Regret Minimization in Heavy-Tailed Bandits

Shubhada Agrawal (TIFR, Mumbai)

With Sandeep Juneja (TIFR) and Wouter M. Koolen (CWI)

COLT 2021

August, 2021
Outline of the talk

- Problem formulation
- UCB algorithms
  1. UCB-1 algorithm
  2. Robust-UCB algorithm
- Lower bound
- Gap in literature
- Our results
  1. A key idea that gives optimal algorithm for regret-minimization MAB, possibly more generally
  2. A method for proving concentration of a solution of an optimization problem
  3. Exactly where the idea in 1. gains over the existing algorithms
- Conclusion
Stochastic multi-armed bandit (MAB)

- Given:
  - Class $\mathcal{L}$ of probability distributions
    - e.g., Gaussian with known variance, distributions with support in $[0, 1]$, etc.
  - $K$ arms ($= K$ probability distributions, $\mu_a \in \mathcal{L}$ for $a \in \{1, \ldots, K\}$).

- At each time $n$, agent
  - chooses an arm $A_n = f_n(A_1, X_1, \ldots, A_{n-1}, X_{n-1})$,
  - observes a sample $X_n \sim \mu_{A_n}$, independently.

- Aim: learn something about the arm-distributions.
Regret-minimization

- **Given:**
  - Class $\mathcal{L}$ of probability distributions
    - e.g., Gaussian with known variance, distributions with support in $[0, 1]$, etc.
  - $K$ arms ($= K$ probability distributions, $\mu_a \in \mathcal{L}$ for $a \in \{1, \ldots, K\}$).

- **At each time** $n$, agent
  - chooses an arm $A_n = f_n(A_1, X_1, \ldots, A_{n-1}, X_{n-1})$,
  - observes a reward $X_n \sim \mu_{A_n}$.

- **Aim:** maximize expected sum of rewards over time: $\max \sum_{i=1}^{n} \mathbb{E}(X_i)$. 
Regret-minimization

• Given:
  • Class \( \mathcal{L} \) of probability distributions
    • e.g., Gaussian with known variance, distributions with support in \([0, 1]\), etc.
  • \( K \) arms (= \( K \) probability distributions, \( \mu_a \in \mathcal{L} \) for \( a \in \{1, \ldots, K\} \)).

• At each time \( n \), agent
  • chooses an arm \( A_n = f_n(A_1, X_1, \ldots, A_{n-1}, X_{n-1}) \), Exploration-exploitation dilemma
  • observes a reward \( X_n \sim \mu_{A_n} \).

• Aim: maximize expected sum of rewards over time: \( \max \sum_{i=1}^{n} \mathbb{E}(X_i) \).
Regret

- $m(\mu_a)$: mean of $\mu_a$, and $m^*(\mu)$: maximum-mean in $\mu$.
- $N_a(n)$: number of samples generated from $\mu_a$ till $n$. 
Regret

- $m(\mu_a)$: mean of $\mu_a$, and $m^*(\mu)$: maximum-mean in $\mu$.
- $N_a(n)$: number of samples generated from $\mu_a$ till $n$.

Aim: maximize $\sum_{i=1}^{n} \mathbb{E}(X_i) \equiv$ minimize $\mathbb{E}(R_n)$,

difference between the expected performance of algorithm and the oracle policy.
Regret

- $m(\mu_a)$: mean of $\mu_a$, and $m^*(\mu)$: maximum-mean in $\mu$.
- $N_a(n)$: number of samples generated from $\mu_a$ till $n$.

Aim: maximize $\sum_{i=1}^{n} \mathbb{E}(X_i) \equiv$ minimize $\mathbb{E}(R_n)$, difference between the expected performance of algorithm and the oracle policy.

$$
\mathbb{E}(R_n) = \sum_{a=1}^{K} (m^*(\mu) - m(\mu_a)) \mathbb{E}(N_a(n)) \\
\quad \quad := \Delta_a \\
= \sum_{a=1}^{K} \Delta_a \mathbb{E}(N_a(n)).
$$
• Agent
  • selects a treatment $A_n$ based on observations till time $n$,
  • observes the outcome $X_n \in \{0, 1\}$.

