Minimax Number of Strata for Online Stratified Sampling: the Case of Noisy Samples

Alexandra Carpentier*, Rémi Munos**

(*) Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, UK
(**) SequeL team, INRIA Lille - Nord Europe, France and *Microsoft Research New England, Cambridge, MA, USA*

Abstract

We consider online stratified sampling for Monte Carlo estimation of the integral of a function given a finite budget n of noisy evaluations to the function. In this paper we address the problem of choosing the best number K of strata as a function of n. A large K provides a high quality stratification where an accurate estimate of the integral of f could be computed by an optimal oracle allocation if the variances within each stratum were known. However the performance of an adaptive allocation (which does not know the variance within the strata) compared to the oracle one deteriorates with K. This defines a trade-off between the *stratification quality* and the *pseudo-regret* of an adaptive strategy.

First we provide an improved pseudo-regret upper-bound of order $\tilde{O}(K^{1/3}n^{-4/3})$ for the adaptive allocation MC-UCB introduced in [1]. Then we prove a *lower-bound* on the pseudo-regret of same order, both in terms of K and n, up to a logarithmic factor. Finally we explain how to choose the best value of K given the budget n and deduce a tight minimax (on the class of Hölder continuous functions) optimal bound on the difference between the performance of the adaptive allocation MC-UCB, and the performance of the estimate returned by the optimal oracle strategy.

1. Introduction

The objective of this paper is to provide an efficient strategy for Monte-Carlo integration of a function f over a domain $[0,1]^d$. We assume that we can query the function n times. Querying the function at a time t and at a point $x_t \in [0,1]^d$ provides a noisy sample:

$$f(x_t) + s(x_t)\epsilon_t, \tag{1}$$

where ϵ_t is an independent sample drawn from ν_{x_t}. Here ν_x is a distribution with mean 0, variance 1 and whose shape may depend on x[1]. This model is actually very general (see Section 2), it is the usual model for regression in thermoelastic noise.

[1] It is the usual model for functions in heteroscedastic noise where we emphasize the standard deviation $s(x)$ of the noise at any point x.

Stratified sampling is a well-known variance reduction technique for estimating the integral of f based on Monte-Carlo method. This method consists in partitioning the domain in K subsets called *strata* and then select the number of samples that will be assigned to each stratum (see [12][Subsection 5.5] or [7]). If the variances of the strata were known, the optimal static allocation would allocate a number of samples proportional to the measure of the stratum times the standard deviation of a sample collected from the stratum (see Equation 3 below). We refer to this allocation as optimal oracle strategy for a given partition. However, in situations (which we consider in the paper) where the variations of f and s are unknown, it is not possible to implement this strategy.

Consider first a fixed partition (stratification) of the space. A way to circumvent this problem consists in estimating the variations of f as well as the noise within the strata *online* (exploration) while allocating the samples according to the estimated optimal oracle proportions (exploitation). This approach is considered in [4, 9, 1]. In the long version [3] of the last paper, the authors propose the so-called MC-UCB algorithm which is based on Upper-Confidence-Bounds (UCB) on the standard deviation. They provide upper bounds for the difference between the mean-squared error[2] of the estimate provided by MC-UCB and the mean-squared error of the estimate provided by the optimal oracle strategy (optimal oracle variance) and prove that the adaptive allocation performs almost as well as the optimal oracle strategy.

However, the authors of [3] do not discuss the optimality of their algorithm. As a matter of fact, and to the best of our knowledge, there is no existing lower bound on the regret of any adaptive strategy compared to the optimal oracle one.

Still in the same paper [3], the authors do not discuss the problem of how to *stratify* the space. In particular, they consider that a partition is provided before-hand. However, in many applications, the learner is allowed to construct the partition. In particular, one can often choose the degree of refinement of the partition to use, such as the number of strata of this partition, assuming that all strata have the same size. The question that naturally arises is: what is the best possible number K of strata to use for a given budget n?

This problem has not been thoroughly investigated although some relevant papers are [8, 11, 5]. The recent, state of the art, work of [5] describes a strategy that samples *asymptotically* almost as efficiently as the optimal oracle strategy, and at the same time adapts the size and number of the strata incrementally. Providing theoretical guarantees for the estimate mentioned in this paper appears very difficult, and the authors do not pursue this aim. They however study theoretical properties of stratified estimates computed using *static allocations*, in the limiting case when the diameter of the largest stratum in the partition $\mathcal{N}$ converges to zero. They provide in this case convergence results for n times the mean squared error (MSE) of the estimate under the optimal oracle allocation strategy. They prove that if the diameter of the largest

[2]The mean squared error is measured with respect to the quantity of interest, i.e. the integral of f.

stratum converges to zero, n times this MSE converges to *a constant that is independent of the shape of the strata in* $\mathcal{N}$. The diameter of the largest stratum is the quantity that determines the rate of convergence of n times the MSE. The intuition behind this result is that the oracle strategy consists in selecting the *number of points* in each stratum adaptively to the specific shape (and thus variance) of the stratum : this allocation adapts to the function and if the partition is refined enough, the shape and positioning of the strata does not really matter. This result highlights the fact that the shape or position of the strata is not the important factor in this problem, but rather the refinement of the partition (i.e. the diameter of the largest stratum). This also gives the intuition that partitioning the space into small balls (as e.g. hypercubes for the l_∞ norm) of same diameter is optimal : this partitioning is indeed the one that minimises the diameter of the largest stratum. The main question that remains is then on the *number of strata* (the balls of same diameter) one should use to partition the space.

Contributions. The more refined a partition is, the smaller the variance of the resulting optimal oracle strategy, but also the more difficult it is to estimate the variance within each of the strata of the partition, and thus the more difficult it is to perform almost as well as the optimal oracle strategy. This defines a trade-off which is comparable to the one in model selection (and in all its variants, e.g. density estimation, regression...): the larger the class of models considered, i.e. the larger the number of strata, the smaller the distance between the true model and the best model of the class, i.e. the approximation error, but the larger the estimation error. Selecting the number of strata is thus crucial and this is the problem we address in this paper.

Although the work [5] does not provide finite-time bounds, it introduces very interesting ideas for bounding the approximation error term. Now, as shown in [1] it is possible to build algorithms that have a small estimation error. Thus by constructing tight and finite-time bounds for the approximation error, it is possible to select the number of strata that minimizes an upper bound on the performance. It is however not clear how consistent this choice is, i.e. if a more efficient strategy exists. Thus the last essential ingredient for making the best choice of the number of strata is to derive lower bounds both on the estimation error and on the approximation error.

The objective of this paper is to propose a method for choosing the minimax-optimal number of strata. This is the extended version of the paper [2], where we present the detailed proofs of the results in this paper. Our contributions are the following.

- We first present results on what we call the *quality* $Q_{n,\mathcal{N}}$ *of a partition* $\mathcal{N}$ composed of K strata (i.e., using the previous analogy to model selection, this would represent the approximation error). Using very mild assumptions we derive a lower bound on the variance of the estimate given by the optimal oracle strategy on the optimal oracle partition. We show that if the function and the standard deviation of the noise are $\alpha-$Hölder continuous functions, then (under some assumptions specified later) $Q_{n,\mathcal{N}} = O(\frac{K^{\alpha/d}}{n})$. and this bound is minimax optimal on the class of $\alpha-$Hölder functions.

- We then report regret bounds on the estimation error for the estimate outputted by algorithm MC-UCB of [1]. Actually we are able to improve the analysis of MC-UCB over [1] in terms of the dependence on K. We show a problem independent bound on the pseudo-regret of order[3] $\tilde{O}(K^{1/3}n^{-4/3})$ (whereas it was only of order $\tilde{O}(Kn^{-4/3})$ in [1]).

- We provide a *lower bound* on the pseudo-regret of order $\Omega(K^{1/3}n^{-4/3})$ which *matches the upper-bound of MC-UCB both in terms of the number of strata and the number of samples* (up to a logarithmic factor). This is the main contribution of the paper, and we believe that the proof technique for this bound is original.

- Finally, we combine the results on the quality of a partition and on the pseudo-regret of MC-UCB to provide a value on the number of strata leading to a minimax-optimal trade-off (up to a $\sqrt{\log(n)}$) on the class of $\alpha-$Hölder functions.

The rest of the paper is organized as follows. In Section 2 we formalize the problem and introduce the notations used throughout the paper. Section 3 states the results on the quality of a partition. Section 4 improves the analysis of the MC-UCB algorithm, and establishes the lower bound on the pseudo-regret. Section 5 reports the best trade-off for choosing the minimax optimal number of strata. And in Section 6, we illustrate how important it is to choose carefully the number of strata with some numerical experiments. We finally conclude the paper and suggest future works. The proofs of the statements presented in these sections are in the appendices.

