A PAC algorithm in relative precision for bandit problem with costly sampling

M. Billaud-Friess* A. Macherey*† A. Nouy* C. Prieur†

Abstract

This paper considers the problem of maximizing an expectation function over a finite set, or finite-arm bandit problem. We first propose a naive stochastic bandit algorithm for obtaining a probably approximately correct (PAC) solution to this discrete optimization problem in relative precision, that is a solution which solves the optimization problem up to a relative error smaller than a prescribed tolerance, with high probability. We also propose an adaptive stochastic bandit algorithm which provides a PAC-solution with the same guarantees. The adaptive algorithm outperforms the mean complexity of the naive algorithm in terms of number of generated samples and is particularly well suited for applications with high sampling cost.

1 Introduction

We consider an optimization problem

$$\max_{\xi \in \Xi} \mathbb{E}[Z(\xi)], \tag{1}$$

where $\mathbb{E}[Z(\xi)]$ is the expectation of a random variable $Z(\xi)$, and where we assume that the set Ξ is finite. Such a problem is encountered in different fields such as reinforcement learning [19] or robust optimization [3].

To solve (1), classical optimization methods include random search algorithms [8, 21], stochastic approximation methods [5, 18] and bandit algorithms [13, 9, 1, 7]. In this paper, we focus on unstructured stochastic bandit problems with a finite number of arms where "arms" stands for "random variables" and corresponds here to the $Z(\xi)$, $\xi \in \Xi$ (see, e.g., [13, Section 4]). Stochastic means that the only way to learn about the probability distribution of arms $Z(\xi)$, $\xi \in \Xi$ is to generate i.i.d. samples from it. Unstructured means that knowledge about the probability distribution of one arm $Z(\xi)$ does not restrict the range of possibilities for other arms $Z(\xi^{'})$, $\xi^{'} \neq \xi$.

Additionally, we suppose here it is numerically costly to sample the random variables $Z(\xi)$, $\xi \in \Xi$. Our aim is thus to solve (1) by sampling as few as possible the random variables $Z(\xi)$, $\xi \in \Xi$. However, it is not feasible to solve (1) almost surely using only a finite number of samples from the random variables $Z(\xi)$, $\xi \in \Xi$. Thus, it is relevant to adopt a Probably Approximately Correct (PAC) approach (see e.g. [6]). For a precision τ_{abs} and a probability $\lambda \in (0, 1)$, a

*Centrale Nantes, LMJL, UMR CNRS 6629, 1 rue de la Noë, 44321 Nantes
†Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP*, LJK, 38000 Grenoble, France

(τ_{abs}, λ)-PAC algorithm returns $\hat{\xi}$ such that

$$\mathbb{P}\left(\mathbb{E}[Z(\xi^\star)] - \mathbb{E}[Z(\hat{\xi})] \leq \tau_{abs}\right) \geq 1 - \lambda, \quad \xi^\star \in \arg\max_{\xi \in \Xi} \mathbb{E}[Z(\xi)]. \tag{2}$$

Until recently, one of the main focus of bandit algorithms was the best arm (random variable) identification [9], through the use of Successive Rejects algorithm or Upper Confidence Bounds algorithms. Such algorithms are $(0, \lambda)$-PAC algorithms, as stated in [6]. Racing algorithms [1] were designed to solve the best arm identification problem too and are mainly analyzed in a finite budget setting, which consists in fixing a maximum number of samples that can be used. While trying to identify the best arm, bandit algorithms also aim at minimizing the regret [2, 4, 7]. More recently, other focuses have emerged, such as the identification of the subset of Ξ containing the m best arms [10, 12] or the identification of "good arms" (also known as thresholding bandit problem) that are random variables whose expectation is greater or equal to a given threshold [11, 20, 17, 14].

The (τ_{abs}, λ)-PAC algorithms mentioned above measure the error in absolute precision. However, without knowing $\mathbb{E}[Z(\xi^\star)]$, providing in advance a relevant value for τ_{abs} is not an easy task. In this work, we rather consider (τ, λ)-PAC algorithms in relative precision that return $\hat{\xi} \in \Xi$ such that

$$\mathbb{P}\left(\mathbb{E}[Z(\xi^\star)] - \mathbb{E}[Z(\hat{\xi})] \leq \tau |\mathbb{E}[Z(\xi^\star)]|\right) \geq 1 - \lambda, \tag{3}$$

where τ and λ are set in advance in $(0, 1)$. We introduce two algorithms that yield a solution $\hat{\xi}$ satisfying (3). The first algorithm builds an estimate precise enough for each expectation $\mathbb{E}[Z(\xi)]$. This naive approach drives a majority of the budget on the random variables with the lowest expectations in absolute value. In order to avoid this drawback and thus to reduce the number of samples required to reach the prescribed relative precision, we propose a second algorithm which adaptively samples random variables exploiting confidence intervals obtained from an empirical Berstein concentration inequality.

The outline of the paper is as follows. In Section 2, we present a Monte-Carlo estimate for the expectation of a single random variable that has been proposed in [15]. It provides an estimation of the expectation with guaranteed relative precision, with high probability. In section 3, we introduce two new algorithms that rely on these Monte-Carlo estimates and yield a solution to (3). Then, we study numerically the performance of our algorithms and compare them to algorithms from the literature, possibly adapted to solve (3).

2 Monte-Carlo estimate with guaranteed relative precision

In what follows, we consider a random variable Z defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote by $\overline{Z}_m$ the empirical mean of Z and by $\overline{V}_m$ its empirical variance, respectively defined by

$$\overline{Z}_m = \frac{1}{m} \sum_{i=1}^{m} Z_i \quad \text{and} \quad \overline{V}_m = \frac{1}{m} \sum_{i=1}^{m} \left(Z_i - \overline{Z}_m\right)^2,$$

where $(Z_i)_{i \geq 1}$ is a sequence of i.i.d. copies of Z. The aim is to provide an estimate $\hat{\mathbb{E}}[Z]$ of $\mathbb{E}[Z]$ which satisfies

$$\mathbb{P}\left(|\hat{\mathbb{E}}[Z] - \mathbb{E}[Z]| \leq \epsilon |\mathbb{E}[Z]|\right) \geq 1 - \delta, \tag{4}$$

with $(\epsilon, \delta) \in (0, 1)^2$ given a priori. For that, we will rely on Theorem 2.1 hereafter.

Theorem 2.1. *If Z takes its values in a bounded interval $[a,b]$, for any $m \in \mathbb{N}$ and $x \in (0,1)$, we have*

$$\mathbb{P}\left(\left|\overline{Z}_m - \mathbb{E}[Z]\right| \leq \sqrt{\frac{2\overline{V}_m \log(3/x)}{m}} + \frac{3(b-a)\log(3/x)}{m}\right) \geq 1 - x. \tag{5}$$

Proof. We simply apply [2, Theorem 1] to $Z - a$ which is a positive random variable whose values are lower than $b - a$. □

Based on Theorem 2.1, several estimates for $\mathbb{E}(Z)$ have been proposed in [15, 16]. We focus in this paper on the estimate introduced in [15, Equation (3.7)].

2.1 Monte-Carlo estimate

Considering a sequence $(d_m)_{m\geq 1}$ in $(0,1)$, we introduce the sequence $(c_m)_{m\geq 1}$ defined, for all $m \geq 1$, by

$$c_m = \sqrt{\frac{2\overline{V}_m \log(3/d_m)}{m}} + \frac{3(b-a)\log(3/d_m)}{m}. \tag{6}$$

Using Theorem 2.1, we see that c_m stands for the half-length of a confidence interval of level $1 - d_m$ for $\mathbb{E}[Z]$, i.e.

$$\mathbb{P}(|\overline{Z}_m - \mathbb{E}[Z]| \leq c_m) \geq 1 - d_m. \tag{7}$$

Let M be an integer-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, such that

$$c_M \leq \epsilon |\overline{Z}_M|, \tag{8}$$

with $\epsilon \in (0,1)$. Then, we define the following estimate

$$\hat{\mathbb{E}}_M[Z] = \overline{Z}_M - \epsilon \ \mathrm{sign}(\overline{Z}_M) c_M. \tag{9}$$

Proposition 2.2. *Let $\epsilon, \delta \in (0,1)$. Assume that Z takes its values in a bounded interval $[a,b]$ and that $(d_m)_{m\geq 1}$ satisfies*

$$\sum_{m\geq 1} d_m \leq \delta. \tag{10}$$

Then the estimate $\hat{\mathbb{E}}_M[Z]$ defined by (9), with M satisfying (8), is such that

$$\mathbb{P}\left(\left|\hat{\mathbb{E}}_M[Z] - \mathbb{E}[Z]\right| \leq \epsilon |\mathbb{E}[Z]|\right) \geq 1 - \delta. \tag{11}$$