• Aim: maximize the expected number of patients cured.
Motivation

• Recommender systems
• Online advertisement placement
• Routing over congested networks
• Investing in stock-market
• ...


UCB Algorithms

1. Construct upper confidence intervals for true-mean using the available samples.
2. Pull the arm with the highest upper confidence bound.
1. Construct upper confidence intervals for true-mean using the available samples.

2. Pull the arm with the highest upper confidence bound.

\[ m_1 > m_2 > m_3. \]
1. Construct upper confidence intervals for true-mean using the available samples.

2. Pull the arm with the highest upper confidence bound.

\[ m_1 > m_2 > m_3. \]

Pull each arm once;
1. Construct **upper confidence intervals for true-mean** using the available samples.

2. Pull the arm with the **highest upper confidence bound**.

\[ m_1 > m_2 > m_3. \]

Pull each arm once; compute UCB-index.

*Shaded regions typically correspond to high probability confidence intervals for true mean.*
1. Construct upper confidence intervals for true-mean using the available samples.

2. Pull the arm with the highest upper confidence bound.

\[ m_1 > m_2 > m_3. \]

Pull each arm once; compute UCB-index.

*Shaded regions typically correspond to high probability confidence intervals for true mean.*

Pull arm 2; update UCB-index.
1. Construct upper confidence intervals for true-mean using the available samples.

2. Pull the arm with the highest upper confidence bound.

\[ m_1 > m_2 > m_3. \]

Pull each arm once; compute UCB-index. 

Shaded regions typically correspond to high probability confidence intervals for true mean.

Pull arm 2; update UCB-index.
Pull arm 2; update UCB-index.
Pull arm 1; update UCB-index.
1. Construct upper confidence intervals for true-mean using the available samples.

2. Pull the arm with the highest upper confidence bound.

$m_1 > m_2 > m_3$.

Pull each arm once; compute UCB-index. **Shaded regions typically correspond to high probability confidence intervals for true mean.**

Pull arm 2; update UCB-index.
Pull arm 2; update UCB-index.
Pull arm 1; update UCB-index.
1. Construct upper confidence intervals for true-mean using the available samples.

2. Pull the arm with the highest upper confidence bound.

\[ m_1 > m_2 > m_3. \]

Pull each arm once; compute UCB-index.

*Shaded regions typically correspond to high probability confidence intervals for true mean.*

Pull arm 2; update UCB-index.

Pull arm 2; update UCB-index.

Pull arm 1; update UCB-index.

Pull arm 1; update UCB-index.
1. Construct upper confidence intervals for true-mean using the available samples.

2. Pull the arm with the highest upper confidence bound.

\[ m_1 > m_2 > m_3. \]

Pull each arm once; compute UCB-index.

*Shaded regions typically correspond to high probability confidence intervals for true mean.*

Pull arm 2; update UCB-index.
Pull arm 2; update UCB-index.
Pull arm 1; update UCB-index.
Pull arm 1; update UCB-index.
Pull arm 3; update UCB-index.
$\mathcal{L} = \{\text{distributions supported in } [0, 1]\}$.

At each time $t$:

1. compute $U_a(t) = m(\hat{\mu}_a(t)) + \sqrt{\frac{2 \log t}{N_a(t)}}$, \hspace{1em} // UCB index for arm $a$ based on Hoeffding's inequality

2. sample $\arg \max_{a \in [K]} U_a(t)$. 
$\mathcal{L} = \{\text{distributions supported in } [0, 1]\}$.

At each time $t$:

1. compute $U_a(t) = m(\hat{\mu}_a(t)) + \sqrt{\frac{2 \log t}{N_a(t)}}$, \hspace{2cm} // UCB index for arm $a$ based on Hoeffding's inequality
   
   \begin{align*}
   & \text{Exploitation} \\
   & \text{Exploration}
   \end{align*}

2. sample $\arg\max_{a \in [K]} U_a(t)$.