2. Setting

We consider the problem of numerical integration of a function $f : [0,1]^d \to \mathbb{R}$ with respect to the uniform (Lebesgue) measure. We are allowed to query the function n times (we refer to n as the budget) in a sequential way. At round t, when querying the function at a point x_t, we receive a noisy sample $X(t)$ of the form described in Equation 1.

We now assume that the space is stratified in K Lebesgue measurable strata that form a partition $\mathcal{N}$ of the space. We index these strata, that we write Ω_k, with indexes $k \in \{1, \dots, K\}$, and write w_k for their measures (according to the Lebesgue measure). We write $\mu_k = \frac{1}{w_k}\int_{\Omega_k}\mathbb{E}_{\epsilon\sim\nu_x}[f(x)+s(x)\epsilon]dx = \frac{1}{w_k}\int_{\Omega_k}f(x)dx$ for their mean and $\sigma_k^2 = \frac{1}{w_k}\int_{\Omega_k}\mathbb{E}_{\epsilon\sim\nu_x}[(f(x)+s(x)\epsilon-\mu_k)^2]dx$ for their variance. These correspond to the mean and variance of the random variable $X(t)$ conditioned on the fact that the associated state x_t is drawn uniformly at random from the stratum Ω_k.

[3] Here $\tilde{O}$ stands for a O notation up to a polynomial $\log(n)$ factor.

We now state the following assumption, i.e. the Hölder assumption.

Assumption 1 (Hölder assumption)**.** *The functions f and s are $(M, \alpha)-$Hölder continuous, i.e., for $g \in \{f, s\}$, for any x and $y \in [0, 1]^d$, $|g(x) - g(y)| \leq M||x - y||_2^\alpha$.*

We are not going to make this assumption for all the results in this paper, but it is going to be useful for several of them.

We write $\mathcal{A}$ for an algorithm that allocates online the samples x_t by selecting at each time step $1 \leq t \leq n$ the index $k_t \in \{1, \ldots, K\}$ of a stratum and then sampling uniformly x_t in the corresponding stratum Ω_{k_t}. The objective is to return the best possible estimate $\hat{\mu}_n$ of the integral of the function f. We write $T_{k,n} = \sum_{t \leq n} \mathbb{I}\{k_t = k\}$ the number of samples in stratum Ω_k at time n. We denote by $\big(X_{k,t}\big)_{1 \leq k \leq K, 1 \leq t \leq T_{k,n}}$ the samples collected from stratum Ω_k, and we define $\hat{\mu}_{k,n} = \frac{1}{T_{k,n}} \sum_{t=1}^{T_{k,n}} X_{k,t}$ the empirical mean per stratum. Finally, we estimate the integral of f by $\hat{\mu}_n = \sum_{k=1}^K w_k \hat{\mu}_{k,n}$.

If we allocate a deterministic number of samples T_k to each stratum Ω_k and if the samples are independent and chosen uniformly on each stratum Ω_k, we have

$$\mathbb{E}(\hat{\mu}_n) = \sum_{k \leq K} w_k \mathbb{E}[\hat{\mu}_{k,n}] = \sum_{k \leq K} w_k \mu_k = \sum_{k \leq K} \int_{\Omega_k} f(u)du = \int_{[0,1]^d} f(u)du,$$

and also

$$\mathbb{V}(\hat{\mu}_n) = \sum_{k \leq K} \frac{w_k^2 \sigma_k^2}{T_k},$$

where the expectation and the variance are computed according to all the samples that the algorithm collected.

2.1. Pseudo-risk of an algorithm, and oracle pseudo-risk

For a given algorithm $\mathcal{A}$ allocating $T_{k,n}$ samples drawn *uniformly* within stratum Ω_k, we denote by *pseudo-risk* the quantity

$$L_{n,\mathcal{N}}(\mathcal{A}) = \sum_{k \leq K} \frac{w_k^2 \sigma_k^2}{T_{k,n}}. \tag{2}$$

Note that if an algorithm $\mathcal{A}^*$ has access the variances σ_k^2 of the strata, it can choose to allocate the budget in order to minimize the pseudo-risk. The theoretical allocation that minimizes the pseudo-risk is, for stratum k, $T_k^* = \frac{w_k \sigma_k}{\sum_{i \leq K} w_i \sigma_i} n$ (this is the so-called oracle allocation). Defined as such, these optimal numbers of samples can be non-integer values. In this case the proposed optimal allocation is not realizable. But we still use its performance as a benchmark. Indeed, its theoretical pseudo-risk (which is also the variance of the estimate here since the sampling strategy is deterministic) is then

$$L_{n,\mathcal{N}}(\mathcal{A}^*) = \frac{\Big(\sum_{k \leq K} w_k \sigma_k\Big)^2}{n} = \frac{\Sigma_{\mathcal{N}}^2}{n}, \tag{3}$$

where $\Sigma_{\mathcal{N}} = \sum_{k \leq K} w_k \sigma_k$, and is smaller than any realizable pseudo-risk. We also refer in the sequel as optimal proportion to $\lambda_k = \frac{w_k \sigma_k}{\sum_{i \leq K} w_i \sigma_i}$, and as optimal oracle strategy to this allocation strategy. Although, as already mentioned, the optimal allocations (and thus the optimal pseudo-risk) might not be realizable, it is still useful to characterize the smallest possible pseudo-risk on partition $\mathcal{N}$. No static algorithm (even oracle) has a pseudo-risk lower than $L_{n,\mathcal{N}}(\mathcal{A}^*)$ on a given partition $\mathcal{N}$.

2.2. Quality of a partition

The more refined the partition $\mathcal{N}$ the smaller $L_{n,\mathcal{N}}(\mathcal{A}^*)$ (and thus $\Sigma_{\mathcal{N}}$). For the same reason we deduce that stratifying the space is always at least as efficient as not stratifying it, see e.g. [8]. Indeed, consider a partition $\mathcal{N}$ and a more refined partition $\mathcal{N}'$, where by more refined we mean that the strata of $\mathcal{N}'$ are included in those of $\mathcal{N}$. Let $(\Omega_k)_{k \in \mathcal{N}}$ and $(\Omega'_k)_{k \in \mathcal{N}'}$ (respectively $(\sigma_k)_{k \in \mathcal{N}}$, $(\sigma'_k)_{k \in \mathcal{N}'}$ and $(w_k)_{k \in \mathcal{N}}$ and $(w'_k)_{k \in \mathcal{N}'}$) be the strata (resp. the standard deviation and the measure) of $\mathcal{N}$ and $\mathcal{N}'$. By definition of the standard deviation in the strata, we have for any stratum $k \in \mathcal{N}$ that

$$w_k \sigma_k^2 \geq \sum_{l \in \mathcal{N}' : \Omega'_l \subset \Omega_k} w'_l (\sigma'_l)^2,$$

and we thus have by Cauchy-Schwartz

$$w_k \sigma_k^2 \geq \frac{1}{\sum_{l \in \mathcal{N}' : \Omega'_l \subset \Omega_k} w'_l} \Big(\sum_{l \in \mathcal{N}' : \Omega'_l \subset \Omega_k} w'_l \sigma'_l \Big)^2 = \frac{1}{w_k} \Big(\sum_{l \in \mathcal{N}' : \Omega'_l \subset \Omega_k} w'_l \sigma'_l \Big)^2.$$

This implies (by taking the square-root of both sides), and summing over $k \in \mathcal{N}$, that (by definition of $\Sigma_{\mathcal{N}}, \Sigma_{\mathcal{N}'}$)

$$\Sigma_{\mathcal{N}} \geq \Sigma_{\mathcal{N}'}.$$

We immediately deduce that the optimal oracle pseudo-risk using $\mathcal{N}'$ is smaller than the one using $\mathcal{N}$. We thus define the *quality of a partition* $Q_{n,\mathcal{N}}$ as the difference between the variance $L_{n,\mathcal{N}}(\mathcal{A}^*)$ of the estimate provided by the optimal oracle strategy on partition $\mathcal{N}$, and the infimum of the variance of the optimal oracle strategy on *any* partition (optimal oracle partition) (with an arbitrary number of strata):

$$Q_{n,\mathcal{N}} = L_{n,\mathcal{N}}(\mathcal{A}^*) - \inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*). \tag{4}$$

The term $\inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*)$ is an important quantity since it is the optimal, oracle pseudo-risk, on the best possible partition. We will study this object more in depth in Section 3, and we will prove in particular that (see Proposition 1)

$$\inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*) = \inf_{\mathcal{N}' measurable} \frac{\Sigma_{\mathcal{N}}^2}{n} = \frac{1}{n} \Big(\int_{[0,1]^d} s(x) dx \Big)^2.$$

No stratified sampling algorithms, even oracle ones (i.e. that have access to the complete structure of the noise s), can have a smaller pseudo-risk than this benchmark. Additional to this, another important information that Proposition 1 contains is that any sequence of partition $(\mathcal{N}_p)_p$ such that the diameter of its strata goes uniformly to 0 is such that $\lim_p \Sigma_{\mathcal{N}_p} \to \int_{[0,1]^d} s(x)dx$ - in other words, it converges to an optimal partition. This implies in particular that the optimal, oracle pseudo-risk on this sequence of partition converges to $\inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*)$. Another important implication of this result is that *the shape of the strata does not really matter*. Indeed, a partition will have a near-optimal quality when the maximum diameter of its strata is small, and typical such partitions are the uniform ones.

