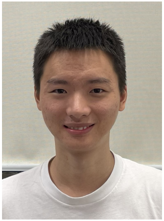
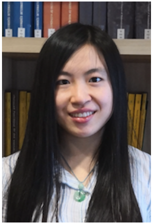
Almost Optimal Variance-Constrained Best Arm Identification

Yunlong Hou Zixin Zhong **Vincent Y. F. Tan**





National University of Singapore

S3 Optimization Day July 2022
IORA, NUS

Introduction to Stochastic Multi-Armed Bandits in BAI




Efficacies: $\mu_1 = 0.8$ $\mu_2 = 0.9$ $\mu_3 = 0.2$ $\mu_4 = 0.5$

- In clinical trials, there are N potential treatments for a disease.
- At each time, the scientist prescribes one of them to each lab animal and the efficacies of the treatments can be observed.
- Goal: Find the best treatment using the smallest number of trials.

Formulation of Stochastic Multi-Armed Bandits in BAI

- N arms: $[N] = \{1, 2, \ldots, N\}$ with unknown distributions $\{\nu_i\}_{i=1}^N$

Formulation of Stochastic Multi-Armed Bandits in BAI

- N arms: $[N] = \{1, 2, \ldots, N\}$ with unknown distributions $\{\nu_i\}_{i=1}^N$
- Sampling strategy: At each round r, select arm $i_r \in [N]$ based on the observation history

$$\mathcal{H}_r = ((i_1, X_{1,i_1}), \ldots, (i_{r-1}, X_{r-1,i_{r-1}}))$$

and observe the reward $X_{r,i_r} \sim \nu_{i_r}$

Formulation of Stochastic Multi-Armed Bandits in BAI

- N arms: $[N] = \{1, 2, \ldots, N\}$ with unknown distributions $\{\nu_i\}_{i=1}^N$
- Sampling strategy: At each round r, select arm $i_r \in [N]$ based on the observation history

$$\mathcal{H}_r = ((i_1, X_{1,i_1}), \ldots, (i_{r-1}, X_{r-1,i_{r-1}}))$$

  and observe the reward $X_{r,i_r} \sim \nu_{i_r}$
- The sequence of random variables $\{X_{r,i}\}_{r=1}^\infty$ is assumed to be i.i.d. across rounds $r \in \mathbb{N}$ and arms $i \in [N]$

Formulation of Stochastic Multi-Armed Bandits in BAI

- N arms: $[N] = \{1, 2, \ldots, N\}$ with unknown distributions $\{\nu_i\}_{i=1}^N$
- Sampling strategy: At each round r, select arm $i_r \in [N]$ based on the observation history

$$\mathcal{H}_r = ((i_1, X_{1,i_1}), \ldots, (i_{r-1}, X_{r-1,i_{r-1}}))$$

  and observe the reward $X_{r,i_r} \sim \nu_{i_r}$
- The sequence of random variables $\{X_{r,i}\}_{r=1}^\infty$ is assumed to be i.i.d. across rounds $r \in \mathbb{N}$ and arms $i \in [N]$
- Goal: Design a policy to find the arm with the highest expectation (best arm) in the smallest number of rounds.

Formulation of Stochastic Multi-Armed Bandits in BAI

- N arms: $[N] = \{1, 2, \ldots, N\}$ with unknown distributions $\{\nu_i\}_{i=1}^N$
- Sampling strategy: At each round r, select arm $i_r \in [N]$ based on the observation history

$$\mathcal{H}_r = ((i_1, X_{1,i_1}), \ldots, (i_{r-1}, X_{r-1,i_{r-1}}))$$

  and observe the reward $X_{r,i_r} \sim \nu_{i_r}$
- The sequence of random variables $\{X_{r,i}\}_{r=1}^{\infty}$ is assumed to be i.i.d. across rounds $r \in \mathbb{N}$ and arms $i \in [N]$
- Goal: Design a policy to find the arm with the highest expectation (best arm) in the smallest number of rounds.
- Probably approximately correct (PAC) framework (Even-Dar et al., 2006). Given a fixed confidence parameter δ, find any $i \in [N]$ s.t.

$$\mathbb{P}\Big[i \in \arg\max_{j \in [N]} \mu_j\Big] \geq 1 - \delta.$$

Introduction to BAI

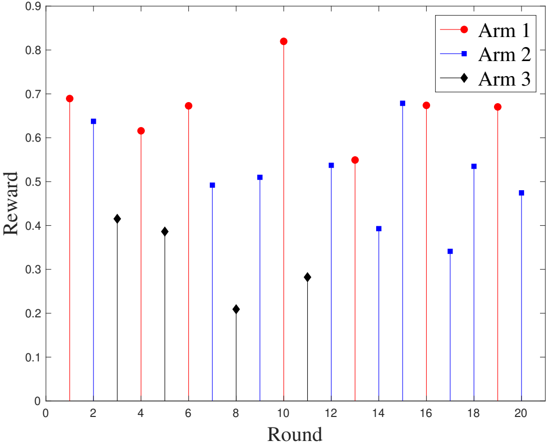


Figure 1: An illustration of a 3-arm BAI problem. At each round, only one arm is sampled and the reward is observed.