$\mathbb{E}(N_a(n)) \leq 8 \frac{\log n}{\Delta_a^2}$ for all sub-optimal arms $a$.

Recall,

$\Delta_a = m^*(\mu) - m(\mu_a)$. 
Fix $1 > \epsilon > 0$, $B > 0$, and let

$$\mathcal{L} = \left\{ \text{probability distributions, } \eta, \text{ satisfying } \mathbb{E}_{X \sim \eta} \left( |X|^{1+\epsilon} \right) \leq B \right\}.$$ 

$\mathcal{L}$ includes many heavy-tailed distributions.
Robust-UCB (Bubeck et al., 2013)

Fix $1 > \epsilon > 0$, $B > 0$, and let

$$\mathcal{L} = \left\{ \text{probability distributions, } \eta, \text{ satisfying } \mathbb{E}_{X \sim \eta} (|X|^{1+\epsilon}) \leq B \right\}.$$ 

$\mathcal{L}$ includes many heavy-tailed distributions.

$$U_a(t) = \tilde{m}(\hat{\mu}_a(t)) + 4B^\frac{1}{1+\epsilon} \left( \frac{2 \log t}{N_a(t)} \right)^{\frac{\epsilon}{1+\epsilon}} , \quad \text{// based on MGF-based Bernstein-like inequality}$$

where $\tilde{m}(\hat{\mu}_a(t)) :$ empirical mean of truncated samples, $X \mathbb{1} (|X| \leq u_t)$, for well-chosen $u_t$. 


Robust-UCB \cite{Bubeck2013} 

Fix $1 > \epsilon > 0$, $B > 0$, and let

$$\mathcal{L} = \left\{ \text{probability distributions, } \eta, \text{ satisfying } \mathbb{E}_{X \sim \eta} (|X|^{1+\epsilon}) \leq B \right\}.$$ 

$\mathcal{L}$ includes many heavy-tailed distributions.

$$U_a(t) = \tilde{m}(\hat{\mu}_a(t)) + 4B \frac{1}{1+\epsilon} \left( \frac{2 \log t}{N_a(t)} \right)^{\frac{\epsilon}{1+\epsilon}}, \quad \text{based on MGF-based Bernstein-like inequality}$$

where $\tilde{m}(\hat{\mu}_a(t))$: empirical mean of truncated samples, $X \mathbb{1} \left( |X| \leq u_t \right)$, for well-chosen $u_t$.

$$\mathbb{E}(N_a(n)) \leq 8(4B)^{\frac{1}{\epsilon}} \frac{\log(n)}{\Delta_a^{1+\frac{1}{\epsilon}}}, \quad \text{for all sub-optimal arms } a.$$
For a given class $\mathcal{L}$, uniformly efficient algorithms satisfy:

$$\forall \mu \in \mathcal{L}^K, \forall \alpha \in (0, 1), \mathbb{E}(R_n) = o(n^\alpha).$$
For a given class $\mathcal{L}$, uniformly efficient algorithms satisfy:

$$\forall \mu \in \mathcal{L}^K, \forall \alpha \in (0, 1), \mathbb{E}(R_n) = o(n^\alpha).$$

For uniformly efficient algorithms, for $\mu \in \mathcal{L}^K$ and each sub-optimal arm $a$,

$$\liminf_{n \to \infty} \frac{\mathbb{E}(N_a(n))}{\log(n)} \geq \frac{1}{KL_{\inf}(\mu_a, m^*(\mu))},$$

where for a probability measure $\eta, x \in \mathcal{R}$,

$$KL_{\inf}(\eta, x) := \min \{KL(\eta, \kappa) : \kappa \in \mathcal{L}, m(\kappa) \geq x\}.$$
Existing literature

- Asymptotic lower bound: (Lai and Robbins, 1985) and (Burnetas and Katehakis, 1996).
- Algorithms for bounded-support / sub-Gaussian distributions: (Auer et al., 2002), (Audibert et al., 2009, 2010), (Bubeck et al., 2012), ...