2.3. Pseudo-regret

We also define the *pseudo-regret* of an algorithm $\mathcal{A}$ on a given partition $\mathcal{N}$, as the difference between its pseudo-risk and the variance of the optimal oracle strategy:

$$R_{n,\mathcal{N}}(\mathcal{A}) = L_{n,\mathcal{N}}(\mathcal{A}) - L_{n,\mathcal{N}}(\mathcal{A}^*). \tag{5}$$

2.4. Performance of a strategy: trade-off between quality and pseudo-regret

We assess the performance of an algorithm $\mathcal{A}$ by comparing its pseudo risk to the minimum possible variance of an optimal oracle strategy on the optimal oracle partition:

$$L_{n,\mathcal{N}}(\mathcal{A}) - \inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*) = R_{n,\mathcal{N}}(\mathcal{A}) + Q_{n,\mathcal{N}}. \tag{6}$$

Using the analogy of model selection mentioned in the Introduction, the quality $Q_{n,\mathcal{N}}$ is similar to the approximation error and the pseudo-risk $R_{n,\mathcal{N}}(\mathcal{A})$ to the estimation error.

2.5. Motivation for the model $f(x_t) + s(x_t)\epsilon_t$

Assume that at any time t, the learner can select a point x and collect an observation $F(x, W_t)$, where W_t is an independent noise, that may depend on x. It is the general model for representing evaluations of a noisy function. Set $f(x) = \mathbb{E}_{W_t}[F(x, W_t)]$, and $s(x)\epsilon_t = F(x, W_t) - f(x)$. Since by definition ϵ_t is of mean 0 and variance 1, we have in fact $s(x) = \sqrt{\mathbb{E}_{\nu_x}[(F(x, W_t) - f(x))^2]}$ and $\epsilon_t = \frac{F(x,W_t)-f(x)}{s(x)}$. Observing $F(x, W_t)$ is equivalent to observing $f(x) + s(x)\epsilon_t$, and this implies that the model that we choose is also very general.

There is also an important setting where this model is relevant, and this is for the integration of a function F in high dimension d^*. Stratifying in dimension d^* seems hopeless, since the budget n has to be exponential with d^* if one wants to discretise every direction of the domain: this is the curse of dimensionality. It is necessary to reduce the dimension of stratification by choosing *a small amount* of directions $(1, \ldots, d)$ that are particularly relevant, and control/stratify in these d directions only[4]. Then the control/stratification considers those d directions only, so when sampling at a time t, one chooses $x = (x_1, \ldots, x_d)$, and the other

[4] This is actually a very common technique for computing the price of options, see [7].

$d^* - d$ coordinates $U(t) = (U_{d+1}(t), \ldots, U_{d^*}(t))$ are uniform random variables on $[0,1]^{d^*-d}$ (without any control). When sampling x at time t, we observe $F(x, U(t))$. Thus our model is still valid in this setting with $f(x) = \mathbb{E}_{U(t)\sim\mathcal{U}([0,1]^{d^*-d})}[F(x, U(t))]$, and $s(x)\epsilon_t = F(x, U(t)) - f(x)$.

3. The quality of a partition: Analysis of the term $Q_{n,\mathcal{N}}$.

In this Section, we focus on the *quality of a partition* defined in Section 2.

3.1. Convergence under very mild assumptions

As mentioned in Section 2 (see the extended discussion about the quality of a partition in Subsection 2.2), the more refined the partition $\mathcal{N}$, the smaller $L_{n,\mathcal{N}}(\mathcal{A}^*)$, and thus $\Sigma_{\mathcal{N}}$. Through this monotonicity property, we know that $\inf_{\mathcal{N}} \Sigma_{\mathcal{N}}$ is also the limit of the $(\Sigma_{\mathcal{N}_p})_p$ of a sequence of partitions $(\mathcal{N}_p)_p$ such that the diameter of each stratum goes to 0. We state in the following Proposition that for *any* such sequence, $\lim_{p\to+\infty} \Sigma_{\mathcal{N}_p} = \int_{[0,1]^d} s(x)dx$. Consequently $\inf_{\mathcal{N}} \Sigma_{\mathcal{N}} = \int_{[0,1]^d} s(x)dx$.

Proposition 1. *Let $(\mathcal{N}_p)_p = (\Omega_{k,p})_{k\in\{1,\ldots,K_p\},p\in\{1,\ldots,+\infty\}}$ be a sequence of measurable partitions (where K_p is the number of strata of partition $\mathcal{N}_p$) such that*

- AS1: *For all $p \geq 1$, and $1 \leq k \leq K_p$, $0 < w_{k,p} \leq \upsilon_p$, for some sequence $(\upsilon_p)_{p\geq 1}$, where $\upsilon_p \to 0$ for $p \to +\infty$.*
- AS2: *The diameters according to the $||.||_2$ norm on $\mathbb{R}^d$ of the strata are such that* $\max_k \mathrm{Diam}(\Omega_{k,p}) \leq$ *$D(w_{k,p})$, for some real valued function $D(\cdot)$, such that $D(w) \to 0$ for $w \to 0$.*

If the functions f and s are in $\mathbb{L}_2([0,1]^d)$, then

$$\lim_{p\to+\infty} \Sigma_{\mathcal{N}_p} = \inf_{\mathcal{N}\text{measurable}} \Sigma_{\mathcal{N}} = \int_{[0,1]^d} s(x)dx,$$

which implies that

$$n \times Q_{n,\mathcal{N}_p} = n\Big(L_{n,\mathcal{N}_p}(\mathcal{A}^*) - \inf_{\mathcal{N}'measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*)\Big) \to 0, \ \textit{when } p \to +\infty,$$

where $\inf_{\mathcal{N}'measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*) = \frac{1}{n}\Big(\int_{[0,1]^d} s(x)dx\Big)^2$.

Sketch of Proof. The full proof is in Appendix B. The form of the model and the definition of $\sigma_{k,p}$ imply that

$$\sigma_{k,p}^2 = \frac{1}{w_{k,p}} \int_{\Omega_{k,p}} \big(f(x) - \frac{1}{w_{k,p}} \int_{\Omega_{k,p}} f(u)du\big)^2 dx + \frac{1}{w_{k,p}} \int_{\Omega_k} s(x)^2 dx, \tag{7}$$

where in the entire proof, we use a double index (k,p) for referring to quantities defined on partition $\mathcal{N}_p$.

We first prove that the result holds for uniformly continuous functions, and then generalize to $\mathbb{L}_2$ functions based on a density argument.

Step 1: Convergence when f and s are uniformly continuous: Assume that f and s are uniformly continuous with respect to the $||.||_2$ norm. For any $\upsilon > 0$, there exists η *s.t.* $\forall x, |s(x+u) - s(x)| \leq \upsilon$ and $|f(x+u) - f(x)| \leq \upsilon$ where $u \in \mathcal{B}_{2,d}(\eta)$. We choose p large enough so that the size of the strata is smaller than υ, and their diameter is smaller than η (it is possible to do so since the diameter of the strata shrinks to 0 as $K_p \rightarrow_p \infty$). From Equation 7 we deduce that

$$\sigma_{k,p}^2 - (\frac{1}{w_{k,p}} \int_{\Omega_{k,p}} s)^2 \leq 2\upsilon^2,$$

and using the concavity of the square-root function, we have $\sum_k w_{k,p}\sigma_{k,p} - \int_{[0,1]^d} s \leq \sqrt{2}\upsilon$, which concludes the proof for uniformly continuous functions.

Step 2: Generalization to the case where f and s are in $\mathbb{L}_2([0,1)^d)$: From the density property of the uniformly continuous functions in $\mathbb{L}_2([0,1]^d)$ (with respect to the $||.||_2$ norm), we deduce that for any p and υ, there exists two uniformly continuous function f_υ and s_υ such that:

$$\left| \sum_{k=1}^{K_p} w_{k,p}\sigma_{k,p} - \sum_{k=1}^{K_p} \sqrt{w_{k,p}} \sqrt{\int_{\Omega_{k,p}} \Big(f_\upsilon(x) - \frac{1}{w_{k,p}} \int_{\Omega_{k,p}} f_\upsilon(u)du \Big)^2 dx - \int_{\Omega_{k,p}} s_\upsilon^2(x)dx} \right| \leq \upsilon,$$

and also that $\int_\Omega |s(x) - s_\upsilon(x)|dx \leq \sqrt{\frac{\upsilon}{2}}$. One concludes by combining those two inequalities with Step 1. □

In Proposition 1, even though the optimal oracle allocation might not be realizable (in particular if the number of strata is larger than the budget), we can still compute the quality of a partition, as defined in 4. It does not correspond to any reachable pseudo-risk, but rather to a lower bound on any (even oracle) static allocation.

