# Best Arm Identification: A Unified Approach to Fixed Budget & Fixed Confidence

Victor Gabillon, Mohammad Ghavamzadeh, Alessandro Lazaric

Inria informatics mathematics

## Abstract

The problem of identifying the best arm(s) in the stochastic multi-armed bandit setting has been studied in the literature from two different perspectives: ***fixed budget*** and ***fixed confidence***.

We propose a unifying approach that leads to a meta-algorithm with **1)** a common structure and **2)** similar theoretical analysis for these two settings.

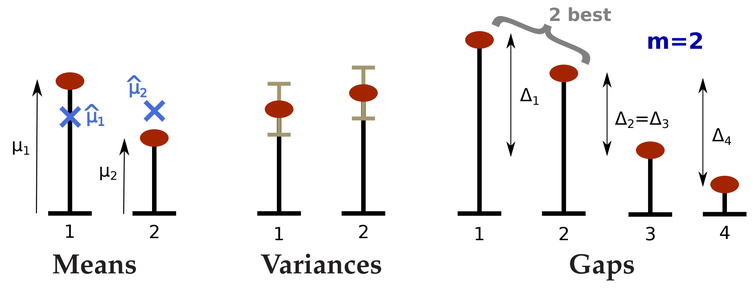
## Problem Formulation

$\boldsymbol{K}$ arms: $\nu_1, \nu_2, \dots \nu_K$ distributions (arms) in $[0, \boldsymbol{b}]$.
At each time step $t$ the player chooses an arm $\boldsymbol{I(t)}$ to pull and receives a sample from $\nu_{I(t)}$.
In arm $k$:
**Number of pulls**: $\mathbf{T_k(t)}$.
**Mean-variance**: $\boldsymbol{\mu_k}$ and $\boldsymbol{\sigma_k^2}$.
**Gap**: $\boldsymbol{\Delta_k} = |\max_{j \neq k} \mu_j - \mu_k|$.
$X_{k,s}$ : $s^{th}$ sample from the arm.

**Estimated mean & variance**:

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{k}}(t) = \frac{1}{T_k(t)} \sum_{s=1}^{T_k(t)} X_{k,s}$$

$$\widehat{\boldsymbol{\sigma^2}}_{\boldsymbol{k}}(\boldsymbol{t}) = \sum_{s=1}^{T_k(t)} \frac{(X_{k,s} - \widehat{\mu}_k(t))^2}{T_k(t)-1}$$




$$\mu_{(m)} = \max_{k \in A}^{m} \mu_k \qquad \text{and} \qquad (m) = \arg\max_{k \in A}^{m} \mu_k \ ,$$

$(m)$ denotes the index of the $m$-th best arm in $A = \{1, \dots, K\}$

$$\{\mu_{(1)}, \dots, \mu_{(m)}\} = \max_{k \in A}^{1..m} \mu_k \qquad \text{and} \qquad S^* = \arg\max_{k \in A}^{1..m} \mu_k \ .$$

Given an accuracy $\epsilon$ and a number of arms $m$, we say that an arm $k$ is ($\epsilon$,$m$)-optimal if $\mu_k \geq \mu_{(m)} - \epsilon$.

**GOAL: Identify a set $S$ of $\boldsymbol{m}$ ($\epsilon$,$m$)-optimal arms.**

**Regret:**

The *arm simple regret*: $r_k = \mu_{(m)} - \mu_k$,

The *simple regret* of $S$, a set of $m$ arms:

$$r_S = \max_{k \in S} r_k = \mu_{(m)} - \min_{k \in S} \mu_k.$$

**Return Strategy:** $\Omega(t)$ is the set of $m$ arms returned after $t$ rounds with simple regret: $r_{\Omega(t)}$

## The Two Settings

**Fixed budget:** The objective is to return a set of $m$ ($\epsilon$,$m$)-optimal arms with the largest possible confidence using a fixed budget of $\boldsymbol{n}$ rounds. The performance is measured by the probability $\widetilde{\delta}$ of not meeting the ($\epsilon$,$m$) requirement,

$$\widetilde{\delta} = \mathbb{P}\big[r_{\Omega(n)} \geq \epsilon\big].$$

**Fixed confidence:** The goal is to stop as soon as possible and return a set of $m$ ($\epsilon$,$m$)-optimal arms with a fixed confidence. Let $\widetilde{n}$ be the time when the algorithm stops. Given a confidence level $\delta$, the forecaster has to guarantee that

$$\mathbb{P}\big[r_{\Omega(\widetilde{n})} \geq \epsilon\big] \leq \delta.$$

The performance of the forecaster is then measured by the number of rounds $\widetilde{n}$ either in expectation or high probability.