Proof. We have

$$\begin{aligned} \mathbb{P}\left(|\overline{Z}_M - \mathbb{E}[Z]| \leq c_M\right) &\geq \mathbb{P}\left(\cap_{m\geq 1}\left\{|\overline{Z}_m - \mathbb{E}[Z]| \leq c_m\right\}\right) \\ &\geq 1 - \sum_{m=1}^{+\infty} \mathbb{P}\left(|\overline{Z}_m - \mathbb{E}[Z]| > c_m\right). \end{aligned} \tag{12}$$

Then using (7) and (10), we deduce that

$$\mathbb{P}\left(|\overline{Z}_M - \mathbb{E}[Z]| \leq c_M\right) \geq 1 - \sum_{m=1}^{+\infty} d_m \geq 1 - \delta. \tag{13}$$

It remains to prove that $\{|\overline{Z}_M - \mathbb{E}[Z]| \leq c_M\}$ implies $\{|\hat{\mathbb{E}}_M[Z] - \mathbb{E}[Z]| \leq \epsilon|\mathbb{E}[Z]|\}$. In the rest of the proof, we assume that $|\overline{Z}_M - \mathbb{E}[Z]| \leq c_M$ holds. Let us recall that $c_M \leq \epsilon|\overline{Z}_M|$. Then, since $\epsilon < 1$, we have

$$|\overline{Z}_M - \mathbb{E}[Z]| \leq c_M \leq \epsilon|\overline{Z}_M| < |\overline{Z}_M|,$$

which implies that $\mathbb{E}[Z]$, $\overline{Z}_M$ and $\hat{\mathbb{E}}_M[Z]$ have the same sign. Therefore,

$$|\hat{\mathbb{E}}_M[Z] - \mathbb{E}[Z]| = \left||\overline{Z}_M| - \epsilon c_M - |\mathbb{E}[Z]|\right|.$$

It suffices to consider the case $\mathbb{E}[Z] > 0$ and we have

$$\mathbb{E}[Z] \geq \overline{Z}_M - c_M \geq \frac{1-\epsilon}{\epsilon} c_M.$$

Therefore

$$\hat{\mathbb{E}}_M[Z] - \mathbb{E}[Z] = \overline{Z}_M - \mathbb{E}[Z] - \epsilon c_M \leq (1-\epsilon) c_M \leq \epsilon \mathbb{E}[Z].$$

Also

$$\mathbb{E}[Z] \leq \overline{Z}_M + c_M \leq (1+\epsilon)\overline{Z}_M,$$

and

$$\hat{\mathbb{E}}_M[Z] - \mathbb{E}[Z] \geq (1-\epsilon^2)\overline{Z}_M - \mathbb{E}[Z] \geq (1-\epsilon)\mathbb{E}[Z] - \mathbb{E}[Z] = -\epsilon\mathbb{E}[Z],$$

which concludes the proof.

□

In practice, the computation of the estimate given by (9) requires a particular choice for the random variable M and for the sequence $(d_m)_{m\geq 1}$. A natural choice for M which satisfies (8) is

$$M = \min\left\{m \in \mathbb{N}^\star \ : \ c_m \leq \epsilon|\overline{Z}_m|\right\}. \tag{14}$$

If the sequence $(d_m)_{m\geq 1}$ is such that

$$\log(3/d_m)/m \underset{m\to+\infty}{\longrightarrow} 0, \tag{15}$$

we have that c_m converges to 0 almost surely. Moreover if $\mathbb{E}[Z] \neq 0$, it is sufficient to ensure that $M < +\infty$ almost surely.

Remark 2.3. *When choosing M as in* (14)*, the estimate defined by* (9) *is the one proposed in [15, equation (3.7)]. A variant of this estimate can be found in [16].*

2.2 Complexity analysis

In this section, we state a complexity result. Following [15], we focus on a particular sequence $(d_m)_{m\geq 1}$ defined by

$$d_m = \delta c m^{-p}, \quad c = \frac{p-1}{p}, \tag{16}$$

which satisfies (10) and (15), for any $p > 1$. The following result extends the result of [15, Theorem 2] stated for random variables Z with range in $[0,1]$.

Proposition 2.4. *Let $0 < \delta \leq 3/4$ and let Z be a random variable taking values in a bounded interval $[a, b]$, with expectation $\mu = \mathbb{E}[Z]$ and variance $\sigma^2 = \mathbb{V}[Z]$. If $\mu \neq 0$ and $(d_m)_{m\geq 1}$ satisfies* (16)*, then M defined by* (14) *satisfies $M < +\infty$ almost surely and*

$$\mathbb{P}\left(M > \left\lceil \frac{2}{\nu}\left(p\log\left(\frac{2p}{\nu}\right) + \log\left(\frac{3}{c\delta}\right)\right)\right\rceil\right) \leq 4\delta/3, \tag{17}$$

where $\lceil\cdot\rceil$ denotes the ceil function and

$$\nu = \min\left(\frac{\max(\sigma^2, \epsilon^2\mu^2)}{(b-a)^2}, \frac{\epsilon^2\mu^2}{(1+\epsilon)^2\max(\sigma^2, \epsilon^2\mu^2)\gamma}\right),$$

with $\gamma = (\sqrt{2 + 2\sqrt{2} + 2/3} + 3)^2$. Moreover,

$$\mathbb{E}(M) \leq \left\lceil \frac{2}{\nu}\left(p\log\left(\frac{2p}{\nu}\right) + \log\left(\frac{3}{c\delta}\right)\right)\right\rceil + 4\delta/3.$$

Proof. See Appendix. □

Remark 2.5. *The result from Proposition 2.4 helps in understanding the influence of parameters (ε, δ) appearing in* (4) *on M. Indeed, we deduce from this result that for $\delta < 1/2$,*

$$\mathbb{E}(M) \lesssim \nu^{-1}\log(\nu^{-1}) + (\nu^{-1} + 1)\log(\delta^{-1}).$$

We first observe a weak impact of δ on the average complexity. When $\epsilon \to 0$, we have $\nu \sim \epsilon^2\frac{\mu^2}{\sigma^2\gamma}$. Then for fixed δ and $\epsilon \to 0$, the bound for $\mathbb{E}(M)$ is in $O(\epsilon^{-2}\frac{\sigma^2}{\mu^2})$. As expected, the relative precision ϵ has a much stronger impact on the average complexity.

3 Optimization algorithms with guaranteed relative precision

In this section we consider a finite collection of bounded random variables $Z(\xi)$ on $(\Omega, \mathcal{F}, \mathbb{P})$, indexed by $\xi \in \Xi$, and such that $\mathbb{E}[Z(\xi)] \neq 0$. Each random variable $Z(\xi)$ takes its values in a bounded interval $[a(\xi), b(\xi)]$, which is assumed to be known. We denote by $\overline{Z(\xi)}_m$ the empirical mean of $Z(\xi)$ and $\overline{V(\xi)}_m$ its empirical variance, respectively defined by

$$\overline{Z(\xi)}_m = \frac{1}{m}\sum_{i=1}^{m} Z(\xi)_i \quad \text{and} \quad \overline{V(\xi)}_m = \frac{1}{m}\sum_{i=1}^{m}\left(Z(\xi)_i - \overline{Z(\xi)}_m\right)^2,$$

where $\{(Z(\xi)_i)_{i\geq 1} : \xi \in \Xi\}$ are independent i.i.d. copies of $\{Z(\xi) : \xi \in \Xi\}$. We also introduce $\#\Xi$ different sequences

$$c_{\xi,m} = \sqrt{\frac{2\overline{V(\xi)}_m\log(3/d_m)}{m}} + \frac{3\,(b(\xi) - a(\xi))\log(3/d_m)}{m},$$

where $(d_m)_{m\geq 1}$ is a positive sequence, independent from ξ, such that $\sum_{m\geq 1} d_m \leq \delta$.
Taking ϵ in $(0, 1)$, for each ξ in Ξ, we define, as in (14),

$$m(\xi) = \min\left\{m \in \mathbb{N}^\star \;:\; c_{\xi,m} \leq \epsilon|\overline{Z(\xi)}_m|\right\}. \tag{18}$$

Then defining $s(\xi) := \operatorname{sign}(\overline{Z(\xi)}_{m(\xi)})$, we propose the following estimate for $\mathbb{E}[Z(\xi)]$:

$$\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)] = \overline{Z(\xi)}_{M(\xi)} - \epsilon\, s(\xi) c_{\xi,m(\xi)}. \tag{19}$$

These notation being introduced, we propose below two algorithms returning $\hat{\xi}$ in Ξ such that

$$\mathbb{P}\left(\mathbb{E}[Z(\xi^\star)] - \mathbb{E}[Z(\hat{\xi})] \leq \tau|\mathbb{E}[Z(\xi^\star)]|\right) \geq 1-\lambda, \quad \xi^\star \in \arg\max_{\xi\in\Xi} \mathbb{E}[Z(\xi)], \tag{20}$$

for given (τ, λ) in $(0,1)^2$.