When f and s are in $\mathbb{L}_2([0,1]^d)$, for any appropriate sequence of partitions $(\mathcal{N}_p)_p$, $\Sigma_{\mathcal{N}_p}$ (which is the principal ingredient of the variance of the optimal oracle allocation) converges to the smallest possible $\Sigma_{\mathcal{N}}$ for given f and s. This smallest $\Sigma_{\mathcal{N}}$ is actually expressed only in function of s - the term in f vanishes - and we have

$$\inf_{\mathcal{N}\text{measurable}} \Sigma_{\mathcal{N}} = \int_{[0,1]^d} s(x)dx,$$

which implies

$$\inf_{\mathcal{N} measurable} L_{n,\mathcal{N}}(\mathcal{A}^*) = \frac{1}{n}\Big(\int_{[0,1]^d} s(x)dx \Big)^2.$$

This is why, as we already mentioned, the quantity $\int_{[0,1]^d} s(x)dx$ is important. Note however that the assumptions required above are not sufficient to obtain a *rate of convergence*. We now make a stronger assumption under which a finite-time analysis can be considered.

3.2. Finite-Time analysis under a Hölder assumption:

The Hölder assumption enables to consider arbitrarily non-smooth functions (for small α, the function can vary arbitrarily fast), and is thus a fairly general assumption. We also consider the following partitions in K hyper-cubic strata.

Definition 1. *We write $\mathcal{N}_K$ the partition of $[0,1]^d$ in K hyper-cubic strata of measure $w_k = w = \frac{1}{K}$ and side length $(\frac{1}{K})^{1/d}$: we assume for simplicity that there exists an integer l such that $K = l^d$.*

The following Proposition holds.

Proposition 2. *Under Assumption 1 we have for any partition $\mathcal{N}_K$ as defined in Definition 1 that*

$$\Sigma_{\mathcal{N}_K} - \int_{[0,1]^d} s(x)dx \leq \sqrt{2d}M\Big(\frac{1}{K}\Big)^{\alpha/d}, \tag{8}$$

which implies

$$Q_{n,\mathcal{N}_K} \leq \frac{2\sqrt{2d}M\Sigma_{\mathcal{N}_1}}{n}\Big(\frac{1}{K}\Big)^{\alpha/d},$$

where $\mathcal{N}_1$ stands for the "partition" with one stratum.

Sketch of Proof for Proposition 2. (the full proof is provided in Appendix C). We deduce from Assumption 1 that

$$\frac{1}{w_k}\int_{\Omega_k}\Big(f(x) - \frac{1}{w_k}\int_{\Omega_k} f(u)du\Big)^2 dx + \frac{1}{w_k}\int_{\Omega_k} s^2(x)dx - (\frac{1}{w_k}\int_{\Omega_k} s(u)du)^2 \leq 2M^2 d\Big(\frac{1}{K}\Big)^{2\alpha/d}.$$

Then, by using Equation 7 and by summing over all strata, we deduce Equation 8. Now the result on the quality follows from the fact that

$$\Sigma^2_{\mathcal{N}_K} - (\int_{[0,1]^d} s(x)dx)^2 = (\Sigma_{\mathcal{N}_K} - \int_{[0,1]^d} s(x)dx)(\Sigma_{\mathcal{N}_K} + \int_{[0,1]^d} s(x)dx) \leq 2\Sigma_{\mathcal{N}_1}(\Sigma_{\mathcal{N}_K} - \int_{[0,1]^d} s(x)dx).$$

□

Here is the corresponding lower bound.

Proposition 3. *Let $0 < \alpha \leq 1$. There exists two functions (f, s) that are $(1,\alpha)-$Hölder (satisfying Assumption 1) and such that, for any partition $\mathcal{N}$ in less than K convex and measurable strata, the associated $\Sigma_{\mathcal{N}}$ verifies*

$$\inf_{\mathcal{N} \text{ measurable, } \mathcal{N} \text{ has } K \text{ convex strata}} \Sigma_{\mathcal{N}} - \int_{[0,1]^d} s(x)dx \geq c(d)\Big(\frac{1}{K}\Big)^{\alpha/d},$$

where $c(d) = \sqrt{\frac{\pi^{d/2}}{2\times 32^2 8^d \Gamma(d/2+1)}} > 0$ is a fixed constant that depends on d only. This implies for these functions that the associated quality satisfies

$$\inf_{\mathcal{N} \text{ measurable, } \mathcal{N} \text{ has } K \text{ convex strata}} Q_{n,\mathcal{N}_K} \geq \frac{c(d)^2}{n}\Big(\frac{1}{K}\Big)^{\alpha/d}.$$

Sketch of Proof for Proposition 3. For any K, it is possible to construct two functions such that $\mathcal{N}_K$ is an optimal partition in K strata for these functions, such that (f, s) are $(1, \alpha)$ Hölder, such that $s = 0$ and f attains the Hölder exponent in many points, and such that it is not well approximated by functions that are constant by parts on the strata of $\mathcal{N}_K$. For these functions, the lower bound holds. □

The full proof for this proposition is provided in Appendix D.

3.3. General comments

The impact of α and d. The quantity $Q_{n,\mathcal{N}_K}$ increases with the dimension d, because the Hölder assumption becomes less constraining with d. This can easily be seen since a hyper-cubic stratum of measure w has a diameter of order $w^{1/d}$. $Q_{n,\mathcal{N}_K}$ decreases with the smoothness α of the function, which is a natural effect of the Hölder assumption. Note also that when defining the partitions $\mathcal{N}_K$ in Definition 1, we made the important assumption that $K^{1/d}$ is an integer. This fact is of little importance in small dimension, but matters in high dimension, as illustrated in the last remark of Section 5.

Minimax optimality of this rate. The rate $n^{-1}K^{-\alpha/d}$ is minimax optimal on the class of $\alpha-$Hölder functions since for any n and K one can build a function with Hölder exponent α such that the corresponding $\Sigma_{\mathcal{N}_K}$ is at least $\int_{[0,1]^d} s(x)dx + cK^{-\alpha/d}$ for some constant c (see Proposition 3). Also, choosing small strata of same shape and size is also minimax optimal on the class of Hölder functions, since the Hölder assumption provides no information on the localization of the irregularities of the function, but just a bound on their magnitude.

Discussion on the shape of the strata. Whatever the shape of the strata, as long as the diameter of the largest stratum goes to zero[5], $\Sigma_{\mathcal{N}}$ converges to $\int_{[0,1]^d} s(x)dx$, see Proposition 1. Thus the shape of the strata has an influence on the second order term only. This result was already made explicit, in a different setting and under different assumptions, in [5]. Moreover, the important point for making sure that the second order term vanishes to 0 at a rate faster than $1/n$ is that the diameter of the largest stratum goes to 0. Intuitively, the optimal partition in K strata is then a partition in K balls of equal diameter (we consider in this work hypercubes which are l_∞ balls). This intuition is confirmed formally in this paper, where we proved that choosing small strata of same shape and size is optimal on the class of Hölder functions, see Propositions 3 and 2. The important question that is remaining is then on the optimal number of strata that one should consider.

[5]And note that in this *noisy* setting, if the diameter of the strata does not go to 0 on non homogeneous part of f and s, then the standard deviation corresponding to the allocation is larger than $\int_{[0,1]^d} s(u)du$.

The decomposition of the variance. Note that the variance σ_k^2 within each stratum Ω_k comes from two sources. First, from the noise $s(x)$ which contributes by $\frac{1}{w_k}\int_{\Omega_k} s(x)^2 dx$. Second, from the mean f -which is not a constant function- which contributes by $\frac{1}{w_k}\int_{\Omega_k}\big(f(x)-\frac{1}{w_k}\int_{\Omega_k} f(u)du\big)^2 dx$. Note that when the size of Ω_k goes to 0, this later contribution vanishes, and the optimal allocation is then proportional to $\sqrt{w_k\int_{\Omega_k} s(x)^2 dx + o(1)} = \int_{\Omega_k} s(x)dx + o(1)$. This means that for small strata, the variation in the mean are negligible when compared to the contribution coming from the noise.