Motivation of Risk-Aware BAI

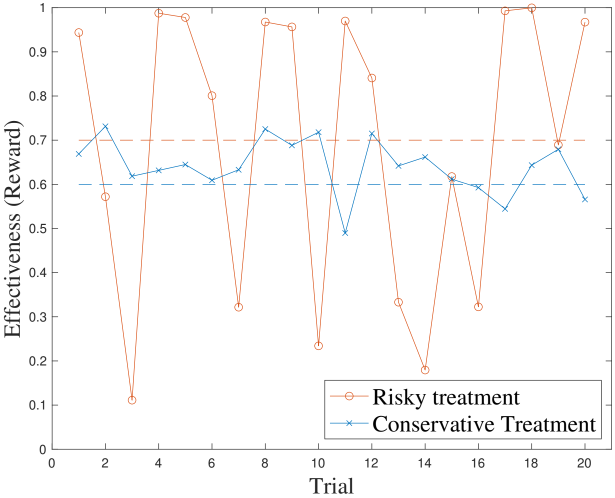
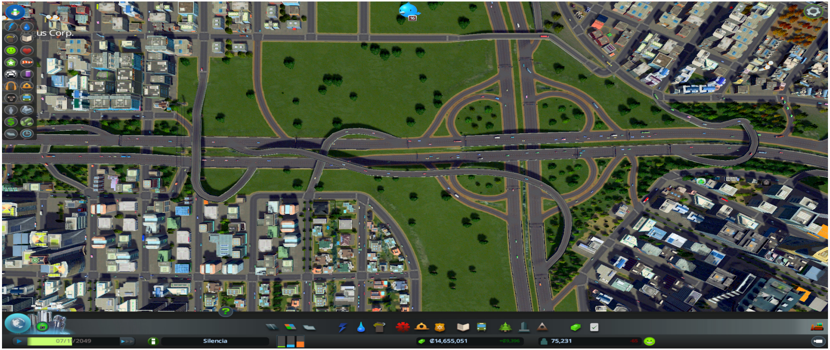


Figure 2: An example of risk-aware BAI: Clinical trials

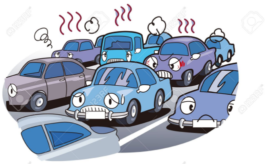
Another motivation: Going to office bandit style



On every day

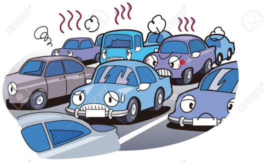
1. Pick a route to office
2. Reach office and record (suffered) delay

Why consider risk?



$\mathbb{E}[\text{time}] = 10$ mins, $\text{Var}(\text{time}) = 10$ $\quad$ $\mathbb{E}[\text{time}] = 11$ mins, $\text{Var}(\text{time}) = 0.1$

Why consider risk?



$\mathbb{E}[\text{time}] = 10$ mins, $\text{Var}(\text{time}) = 10$

$\mathbb{E}[\text{time}] = 11$ mins, $\text{Var}(\text{time}) = 0.1$

- Delays are stochastic.
- In choosing between routes, we need not necessarily want to minimize expected delay.
- Two route scenario: Average delay of Route 1 slightly below that of Route 2.
- Route 1 has a small chance of very high delay, e.g., jams.
- I might prefer Route 2.

Introduction to Risk-Aware Bandits

- Incorporate the risk into the quality measure:
  - Mean-variance: Sani et al. (2012), Vakili and Zhao (2016), and Zhu and Tan (2020).
  - Value-at-Risk or α-quantile: David and Shimkin (2016)
  - Conditional Value-at-risk (CVaR): Kagrecha et al. (2020) and Baudry et al. (2021)

Introduction to Risk-Aware Bandits

- Incorporate the risk into the quality measure:
  - Mean-variance: Sani et al. (2012), Vakili and Zhao (2016), and Zhu and Tan (2020).
  - Value-at-Risk or α-quantile: David and Shimkin (2016)
  - Conditional Value-at-risk (CVaR): Kagrecha et al. (2020) and Baudry et al. (2021)
- Risk-constrained problem, i.e., conventional BAI with constraints
  - Variance
  - α-quantile: David et al. (2018)
  - Safe bandits: Wu et al. (2016); Amani et al. (2019)

Introduction to Variance-Constrained BAI

An instance $(\nu = \{\nu_i\}_{i=1}^N, \bar{\sigma}^2)$ consists of

- N arms with associated with unknown reward distributions $\{\nu_i\}_{i=1}^N$, where arm i follows ν_i with expectation μ_i and variance σ_i^2.
- permissible upper bound on the variance: $\bar{\sigma}^2$.

Introduction to Variance-Constrained BAI

An instance $(\nu = \{\nu_i\}_{i=1}^N, \bar{\sigma}^2)$ consists of

- N arms with associated with unknown reward distributions $\{\nu_i\}_{i=1}^N$, where arm i follows ν_i with expectation μ_i and variance σ_i^2.
- permissible upper bound on the variance: $\bar{\sigma}^2$.

Based on the means and variances, define

- Feasible set $\mathcal{F} := \{i \in [N] : \sigma_i^2 \leq \bar{\sigma}^2\}$

Introduction to Variance-Constrained BAI

An instance $(\nu = \{\nu_i\}_{i=1}^N, \bar{\sigma}^2)$ consists of

- N arms with associated with unknown reward distributions $\{\nu_i\}_{i=1}^N$, where arm i follows ν_i with expectation μ_i and variance σ_i^2.
- permissible upper bound on the variance: $\bar{\sigma}^2$.

Based on the means and variances, define

- Feasible set $\mathcal{F} := \{i \in [N] : \sigma_i^2 \leq \bar{\sigma}^2\}$
- Infeasible set $\bar{\mathcal{F}}^c := [N] \backslash \mathcal{F}$

Introduction to Variance-Constrained BAI

An instance $(\nu = \{\nu_i\}_{i=1}^N, \bar{\sigma}^2)$ consists of

- N arms with associated with unknown reward distributions $\{\nu_i\}_{i=1}^N$, where arm i follows ν_i with expectation μ_i and variance σ_i^2.
- permissible upper bound on the variance: $\bar{\sigma}^2$.