3.1 Non-adaptive algorithm

We first propose a non-adaptive algorithm that provides a parameter $\hat{\xi}$ satisfying (20), by selecting the maximizer of independent estimates $\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]$ of $\mathbb{E}[Z(\xi)]$ over Ξ.

Algorithm 1 (Non-adaptive)

Input: τ, λ, $\{Z(\xi)\}_{\xi\in\Xi}$.
Output: $\hat{\xi}$
1: Set $\epsilon = \frac{\tau}{2+\tau}$ and $\delta = \lambda/\#\Xi$.
2: **for all** $\xi \in \Xi$ **do**
3: Build an estimate $\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]$ of $\mathbb{E}[Z(\xi)]$ using (19) with ϵ and δ as above.
4: **end for**
5: Select $\hat{\xi}$ such that

$$\hat{\xi} \in \arg\max_{\xi\in\Xi} \hat{\mathbb{E}}_{m(\xi)}\left[Z(\xi)\right].$$

Proposition 3.1. *Let $(\tau, \lambda) \in (0,1)^2$. We assume that, for all $\xi \in \Xi$, $Z(\xi)$ is a bounded random variable with $\mathbb{E}[Z(\xi)] \neq 0$. Moreover we assume that the sequence $(d_m)_{m\geq 1}$ is such that*

$$\sum_{m=1}^{+\infty} d_m \leq \frac{\lambda}{\#\Xi} := \delta \quad \text{and} \quad \log(3/d_m)/m \underset{m\to+\infty}{\to} 0. \tag{21}$$

Then, for all ξ in Ξ, the estimate $\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]$ is well defined and satisfies

$$\mathbb{P}\left(\left|\mathbb{E}[Z(\xi)] - \hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]\right| \leq \epsilon\,|\mathbb{E}[Z(\xi)]|\right) \geq 1-\delta, \tag{22}$$

with $\epsilon = \frac{\tau}{2+\tau}$. Moreover, the output $\hat{\xi}$ of Algorithm 1 satisfies (20).

Proof. The assumptions on $(d_m)_{m\geq 1}$ in (21) combined with $\mathbb{E}[Z(\xi)] \neq 0$ ensure that for all ξ in Ξ , $M(\xi)$ is almost surely finite. Then, for all ξ in Ξ, the estimate $\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]$ is well defined. Applying Proposition 2.2 for each $Z(\xi)$ with $\delta = \lambda/\#\Xi$ and $\epsilon = \frac{\tau}{2+\tau}$, we obtain (22).
Now let $A(\xi) = \left\{\left|\mathbb{E}\left[Z(\xi)\right] - \hat{\mathbb{E}}_{m(\xi)}\left[Z(\xi)\right]\right| \leq \epsilon\,|\mathbb{E}\left[Z(\xi)\right]|\right\}$. By (22), $\mathbb{P}(A(\xi)) \geq 1 - \frac{\lambda}{\#\Xi}$ and by a union bound argument, $\mathbb{P}(\cap_{\xi\in\Xi} A(\xi)) \geq 1-\lambda$. To prove that $\hat{\xi}$ satisfies (20), it remains to prove that $\cap_{\xi\in\Xi} A(\xi)$ implies $\mathbb{E}\left[Z(\xi^\star)\right] - \mathbb{E}[Z(\hat{\xi})] \leq \tau\,|\mathbb{E}\left[Z(\xi^\star)\right]|$. In what follows we suppose that

$\cap_{\xi\in\Xi}A(\xi)$ holds. Since $\epsilon < 1$, $\mathbb{E}[Z(\xi)]$, $\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]$ and $\overline{Z(\xi)}_{m(\xi)}$ have the same sign, that we denote by $s(\xi)$. Since $A(\xi^\star) \cap A(\hat{\xi})$ holds, we have

$$\begin{aligned}\mathbb{E}\left[Z(\xi^\star)\right] - \mathbb{E}[Z(\hat{\xi})] &\leq \mathbb{E}\left[Z(\xi^\star)\right] - \frac{\hat{\mathbb{E}}_{m(\hat{\xi})}[Z(\hat{\xi})]}{1+s(\hat{\xi})\epsilon} \leq \mathbb{E}\left[Z(\xi^\star)\right] - \frac{\hat{\mathbb{E}}_{m(\xi^\star)}\left[Z(\xi^\star)\right]}{1+s(\hat{\xi})\epsilon} \\ &\leq \mathbb{E}\left[Z(\xi^\star)\right] - \frac{1-s(\xi^\star)\epsilon}{1+s(\hat{\xi})\epsilon}\mathbb{E}\left[Z(\xi^\star)\right] = \frac{\epsilon(s(\xi^\star)+s(\hat{\xi}))}{1+s(\hat{\xi})\epsilon}\mathbb{E}\left[Z(\xi^\star)\right].\end{aligned}$$

Then we deduce

$$\mathbb{E}\left[Z(\xi^\star)\right] - \mathbb{E}[Z(\hat{\xi})] \leq \frac{2\epsilon}{1-\epsilon}|\mathbb{E}[Z(\xi^\star)]| = \tau|\mathbb{E}[Z(\xi^\star)]|, \tag{23}$$

which ends the proof.

□

Remark 3.2. *If $\mathbb{E}[Z(\xi^\star)] > 0$, we can prove that $s(\hat{\xi}) = s(\xi^\star) = 1$, so that the inequality* (23) *becomes*

$$\mathbb{E}\left[Z(\xi^\star)\right] - \mathbb{E}[Z(\hat{\xi})] \leq \frac{2\epsilon}{1+\epsilon}|\mathbb{E}[Z(\xi^\star)]|.$$

Therefore, we can set $\epsilon = \frac{\tau}{2-\tau}$ in Algorithm 1 to lower the complexity and still guarantee that $\hat{\xi}$ satisfies (20).

Algorithm 1 provides for each random variable an estimate $\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]$ that satisfies (22). However, as will be illustrated later, this algorithm tends to use many samples for variables with a low expectation in absolute value. We propose in the next subsection an adaptive algorithm avoiding this drawback by using confidence intervals, which results in a lower overall complexity.

3.2 Adaptive algorithm

The idea of the adaptive algorithm is to successively increase the number of samples $m(\xi)$ of a subset of random variables $Z(\xi)$ that are selected based on confidence intervals of $\mathbb{E}[Z(\xi)]$ deduced from the concentration inequality of Theorem 2.1. This algorithm follows the main lines of the racing algorithms [16, section 4]. However racing algorithms do not allow to sample again a random variable discarded in an earlier step of the algorithm. The adaptive algorithm presented hereafter allow it.

In order to present this adaptive algorithm, for each ξ, we introduce the confidence interval $[\beta^-_{m(\xi)}(\xi), \beta^+_{m(\xi)}(\xi)]$, with

$$\beta^-_{\xi,m(\xi)} = \overline{Z(\xi)}_{m(\xi)} - c_{\xi,m(\xi)} \quad \text{and} \quad \beta^+_{\xi,m(\xi)} = \overline{Z(\xi)}_{m(\xi)} + c_{\xi,m(\xi)}. \tag{24}$$

From Equation (7), we have that

$$\mathbb{P}\left(\beta^-_{\xi,m(\xi)} \leq \mathbb{E}(Z(\xi)) \leq \beta^+_{\xi,m(\xi)}\right) \geq 1 - d_{m(\xi)}. \tag{25}$$

We define $\epsilon_{\xi,m(\xi)}$ by

$$\epsilon_{\xi,m(\xi)} = \frac{c_{\xi,m(\xi)}}{|\overline{Z(\xi)}_{m(\xi)}|}$$

if $\overline{Z(\xi)}_{m(\xi)} \neq 0$, or $\epsilon_{\xi,m(\xi)} = +\infty$ otherwise. Letting $s(\xi) := \mathrm{sign}(\overline{Z(\xi)}_{m(\xi)})$, we use as an estimate for $\mathbb{E}[Z(\xi)]$

$$\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)] = \begin{cases} \overline{Z(\xi)}_{m(\xi)} - \epsilon_{\xi,m(\xi)}\ s(\xi)c_{\xi,m(\xi)} & \text{if } \epsilon_{\xi,m(\xi)} < 1, \\ \overline{Z(\xi)}_{m(\xi)} & \text{otherwise.} \end{cases} \tag{26}$$

If $\epsilon_{\xi,m(\xi)} < 1$, we note that

$$\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)] = (\overline{Z(\xi)}_{m(\xi)} \mp c_{\xi,m(\xi)}) \left(1 \pm s(\xi)\epsilon_{\xi,m(\xi)}\right),$$

so that

$$\frac{\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]}{1 \pm s(\xi)\epsilon_{\xi,m(\xi)}} = \beta^{\mp}_{\xi,m(\xi)}. \tag{27}$$

The adaptive algorithm is described in Algorithm 2. At each iteration n, one sample of $Z(\xi)$ is drawn for each ξ in a subset Ξ_n selected according to Equation (28).

