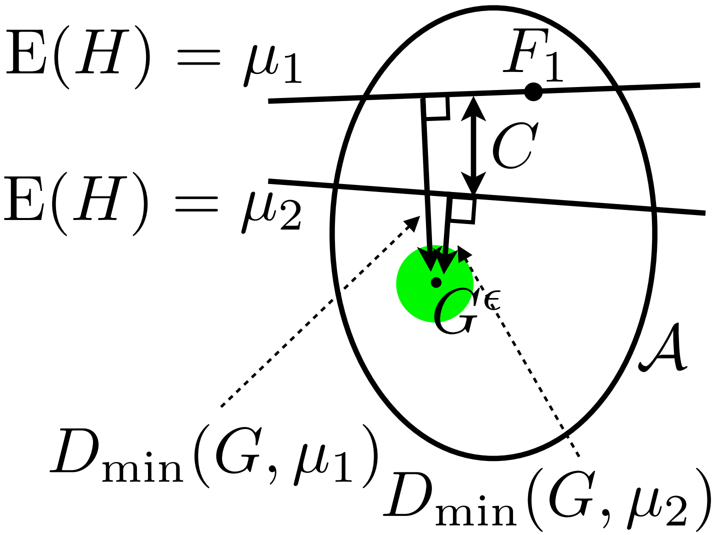


Before the convergence (4)

$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n \cap \{\hat{F}_1(n) \in G^{\epsilon}\} \cap \{T_1(n) = t\}]\right]$$

$$\leq \exp\Big(- t\big(D_{\min}(G, \mu_1) - D_{\min}(G, \mu_2)\big)\Big)$$

$$\leq \exp(-t\, C)$$



Before the convergence (4)

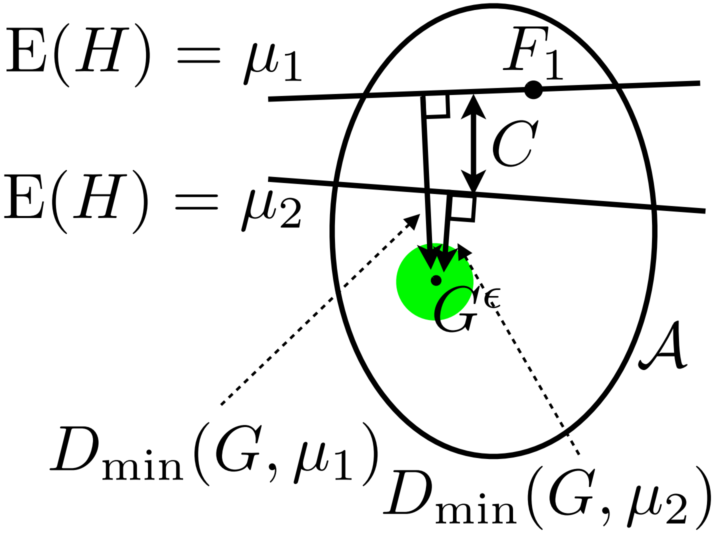
$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n \cap \{\hat{F}_1(n) \in G^{\epsilon}\} \cap \{T_1(n) = t\}]\right]$$

$$\leq \exp\Big(- t\big(D_{\min}(G, \mu_1) - D_{\min}(G, \mu_2)\big)\Big)$$

$$\leq \exp(-t\,C)$$

- By taking the summation over t,

$$\mathrm{E}\left[\sum_{n=1}^{N} \mathrm{I}[B_n \cap \{\hat{F}_1(n) \in G^{\epsilon}\}]\right] = \mathrm{O}(1)$$



Outline

- Introduction
- DMED policy
  - Proof of the optimality
  - Efficient computation
- Simulation results
- Conclusion

Computation of $D_{\min}$

- $D_{\min}(\hat{F}_i(n), \hat{\mu}^*(n))$ has to be computed at each round
- $D_{\min}$ is represented as

$$D_{\min}(F,\mu) \equiv \min_{H \in \mathcal{A}:\mathrm{E}(H) \geq \mu} D(F||G)$$

$$= \max_{0 \leq \nu \leq \frac{1}{1-\mu}} \mathrm{E}_F[\log(1-(X-\mu)\nu)]$$

  - univariate convex optimization problem
  - efficiently computable by e.g. Newton's method
  - ν^*_{n-1} is a good approximation of current ν^*_n