Do not match the constants.
Existing literature

- Asymptotic lower bound: (Lai and Robbins, 1985) and (Burnetas and Katehakis, 1996).
- Algorithms for bounded-support / sub-Gaussian distributions: (Auer et al., 2002), (Audibert et al., 2009, 2010), (Bubeck et al., 2012), ...

  Do not match the constants.

- Asymptotically optimal algorithms for finite/bounded-support distributions: (Honda et al., 2010, 2011, 2015).
- Asymptotically optimal algorithm for parametric family: (Cappé et al., 2011, 2013), (Maillard et al., 2011).
Existing literature

- Asymptotic lower bound: (Lai and Robbins, 1985) and (Burnetas and Katehakis, 1996).

- Algorithms for bounded-support / sub-Gaussian distributions: (Auer et al., 2002), (Audibert et al., 2009, 2010), (Bubeck et al., 2012), ...

  Do not match the constants.

- Asymptotically optimal algorithms for finite/bounded-support distributions: (Honda et al., 2010, 2011, 2015).

- Asymptotically optimal algorithm for parametric family: (Cappé et al., 2011, 2013), (Maillard et al., 2011).

- Algorithms for heavy-tailed setting: (Bubeck et al., 2013), (Lattimore T., 2017).

  Do not match the constants.
Recall,\[ \liminf_{n \to \infty} \frac{\mathbb{E}(N_a(n))}{\log(n)} \geq \frac{1}{\text{KL}_{\infty}(\mu_a, m^*(\mu))}, \]

where For a probability measure $\eta, x \in \mathbb{R}$,

\[ \text{KL}_{\infty}(\eta, x) = \inf_{\kappa \in \mathcal{L}: m(\kappa) \geq x} \text{KL}(\eta, \kappa). \]
Recall,\[
\liminf_{n \to \infty} \frac{\mathbb{E}(N_a(n))}{\log(n)} \geq \frac{1}{\text{KL}_{\inf}(\mu_* m^*(\mu))},
\]
where For a probability measure $\eta, x \in \mathbb{R}$,
\[
\text{KL}_{\inf}(\eta, x) = \inf_{\kappa \in \mathcal{L}: m(\kappa) \geq x} \text{KL}(\eta, \kappa).
\]
Recall,
\[
\liminf_{n \to \infty} \frac{\mathbb{E}(N_a(n))}{\log(n)} \geq \frac{1}{\text{KL}_{\inf}(\mu_a, m^*(\mu))},
\]
where For a probability measure \( \eta, x \in \mathbb{R} \),
\[
\text{KL}_{\inf}(\eta, x) = \inf_{\kappa \in \mathcal{L} : m(\kappa) \geq x} \text{KL}(\eta, \kappa).
\]

1. For \( \eta \in \mathcal{L} \), \( \text{KL}_{\inf}(\eta, m(\eta)) = 0 \).

2. \( \text{KL}_{\inf}(\eta, x) \) is non-decreasing and convex in \( x \).
Our setup

Given $\epsilon > 0$, $B > 0$ (known to the algorithm),

$$\mathcal{L} = \left\{ \text{probability distributions, } \nu, \text{ satisfying } \mathbb{E}_{X \sim \nu} \left( |X|^{1+\epsilon} \right) \leq B \right\}.$$

$\mathcal{L}$ includes many heavy-tailed distributions.
Algorithm: At time $t$,

- Compute index $U_a(t)$ for all the arms.
- Select the arm with maximum index.
KL_{\text{inf}}-UCB Algorithm

\[ U_a(t) = \max \left\{ m(\kappa) : \kappa \in \mathcal{L}, \ \text{KL}(\hat{\mu}_a(t), \kappa) \leq \frac{g(t, N_a(t))}{N_a(t)} \right\}, \]

\[ g(t, N) \approx \log(t) + \log(N). \]