Based on the means and variances, define

- Feasible set $\mathcal{F} := \{i \in [N] : \sigma_i^2 \leq \bar{\sigma}^2\}$
- Infeasible set $\bar{\mathcal{F}}^c := [N] \backslash \mathcal{F}$
- Best feasible arm $i^\star := \mathrm{argmax}\{\mu_i : i \in \mathcal{F}\}$

Introduction to Variance-Constrained BAI

An instance $(\nu = \{\nu_i\}_{i=1}^N, \bar{\sigma}^2)$ consists of

- N arms with associated with unknown reward distributions $\{\nu_i\}_{i=1}^N$, where arm i follows ν_i with expectation μ_i and variance σ_i^2.
- permissible upper bound on the variance: $\bar{\sigma}^2$.

Based on the means and variances, define

- Feasible set $\mathcal{F} := \{i \in [N] : \sigma_i^2 \leq \bar{\sigma}^2\}$
- Infeasible set $\bar{\mathcal{F}}^c := [N] \backslash \mathcal{F}$
- Best feasible arm $i^\star := \text{argmax}\{\mu_i : i \in \mathcal{F}\}$
- Suboptimal set $\mathcal{S} := \{i \in [N] : \mu_i < \mu_{i^\star}\}$

Introduction to Variance-Constrained BAI

An instance $(\nu = \{\nu_i\}_{i=1}^N, \bar{\sigma}^2)$ consists of

- N arms with associated with unknown reward distributions $\{\nu_i\}_{i=1}^N$, where arm i follows ν_i with expectation μ_i and variance σ_i^2.
- permissible upper bound on the variance: $\bar{\sigma}^2$.

Based on the means and variances, define

- Feasible set $\mathcal{F} := \{i \in [N] : \sigma_i^2 \le \bar{\sigma}^2\}$
- Infeasible set $\bar{\mathcal{F}}^c := [N] \backslash \mathcal{F}$
- Best feasible arm $i^\star := \operatorname{argmax}\{\mu_i : i \in \mathcal{F}\}$
- Suboptimal set $\mathcal{S} := \{i \in [N] : \mu_i < \mu_{i^\star}\}$
- Risky set $\mathcal{R} := [N] \setminus (\mathcal{S} \cup \{i^\star\})$

Introduction to Variance-Constrained BAI

An instance $(\nu = \{\nu_i\}_{i=1}^N, \bar{\sigma}^2)$ consists of

- N arms with associated with unknown reward distributions $\{\nu_i\}_{i=1}^N$, where arm i follows ν_i with expectation μ_i and variance σ_i^2.
- permissible upper bound on the variance: $\bar{\sigma}^2$.

Based on the means and variances, define

- Feasible set $\mathcal{F} := \{i \in [N] : \sigma_i^2 \leq \bar{\sigma}^2\}$
- Infeasible set $\bar{\mathcal{F}}^c := [N] \backslash \mathcal{F}$
- Best feasible arm $i^\star := \mathrm{argmax}\{\mu_i : i \in \mathcal{F}\}$
- Suboptimal set $\mathcal{S} := \{i \in [N] : \mu_i < \mu_{i^\star}\}$
- Risky set $\mathcal{R} := [N] \setminus (\mathcal{S} \cup \{i^\star\})$
- Mean gap $\Delta_i = \mu_{i^\star} - \mu_i \geq 0$

Introduction to Variance-Constrained BAI

An instance $(\nu = \{\nu_i\}_{i=1}^N, \bar{\sigma}^2)$ consists of

- N arms with associated with unknown reward distributions $\{\nu_i\}_{i=1}^N$, where arm i follows ν_i with expectation μ_i and variance σ_i^2.
- permissible upper bound on the variance: $\bar{\sigma}^2$.

Based on the means and variances, define

- Feasible set $\mathcal{F} := \{i \in [N] : \sigma_i^2 \le \bar{\sigma}^2\}$
- Infeasible set $\bar{\mathcal{F}}^c := [N] \backslash \mathcal{F}$
- Best feasible arm $i^\star := \mathrm{argmax}\{\mu_i : i \in \mathcal{F}\}$
- Suboptimal set $\mathcal{S} := \{i \in [N] : \mu_i < \mu_{i^\star}\}$
- Risky set $\mathcal{R} := [N] \setminus (\mathcal{S} \cup \{i^\star\})$
- Mean gap $\Delta_i = \mu_{i^\star} - \mu_i \ge 0$
- Variance gap $\Delta_i^{\mathrm{v}} = |\sigma_i^2 - \bar{\sigma}^2|$

Introduction to Variance-Constrained BAI

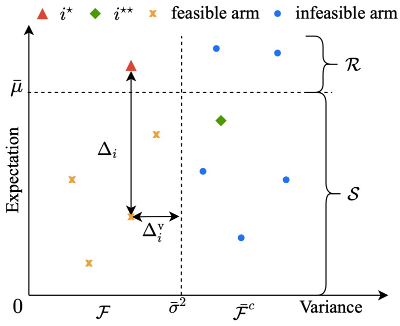


Figure 3: An illustration of an instance

Introduction to Variance-Constrained BAI

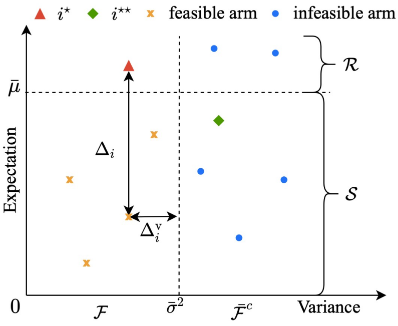


Figure 3: An illustration of an instance

Goal: Minimize the number of arm pulls needed for

- Ascertaining the feasibility of the instance $(\nu, \bar{\sigma}^2)$;
- Finding the best feasible arm $i^\star$ if $\mathcal{F} \neq \emptyset$

with probability $\geq 1 - \delta$

Main Result: Upper Bound

Theorem 1 (Upper bound)