The optimal solution for the $n-1$-st round

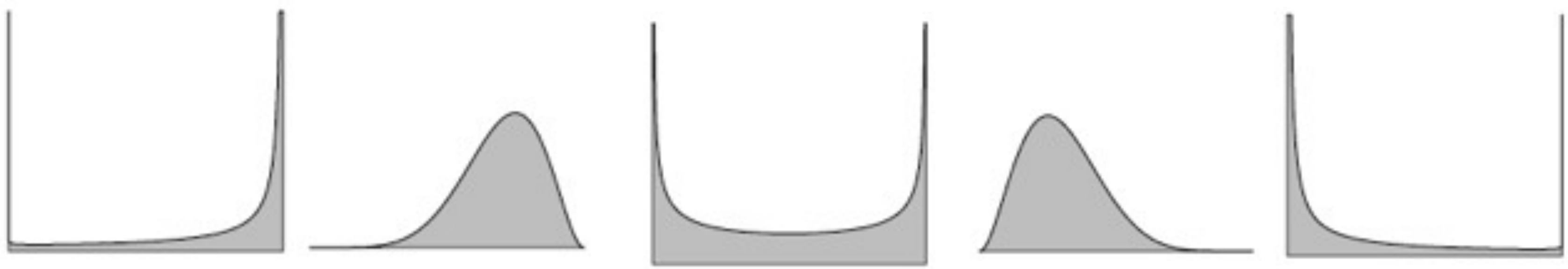
Outline

- Introduction
- DMED policy
  - Proof of the optimality
  - Efficient computation
- Simulation results
- Conclusion

Simulation 1

- $K = 5$, beta distributions

F_1	F_2	F_3	F_4	F_5
$\text{Be}(0.9, 0.1)$	$\text{Be}(7, 3)$	$\text{Be}(0.5, 0.5)$	$\text{Be}(3, 7)$	$\text{Be}(0.1, 0.9)$



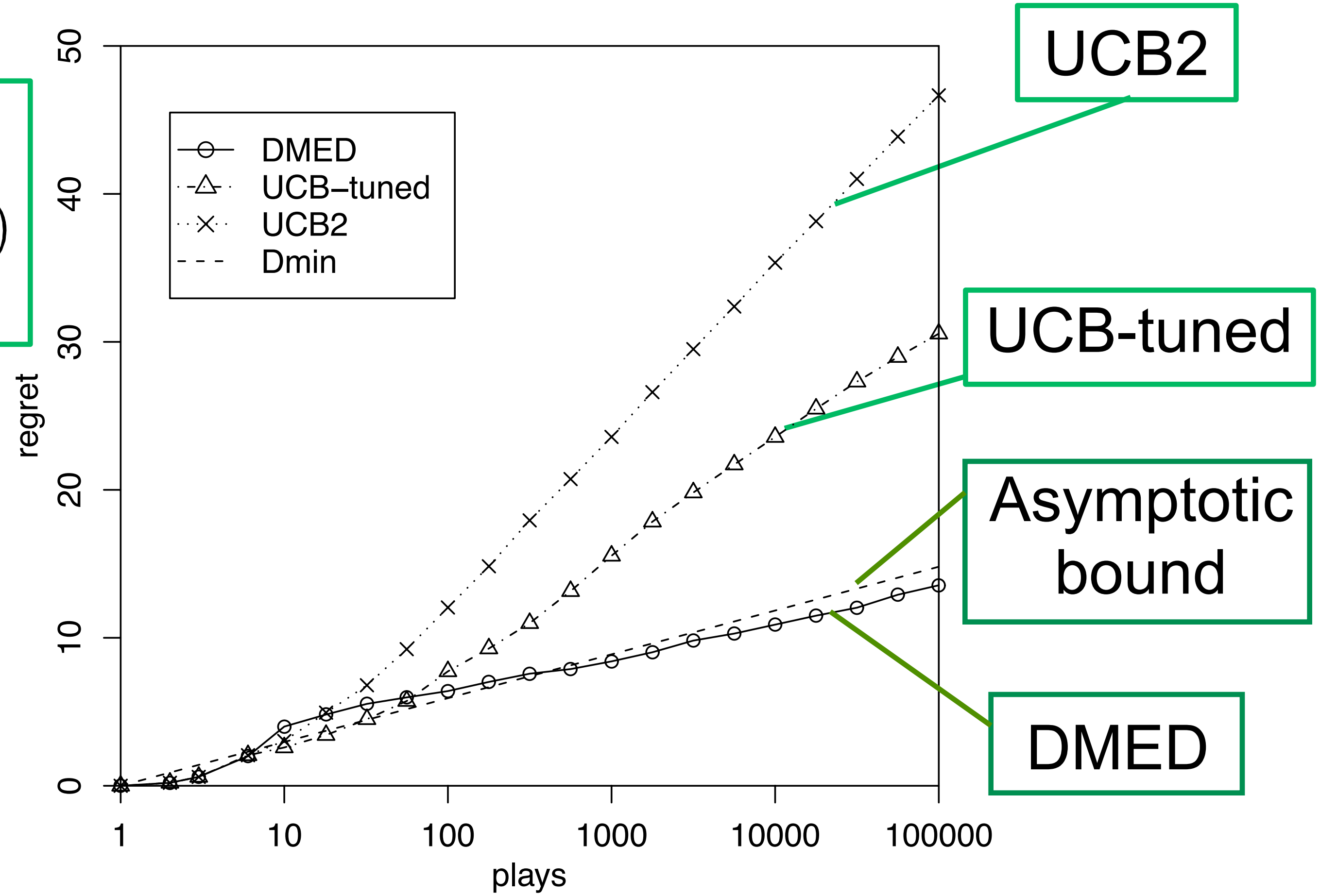
$\mu_1 = 0.9$	$\mu_2 = 0.7$	$\mu_3 = 0.5$	$\mu_4 = 0.3$	$\mu_5 = 0.1$