**Algorithm:** At time $t$,
- Compute index $U_a(t)$ for all the arms.
- Select the arm with maximum index.

![Diagram](attachment:image.png)
Regret bound

**Theorem**

For \( n \geq K \) and \( g(x, N) = \log(x) + 2 \log \log(x) + 2 \log(1 + N) + 1 \),

\[
\mathbb{E}(N_a(n)) \leq \frac{\log n}{\text{KL}_{\inf}(\mu_a, m^*(\mu))} + O\left((\log n)^{\frac{2}{3}}\right), \quad \forall a \neq 1.
\]

**Corollary**

\[
\limsup_{n \to \infty} \frac{\mathbb{E}(N_a(n))}{\log n} \leq \frac{1}{\text{KL}_{\inf}(\mu_a, m^*(\mu))}, \quad \text{for a suboptimal arm } a.
\]

**Recall**

\[
\liminf_{n \to \infty} \frac{\mathbb{E}(N_a(n))}{\log n} \geq \frac{1}{\text{KL}_{\inf}(\mu_a, m^*(\mu))}, \quad \text{for a suboptimal arm } a.
\]
Is $\text{KL}_{\text{inf}}$-UCB Index a high probability upper bound?

Recall,

$\eta$,

$\text{KL}_{\text{inf}}(\eta, x) = \text{KL}(\eta, \kappa^*)$

$\kappa : m(\kappa) \geq x$

$\mathcal{L}$
Is $\text{KL}_{\inf}$-UCB Index a high probability upper bound?

$U_a(t) = \max \{ m(\kappa) : \kappa \in \mathcal{L}, \ KL(\hat{\mu}_a(t), \kappa) \leq C \} = \max \{ x \in \mathbb{R} : \ KL_{\inf}(\hat{\mu}_a(t), x) \leq C \}.$

Recall,

$\eta \Rightarrow \text{KL}_{\inf}(\eta, x) = \text{KL}(\eta, \kappa^*)$

$\kappa : m(\kappa) \geq x$

$\mathcal{L}$
Is $KL_{\inf}$-UCB Index a high probability upper bound?

$U_a(t) = \max \{ m(\kappa) : \kappa \in \mathcal{L}, \ KL(\hat{\mu}_a(t), \kappa) \leq C \}$

$= \max \{ x \in \mathbb{R} : \ KL_{\inf}(\hat{\mu}_a(t), x) \leq C \}.

\{ U_a(t) \leq m(\mu_a) \} \equiv \{ KL_{\inf}(\hat{\mu}_a(t), m(\mu_a)) > C \}.$

Recall,

$KL_{\inf}(\eta, x) = KL(\eta, \kappa^*)$
Is $\text{KL}_{\text{inf}}$-UCB Index a high probability upper bound?

$$U_a(t) = \max \{ m(\kappa) : \kappa \in \mathcal{L}, \ KL(\hat{\mu}_a(t), \kappa) \leq C \}$$

$$= \max \{ x \in \mathbb{R} : KL_{\text{inf}}(\hat{\mu}_a(t), x) \leq C \}.$$

$$\{ U_a(t) \leq m(\mu_a) \} \equiv \{ KL_{\text{inf}}(\hat{\mu}_a(t), m(\mu_a)) > C \}.$$

Setting $C = \frac{g_a(t, N_a(t))}{N_a(t)}$, sufficient to bound

$$\mathbb{P}[N_a(t) KL_{\text{inf}}(\hat{\mu}_a(t), m(\mu_a)) \geq g_a(t, N_a(t))].$$

Recall,

$$\eta$$

$KL_{\text{inf}}(\eta, x) = KL(\eta, \kappa^*)$

$\kappa^* : m(\kappa) \geq x$
An anytime concentration inequality

Recall, \( g(t, N) = \log(t) + 2\log\log(t) + 2\log(1 + N) + 1. \)

**Proposition**

For \( x \geq 0, \ a \in [K], \)

\[
P(\exists t \in \mathbb{N} : N_a(t) \text{KL}_{\text{inf}}(\hat{\mu}_a(t), m(\mu_a)) - (2\log(1 + N_a(t)) + 1) \geq x) \leq e^{-x}.
\]
An anytime concentration inequality

Recall, \( g(t, N) = \log(t) + 2 \log \log(t) + 2 \log(1 + N) + 1 \).