Define the hardness parameter

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta^{\mathrm{v}}_{i^\star}\}^2} + \sum_{i\in\mathcal{F}\cap\mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i\in\bar{\mathcal{F}}^c\cap\mathcal{R}} \frac{1}{(\Delta^{\mathrm{v}}_i)^2} + \sum_{i\in\bar{\mathcal{F}}^c\cap\mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta^{\mathrm{v}}_i\}^2}.$$

Main Result: Upper Bound

Theorem 1 (Upper bound)

Define the hardness parameter

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta^{\mathrm{v}}_{i^\star}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{R}} \frac{1}{(\Delta^{\mathrm{v}}_i)^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta^{\mathrm{v}}_i\}^2}.$$

Given an instance $(\nu, \bar{\sigma}^2)$ with probability at least $1-\delta$, our proposed algorithm VA-LUCB *succeeds and terminates in*

$$O\left(H_{\mathrm{VA}} \ln \frac{H_{\mathrm{VA}}}{\delta}\right)$$

time steps.

Main Result: Lower Bound and Almost-Tightness

Theorem 2 (Lower bound)

Given any instance $(\nu, \bar{\sigma}^2)$ *with* $\bar{\sigma}^2 \in (0, 1/4)$, *the optimal expected time complexity* $\tau_\delta^\star$ *satisfies*

$$\tau_\delta^\star = \Omega\left(H_{\mathrm{VA}} \ln \frac{1}{\delta} \right).$$

Main Result: Lower Bound and Almost-Tightness

Theorem 2 (Lower bound)

Given any instance $(\nu, \bar{\sigma}^2)$ *with* $\bar{\sigma}^2 \in (0, 1/4)$*, the optimal expected time complexity* $\tau_\delta^\star$ *satisfies*

$$\tau_\delta^\star = \Omega\left(H_{\mathrm{VA}} \ln \frac{1}{\delta} \right).$$

Corollary 3 (Almost optimality of VA-LUCB)

Given any instance $(\nu, \bar{\sigma}^2)$ *and variance threshold* $\bar{\sigma}^2 \in (0, \frac{1}{4})$*,*

$$\tau_\delta^\star = \widetilde{\Theta}\left(H_{\mathrm{VA}} \ln \frac{1}{\delta} \right),$$

which is achieved by VA-LUCB.

Interpretation of Hardness Parameter I

Hardness parameter:

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta_{i^\star}^{\mathrm{v}}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{R}} \frac{1}{(\Delta_i^{\mathrm{v}})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta_i^{\mathrm{v}}\}^2}.$$

Interpretation of Hardness Parameter I

Hardness parameter:

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta^{\mathrm{v}}_{i^\star}\}^2} + \sum_{i\in\mathcal{F}\cap\mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i\in\bar{\mathcal{F}}^c\cap\mathcal{R}} \frac{1}{(\Delta^{\mathrm{v}}_i)^2} + \sum_{i\in\bar{\mathcal{F}}^c\cap\mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta^{\mathrm{v}}_i\}^2}.$$

- Ascertain optimality and feasibility of $i^\star$;
-
-
-

Interpretation of Hardness Parameter I

Hardness parameter:

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta^{\mathrm{v}}_{i^\star}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{R}} \frac{1}{(\Delta^{\mathrm{v}}_i)^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta^{\mathrm{v}}_i\}^2}.$$

- Ascertain optimality and feasibility of $i^\star$;
- Ascertain suboptimality of arms in the feasible and suboptimal set;
-
-

Interpretation of Hardness Parameter I

Hardness parameter:

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta^{\mathrm{v}}_{i^\star}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{R}} \frac{1}{(\Delta^{\mathrm{v}}_i)^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta^{\mathrm{v}}_i\}^2}.$$

- Ascertain optimality and feasibility of $i^\star$;
- Ascertain suboptimality of arms in the feasible and suboptimal set;
- Ascertain infeasibility of risky and infeasible arms;
-

Interpretation of Hardness Parameter I

Hardness parameter:

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta^{\mathrm{v}}_{i^\star}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{R}} \frac{1}{(\Delta^{\mathrm{v}}_i)^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta^{\mathrm{v}}_i\}^2}.$$

- Ascertain optimality and feasibility of $i^\star$;
- Ascertain suboptimality of arms in the feasible and suboptimal set;
- Ascertain infeasibility of risky and infeasible arms;
- Ascertain either infeasibility or suboptimality of arms in infeasible and suboptimal set.

Interpretation of Hardness Parameter II

Hardness Parameter

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta^{\mathrm{v}}_{i^\star}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{R}} \frac{1}{(\Delta^{\mathrm{v}}_i)^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta^{\mathrm{v}}_i\}^2}.$$

Interpretation of Hardness Parameter II

Hardness Parameter

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta^{\mathrm{v}}_{i^\star}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{R}} \frac{1}{(\Delta^{\mathrm{v}}_i)^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta^{\mathrm{v}}_i\}^2}.$$

As $\bar{\sigma}^2 \to \infty$,

Interpretation of Hardness Parameter II

Hardness Parameter

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta^{\mathrm{v}}_{i^\star}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2}$$

$$+ \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{R}} \frac{1}{(\Delta^{\mathrm{v}}_i)^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta^{\mathrm{v}}_i\}^2}.$$

As $\bar{\sigma}^2 \to \infty$,

$$H_{\mathrm{VA}} = H_{\mathrm{VA}}(\bar{\sigma}^2) \longrightarrow \widetilde{H}_1 = \sum_{i \in [N]} \frac{4}{\Delta_i^2} \quad \text{and} \quad 4\,H_1 \le \widetilde{H}_1 \le 8\,H_1$$

where

$$H_1 := \sum_{i \ne i^\star} \frac{1}{\Delta_i^2}.$$

Particularizes to classical unconstrained result (Even-Dar et al., 2006).