simple distributions on [0,1]

Simulation result 1

regret:

$$\sum_{i:\mu_i<\mu^*} (\mu^* - \mu_i) T_i(n)$$

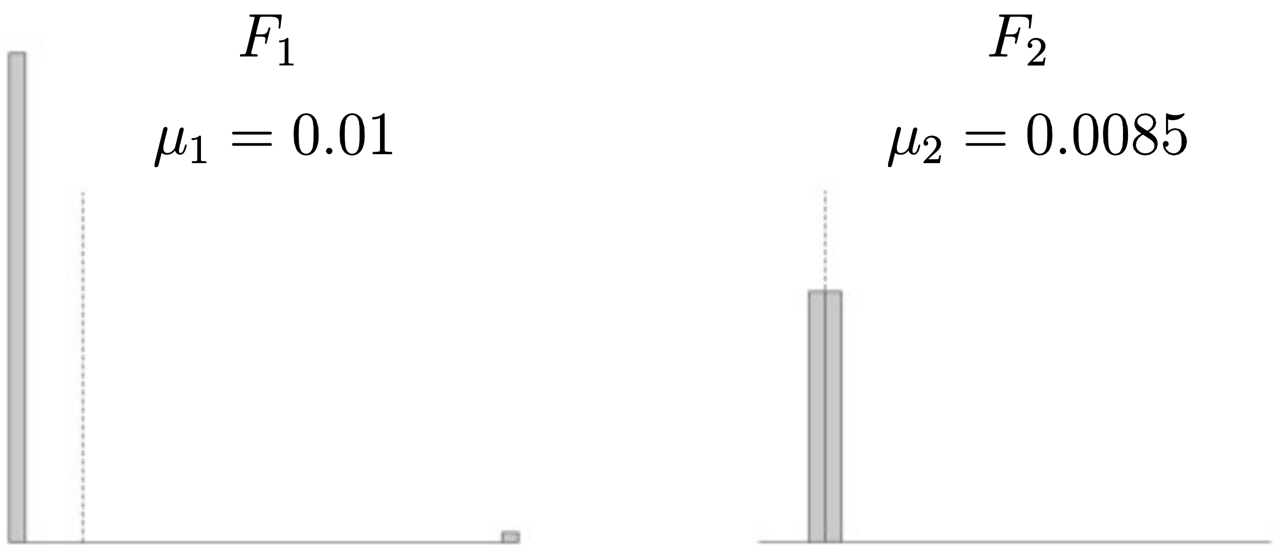


- Asymptotic slope of the regret is always larger than or equal to that of “Asymptotic bound”
- DMED seems to be achieving the asymptotic bound