4. MC-UCB and a matching lower bound

4.1. Improved analysis of MC-UCB

In this subsection, we describe a slight modification of the algorithm $MC-UCB$ introduced in [1]. The algorithm computes a high-probability bound on the standard deviation of each arm and samples the arms proportionally to their upper-bounds times the corresponding weights. The only difference with the algorithm described in [1] is that we change the form of the high-probability upper confidence bound on the standard deviations, in order to simplify the proofs, and improve the analysis. The algorithm takes as input four parameters: b and $f_{\max}$, which are linked to the distribution of the arms, δ which is a (small) probability, and the partition $\mathcal{N}_K$. We remind in Figure 1 the algorithm MC-UCB.

Input: b, $f_{\max}$, δ, $\mathcal{N}_K$.
Set $A = 2(2f_{\max}+1)\sqrt{(2f_{\max}+3b+12f_{\max}^2)\log(6nK/\delta)}$
Initialize: Sample 2 points in each stratum.
for $t = 2K+1,\ldots,n$ **do**
  Compute $B_{k,t} = \frac{w_k}{T_{k,t-1}}\Big(\hat{\sigma}_{k,t-1} + A\sqrt{\frac{1}{T_{k,t-1}}}\Big)$ for each stratum $1 \le k \le K$
  Sample a point in stratum $k_t \in \arg\max_{1\le k\le K} B_{k,t}$
end for
Output: $\hat{\mu}_n = \sum_{k=1}^K w_k \hat{\mu}_{k,n}$

Figure 1: The pseudo-code of the MC-UCB algorithm. The empirical standard deviations and means $\hat{\sigma}_{k,t}^2$ and $\hat{\mu}_{k,t}$ are computed using Equations 9 and 10.

The estimates of $\hat{\sigma}_{k,t-1}^2$ and $\hat{\mu}_{k,t-1}$ are computed according to

$$\hat{\sigma}_{k,t-1}^2 = \frac{1}{T_{k,t-1}}\sum_{i=1}^{T_{k,t-1}}\left(X_{k,i}-\hat{\mu}_{k,t-1}\right)^2 , \tag{9}$$

and

$$\hat{\mu}_{k,t-1} = \frac{1}{T_{k,t-1}}\sum_{i=1}^{T_{k,t-1}} X_{k,i} . \tag{10}$$

4.2. Upper bound on the pseudo-regret of algorithm MC-UCB.

We first state a sub-Gaussian assumption on the noise ϵ_t, satisfied for e.g., Gaussian as well as bounded distributions.

Assumption 2. *There exist $b > 0$ such that $\forall x \in [0,1]^d$, $\forall t$, and $\forall \lambda < \frac{1}{b}$,*

$$\mathbb{E}_{\nu_x}\Big[\exp(\lambda \epsilon_t)\Big] \leq \exp\Big(\frac{\lambda^2}{2(1-\lambda b)}\Big), \quad \text{and} \quad \mathbb{E}_{\nu_x}\Big[\exp(\lambda \epsilon_t^2 - \lambda)\Big] \leq \exp\Big(\frac{\lambda^2}{2(1-\lambda b)}\Big).$$

We also state an assumption on f and s.

Assumption 3. *The functions $|f|$ and s are bounded by $f_{\max}$.*

Note that since the functions f and s are defined on $[0,1]^d$, if Assumption 1 is satisfied, then Assumption 3 holds for any constant $f_{\max}$ larger than $\max(|f(0)|, s(0)) + Md^{\alpha/2}$.

We now prove the following bound on the pseudo-regret. Note that we state it on partitions $\mathcal{N}_K$, but it actually holds for any partition in K strata.

Proposition 4. *Under Assumptions 2 and 3 (where we choose $f_{\max} \geq 1$), the pseudo-regret of MC-UCB run on a partition $\mathcal{N}_K$ with $n \geq \max(4K, 2b\log(2n))$ and $\delta = n^{-2}$, satisfies*

$$\mathbb{E}[R_{n,\mathcal{N}_K}(\mathcal{A}_{MC-UCB})] \leq 24\sqrt{2}\Sigma_{\mathcal{N}_K}\sqrt{(2f_{\max} + 3b + 12f_{\max}^2)}\Big(2f_{\max} + 1\Big)^{4/3}\frac{K^{1/3}}{n^{4/3}}\sqrt{\log(nK)} + \frac{14K\Sigma_{\mathcal{N}_K}^2}{n^2}.$$

The proof, provided in Appendix A, is similar to the one of MC-UCB in [1]. But an improved analysis leads to a better dependency in terms of the number of strata K. We remind that in paper [1], the bound is of order $\tilde{O}(Kn^{-4/3})$. This improvement is crucial here since the larger K is, the closer $\Sigma_{\mathcal{N}_K}$ is from $\int_{[0,1]^d} s(x)dx$. The next Subsection states that the rate $K^{1/3}\tilde{O}(n^{-4/3})$ of MC-UCB is optimal both in terms of K and n.

4.3. Lower Bound

We now study the minimax rate for the pseudo-regret of any algorithm on a given partition $\mathcal{N}_K$. Note that we state it for partitions $\mathcal{N}_K$, but it actually holds for any partition in K strata of equal measure.

Theorem 1. *Let $K \in \mathbb{N}$. Let* inf *be the infimum taken over all online stratified sampling algorithms on $\mathcal{N}_K$ and* sup *represent the supremum taken over all environments, then:*

$$\inf \sup \mathbb{E}[R_{n,\mathcal{N}_K}] \geq C\frac{K^{1/3}}{n^{4/3}},$$

where C is a numerical constant.

Sketch of proof (The full proof is reported in Appendix F). For $2K$ smaller than some constant, we prove a lower bound in $n^{-4/3}$ with a two-armed bandit. For $2K$ larger than this constant, we consider a partition in $2K$ strata. On the K first strata, the samples are drawn from Bernoulli distributions of parameter μ_k where $\mu_k \in \{\frac{\mu}{2}, \mu, 3\frac{\mu}{2}\}$, and on the K last strata, the samples are drawn from a Bernoulli of parameter 1/2. We write $\sigma = \sqrt{\mu(1-\mu)}$ the standard deviation of a Bernoulli of parameter μ. We index by υ a set of 2^K possible environments, where $\upsilon = (\upsilon_1, \ldots, \upsilon_K) \in \{-1, +1\}^K$, and the K first strata are defined by $\mu_k = \mu + \upsilon_k \frac{\mu}{2}$. Write $\mathbb{P}_\sigma$ the probability under such an environment, also consider $\mathbb{P}_\sigma$ the probability under which all the K first strata are Bernoulli with mean μ.

We define Ω_υ the event on which there are less than $\frac{K}{3}$ arms not pulled correctly for environment υ (i.e. for which $T_{k,n}$ is larger than the optimal allocation corresponding to μ when actually $\mu_k = \frac{\mu}{2}$, or smaller than the optimal allocation corresponding to μ when $\mu_k = 3\frac{\mu}{2}$). See Appendix F for a precise definition of these events. Then, the idea is that there are so many such environments that any algorithm will be such that for at least one of them we have $\mathbb{P}_\sigma(\Omega_\upsilon) \leq \exp(-K/72)$. Then we derive by a variant of Pinsker's inequality applied to an event of small probability that $\mathbb{P}_\upsilon(\Omega_\upsilon) \leq \frac{KL(\mathbb{P}_\sigma, \mathbb{P}_\upsilon)}{K} = O(\frac{\sigma^{3/2}n}{K})$. Finally, by choosing σ of order $(\frac{K}{n})^{1/3}$, we have that $\mathbb{P}_\upsilon(\Omega_\upsilon^c)$ is bigger than a constant, and on Ω_υ^c we know that there are more than $\frac{K}{3}$ arms not pulled correctly. This leads to an expected pseudo-regret in environment υ of order $\Omega(\frac{K^{1/3}}{n^{4/3}})$. □

This is the first lower-bound for the problem of online stratified sampling for Monte-Carlo. Note that this bound is of same order as the upper bound for the pseudo-regret of algorithm MC-UCB. It means that this algorithm is, up to a logarithmic factor, minimax optimal, both in terms of the number of samples and in terms of the number of strata. It is proven here on the partitions $\mathcal{N}_K$ only but we conjecture that a similar result holds for *any* measurable partition $\mathcal{N}$ with a bound of order $\Omega\Big(\sum_{x\in\mathcal{N}} \frac{w_x^{2/3}}{n^{4/3}}\Big)$.