**Proposition**

For \( x \geq 0, \ a \in [K] \),

\[
P( \exists t \in \mathbb{N} : N_a(t) \text{KL}_{\text{inf}}(\hat{\mu}_a(t), m(\mu_a)) - (2 \log(1 + N_a(t)) + 1) \geq x ) \leq e^{-x}.
\]

This gives:

\[
P(N_a(t) \text{KL}_{\text{inf}}(\hat{\mu}_a(t), m(\mu_a)) \geq g_a(t, N_a(t))) \leq \frac{1}{t(\log(t))^2}.
\]
Recall, $g(t, N) = \log(t) + 2 \log \log(t) + 2 \log(1 + N) + 1$.

Proposition

For $x \geq 0$, $a \in [K]$, 

$$\mathbb{P}(\exists t \in \mathbb{N}: N_a(t) \infKL(\hat{\mu}_a(t), m(\mu_a)) - (2 \log(1 + N_a(t)) + 1) \geq x) \leq e^{-x}.$$ 

This gives:

$$\mathbb{P}(N_a(t) \infKL(\hat{\mu}_a(t), m(\mu_a)) \geq g_a(t, N_a(t))) \leq \frac{1}{t(\log(t))^2}.$$

Two key ideas:

- **Dual** formulation for $\infKL$.
- **Mixtures of super-martingales** dominating L.H.S.
Key proof ideas

Dual formulation (A., Juneja, S., Glynn, P., 2020):

\[ N_a(t) \, \text{KL}_{\inf}(\hat{\mu}_a(t), m(\mu_a)) = \max_{\lambda \in S} \log \prod_{i=1}^{N_a(t)} Y(X_i, \lambda), \]

where for \( \lambda \in S \), \( Y(X_i, \lambda) \) are

• i.i.d.
• non-negative
• mean bounded by 1.

\( N_a(t) \, \prod_{i=1}^{N_a(t)} Y(X_i, \lambda) \) is a super-martingale.

Mix these over \( \lambda \) in \( S \) to dominate

\[ \max_{\lambda \in S} \log \prod_{i=1}^{N_a(t)} Y(X_i, \lambda) - (2 \log(1 + N_a(t)) + 1). \]
Key proof ideas

Dual formulation (A., Juneja, S., Glynn, P., 2020):

$$N_a(t) \text{KL}_{\inf}\left(\hat{\mu}_a(t), m(\mu_a)\right) = \max_{\lambda \in S} \log \prod_{i=1}^{N_a(t)} Y(X_i; \lambda),$$

where for $\lambda \in S$, $Y(X_i; \lambda)$ are

- i.i.d.
- non-negative
- mean bounded by 1.
Key proof ideas

Dual formulation (A., Juneja, S., Glynn, P., 2020):

\[ N_a(t) \text{KL}_\text{inf}(\hat{\mu}_a(t), m(\mu_a)) = \max_{\lambda \in S} \log \prod_{i=1}^{N_a(t)} Y(X_i, \lambda), \]

where for \( \lambda \in S \), \( Y(X_i, \lambda) \) are

• i.i.d.
• non-negative
• mean bounded by 1.

\[ \prod_{i=1}^{N_a(t)} Y(X_i, \lambda) \] is a super-martingale.
Key proof ideas

Dual formulation (A., Juneja, S., Glynn, P., 2020):

\[ N_a(t) \text{KL}_{\inf}(\hat{\mu}_a(t), m(\mu_a)) = \max_{\lambda \in S} \log \prod_{i=1}^{N_a(t)} Y(X_i, \lambda), \]

where for \( \lambda \in S \), \( Y(X_i, \lambda) \) are

- i.i.d.
- non-negative
- mean bounded by 1.