Algorithm 2 (Adaptive)

Input: τ, λ, $\{Z(\xi)\}_{\xi\in\Xi}$.
Output: $\hat{\xi}$.

Set $n = 0$, $\Xi_0 = \Xi$, $\epsilon_{\xi,0} = +\infty$ and $m(\xi) = 0$ for all $\xi \in \Xi$.
while $\#\Xi_n > 1$ **and** $\max_{\xi\in\Xi_n} \epsilon_{\xi,m(\xi)} > \dfrac{\tau}{2+\tau}$ **do**
  for all $\xi \in \Xi_n$ **do**
    Sample $Z(\xi)$, increment $m(\xi)$ and update $\epsilon_{\xi,m(\xi)}$.
    Build the estimate $\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]$ using (26).
  end for
  Increment n and put in Ξ_n every $\xi \in \Xi$ such that

$$\beta^{+}_{\xi,m(\xi)} \geq \max_{\nu\in\Xi} \beta^{-}_{\nu,m(\nu)}(\nu). \tag{28}$$

end while
Select $\hat{\xi}$ such that

$$\hat{\xi} \in \arg\max_{\xi\in\Xi_n} \hat{\mathbb{E}}_{m(\xi)}\left[Z(\xi)\right].$$

In the next proposition, we prove that the algorithm returns a solution to (20) under suitable assumptions.

Proposition 3.3. *Let $(\tau, \lambda) \in (0,1)^2$. We assume that $(d_m)_{m\geq 1}$ is a positive sequence that satisfies*

$$\sum_{m=1}^{+\infty} d_m \leq \frac{\lambda}{\#\Xi} \quad \textit{and} \quad \log(3/d_m)/m \underset{m\to+\infty}{\to} 0. \tag{29}$$

Moreover, we assume that, for all ξ in Ξ, $Z(\xi)$ is a bounded random variable with $\mathbb{E}[Z(\xi)] \neq 0$. Then, it holds almost surely that Algorithm 2 stops and $\hat{\xi}$ satisfies (20).

Proof. Let $m_n(\xi)$ denote the number of samples of $Z(\xi)$ at iteration n of the algorithm. We first prove by contradiction that Algorithm 2 stops almost surely. Let us suppose that Algorithm 2 does not stop with probability $\eta > 0$, that means

$$\mathbb{P}\left(\forall n > 0,\ \#\Xi_n > 1 \quad \text{and} \quad \max_{\xi\in\Xi_n} \epsilon_{\xi,m_n(\xi)} > \frac{\tau}{2+\tau}\right) = \eta > 0. \tag{30}$$

Since $(\Xi_n)_{n\geq 1}$ is a sequence from a finite set, we can extract a constant subsequence, still denoted $(\Xi_n)_{n\geq 1}$, equal to $\Xi^\star \subset \Xi$, with $\Xi^\star \neq \emptyset$ such that

$$\mathbb{P}\left(\forall n > 1,\ \max_{\xi\in\Xi^\star} \epsilon_{\xi,m_n(\xi)} > \frac{\tau}{2+\tau}\right) \geq \eta > 0. \tag{31}$$

Since $\log(3/d_m)/m \to 0$ as $m \to +\infty$ and $\mathbb{E}[Z(\xi)] \neq 0$ for all ξ, we have that $\epsilon_{\xi,m} \underset{m\to+\infty}{\overset{a.s}{\to}} 0$ for all ξ in Ξ. Yet, since at each iteration n (from the subsequence), we increase $m_n(\xi)$ for all ξ in $\Xi^\star$, we have that $m_n(\xi) \to \infty$ as $n \to \infty$ for all $\xi \in \Xi^\star$. Therefore, $\lim_{n\to+\infty} \max_{\xi\in\Xi^\star} \epsilon_{\xi,m_n(\xi)} = 0$ holds almost surely, which contradicts (31).

We now prove that $\hat{\xi}$ satisfies (20). For clarity, we remove the index n from $m_n(\xi)$. Defining $A(\xi) = \left\{|\overline{Z(\xi)}_{m(\xi)} - \mathbb{E}[Z(\xi)]| \leq c_{\xi,m(\xi)}\right\}$ for all ξ in Ξ, we proceed as in (13) to obtain

$$\mathbb{P}\left(A(\xi)\right) \geq 1 - \lambda/\#\Xi.$$

Thus, by a union bound argument

$$\mathbb{P}\left(\cap_{\xi\in\Xi} A(\xi)\right) \geq 1 - \lambda.$$

It remains to prove that $\cap_{\xi\in\Xi} A(\xi)$ implies $\mathbb{E}[Z(\xi^\star)] - \mathbb{E}[Z(\hat{\xi})] \leq \tau|\mathbb{E}[Z(\xi^\star)]|$ in order to prove that $\hat{\xi}$ satisfies (20). In the rest of the proof, we suppose that $\cap_{\xi\in\Xi} A(\xi)$ holds. First, for all $\xi \notin \Xi_n$, using (24), we have

$$\mathbb{E}[Z(\xi)] \leq \beta^+_{\xi,m(\xi)} < \max_{\nu\in\Xi} \beta^-_{\nu,m(\nu)} \leq \max_{\nu\in\Xi} \mathbb{E}[Z(\nu)] = \mathbb{E}[Z(\xi^\star)], \tag{32}$$

that implies $\xi^\star \in \Xi_n$. If the stopping condition is $\#\Xi_n = 1$, we then have $\hat{\xi} = \xi^\star$. If the stopping condition is $\max_{\xi\in\Xi_n} \epsilon_{\xi,m(\xi)} \leq \frac{\tau}{2+\tau} < 1$, it means that, for all ξ in Ξ_n, $\epsilon_{\xi,m(\xi)} \leq \frac{\tau}{2+\tau} < 1$. Then for all $\xi \in \Xi_n$, using Proposition 2.2 with $\epsilon = \epsilon_{\xi,m(\xi)} < 1$ and $\delta = \lambda/\#\Xi < 1$ and the fact that $\cap_{\xi\in\Xi} A(\xi)$ holds, we obtain that the estimate $\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]$ satisfies

$$\left|\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)] - \mathbb{E}[Z(\xi)]\right| \leq \epsilon_{\xi,m(\xi)} |\mathbb{E}[Z(\xi)]|. \tag{33}$$

We have that $\epsilon_{\xi,m(\xi)} < 1$ and (27) hold for all $\xi \in \Xi_n$. In particular, since $\hat{\xi}, \xi^\star \in \Xi_n$ we get

$$\begin{aligned}
\mathbb{E}\left[Z(\xi^\star)\right] - \mathbb{E}[Z(\hat{\xi})] &\leq \mathbb{E}\left[Z(\xi^\star)\right] - \frac{\hat{\mathbb{E}}_{m(\hat{\xi})}[Z(\hat{\xi})]}{1 + s(\hat{\xi})\epsilon_{\hat{\xi},m(\hat{\xi})}} \leq \mathbb{E}\left[Z(\xi^\star)\right] - \frac{\hat{\mathbb{E}}_{m(\xi^\star)}\left[Z(\xi^\star)\right]}{1 + s(\hat{\xi})\epsilon_{\hat{\xi},m(\hat{\xi})}} \\
&\leq \mathbb{E}\left[Z(\xi^\star)\right] - \frac{1 - s(\xi^\star)\epsilon_{\xi^\star,m(\xi^\star)}}{1 + s(\hat{\xi})\epsilon_{\hat{\xi},m(\hat{\xi})}} \mathbb{E}\left[Z(\xi^\star)\right] \\
&= \frac{s(\xi^\star)\epsilon_{\xi^\star,m(\xi^\star)} + s(\hat{\xi})\epsilon_{\hat{\xi},m(\hat{\xi})}}{1 + s(\hat{\xi})\epsilon_{\hat{\xi},m(\hat{\xi})}} \mathbb{E}\left[Z(\xi^\star)\right].
\end{aligned}$$