5. Best trade-off between $Q_{n,\mathcal{N}_K}$ and $R_{n,\mathcal{N}_K}(\mathcal{A}_{MC-UCB})$

5.1. *Best trade-off*

We consider in this Section the hyper-cubic partitions $\mathcal{N}_K$ defined in Definition 1, and we want to find the optimal number of strata K_n as a function of n. Using the results in Section 3 and Subsection 4.1, it is possible to deduce an optimal number of strata K to assign as parameter to the MC-UCB algorithm. Note that since the performance of the algorithm is defined as the sum of the quality of partition $\mathcal{N}_K$, i.e. $Q_{n,\mathcal{N}_K}$ and of the pseudo-regret of the algorithm MC-UCB, namely $R_{n,\mathcal{N}_K}(\mathcal{A}_{MC-UCB})$, one wish to (i) on the one hand, use many strata so that $Q_{n,\mathcal{N}_K}$ is small but (ii) on the other hand, pay attention to the impact that this number of strata has on the pseudo-regret $R_{n,\mathcal{N}_K}(\mathcal{A}_{MC-UCB})$. A good way to do that is to choose K_n as a function of n such that $Q_{n,\mathcal{N}_{K_n}}$ and $R_{n,\mathcal{N}_{K_n}}(\mathcal{A}_{MC-UCB})$ are of the same order.

Theorem 2. *Under Assumptions 1 and 2 (since on* $[0,1]^d$*, Assumption 1 implies Assumption 3, by setting* $f_{\max} = \max\Big(1, \max(|f(0)|, s(0)) + Md^{\alpha/2}\Big)$*),* $\delta = n^{-2}$*, choosing* $K_n = \Big(\lfloor (n^{\frac{d}{d+3\alpha}})^{1/d} \rfloor\Big)^d (\leq n^{\frac{d}{d+3\alpha}} \leq n)$*, we* *have if* $n \geq \max(4K, b\log(2/\delta), 2^{d+3\alpha})$ *that*

$$\mathbb{E}[L_n(\mathcal{A}_{MC-UCB})] - \inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*)$$
$$\leq 112\sqrt{d}(M+1)\sqrt{(2f_{\max}+3b+12f_{\max}^2)}\Big(2f_{\max}+1\Big)^{4/3} f_{\max} n^{-\frac{d+4\alpha}{d+3\alpha}}\sqrt{\log(n)} + 56 f_{\max}^2 n^{-\frac{d+6\alpha}{d+3\alpha}}.$$

If $d \ll n$*, the simplified bound is*

$$\mathbb{E}[L_n(\mathcal{A}_{MC-UCB})] - \inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*) = \tilde{O}(n^{-\frac{d+4\alpha}{d+3\alpha}}).$$

We remind that $\inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*) = \frac{1}{n}\Big(\int_{[0,1]^d} s(x)dx\Big)^2$.

Sketch of proof of Theorem 2. (The full proof is in Appendix E) The definition of K_n implies that $K_n \geq \Big(n^{\frac{1}{d+3\alpha}} - 1\Big)^d \geq n^{\frac{d}{d+3\alpha}}\Big(1 - \frac{d}{n^{\frac{1}{d}(\frac{d}{d+3\alpha})}}\Big)$. Also, by definition, $K_n \leq n^{\frac{d}{d+3\alpha}}$. By plugging these lower and upper bounds, in respectively $Q_{n,\mathcal{N}_{K_n}}$ and $R_{n,\mathcal{N}_{K_n}}$ (Propositions 2 and 4), we obtain the the final bound. □

We can also prove a matching minimax lower bound using the results in Theorem 1.

Theorem 3. *Let* sup *represent the supremum taken over all* $\alpha-$*Hölder functions* (f,s) *and* inf *be the infimum taken over all algorithms that consider partitions in convex strata of same size (i.e. the* $(\mathcal{N}_K)_K$*), then the following holds true:*

$$\inf\sup \mathbb{E}L_n(\mathcal{A}) - \inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*) = \Omega(n^{-\frac{d+4\alpha}{d+3\alpha}}).$$

Proof of Theorem 3. (The full proof is in Appendix G) This is a consequence of Theorem 1 and of Proposition 3, dividing the domain in parts in order to fit the conditions of these two results. □

5.2. Discussion

Optimal pseudo-risk. The dominant term in the pseudo-risk of MC-UCB with proper number of strata is $\frac{(\inf_{\mathcal{N}} \Sigma_{\mathcal{N}})^2}{n} = \frac{1}{n}\big(\int_{[0,1]^d} s(x)dx\big)^2$ (the other term is negligible). This means that MC-UCB is almost as efficient as the optimal oracle strategy on the optimal oracle partition. In comparison, the variance of the estimate given by crude Monte-Carlo is $\int_{[0,1]^d}\big(f(x) - \int_{[0,1]^d} f(u)du\big)^2 dx + \int_{[0,1]^d} s(x)^2 dx$. Thus MC-UCB enables to have the term coming from the variations in the mean vanish, and the noise term decreases (since by Cauchy-Schwarz, $\big(\int_{[0,1]^d} s(x)dx\big)^2 \leq \int_{[0,1]^d} s(x)^2 dx$).

Minimax-optimal trade-off for algorithm MC-UCB. The optimal trade-off on the number of strata K_n of order $n^{\frac{d}{d+3\alpha}}$ depends on the dimension and the smoothness of the function. The higher the dimension, the more strata are needed in order to have a decent speed of convergence for $\Sigma_{\mathcal{N}_K}$. The smoother the function, the less strata are needed.

It is also important to notice that this trade-off is not perfect and a $\sqrt{\log(n)}$ factor remains between the lower and upper bound.

Link between risk and pseudo-risk. It is important to compare the pseudo-risk $L_n(\mathcal{A}) = \sum_{k=1}^{K} \frac{w_k^2\sigma_k^2}{T_{k,n}}$ and the true risk $\mathbb{E}[(\hat{\mu}_n - \mu)^2]$. These quantities are in general not equal for an algorithm $\mathcal{A}$ that allocates the samples in a dynamic way: indeed, the quantities $T_{k,n}$ are in that case stopping times and the variance of the estimate $\hat{\mu}_n$ is not equal to the pseudo-risk. However, in the work [3], some links between the risk and the pseudo-risk were discussed. Links between $L_n(\mathcal{A})$ and $\sum_{k=1}^{K} w_k^2 \mathbb{E}[(\hat{\mu}_{k,n} - \mu_k)^2]$ were established since $\mathbb{E}[(\hat{\mu}_{k,n} - \mu_k)^2] \leq \frac{w_k^2\sigma_k^2}{\underline{T}_{k,n}^2}\mathbb{E}[T_{k,n}]$, where $\underline{T}_{k,n}$ is a lower-bound on the number of pulls $T_{k,n}$ on a high probability event. A bound on the cross products $\mathbb{E}[(\mu_{k,n} - \mu_k)(\hat{\mu}_{p,n} - \mu_p)]$ was also deduced. A tight analysis of these terms as a function of the number of strata K remains to be investigated.

Knowledge of the Hölder exponent. In order to be able to choose properly the number of strata in order to achieve the rate in Theorem 2, it is needed to know a correct lower bound on the Hölder exponent of the function: indeed, the rougher the function is, the more strata are required. On the other hand, such a knowledge on the function is not always available and an interesting question is whether it is possible to estimate this exponent simultaneously as sampling the function. There are interesting papers on that subject like [10] where the authors tackle the problem of regression and prove that it is possible, up to a certain extent, to adapt to the unknown smoothness of the function. The authors in [6] further prove (in the case of density estimation) that it is even possible, under the assumption that the function attain its Hölder exponent, to have a proper estimation of this exponent and thus adaptive confidence bands. An idea would be to try to adapt those results in the case of finite sample.

MC-UCB On a noiseless function.. Consider the case where $s = 0$ almost surely, i.e. the samples collected are noiseless. Proposition 1 ensures that $\inf_{\mathcal{N}} \Sigma_{\mathcal{N}} = 0$: it is thus possible to achieve a pseudo-risk that has a faster rate than $O(\frac{1}{n})$. If the function f is smooth, e.g. Hölder with a not too low exponent α, it may be efficient to use low discrepancy methods to integrate the functions. An idea would consists in stratifying the domain in n hyper-rectangular strata of minimal diameter, and select at random one sample per stratum. The variance of the resulting estimate is of order $O(\frac{1}{n^{1+2\alpha/d}})$. Algorithm MC-UCB is not as efficient as a low discrepancy scheme: it needs a number of strata $K < n$ in order to be able to estimate the variance of each stratum. Its pseudo-risk is then of the order $O(\frac{1}{nK^{2\alpha/d}})$. It is however only true when the observations are noiseless. Otherwise, the order for the variance of the estimate is in $O(1/n)$, no matter what strategy the learner chooses.