VA-LUCB

Concentration inequalities (for distributions bounded on $[0,1]$):

Lemma 4 (Implication of Hoeffding's and McDiarmid's Inequalities)

Given an instance $(\nu, \bar{\sigma}^2)$*, for any arm* i *with* $T_i(t) \geq 2$ *and* $\varepsilon > 0$ *we have*

$$\mathbb{P}\Big[|\hat{\mu}_i(t) - \mu_i| \geq \varepsilon\Big] \leq 2\exp(-2T_i(t)\varepsilon^2)$$

and

$$\mathbb{P}\Big[|\hat{\sigma}_i^2(t) - \sigma_i^2| \geq \varepsilon\Big] \leq 2\exp(-2T_i(t)\varepsilon^2)$$

where

- $T_i(t)$ *is the number of times arm* i *is sampled before time* t*;*
- $\hat{\mu}_i(t)$ *is the sample mean of arm* i *at time* t*;*
- $\hat{\sigma}_i^2(t)$ *is the unbiased sample variance of arm* i *at time* t*.*

VA-LUCB: Confidence Bounds

- With high probability,

$$\mu_i \in [\hat{\mu}_i(t) - \varepsilon, \hat{\mu}_i(t) + \varepsilon] =: [L_i^{\mu}(t), U_i^{\mu}(t)]$$
$$\sigma_i^2 \in [\hat{\sigma}_i^2(t) - \varepsilon, \hat{\sigma}_i^2(t) + \varepsilon] =: [L_i^{\mathrm{v}}(t), U_i^{\mathrm{v}}(t)]$$

VA-LUCB: Confidence Bounds

- With high probability,

$$\mu_i \in [\hat{\mu}_i(t) - \varepsilon, \hat{\mu}_i(t) + \varepsilon] =: [L_i^{\mu}(t), U_i^{\mu}(t)]$$
$$\sigma_i^2 \in [\hat{\sigma}_i^2(t) - \varepsilon, \hat{\sigma}_i^2(t) + \varepsilon] =: [L_i^{\mathrm{v}}(t), U_i^{\mathrm{v}}(t)]$$

- Define good event

$$E = \bigcap_{t \in \mathbb{N}} \bigcap_{i \in [N]} \left\{ \mu_i \in [L_i^{\mu}(t), U_i^{\mu}(t)], \sigma_i^2 \in [L_i^{\mathrm{v}}(t), U_i^{\mathrm{v}}(t)] \right\}$$

VA-LUCB: Confidence Bounds

- With high probability,

$$\mu_i \in [\hat{\mu}_i(t) - \varepsilon, \hat{\mu}_i(t) + \varepsilon] =: [L_i^{\mu}(t), U_i^{\mu}(t)]$$
$$\sigma_i^2 \in [\hat{\sigma}_i^2(t) - \varepsilon, \hat{\sigma}_i^2(t) + \varepsilon] =: [L_i^{\mathrm{v}}(t), U_i^{\mathrm{v}}(t)]$$

- Define good event

$$E = \bigcap_{t \in \mathbb{N}} \bigcap_{i \in [N]} \left\{ \mu_i \in [L_i^{\mu}(t), U_i^{\mu}(t)], \sigma_i^2 \in [L_i^{\mathrm{v}}(t), U_i^{\mathrm{v}}(t)] \right\}$$

- At time step $t \in \mathbb{N}$ and for arm $i \in [N]$, take

$$\varepsilon = \sqrt{\frac{1}{2T_i(t)} \ln \left(\frac{2Nt^4}{\delta} \right)}$$

VA-LUCB: Confidence Bounds

- With high probability,

$$\mu_i \in [\hat{\mu}_i(t) - \varepsilon, \hat{\mu}_i(t) + \varepsilon] =: [L_i^{\mu}(t), U_i^{\mu}(t)]$$
$$\sigma_i^2 \in [\hat{\sigma}_i^2(t) - \varepsilon, \hat{\sigma}_i^2(t) + \varepsilon] =: [L_i^{\mathrm{v}}(t), U_i^{\mathrm{v}}(t)]$$

- Define good event

$$E = \bigcap_{t \in \mathbb{N}} \bigcap_{i \in [N]} \left\{ \mu_i \in [L_i^{\mu}(t), U_i^{\mu}(t)], \sigma_i^2 \in [L_i^{\mathrm{v}}(t), U_i^{\mathrm{v}}(t)] \right\}$$

- At time step $t \in \mathbb{N}$ and for arm $i \in [N]$, take

$$\varepsilon = \sqrt{\frac{1}{2T_i(t)} \ln\left(\frac{2Nt^4}{\delta}\right)}$$

- Event E occurs with probability at least $1 - \frac{\delta}{2}$.

VA-LUCB: Possibly Feasible Set

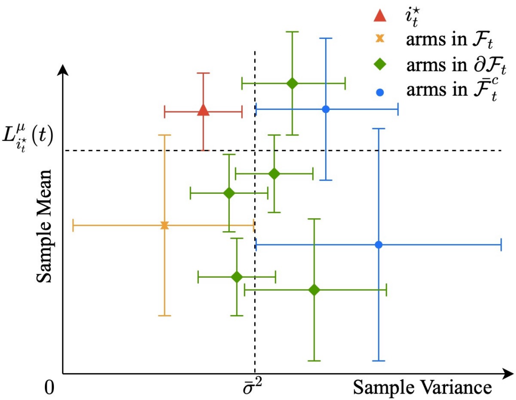
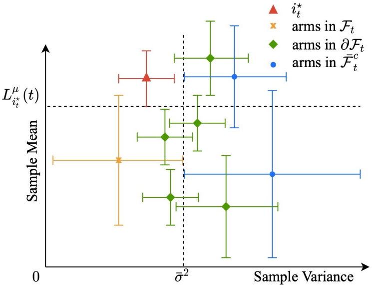


Figure 4: Illustration of the empirical sets. Each dot represents the sample mean and sample variance of each arm at time step t.