\( \prod_{i=1}^{N_a(t)} Y(X_i, \lambda) \) is a super-martingale.

Mix these over \( \lambda \) in \( S \) to dominate

\[ \max_{\lambda \in S} \log \prod_{i=1}^{N_a(t)} Y(X_i, \lambda) - (2 \log(1 + N_a(t)) + 1). \]
Where does KL-based UCB index win?

Our index for a sub-optimal arm \(a\) at time \(t\) is

\[
\max \left\{ \mathbb{E}_\kappa(X) : \kappa \in \mathcal{L}, \quad \text{KL}(\hat{\mu}_a(t), \kappa) \leq C \right\},
\]

where \(C = \frac{g_a(t, N_a(t))}{N_a(t)}\).
Where does KL-based UCB index win?

Our index for a sub-optimal arm \( a \) at time \( t \) is

\[
\max \{ \mathbb{E}_{\kappa}(X) : \kappa \in \mathcal{L}, \ KL(\hat{\mu}_a(t), \kappa) \leq C \},
\]

where \( C = \frac{g_a(t, N_a(t))}{N_a(t)} \). For probability measures \( P, Q \), recall (Donsker and Varadhan):

\[
KL(P, Q) = \sup_{g : \mathbb{E}_Q(e^{g(X)}) < \infty} \left\{ \mathbb{E}_P(g(X)) - \log \mathbb{E}_Q(e^{g(X)}) \right\}.
\]
Where does KL-based UCB index win?

Our index for a sub-optimal arm \( a \) at time \( t \) is

\[
\max \{ \mathbb{E}_\kappa (X) : \kappa \in \mathcal{L}, \ KL(\hat{\mu}_a(t), \kappa) \leq C \},
\]

where \( C = \frac{g_a(t, N_a(t))}{N_a(t)} \). For probability measures \( P, Q \), recall (Donsker and Varadhan):

\[
KL(P, Q) = \sup_{g : \mathbb{E}_Q(g(X)) < \infty} \left\{ \mathbb{E}_P (g(X)) - \log \mathbb{E}_Q \left( e^{g(X)} \right) \right\}.
\]

Using this, for any particular choice of \( g \), our index is at most

\[
\max \{ \mathbb{E}_\kappa (X) : \kappa \in \mathcal{L}, \ \mathbb{E}_{\hat{\mu}_a(t)} (g(X)) - \log \mathbb{E}_\kappa \left( e^{g(X)} \right) \leq C \}.
\]
Where does KL-based UCB index win?

Our index for a sub-optimal arm $a$ at time $t$ is

$$\max \{ \mathbb{E}_\kappa(X) : \kappa \in \mathcal{L}, \ \text{KL}(\hat{\mu}_a(t), \kappa) \leq C \},$$

where $C = \frac{g_a(t, N_a(t))}{N_a(t)}$. Using Donsker-Varadhan representation, our index is at most

$$\max \left\{ \mathbb{E}_\kappa(X) : \kappa \in \mathcal{L}, \ \mathbb{E}_{\hat{\mu}_a(t)}(g(X)) - \log \mathbb{E}_\kappa \left( e^{g(X)} \right) \leq C \right\}.$$
Where does KL-based UCB index win?