Then we deduce

$$\mathbb{E}\left[Z(\xi^\star)\right] - \mathbb{E}[Z(\hat{\xi})] \leq \frac{2\tau/(2+\tau)}{1 - \tau/(2+\tau)} |\mathbb{E}[Z(\xi^\star)]| = \tau|\mathbb{E}[Z(\xi^\star)]|, \tag{34}$$

which ends the proof. □

Remark 3.4. *As for Algorithm 1 (see Remark 3.2), if* $\mathbb{E}[Z(\xi^\star)] > 0$*, we can set* $\epsilon = \frac{\tau}{2-\tau}$ *in Algorithm 2 to lower the complexity and still guarantee that* $\hat{\xi}$ *satisfies* (20)*.*

Remark 3.5. *A variant of Algorithm 2 using batch sampling would generate several samples of* $Z(\xi)$ *at step 4. The result of Proposition 3.3 also holds for the algorithm with batch sampling. An optimal choice of the number of samples should depend on sampling costs.*

4 Numerical results

In this section, we propose a numerical study of the behaviour of our algorithms on a toy example. We consider the set of random variables $Z(\xi) = f(\xi) + U(\xi)$, $\xi \in \Xi$, where $f(\xi) = \sin(\xi) + \sin(10\xi/3)$, the $U(\xi)$ are i.i.d. uniform random variables over $(-1/20, 1/20)$, and $\Xi = \{3 + 4i/100 : 0 \leq i \leq 100\}$. The numerical results are obtained with the sequence $(d_m)_{m\geq 1}$ defined by (16) with $p = 2$. We set $\tau = 0.1$ and $\lambda = 0.1$.

We first compare our algorithms with two existing ones. The first one is the Median Elimination (ME) algorithm (see [6] for a description of the algorithm), that solves problem (2). We take $\tau_{abs} = \tau|\mathbb{E}[Z(\xi^\star)]|$ to ensure ME algorithm provides a solution that also guarantees (3). Of course, this is not feasible in practice without knowing the solution of the optimization problem or at least a bound of $|\mathbb{E}[Z(\xi^\star)]|$. The second algorithm which we compare to our algorithms is the UCB-V Algorithm (see [2, section 3.1]). It consists in only resampling the random variable whose confidence interval has the highest upper bound. To do so, we replace Steps 3 to 6 of Algorithm 2 by:

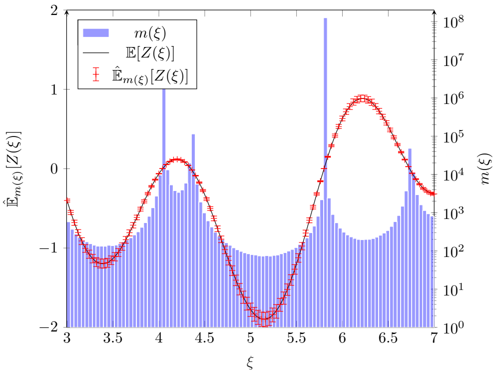
$$\begin{aligned}&\text{Compute } \xi^+ = \arg\max_{\xi\in\Xi_n} \beta^+_{\xi,m(\xi)},\\ &\text{Sample } Z(\xi^+), \text{ increment } m(\xi^+) \text{ and update } \epsilon_{\xi^+,m(\xi^+)}.\end{aligned}$$

We choose these algorithms to perform the comparison because i) ME Algorithm ensures theoretical guarantees similar to ours (although in absolute precision) and ii) the UCB-V Algorithm is optimal, in a sense that we will define later, for solving the optimization problem (1).

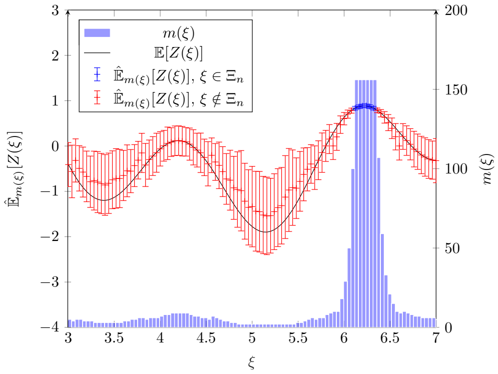
We illustrate on Figure 1 the behavior of algorithms. The results that we show on Figure 1 are the ones of a single run of each algorithm. On the left scale, we plot the estimates $\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]$ as defined in (26) and the associated confidence intervals $[\beta^-_{\xi,m(\xi)}, \beta^+_{\xi,m(\xi)}]$ of level $1 - d_{m(\xi)}$ given by Equation (25). The estimates and confidence intervals for $\xi \in \Xi_n$ are drawn in blue, while the ones for $\xi \notin \Xi_n$ are drawn in red. On the right scale, we plot the number of samples $m(\xi)$ generated for each $\xi \in \Xi$. We observe that Algorithm 1 samples too much the random variables with low expectation in absolute value. This is responsible for the three peaks on $m(\xi)$ observed on Figure 1a. Algorithm 2 avoids this drawback as it does not try to reach the condition $\epsilon_{\xi,m(\xi)} < 1$ for all random variables. The UCB-V algorithm samples mostly the two random variables with highest expectations (more than 99% of the samples are drawn from these random variables). Other random variables are not sufficiently often sampled for reaching rapidly the stopping condition based on confidence intervals. The Median Elimination Algorithm oversamples all random variables in comparison with other algorithms.

Complexity. To perform a quantitative comparison with existing algorithms in the case of costly sampling, a relevant complexity measure is the total number of samples generated after a single run of the algorithm

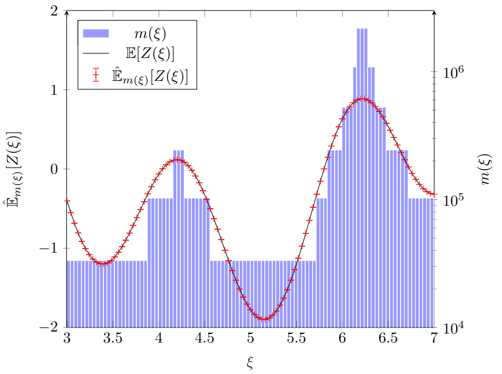
$$\mathcal{M} = \sum_{\xi\in\Xi} m(\xi).$$



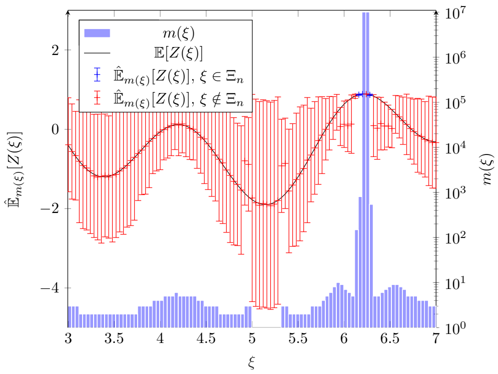
(a) Algorithm 1



(b) Algorithm 2



(c) ME Algorithm



(d) UCB-V Algorithm

Figure 1: Final state of each algorithm after one run with $\tau = 0.1$, $\lambda = 0.1$ and $\tau_{abs} = \tau|\mathbb{E}[Z(\xi^\star)]|$ for ME Algorithm. Left scale : values of the estimates $\hat{\mathbb{E}}_{m(\xi)}[Z(\xi)]$ together with the associated confidence intervals of level $1 - d_{m(\xi)}$. Right scale : values of $m(\xi)$.

Table 1 shows the average complexity $\mathbb{E}(\mathcal{M})$ estimated using 30 independent runs of each algorithm. We observe that the expected complexity of Algorithm 2 is far below the one of the other algorithms. It means that, for the complexity measure $\mathbb{E}(\mathcal{M})$, the adaptive algorithm we have proposed performs the best.