## Open problem

Can we show an equivalence between the fixed budget setting and the fixed confidence setting when we don't assume the knowledge of the complexity in the fixed budget problem?

## Unified Gap-based Exploration Algorithm

The unified gap-based exploration (UGapE) meta-algorithm can be implemented in the fixed-budget (UGapEb) and fixed-confidence (UGapEc) settings.

**Definition of Indices:**

For **arm** $k \in A$, $\quad \boldsymbol{B_k(t)} = \max_{i \neq k}^{m} \boldsymbol{U_i(t)} - \boldsymbol{L_k(t)}$ — h. p. upper bound on the simple regret $r_k$

where $\quad \boldsymbol{U_k(t)} = \widehat{\mu}_k(t-1) + \beta_k(t-1)$ — h. p. upper bound on the mean

$\boldsymbol{L_k(t)} = \widehat{\mu}_k(t-1) - \beta_k(t-1).$ — h. p. lower bound on the mean

For a **set** $S$, $\quad B_S(t) = \max_{i \in S} B_i(t)$ — h. p. upper bound on the simple regret $r_S$.

$\beta_k(t-1)$ is a confidence interval. From the Chernoff-Hoeffding bound:

$$\text{UGapEb: } \beta_k(t) = b\sqrt{\frac{a}{T_k(t)}}$$

$$\text{UGapEc: } \beta_k(t) = b\sqrt{\frac{c \log \frac{4K(t-1)^3}{\delta}}{T_k(t)}}$$

**Fixed-Budget:** **UGapEb**$(\epsilon, n, a)$

**Parameters:** accuracy $\epsilon$, budget $n$, exploration parameter $a$

**for** $t = K+1, \dots, n$ **do**
  SELECT-ARM $(t)$
**end for**

Return $\Omega(n) = \arg\min_{J(t)} B_{J(t)}(t)$

**Fixed-Confidence:** **UGapEc**$(\epsilon, \delta, c)$

**Parameters:** accuracy $\epsilon$, confidence level $\delta$, exploration parameter $c$

**while** $B_{J(t)}(t) \geq \epsilon$ **do**
  SELECT-ARM $(t)$
  $t \leftarrow t+1$
**end while**

Return $\Omega(t) = J(t)$

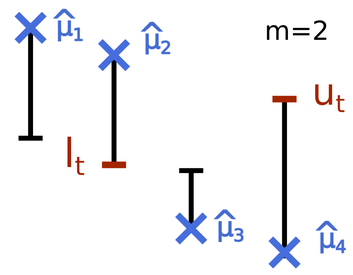
**Arm selection:** UGapEc & GapEb are based on the same sample routine.

$J(t)$ : the set of $m$ estimated best arms

$$J(t) = \arg\min_{k \in A}^{1..m} B_k(t).$$

$$u_t = \arg\max_{j \notin J(t)} U_j(t)$$

$$l_t = \arg\min_{i \in J(t)} L_i(t)$$




**SELECT-ARM** $(t)$

Compute $B_k(t)$ for each arm $k \in A$

Identify the set of $m$ arms $J(t) \in \arg\min_{k \in A}^{1..m} B_k(t)$

Pull the arm $I(t) = \arg\max_{k \in \{l_t, u_t\}} \beta_k(t-1)$

Observe $X_{I(t)}\big(T_{I(t)}(t-1)+1\big) \sim \nu_{I(t)}$

Update $\widehat{\mu}_{I(t)}(t)$ and $T_{I(t)}(t)$

## Analysis

$B_{J(t)}(t)$ **is upper bounded by the quantities of interest (the gaps) independently of the setting**