Possibly feasible set at time t is $\bar{\mathcal{F}}_t := \mathcal{F}_t \cup \partial\mathcal{F}_t$.

VA-LUCB: Sampling Strategy



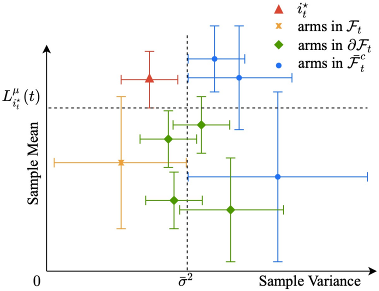


Figure 5: Illustration of the empirical sets.

VA-LUCB: Sampling Strategy

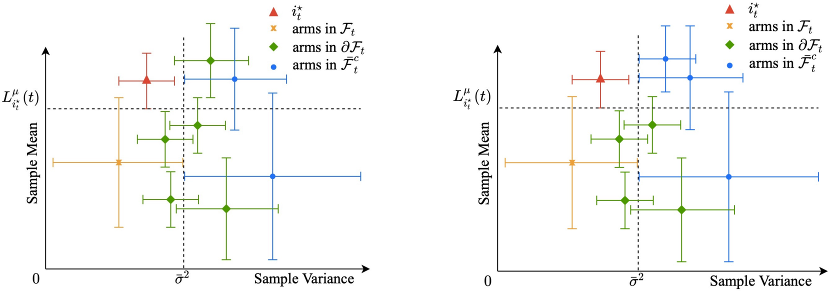


Figure 5: Illustration of the empirical sets.

- Empirical Leader $i_t := \operatorname{argmax}\left\{\hat{\mu}_i(t) : i \in \bar{\mathcal{F}}_t\right\}$

VA-LUCB: Sampling Strategy

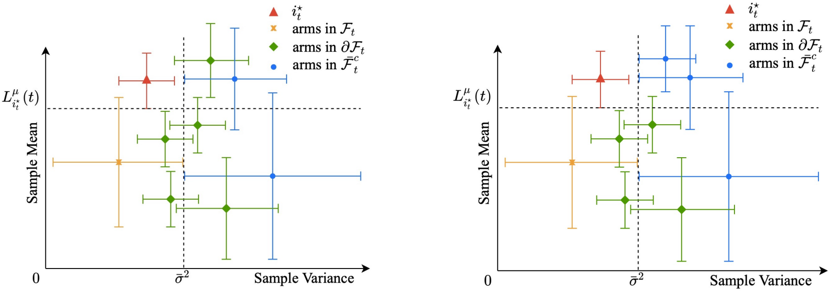


Figure 5: Illustration of the empirical sets.

- Empirical Leader $i_t := \operatorname{argmax}\left\{\hat{\mu}_i(t) : i \in \bar{\mathcal{F}}_t\right\}$
- Empirical Challenger $c_t := \operatorname{argmax}\left\{U_i^\mu(t) : i \in \bar{\mathcal{F}}_t \setminus \{i_t\}\right\}$.

VA-LUCB: Stopping Strategy

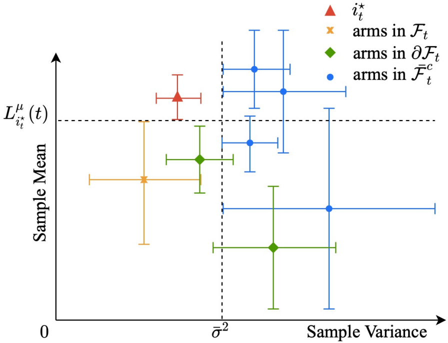


Figure 6: Illustration of the empirical sets.

VA-LUCB: Stopping Strategy

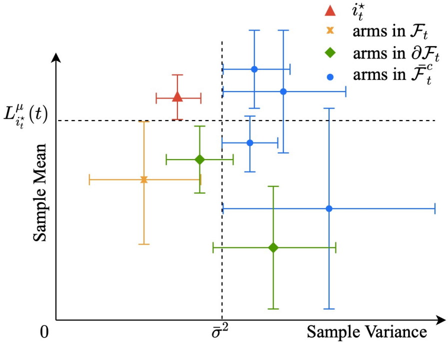


Figure 6: Illustration of the empirical sets.

- Potential set $\mathcal{P}_t := \begin{cases} \{i : L^\mu_{i_t^\star}(t) \le U^\mu_i(t), i \ne i_t^\star\}, & \mathcal{F}_t \ne \emptyset \\ [N], & \mathcal{F}_t = \emptyset \end{cases}$
- Termination condition: $\bar{\mathcal{F}}_t \cap \mathcal{P}_t = \emptyset$.

Time/Sample Complexity of VA-LUCB

Enhance deployment and analysis of LUCB (Kalyanakrishnan et al., 2012)

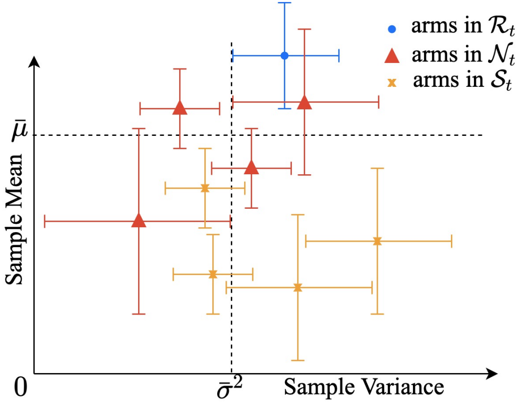


Figure 7: An illustration of the suboptimal and risky set.