In high dimensions. The first bound in Theorem 2 expresses precisely how the performance of the estimate returned by MC-UCB relies on d. The first bound states that the quantity $L_n(\mathcal{A}) - \frac{1}{n}\Big(\int_{[0,1]^d} s(x)dx\Big)^2$ is negligible compared to $1/n$ when n is exponential in d. This is not surprising since our technique aims at stratifying equally in every direction, and it is not possible to stratify in every directions of the domain if the function lies in a high dimensional domain (i.e. such that $n < \exp(d)$). This is however *not* a reason for not using our algorithm in high dimensions. Indeed, stratifying even in a small number of strata already reduces the variance, and in high dimensions, any variance reduction techniques are welcome. As mentioned in the end of Section 2, the model that we propose for the function is suitable for modeling d^* dimensional functions that we only stratify in $d < d^*$ directions (and $\exp(d) < n$). A reasonable trade-off for d can also be inferred from the bound, but we believe that a good choice of d heavily depends on the specific problem. We then believe that it is a good idea to select the number of strata in the minimax way that we propose. Again, having a very high dimensional function that one stratifies in only a few directions is a very common technique in financial mathematics, for pricing options (practitioners stratify an infinite dimensional process in only 1 to 5 carefully chosen dimensions). We illustrate this idea in the next Section.

6. Numerical experiment: influence of the number of strata for the Pricing of an Asian option

We consider the pricing problem of an Asian option introduced in [8] and later considered in [11, 4]. This uses a Black-Scholes model with strike C and maturity T. Let $(W(t))_{0 \le t \le T}$ be a Brownian motion. The discounted payoff of the Asian option is defined as a function of W, by:

$$F((W)_{0 \le t \le T}) = \exp(-rT) \max\Big[\int_0^T S_0 \exp\Big((r - \tfrac{1}{2}s_0^2)t + s_0 W_t\Big) dt - C, 0\Big], \qquad (11)$$

where S_0, r, and s_0 are constants, and the price is defined by the expectation $p = \mathbb{E}_W F(W)$.

We want to estimate the price p by Monte-Carlo simulations (by sampling on W). In order to reduce the variance of the estimated price, we can stratify the space of W. In [8] it is suggested to stratify according to a one dimensional projection of W, i.e., by choosing a time t and stratifying according to the quantiles of W_t (and simulating the rest of the Brownian according to a Brownian Bridge, see [11]). They further argue that the best direction for stratification is to choose $t = T$, i.e., to stratify according to the last time of T. This choice of stratification is also intuitive since W_T has the highest variance, the biggest exponent in the payoff (11), and thus the highest volatility. [11] and [4] also use the same direction of stratification. We stratify according to the quantiles of W_T, i.e. the quantiles of a normal distribution $\mathcal{N}(0, T)$. When stratifying in K strata, we stratify according to the $1/K$-th quantiles (so that the strata are hyper-cubes of same measure).

We choose the same numerical values as in [11]: $S_0 = 100$, $r = 0.05$, $s_0 = 0.30$, $T = 1$ and $d = 16$. Like in [11], we discretise the Brownian motion in 16 equidistant times, so that we are able to simulate it. We choose $C = 120$.

In this paper, we only do experiments for MC-UCB, and exhibit the influence of the number of strata. For a comparison between MC-UCB and other algorithms, see [1]. By studying the range of the $F(W)$, we set the parameter of the algorithm MC-UCB to $A = 150\log(n)$.

For $n = 200$ and $n = 2000$, we observe the influence of the number of strata in Figure 2. We observe the trade-off that we mentioned between pseudo-regret and quality, in the sense that the mean squared error of the estimate returned by MC-UCB (when compared to the true integral of f) first decreases and then increases with K. We also illustrate that for a large n the minimum of the mean squared error is reached with a larger number of strata K. The numerical results corroborate (surprisingly well) the theoretical findings. Finally, note that our technique is never outperformed by a uniform stratified Monte-Carlo: it is thus always a good strategy to use MC-UCB on a relevant number of strata.

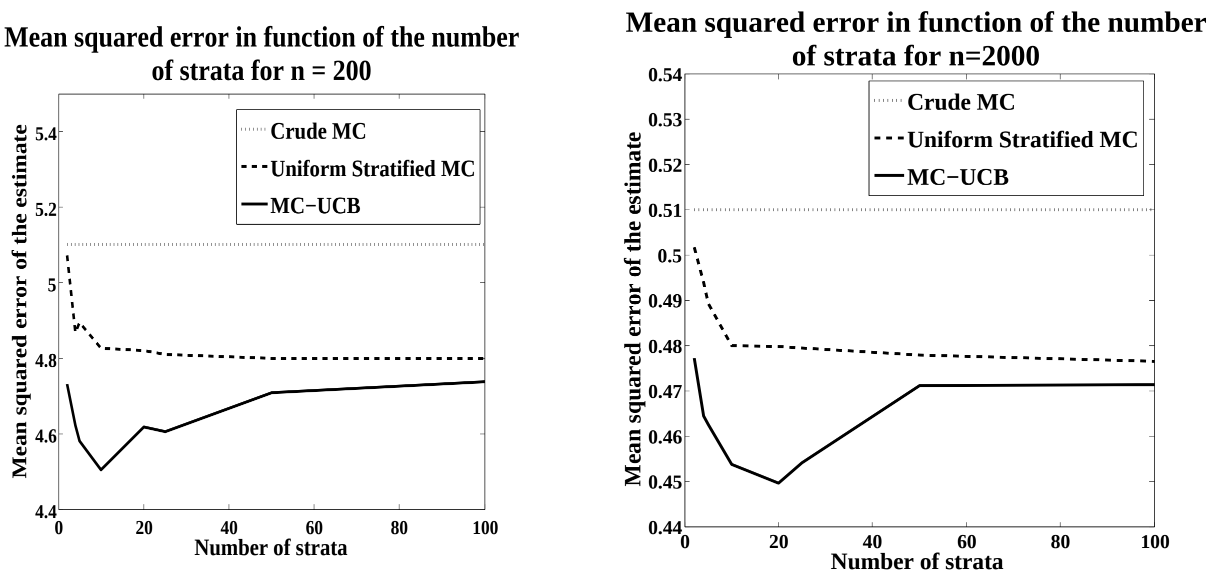

Figure 2: Mean squared error for uniform stratified sampling for different number of strata, for (Left:) n=200 and (Right:) n=2000.

7. Conclusion

In this paper we studied the problem of online stratified sampling for the numerical integration of a function given noisy evaluations, and more precisely we discussed the problem of choosing the *minimax-optimal number of strata*.

We explained why, to our minds, this is a crucial problem when one wants to design an efficient algorithm. We enlightened the fact that there is a trade-off between a large number of strata (in order to have a low approximation error, called the quality of a partition) but not too many, in order to perform almost as well as the optimal oracle allocation on a given partition (small estimation error, called pseudo-regret).

When there is noise to the function, the noise is the dominant quantity in the optimal oracle variance on the optimal oracle partition. Indeed, decreasing the size of the strata does not diminish the (local) variance of the noise. In this case, the pseudo-risk of algorithm MC-UCB is equal, up to negligible terms, to the mean squared error of the estimate outputted by the optimal oracle strategy on the best (oracle) partition, at a rate of $O(n^{-\frac{d+4\alpha}{d+3\alpha}})$ where α is the Hölder exponent of s and f. This rate is minimax optimal on the class of α-Hölder functions: it is not possible, up to a logarithmic factor, to do better on simultaneously all α-Hölder functions.

We believe that there are (at least) three very interesting remaining open questions:

- The first one is to investigate whether it is possible to estimate online the Hölder exponent *fast enough* when this exponent is initially unknown, and perform almost as well as if this exponent where known (and used to compute the best number of strata for MC-UCB).

- The second direction is to build a more efficient algorithm in the noiseless case. We noticed that in this case, MC-UCB is not as efficient as a simple non-adaptive method. The problem comes from the fact that in the case of a noiseless function, it is important to sample the space in a way that guarantees that the points are as spread out as possible. An interesting problem is thus to build an algorithm that mixes ideas from quasi Monte-Carlo and ideas from online stratified Monte-Carlo.

- Another question is the relevance of fixing the strata in advance. Although it is minimax-optimal on the class of $\alpha-$Hölder functions to have hyper-cubic strata of same measure, it might in some cases be more interesting to focus and stratify more finely at places where the function is rough. On that perspective, it could be more efficient to design an adaptive procedure that would also decides *where* to refine the stratification.

Acknowledgements

This research was partially funded by Region Nord-Pas-de-Calais Regional Council and the European Community Seventh Framework Programme (FP7/2007-2013) under grant agreement 270327 (project CompLACS).

References

[1] A. Carpentier and R. Munos. Finite-time analysis of stratified sampling for monte carlo. In *In Neural Information Processing Systems (NIPS)*, 2011.

[2] A. Carpentier and R. Munos. Minimax number of strata for online stratified sampling given noisy samples. In *Algorithmic Learning Theory*, pages 229–244. Springer, 2012.

[3] A. Carpentier, R. Munos, and A. Antos. Minimax strategy for stratified sampling for monte carlo. Technical report, INRIA-00636924, 2011.

[4] P. Etoré and P. Jourdain. Adaptive optimal allocation in stratified sampling methods. *Methodol. Comput. Appl. Probab.*, 12(3):335–360, September 2010.