**Lemma 1** *With high probability, if arm $k$ is pulled at time $t$, $B_{J(t)}(t) \leq \min\big(0, -\Delta_k + 2\beta_k(t-1)\big) + 2\beta_k(t-1)$.*

From this Lemma, we can prove an upper-bound on the simple-regret of both UGapEc & UGapEb:

**Regret Bound for the Fixed-Budget Setting**

**Theorem 1** *If we run UGapEb with parameter $0 < a \leq \frac{n-K}{4H_\epsilon}$, its simple regret $r_{\Omega(n)}$ satisfies*

$$\widetilde{\delta} = \mathbb{P}\big(r_{\Omega(n)} \geq \epsilon\big) \leq 2Kn \exp(-2a),$$

*and this probability is minimized for $a = \frac{n-K}{4H_\epsilon}$.*

**Regret Bound for the Fixed-Confidence Setting Theorem 2**

$$\mathbb{P}\big(r_{\Omega(\widetilde{n}+1)} \leq \epsilon \wedge \widetilde{n} \leq N\big) \geq 1 - \delta,$$

where $N = K + O(H_\epsilon \log \frac{H_\epsilon}{\delta})$ and $c$ has been set to its optimal value $1/2$.

---

$H_\epsilon$ captures the intrinsic difficulty of the ($\epsilon$,$m$)-best arm(s) identification problem independently from the specific setting considered.

$$H_\epsilon = \sum_{i=1}^{K} \frac{b^2}{\max(\frac{\Delta_i + \epsilon}{2}, \epsilon)^2} \ .$$

**Targeted number of pulls in each setting for arm** $i$

**FC**: $T_i(\tilde{n}) = \frac{b^2}{\Delta_i^2}$, **FB**: $T_i(n) = \frac{b^2}{\Delta_i^2} * \frac{n}{H}$

So, if we **assume the knowledge of the complexity** $H$ the two problems have a very similar targeted allocation where the number of pulls of each arm does not depend on the gap of the other arms.

## Extension to variance in the Fixed Confidence setting

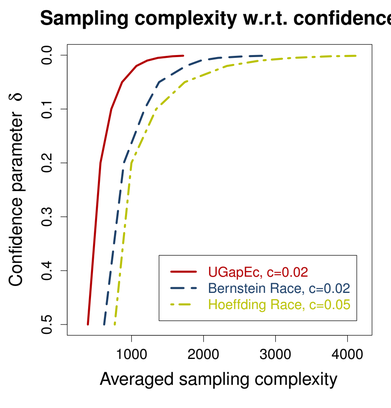
**UGapE-Variance (UGapE-V).** UGapE and its analysis can be naturally and easily extended to take into account the variances of the arms in both cases. We focus here on the Fixed Confidence setting:

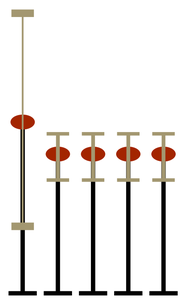
UGapEc-V: $$\beta_k(t) = \sqrt{\frac{2c \log \frac{Kt^3}{\delta}\ \widehat{\sigma}_k^2(t)}{T_k(t)}} + \frac{(7/3)bc \log \frac{Kt^3}{\delta}}{T_k(t) - 1},$$

New definition of complexity,

$$H_\epsilon^\sigma = \sum_{i=1}^{K} \frac{\big(\sigma_i + \sqrt{(13/3)b\Delta_i}\big)^2}{\max\big(\frac{\Delta_i + \epsilon}{2}, \epsilon\big)^2}.$$

**Bernstein Race**: $H_\epsilon^\sigma \approx (\sigma_{(1)}^2 + \sigma_i^2)/\Delta_i^2$, where $\sigma_{(1)}^2$ is the variance of the best arm. In case of $K = 2$, Bernstein Race allocate uniformly, while UGapE-V pulls the arms proportionally to their variances.



For each algorithm, we only consider the results corresponding to the value of $c$ for which the required confidence level $\delta$ is satisfied. We report the results for the value of $c$ with the smallest sample complexity.