Time/Sample Complexity of VA-LUCB

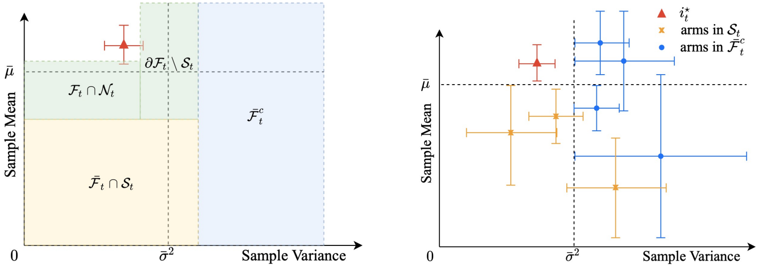


Figure 8: A scenario where all arms have been pulled sufficiently many times.

Time/Sample Complexity of VA-LUCB

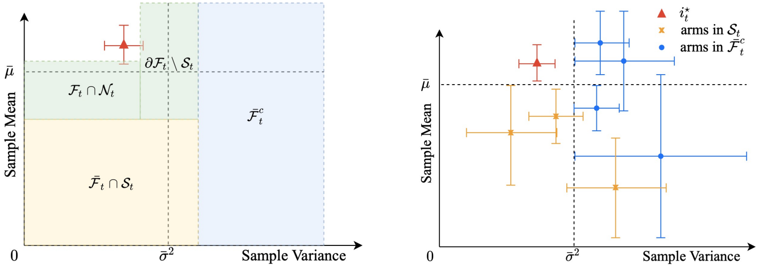


Figure 8: A scenario where all arms have been pulled sufficiently many times.

When **empirically potential best feasible arm set**

$$(\partial\mathcal{F}_t \setminus \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t) = \emptyset,$$

VA-LUCB must stop.

Time/Sample Complexity of VA-LUCB

Lemma: Existence of Arms to Pull

On the event E, if VA-LUCB does not terminate, then at least one of the following statements holds:

- $i_t \in (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t)$.
- $c_t \in (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t)$.

Time/Sample Complexity of VA-LUCB

Lemma: Existence of Arms to Pull

On the event E, if VA-LUCB does not terminate, then at least one of the following statements holds:

- $i_t \in (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t)$.
- $c_t \in (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t)$.

Need to compute the number of pulls needed for each arm i such that

$$i \notin (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t).$$

Time/Sample Complexity of VA-LUCB

Lemma: Existence of Arms to Pull

On the event E, if $\mathrm{VA\text{-}LUCB}$ does not terminate, then at least one of the following statements holds:

- $i_t \in (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t)$.
- $c_t \in (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t)$.

Need to compute the number of pulls needed for each arm i such that

$$i \notin (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t).$$

Lemma: Small Non-Termination Probability

Let $t^\star := 152\ H_{\mathrm{VA}} \ln \frac{H_{\mathrm{VA}}}{\delta}$. For any $t > t^\star$,

$$\mathbb{P}\Big[\mathrm{VA\text{-}LUCB} \text{ does not terminate} \,\Big|\, E\Big] \leq \frac{2\delta}{t^2}.$$

Experiment: VA-LUCB

- Recall the upper bound on $\tau_\delta^\star$ is $O\left(H_{\mathrm{VA}} \ln \frac{H_{\mathrm{VA}}}{\delta}\right)$ where

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta_{i^\star}^{\mathrm{v}}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{R}} \frac{1}{(\Delta_i^{\mathrm{v}})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta_i^{\mathrm{v}}\}^2}.$$

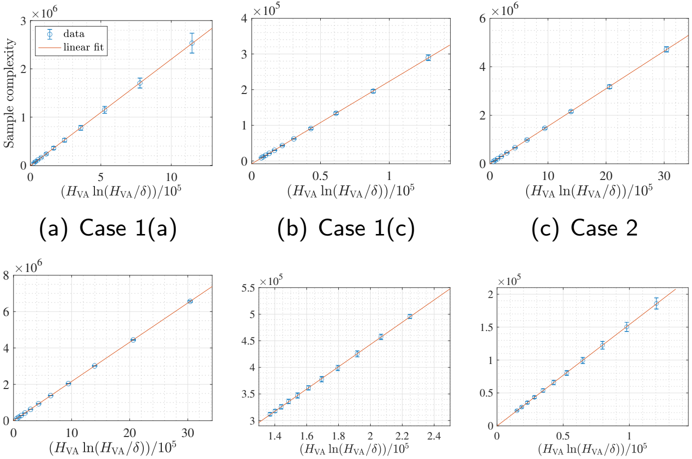
- Test the four terms individually

- Recall the upper bound on $\tau_\delta^\star$ is $O\left(H_{\mathrm{VA}} \ln \frac{H_{\mathrm{VA}}}{\delta}\right)$ where

$$H_{\mathrm{VA}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, \Delta_{i^\star}^{\mathrm{v}}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{R}} \frac{1}{(\Delta_i^{\mathrm{v}})^2} + \sum_{i \in \bar{\mathcal{F}}^c \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, \Delta_i^{\mathrm{v}}\}^2}.$$

- Test the four terms individually
- For example, change $\Delta_{i^\star}$ and observe how sample complexity changes as a function of H_{VA} or $H_{\mathrm{VA}} \ln \frac{H_{\mathrm{VA}}}{\delta}$.