Our index for a sub-optimal arm $a$ at time $t$ is

$$\max \left\{ \mathbb{E}_\kappa (X) : \kappa \in \mathcal{L}, \ KL(\hat{\mu}_a(t), \kappa) \leq C \right\},$$

where $C = \frac{g_a(t, N_a(t))}{N_a(t)}$. Using Donsker-Varadhan representation, our index is at most

$$\max \left\{ \mathbb{E}_\kappa (X) : \kappa \in \mathcal{L}, \ \mathbb{E}_{\hat{\mu}_a(t)} (g(X)) - \log \mathbb{E}_\kappa \left( e^{g(X)} \right) \leq C \right\}.$$  

For $\theta > 0$ (and optimized later), choosing

$$g(X) = -\theta X \mathbb{1} (|X| \leq u),$$

with appropriate truncation level $u$ recovers Robust-UCB index for the sub-optimal arm $a$. 

Where does KL-based UCB index win?

Our index for a sub-optimal arm $a$ at time $t$ is

$$
\max \{ \mathbb{E}_{\kappa} (X) : \kappa \in \mathcal{L}, \ KL(\hat{\mu}_a(t), \kappa) \leq C \},
$$

where $C = \frac{g_a(t, N_a(t))}{N_a(t)}$. Using Donsker-Varadhan representation, our index is at most

$$
\max \left\{ \mathbb{E}_{\kappa} (X) : \kappa \in \mathcal{L}, \ \mathbb{E}_{\hat{\mu}_a(t)} (g(X)) - \log \mathbb{E}_{\kappa} \left( e^{g(X)} \right) \leq C \right\}.
$$

For $\theta > 0$ (and optimized later), choosing

$$
g(X) = -\theta X \mathbb{1} (|X| \leq u),
$$

with appropriate truncation level $u$ recovers Robust-UCB index for the sub-optimal arm $a$.

- Our index for sub-optimal arms is smaller than that for Robust-UCB!.
- Argument does not work for optimal arm as the corresponding threshold ($C$) is higher.
Conclusion

UCB algorithms: typically rely on high probability confidence intervals for true mean.
Conclusion

UCB algorithms: typically rely on high probability confidence intervals for true mean.

Lower bound for regret-minimization MAB: \( \approx \frac{\log(n)}{\text{KL}_{\infty}(\mu_a, m^*(\mu))} \).

Understood the structure of lower-bound optimization problem.
Conclusion

UCB algorithms: typically rely on high probability confidence intervals for true mean.

Lower bound for regret-minimization MAB: $\approx \frac{\log(n)}{\text{KL}_{\text{inf}}(\mu_a, m^*(\mu))}$.

Understood the structure of lower-bound optimization problem.

Heavy-tailed arms:
- Robust-UCB - index derived using MGF-based concentration inequalities.
- Optimal $\text{KL}_{\text{inf}}$-UCB - index derived using $\text{KL}_{\text{inf}}$ concentration.
Conclusion

UCB algorithms: typically rely on high probability confidence intervals for true mean.

Lower bound for regret-minimization MAB: \( \approx \frac{\log(n)}{\text{KL}_{\text{inf}}(\mu_a, m^*(\mu))} \).

Understood the structure of lower-bound optimization problem.

Heavy-tailed arms:

- Robust-UCB - index derived using MGF-based concentration inequalities.
- Optimal \( \text{KL}_{\text{inf}} \)-UCB - index derived using \( \text{KL}_{\text{inf}} \) concentration.

Further work needed to improve the concentration for empirical \( \text{KL}_{\text{inf}} \):

- Cost of \( 2 \log(1 + N_a(n)) \) needed for the martingale construction is too high.
Conclusion

UCB algorithms: typically rely on high probability confidence intervals for true mean.

Lower bound for regret-minimization MAB: \( \approx \frac{\log(n)}{\text{KL}_{\text{inf}}(\mu_a, \mu^*)} \).

Understood the structure of lower-bound optimization problem.

Heavy-tailed arms:
- Robust-UCB - index derived using MGF-based concentration inequalities.
- Optimal KL_{\text{inf}}-UCB - index derived using KL_{\text{inf}} concentration.

Further work needed to improve the concentration for empirical KL_{\text{inf}}:
- Cost of \( 2 \log(1 + N_a(n)) \) needed for the martingale construction is too high.

Thank you!