Simulation 2

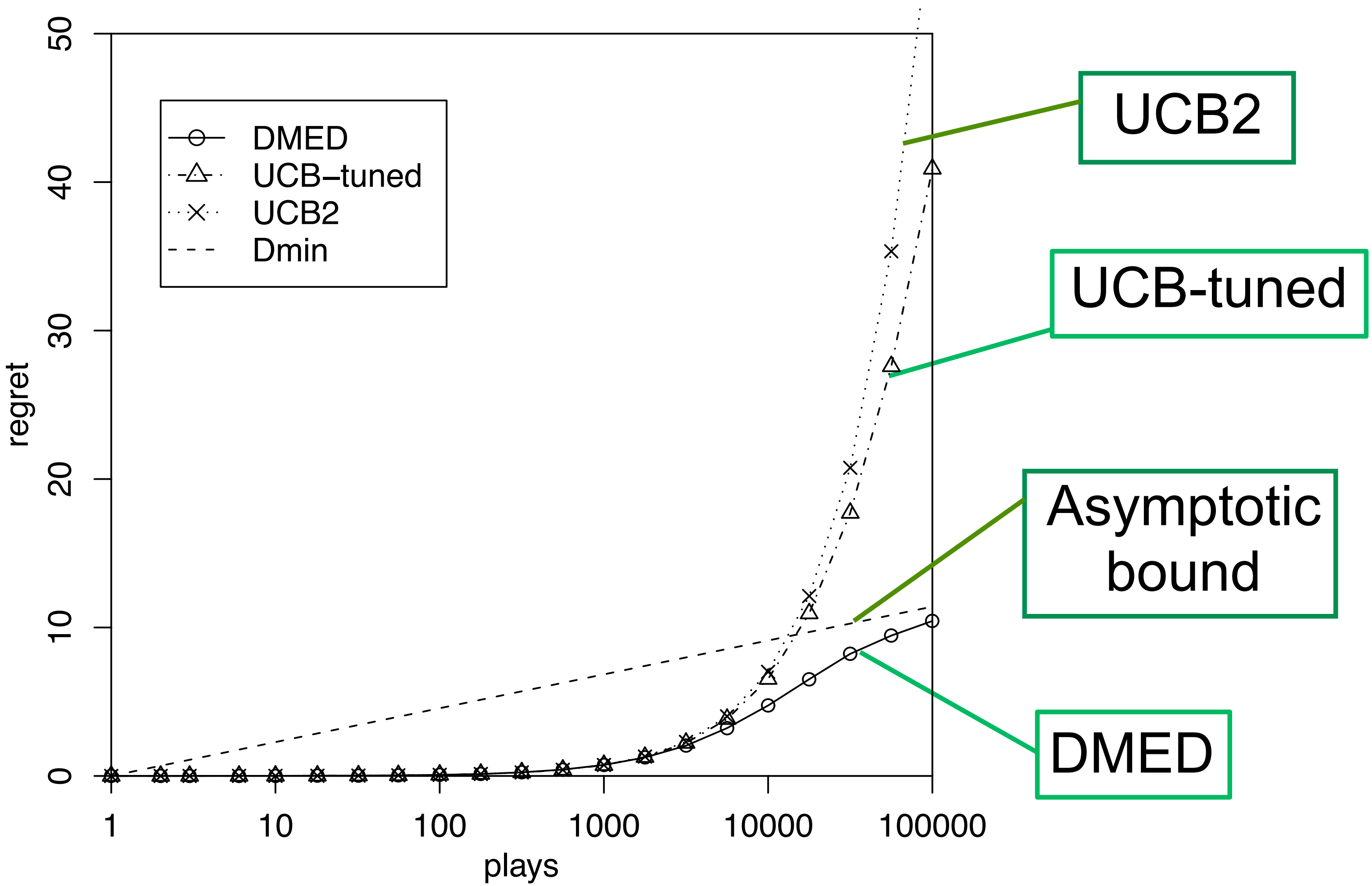
- $K = 2$, example where the best arm is hard to distinguish

$$F_1(0) = 0.99, \quad F_1(1) = 0.01, \quad \mathrm{E}(F_1) = 0.01$$
$$F_2(0.008) = 0.5, \quad F_2(0.009) = 0.5, \quad \mathrm{E}(F_2) = 0.0085$$



(Arm 2 seems to be best with high probability)

Simulation result 2



- DMED distinguishes the best arm quickly

Outline

- Introduction
- DMED policy
  - Proof of the optimality
  - Efficient computation
- Simulation results
- Conclusion

Conclusion

- Proposed DMED policy and proved its asymptotic optimality.
- Showed that the minimization of KL divergence is solvable efficiently by a convex optimization technique.
- Confirmed by simulations that DMED achieves the regret near the asymptotic bound in finite time.

Thank you!