VA-LUCB: Exploring Effects of the Terms in H_{VA}



(a) Case 1(a) (b) Case 1(c) (c) Case 2

(d) Case 3 (e) Case 4(b) (f) Case 4(c)

Figure 9: Time complexities with respect to $H_{\mathrm{VA}} \ln(H_{\mathrm{VA}}/\delta)$ with $\delta = 0.05$.

Experiment: Comparison to Competing Algorithms

- VA-UNIFORM: Randomly and uniformly sample two different arms at each time step.

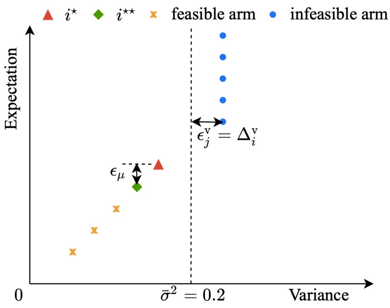
Experiment: Comparison to Competing Algorithms

- VA-UNIFORM: Randomly and uniformly sample two different arms at each time step.
- RISKAVERSE-UCB-BAI: A variant of the algorithm proposed in David et al. (2018):
  - Sample $i_t = \operatorname{argmax}\left\{U_i^{\mu}(t) : i \in \bar{\mathcal{F}}_t\right\}$ (UCB type).
  - Terminate at time t when the confidence radius of the mean of i_t is smaller than ϵ_{v}.

Experiment: Comparison to Competing Algorithms

- VA-UNIFORM: Randomly and uniformly sample two different arms at each time step.
- RISKAVERSE-UCB-BAI: A variant of the algorithm proposed in David et al. (2018):
  - Sample $i_t = \operatorname{argmax}\left\{U_i^{\mu}(t) : i \in \bar{\mathcal{F}}_t\right\}$ (UCB type).
  - Terminate at time t when the confidence radius of the mean of i_t is smaller than ϵ_{v}.
  - Not parameter free: find the ϵ_{v}-approximately feasible and ϵ_{μ}-approximately optimal arm; the confidence radius involves H, the hardness parameter in David et al. (2018).
  - The upper bound is greater than that of VA-LUCB in sample complexity.
  - The lower bound is looser than ours for almost all instances.

Experiment: Comparison to Competing Algorithms



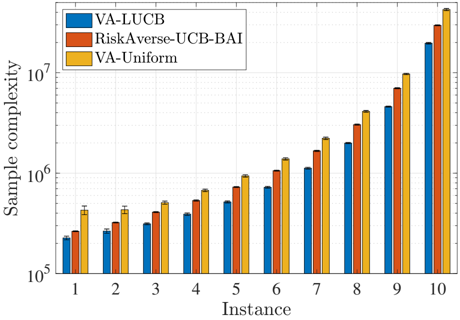


Figure 10: Parameter settings for instance $j \in [10]$. The variance gaps for the infeasible arms $\Delta_i^{\mathrm{v}} = \epsilon_j^{\mathrm{v}} = 0.233 - 0.003 \cdot j$ in instance $j \in [10]$.

Conclusion and Extensions

- Proposed a framework for risk-constrained best arm identification
- Developed an algorithm $\mathrm{VA\text{-}LUCB}$ whose time/sample complexity matches the information-theoretic lower bound (up to constants and log terms)

Conclusion and Extensions

- Proposed a framework for risk-constrained best arm identification
- Developed an algorithm $\mathrm{VA\text{-}LUCB}$ whose time/sample complexity matches the information-theoretic lower bound (up to constants and log terms)
- Future work 1: Development of tracking-based risk-constrained BAI algorithms that can nail down constants
- Future work 2: Other bandit feedback models, e.g., dueling bandits.

Reference I

Amani, S., Alizadeh, M., and Thrampoulidis, C. (2019). Linear stochastic bandits under safety constraints. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, volume 32, pages 9256–9266.

Baudry, D., Gautron, R., Kaufmann, E., and Maillard, O. (2021). Optimal Thompson sampling strategies for support-aware CVaR bandits. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 716–726.

David, Y. and Shimkin, N. (2016). Pure exploration for max-quantile bandits. In *Machine Learning and Knowledge Discovery in Databases*, pages 556–571. Springer.

David, Y., Szörényi, B., Ghavamzadeh, M., Mannor, S., and Shimkin, N. (2018). PAC bandits with risk constraints. In *ISAIM*.

Even-Dar, E., Mannor, S., Mansour, Y., and Mahadevan, S. (2006). Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 7(39):1079–1105.

Hou, Y., Tan, V. Y. F., and Zhong, Z. (2022). Almost optimal variance-constrained best arm identification.

Kagrecha, A., Nair, J., and Jagannathan, K. (2020). Constrained regret minimization for multi-criterion multi-armed bandits. *arXiv preprint arXiv:2006.09649*.

Kalyanakrishnan, S., Tewari, A., Auer, P., and Stone, P. (2012). PAC subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on Machine Learning*, pages 227–234. PMLR.

Sani, A., Lazaric, A., and Munos, R. (2012). Risk-aversion in multi-armed bandits. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, page 3275–3283. Curran Associates Inc.

Vakili, S. and Zhao, Q. (2016). Risk-averse multi-armed bandit problems under mean-variance measure. *IEEE Journal of Selected Topics in Signal Processing*, 10(6):1093–1111.

Wu, Y., Shariff, R., Lattimore, T., and Szepesvári, C. (2016). Conservative bandits. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, pages 1254–1262. PMLR.

Zhu, Q. and Tan, V. Y. F. (2020). Thompson sampling algorithms for mean-variance bandits. In *Proceedings of the 37th International Conference on Machine Learning*, pages 11599–11608. PMLR.