ME Alg.	Alg. 1	Alg. 2	UCB-V Alg.
$2.0 \cdot 10^7$	$1.4 \cdot 10^8$	$1.9 \cdot 10^3$	$1.9 \cdot 10^8$

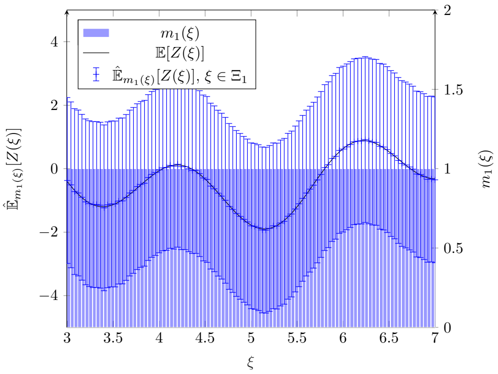
Table 1: Average complexity $\mathbb{E}(\mathcal{M})$, estimated using 30 runs for each algorithm, with $\tau = 0.1$, $\lambda = 0.1$ and $\tau_{abs} = \tau|\mathbb{E}[Z(\xi^\star)]|$ for ME algorithm.

We now compare the four algorithms in terms of expected runtime, that is a measure of complexity taking into account the sampling cost and the cost of all other operations performed by the algorithms. Denoting by $t^\star$ the time (assumed constant) for generating one sample from a distribution, the runtime of an algorithm is a random variable $T = \mathcal{M}t^\star + \mathcal{N}$, where $\mathcal{M}t^\star$ is the sampling time, and $\mathcal{N}$ is the (random) time taken by all other operations. The expected runtime is then $\mathbb{E}(T) = \mathbb{E}(\mathcal{M})t^\star + \mathbb{E}(\mathcal{N})$. From the values of $\mathbb{E}(\mathcal{N})$ and $\mathbb{E}(\mathcal{M})$, estimated over 30 runs of the algorihms, we deduce Table 2, which shows the average runtime $\mathbb{E}(T)$ for different values of $t^\star$. We observe that Algorithm 2 has the smallest average runtime whatever the sampling cost. The first line corresponds to $\mathbb{E}(\mathcal{N})$ and shows that Algorithm 2 performs the best when sampling cost $t^\star = 0$ (or negligible). The impressive gain for large sampling costs $t^\star$ is due to the small value of the average number of samples $\mathbb{E}(\mathcal{M})$ required by the algorithm.

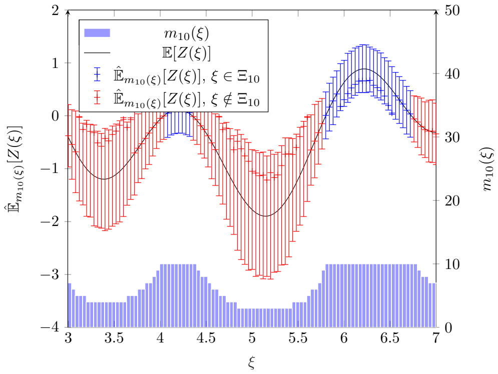
$t^\star$	ME Alg.	Alg. 1	Alg. 2	UCB-V Alg.
0	2.5911	$5.0 \cdot 10^1$	$3 \cdot 10^{-3}$	$1.2 \cdot 10^3$
10^{-6}	$2.0 \cdot 10^1$	$1.9 \cdot 10^2$	$3.8 \cdot 10^{-3}$	$1.4 \cdot 10^3$
10^{-4}	$2.0 \cdot 10^3$	$1.9 \cdot 10^4$	$3.8 \cdot 10^{-1}$	$1.4 \cdot 10^5$
10^{-2}	$2.0 \cdot 10^5$	$1.9 \cdot 10^6$	$3.8 \cdot 10^1$	$1.4 \cdot 10^7$
1	$2.0 \cdot 10^7$	$1.9 \cdot 10^8$	$3.8 \cdot 10^3$	$1.4 \cdot 10^9$

Table 2: Estimated runtime T (in seconds) for different values of $t^\star$, with $\tau = \lambda = 0.1$ and $\tau_{abs} = \tau|\mathbb{E}[Z(\xi^\star)]|$ for ME algorithm. All times are given in seconds.

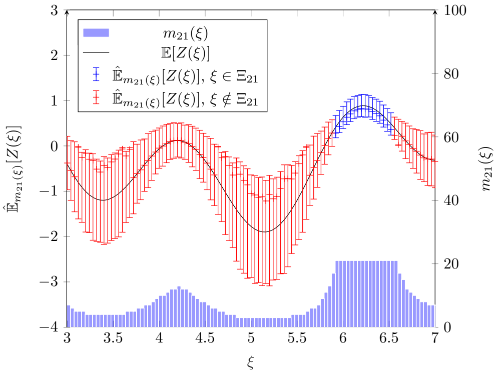
Behavior of Algorithm 2. Now, we illustrate the behavior of Algorithm 2 on Figure 2 and show the evolution with n of Ξ_n and $m_n(\xi)$ for a single run of Algorithm 2, where $m_n(\xi)$ denotes the total number of samples from $Z(\xi)$ generated from iteration 1 to iteration n. When $n = 1$, the algorithm has sampled every random variable once, which is not enough to distinguish some confidence intervals. So Ξ_1 is equal to Ξ. When $n = 10$, some confidence intervals can be distinguished and the algorithm has identified two groups of values where a quasi-maximum could be. These two groups correspond to the two groups of random variables in Ξ_{10}. When $n = 21$, the algorithm has identified the main peak of the function. However, the values of $\epsilon_{\xi,m(\xi)}$ for ξ in Ξ_{21} are not small enough for the algorithm to stop. Then the algorithm continues sampling the random variables in Ξ_n, updating Ξ_n when it is necessary. $\epsilon_{\xi,m(\xi)}$ for ξ in Ξ_n decreases since $m(\xi)$ is increasing for these values of ξ and the algorithm stops at $n = 214$ when $\max_{\xi \in \Xi_{211}} \epsilon_{\xi,m(\xi)} < \frac{\tau}{2+\tau}$.



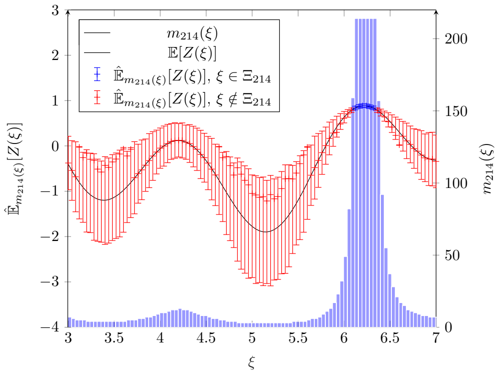
(a) $n = 1$



(b) $n = 10$



(c) $n = 21$



(d) $n = 214$

Figure 2: Evolution of Ξ_n and number of samples $m_n(\xi)$ with n for Algorithm 2 with $\tau = \lambda = 0.1$.

Figure 3 shows the influence of τ and λ on the average complexity $\mathbb{E}(\mathcal{M})$ of Algorithm 2. We observe that τ has a much bigger impact than λ. This observation is consistant with the impact of $\epsilon = \tau/(2+\tau)$ and $\delta = \lambda/\#\Xi$ on the expected number of sampling $\mathbb{E}(M)$ to build an estimate $\hat{\mathbb{E}}_M[Z]$ of $\mathbb{E}[Z(\xi)]$ with relative precision ϵ with probability $1-\delta$ (see Remark 2.5).

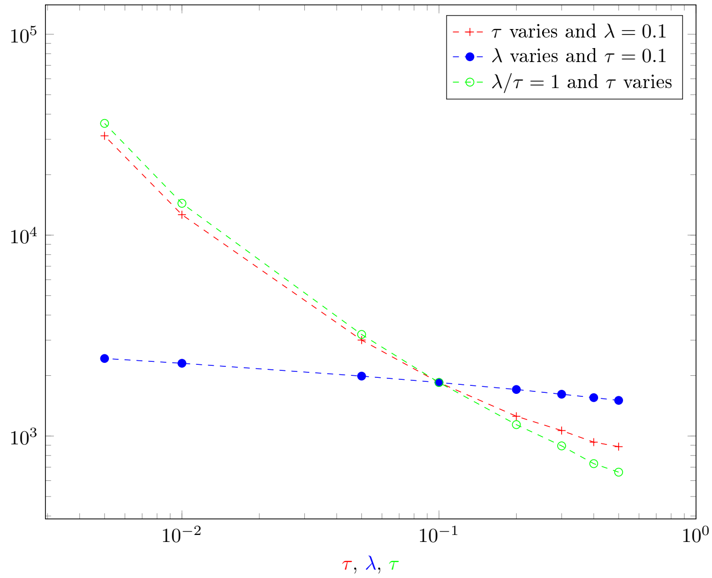


Figure 3: Average complexity $\mathbb{E}(\mathcal{M})$ of Algorithm 2 with respect to τ and λ (in log-log scale).

References

[1] Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Bandit view on noisy optimization. *Optimization for Machine Learning*, page 431, 2011.

[2] Jean-Yves Audibert, Rémi Munos, and Csaba Szepesvári. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876–1902, 2009.

[3] Hans-Georg Beyer and Bernhard Sendhoff. Robust optimization–a comprehensive survey. *Computer methods in applied mechanics and engineering*, 196(33-34):3190–3218, 2007.

[4] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19):1832–1852, 2011.

[5] Vaclav Dupa and Ulrich Herkenrath. Stochastic approximation on a discrete set and the multi-armed. *Sequential Analysis*, 1(1):1–25, 1982.

[6] Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Pac bounds for multi-armed bandit and markov decision processes. In *International Conference on Computational Learning Theory*, pages 255–270. Springer, 2002.

[7] Aurélien Garivier and Olivier Cappé. The kl-ucb algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th annual conference on learning theory*, pages 359–376, 2011.

[8] Wei-Bo Gong, Yu-Chi Ho, and Wengang Zhai. Stochastic comparison algorithm for discrete optimization with estimation. *SIAM Journal on Optimization*, 10(2):384–404, 2000.

[9] R. Munos J.-Y. Audibert, S. Bubeck. Best arm identification in multi-armed bandits. In *Annual Conference on Learning Theory (COLT)*, 2010.

[10] Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *ICML*, volume 12, pages 655–662, 2012.

[11] Hideaki Kano, Junya Honda, Kentaro Sakamaki, Kentaro Matsuura, Atsuyoshi Nakamura, and Masashi Sugiyama. Good arm identification via bandit feedback. *Machine Learning*, 108(5):721–745, 2019.

[12] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

[13] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[14] Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. *arXiv preprint arXiv:1605.08671*, 2016.

[15] Volodymyr Mnih. *Efficient stopping rules*. PhD thesis, University of Alberta, 2008.

[16] Volodymyr Mnih, Csaba Szepesvári, and Jean-Yves Audibert. Empirical bernstein stopping. In *Proceedings of the 25th international conference on Machine learning*, pages 672–679, 2008.

[17] Subhojyoti Mukherjee, Kolar Purushothama Naveen, Nandan Sudarsanam, and Balaraman Ravindran. Thresholding bandits with augmented ucb. *arXiv preprint arXiv:1704.02281*, 2017.

[18] Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on optimization*, 19(4):1574–1609, 2009.

[19] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[20] Chao Tao, Saúl Blanco, Jian Peng, and Yuan Zhou. Thresholding bandit with optimal aggregate regret. In *Advances in Neural Information Processing Systems*, pages 11664–11673, 2019.

[21] Di Yan and H Mukai. Stochastic discrete optimization. *SIAM Journal on control and optimization*, 30(3):594–612, 1992.

A Intermediate results

Here we provide intermediate results used thereafter for the proof of Proposition 2.4 in Appendix B. We first recall a version of Bennett's inequality from [2, Lemma 5].

Lemma A.1. *Let U be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $U \leq b$ almost surely, with $b \in \mathbb{R}$. Let $U_1, \dots, U_m$ be i.i.d. copies of U and $\overline{U}_\ell = \frac{1}{\ell} \sum_{i=1}^{\ell} U_i$. For any $x > 0$, it holds, with probability at least $1 - \exp(-x)$, simultaneously for all $1 \leq \ell \leq m$*

$$\ell \left(\overline{U}_\ell - \mathbb{E}\left[U\right] \right) \leq \sqrt{2m\mathbb{E}\left[U^2\right] x} + b_+ x/3, \tag{35}$$

with $b_+ = \max(0, b)$.

Now, the following result provides a bound with high probability for the estimated variance of an i.i.d. sequence of bounded random variables.

Lemma A.2. *Let X be a bounded random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$, such that $a \leq X \leq b$ almost surely, with $a < b$ two real numbers. Let $X_1, \dots, X_m$ be i.i.d. copies of X and $\overline{V}_m = \dfrac{1}{m}\sum_{i=1}^m (X_i - \overline{X}_m)^2$ where $\overline{X}_m = \frac{1}{m}\sum_{i=1}^m X_i$. Then, for any $x > 0$*

$$\mathbb{P}\left(\overline{V}_m \leq \mathbb{V}[X] + \sqrt{2\mathbb{V}[X]\frac{(b-a)^2 x}{m}} + \frac{x(b-a)^2}{3m}\right) \geq 1 - \exp(-x). \tag{36}$$

Proof. Let us define $U = (X - \mathbb{E}[X])^2$ which satisfies $U \leq (b-a)^2$ almost surely. Applying Lemma A.1 with U defined previously with $\ell = m$ gives for any $x > 0$

$$\mathbb{P}\left(m\left(\overline{U}_m - \mathbb{E}[U]\right) \leq \sqrt{2m\mathbb{E}[U^2]x} + \frac{x(b-a)^2}{3}\right) \geq 1 - \exp(-x).$$

Moreover, as $\overline{U}_m = \overline{V}_m + (\overline{X}_m - \mathbb{E}[X])^2$ and using the boundedness of U we get

$$\mathbb{P}\left(\overline{V}_m \leq \mathbb{E}[U] + \sqrt{2\mathbb{E}[U]\frac{(b-a)^2 x}{m}} + \frac{x(b-a)^2}{3m}\right) \geq 1 - \exp(-x),$$

which ends the proof since $\mathbb{E}[U] = \mathbb{V}[X]$. □

We recall a second result in the line of [15, Lemma 3].

Lemma A.3. *Let q, k be positive real numbers. If $t > 0$ is a solution of*

$$\frac{\log qt}{t} = k, \tag{37}$$

then

$$t \leq \frac{2}{k}\log\frac{2q}{k}. \tag{38}$$

Moreover, if t' is such that

$$t' \geq \frac{2}{k}\log\frac{2q}{k}, \tag{39}$$

then

$$\frac{\log qt'}{t'} \leq k. \tag{40}$$

Proof. Let $t > 0$ be a solution of Equation (37). Since the function log is concave, it holds for all $s > 0$

$$kt = \log(qt) \leq \log(qs) + \frac{t-s}{s}.$$

In particular, for $s = \frac{2}{k} > 0$ we get

$$t \leq \frac{2}{k}\left(\log\frac{2q}{k} - 1\right) \leq \frac{2}{k}\log\frac{2q}{k}, \tag{41}$$

which yields (38).
Now, let $\varphi : s \mapsto \frac{\log(qs)}{s}$ defined for $s > 0$. This function is continuous, strictly increasing on $(0, \frac{e}{q}]$ and strictly decreasing on $[\frac{e}{q}, \infty)$ so it admits a maximum at $t = \frac{e}{q}$. The existence of a solution $t > 0$ of Equation (37) implies $k \leq \frac{q}{e}$. If $k = \frac{q}{e}$ then $t = \frac{e}{q}$ and $\varphi(t)$ is the maximum of φ. For any $t' > 0$, in particular satisfying Equation (39), we have $\varphi(t') \leq \varphi(t) = k$ which is Equation (40). If $0 < k < \frac{q}{e}$, there are two solutions t_1, t_2 to Equation (37) such that $0 < t_1 < \frac{e}{q} < t_2$. By Equations (38) and (39) we have $t' \geq t_2 > \frac{e}{q}$ and since φ is stricly discreasing on $[\frac{e}{q}, \infty)$ it holds $\varphi(t') \leq \varphi(t_2) = k$, that is Equation (40). □

B Proof of Proposition 2.4

Let us define the two following events

$$A = \bigcap_{m \geq 1} A_m \text{ with } A_m = \left\{ \overline{V}_m \leq \sigma^2 + \sqrt{2\sigma^2(b-a)^2 \log(3/d_m)/m} + \log(3/d_m)(b-a)^2/3m \right\},$$

and

$$B = \bigcap_{m \geq 1} B_m \text{ with } B_m = \left\{ |\overline{Z}_m - \mu| \leq c_m \right\}.$$

Applying Lemma A.2 with $x = \log(3/d_m)$ for $A_m, m \geq 1$ together with a union bound argument leads to $\mathbb{P}(A) \geq 1 - \delta/3$. Similarly, using a union bound argument and Theorem 2.1 with $x = \log(3/d_m)$, for $B_m, m \geq 1$, gives $\mathbb{P}(B) \geq 1 - \delta$. By gathering these two results we have

$$\mathbb{P}\left(A \cap B\right) \geq 1 - \left(\mathbb{P}(A^c) + \mathbb{P}(B^c)\right) \geq 1 - \frac{4\delta}{3}, \tag{42}$$

where A^c and B^c correspond respectively to the complementary events of A and B.
It remains to prove that $A \cap B$ implies

$$M \leq \left\lceil \frac{2}{\nu} \left[\log\left(\frac{3}{\delta c}\right) + p \log(\frac{2p}{\nu}) \right] \right\rceil, \tag{43}$$

which will prove (17).In what follows, we suppose that $A \cap B$ holds.
First we derive an upper bound for $\overline{V}_m$. Since A holds, we have

$$\overline{V}_m \leq \sigma^2 + \sqrt{2\sigma^2(b-a)^2 \log(3/d_m)/m} + \log(3/d_m)(b-a)^2/3m. \tag{44}$$

Lemma A.3 with $k = \frac{\sigma^2}{p(b-a)^2}$ and $q = \left(\frac{3}{\delta c}\right)^{1/p}$ gives for any integer $m \geq M_{\sigma^2}$

$$\frac{(b-a)^2}{m} \log \frac{3}{d_m} \leq \sigma^2, \tag{45}$$

where

$$M_{\sigma^2} = \frac{2(b-a)^2}{\sigma^2} \left(p \log\left(\frac{2p(b-a)^2}{\sigma^2}\right) + \log\left(\frac{3}{c\delta}\right) \right).$$

Again, Lemma A.3 with $k = \frac{\epsilon^2\mu^2}{p(b-a)^2}$ and $q = \left(\frac{3}{\delta c}\right)^{1/p}$ gives for any integer $m \geq M_{\epsilon^2\mu^2}$

$$\frac{(b-a)^2}{m} \log \frac{3}{d_m} \leq \epsilon^2\mu^2, \tag{46}$$

where

$$M_{\epsilon^2\mu^2} = \frac{2(b-a)^2}{\epsilon^2\mu^2}\left(p\log\left(\frac{2p(b-a)^2}{\epsilon^2\mu^2}\right) + \log\left(\frac{3}{c\delta}\right)\right).$$

For all $m \geq \min\left(M_{\sigma^2}, M_{\epsilon^2\mu^2}\right)$, i.e. $m \geq M_{\sigma^2}$ or $m \geq M_{\epsilon^2\mu^2}$, we obtain from Equations (44) and (45), or Equations (44) and (46), that

$$\overline{V}_m \leq (1+\sqrt{2}+1/3)\max(\sigma^2, \epsilon^2\mu^2). \tag{47}$$

In what follows, we define $\underline{M} = \min\left(M_{\sigma^2}, M_{\epsilon^2\mu^2}\right)$. Now, we deduce from (47) an upper bound for c_m. By definition,

$$c_m = \sqrt{\frac{2\overline{V}_m \log(3/d_m)}{m}} + \sqrt{\frac{3(b-a)^2\log(3/d_m)^2}{m^2}},$$

then for all integer $m \geq \underline{M}$ and using either Equation (45), or Equation (46), we have

$$c_m \leq \sqrt{\frac{\alpha\log(3/d_m)}{m}}, \tag{48}$$

with $\alpha := (\sqrt{2+2\sqrt{2}+2/3}+3)^2\max(\sigma^2, \epsilon^2\mu^2)$.

Now, using (48), we seek a bound for M, the smallest integer such that $c_M \leq \epsilon|\overline{Z}_M|$. To that aim, let us introduce the integer $M^\star$,

$$M^\star = \min\left\{m \in \mathbb{N}^* : m \geq \underline{M}, \sqrt{\frac{\alpha\log(3/d_m)}{m}} \leq \frac{\epsilon|\mu|}{1+\epsilon}\right\}, \tag{49}$$

and the integer valued random variable M^+

$$M_+ = \min\left\{m \in \mathbb{N}^* : c_m \leq \frac{\epsilon|\mu|}{1+\epsilon}\right\}. \tag{50}$$

If $\underline{M} \geq M_+$ then $M^\star \geq M_+$. Otherwise, $\underline{M} < M_+$ and we have $M_+ = \min\left\{m \geq \underline{M} : c_m \leq \frac{\epsilon|\mu|}{1+\epsilon}\right\}$. Moreover, as Equation (48) holds for all $m \geq \underline{M}$, we get the inclusion

$$\left\{m \in \mathbb{N}^* : m \geq \underline{M}, \sqrt{\frac{\alpha\log(3/d_m)}{m}} \leq \frac{\epsilon|\mu|}{1+\epsilon}\right\} \subset \left\{m \in \mathbb{N}^* : m \geq \underline{M}, c_m \leq \frac{\epsilon|\mu|}{1+\epsilon}\right\}.$$

Taking the min leads again to $M^\star \geq M_+$. Moreover, since B holds, $|\mu| - c_{M_+} \leq |\overline{Z}_{M_+}|$ and using (50) it implies that $c_{M_+} \leq \epsilon|\overline{Z}_{M_+}|$. By definition of M we get $M_+ \geq M$. Hence, we have $M^\star \geq M$. To conclude the proof, it remains to find an upper bound for $M^\star$. Applying again Lemma A.3 with $k = \frac{\epsilon^2\mu^2}{(1+\epsilon)^2\alpha p}$ and $q = \left(\frac{3}{\delta c}\right)^{1/p}$ gives for any integer $m \geq M_f$

$$\frac{\alpha\log(3/d_m)}{m} \leq \frac{\epsilon^2\mu^2}{(1+\epsilon)^2} \tag{51}$$

with

$$M_f = \frac{2(1+\epsilon)^2\alpha}{\epsilon^2\mu^2}\left(p\log\left(\frac{2p(1+\epsilon)^2\alpha}{\epsilon^2\mu^2}\right) + \log\left(\frac{3}{c\delta}\right)\right).$$

If $M_f \leq \underline{M}$, Equations (49) and (51) imply $M^\star = \lceil \underline{M} \rceil$, where $\lceil \cdot \rceil$ denotes the ceil function. Otherwise $M_f > \underline{M}$ and we obtain $M^\star \leq \lceil M_f \rceil$. Thus, it provides the following upper bound

$$M^\star \leq \max\left(\lceil \underline{M} \rceil, \lceil M_f \rceil\right) = \lceil \max\left(\underline{M}, M_f\right) \rceil.$$

Introducing $\nu = \min\left(\frac{\max(\sigma^2, \epsilon^2\mu^2)}{(b-a)^2}, \frac{\epsilon^2\mu^2}{(1+\epsilon)^2\alpha}\right)$ we have from the definition of $M_{\sigma^2}, M_{\epsilon^2\mu^2}$ and M_f

$$M^\star \leq \left\lceil \frac{2}{\nu}\left(p\log\left(\frac{2p}{\nu}\right) + \log\left(\frac{3}{c\delta}\right)\right)\right\rceil. \tag{52}$$

Since $M^\star \geq M$ and $A \cap B$ implies Equation (52), we deduce that $A \cap B$ implies (43), which concludes the proof of the first result.

Let us now prove the result in expectation. Let $K := \left\lceil \frac{2}{\nu}\left(p\log\left(\frac{2p}{\nu}\right) + \log\left(\frac{3}{c\delta}\right)\right)\right\rceil$. We first note that

$$\mathbb{E}(M) = \sum_{k=0}^{\infty} \mathbb{P}(M > k) \leq K + \sum_{k=K}^{\infty} \mathbb{P}(M > k).$$

If $M > k$, then $c_k > \epsilon|\bar{Z}_k|$. For $k \geq K$, we would like to prove that $c_k > \epsilon|\bar{Z}_k|$ implies $(A_k \cap B_k)^c$, or equivalently that $A_k \cap B_k$ implies $c_k \leq \epsilon|\bar{Z}_k|$. For $k \geq K$, A_k implies (48) and (51), and therefore $c_k \leq \frac{\epsilon|\mu|}{1+\epsilon}$. Also, B_k implies $|\bar{Z}_k| \leq |\mu| + c_k$. Combining the previous inequalities, we easily conclude that $A_k \cap B_k$ implies $c_k \leq \epsilon|\bar{Z}_k|$. For $k \geq K$, we then have $\mathbb{P}(M > k) \leq \mathbb{P}(c_k > \epsilon|\bar{Z}_k|) \leq \mathbb{P}((A_k \cap B_k)^c) \leq \mathbb{P}(A_k^c) + \mathbb{P}(B_k^c) \leq 4d_k/3$, and then

$$\mathbb{E}(M) \leq K + \sum_{k=K}^{\infty} 4d_k/3 \leq K + 4\delta/3,$$

which ends the proof.