[5] P. Etoré, G. Fort, B. Jourdain, and E. Moulines. On adaptive stratification. *Annals of Operation Research*, 189:127–154, 2011.

[6] E. Giné and R. Nickl. Confidence bands in density estimation. *The Annals of Statistics*, 38(2):1122–1170, 2010.

[7] P. Glasserman. *Monte Carlo methods in financial engineering*. Springer Verlag, 2004. ISBN 0387004513.

[8] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance*, 9(2):117–152, 1999.

[9] V. Grover. Active learning and its application to heteroscedastic problems. *Department of Computing Science, Univ. of Alberta, MSc thesis*, 2009.

[10] M. Hoffmann and O. Lepski. Random rates in anisotropic regression. *Annals of statistics*, pages 325–358, 2002.

[11] R. Kawai. Asymptotically optimal allocation of stratified sampling with adaptive variance reduction by strata. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 20(2):1–17, 2010. ISSN 1049-3301.

[12] R.Y. Rubinstein and D.P. Kroese. *Simulation and the Monte Carlo method*. Wiley-interscience, 2008. ISBN 0470177942.

Appendix A. Proof of Proposition 4

Let $\mathcal{N}_K$ be the partition of the space that we consider. At each time t, the algorithm can choose between the K strata (arms) and observes a sample $X(t) = X_{k,t}$ if it chooses stratum K. We write ν_k for the distribution of stratum (arm) K, i.e. the distribution of the sample conditional to the fact that the state x_t is chosen in Ω_k. Then the $X_{k,t} \sim \nu_k$ and are independent.

Appendix A.1. Main tool: a high probability bound on the standard deviations

Upper bound on the standard deviation:

Lemma 1. *Let Assumption 2 and 3 hold and $n \geq b\log(2/\delta)$. Define the following event*

$$\xi = \xi_{K,n}(\delta) = \bigcap_{1\leq k\leq K,\ 2\leq t\leq n} \left\{ \left| \sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\Big(X_{k,i} - \frac{1}{t}\sum_{j=1}^{t} X_{k,j}\Big)^2} - \sigma_k \right| \leq A\sqrt{\frac{1}{t}} \right\}, \tag{A.1}$$

where $A = 2(2f_{\max}+1)\sqrt{(2f_{\max}+3b+12f_{\max}^2)\log(6nK/\delta)}$. Then $\Pr(\xi) \geq 1-\delta$.

Note that the first term in the absolute value in Equation A.1 is the empirical standard deviation of arm k computed like in Equation 9 with t samples. The event ξ plays an important role in the proofs of this section and a number of statements will be proved on this event.

Proof. Under Assumption 2 and 3, all assumptions of Lemma 4 are verified and with probability $1-\delta$ we have

$$\left| \sqrt{\frac{1}{t-1}\sum_{i=1}^{t}\Big(X_{k,i} - \frac{1}{t}\sum_{j=1}^{t} X_{k,j}\Big)^2} - \sigma_k \right| \leq 2(2f_{\max}+1)\sqrt{\frac{(2f_{\max}+3b+12f_{\max}^2)\log(6/\delta)}{t}}, \tag{A.2}$$

since $f_{\max} \geq \max_i(\max(|\mathbb{E}X_{k,i}|, \sqrt{\mathbb{V}X_{k,i}})) = \max_i(\max(\mu_i, \sigma_i))$.

Then by doing a simple union bound on (k,t), we obtain the result.

□

We deduce the following corollary when the number of samples $T_{k,t}$ are random.

Corollary 1. *For any $k = 1,\dots,K$ and $t = 2K,\dots,n$, let $\{X_{k,i}\}_i$ be n i.i.d. random variables drawn from ν_k, satisfying Assumption 2. Let $T_{k,t}$ be any random variable taking values in $\{2,\dots,n\}$. Let $\hat{\sigma}_{k,t}^2$ be the empirical variance computed from Equation 9. Then, on the event ξ, we have:*

$$|\hat{\sigma}_{k,t} - \sigma_k| \leq A\sqrt{\frac{1}{T_{k,t}}}\ , \tag{A.3}$$

where $A = 2(2f_{\max}+1)\sqrt{(2f_{\max}+3b+12f_{\max}^2)\log(6nK/\delta)}$.

Appendix A.2. Main Proof

Proof of Proposition 4. **Step 1. Lower bound of order $\widetilde{O}(n^{2/3})$.** Let k be the index of an arm such that $T_{k,n} \geq \frac{n}{K}$ (this implies $T_{k,n} \geq 4$ as $n \geq 4K$, and arm k is thus pulled after the initialization) and let $t+1 \leq n$ be the last time at which it was pulled [6], i.e., $T_{k,t} = T_{k,n} - 1$ and $T_{k,t+1} = T_{k,n}$. From Equation A.3 and the fact that $T_{k,n} \geq \frac{n}{K}$, we obtain on ξ

$$B_{k,t+1} \leq \frac{w_k}{T_{k,t}} \left(\sigma_k + 2A\sqrt{\frac{1}{T_{k,t}}} \right) \leq \frac{Kw_k\Big(\sigma_k + 2A\Big)}{n}, \tag{A.4}$$

where the second inequality follows from the facts that $T_{k,t} \geq 1$, $w_k\sigma_k \leq \Sigma_{\mathcal{N}_K}$, and $w_k \leq \sum_k w_k = 1$. Since at time $t+1$ the arm k has been pulled, then for any arm q, we have

$$B_{q,t+1} \leq B_{k,t+1}. \tag{A.5}$$

From the definition of $B_{q,t+1}$, and also using the fact that $T_{q,t} \leq T_{q,n}$, we deduce on ξ that

$$B_{q,t} \geq \frac{2Aw_q}{T_{q,t}^{3/2}} \geq \frac{2Aw_q}{T_{q,n}^{3/2}} \ . \tag{A.6}$$

Combining Equations A.4–A.6, we obtain on ξ

$$\frac{2Aw_q}{T_{q,n}^{3/2}} \leq \frac{Kw_k\Big(\sigma_k + 2A\Big)}{n}.$$

Finally, this implies on ξ that for any q (since $w_k = w_q$),

$$T_{q,n} \geq \Big(\frac{2A}{\sigma_k + 2A}\frac{n}{K}\Big)^{2/3}. \tag{A.7}$$

Thus $\forall q, T_{q,n} \geq C\Big(\frac{n}{K}\Big)^{2/3}$ where $C = \Big(\frac{2A}{\max_k \sigma_k + 2A}\Big)^{2/3}$.

Step 2. Properties of the algorithm. We first remind the definition of $B_{q,t+1}$ used in the MC-UCB algorithm

$$B_{q,t+1} = \frac{w_q}{T_{q,t}} \left(\hat{\sigma}_{q,t} + A\sqrt{\frac{1}{T_{q,t}}} \right).$$

Using Corollary 1 it follows that, on ξ

$$\frac{w_q\sigma_q}{T_{q,t}} \leq B_{q,t+1} \leq \frac{w_q}{T_{q,t}} \left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,t}}} \right). \tag{A.8}$$

Let $t+1 \geq 2K+1$ be the time at which an arm q is pulled for the last time, thus $T_{q,t} = T_{q,n} - 1$. Note that there is at least one arm such that this happens since $n \geq 4K$. Since at $t+1$ arm q is chosen, then for any other arm p, we have

$$B_{p,t+1} \leq B_{q,t+1} \ . \tag{A.9}$$

[6] Note that such an arm always exists for any possible allocation strategy given the constraint $n = \sum_q T_{q,n}$.

From Equation A.8 and $T_{q,t} = T_{q,n} - 1$, we obtain on ξ

$$B_{q,t+1} \leq \frac{w_q}{T_{q,t}}\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,t}}}\right) = \frac{w_q}{T_{q,n}-1}\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,n}-1}}\right). \quad \text{(A.10)}$$

Furthermore, since $T_{p,t} \leq T_{p,n}$, then on ξ

$$B_{p,t+1} \geq \frac{w_p \sigma_p}{T_{p,t}} \geq \frac{w_p \sigma_p}{T_{p,n}}. \quad \text{(A.11)}$$

Combining Equations A.9–A.11, we obtain on ξ

$$\frac{w_p \sigma_p}{T_{p,n}}(T_{q,n} - 1) \leq w_q\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,n}-1}}\right).$$

Summing over all q such that the previous Equation is verified, i.e. such that $T_{q,n} \geq 3$, on both sides, we obtain on ξ

$$\frac{w_p \sigma_p}{T_{p,n}} \sum_{q|T_{q,n}\geq 3}(T_{q,n} - 1) \leq \sum_{q|T_{q,n}\geq 3} w_q\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,n}-1}}\right).$$

This implies (since $\sum_{q|T_{q,n}\geq 3}(T_{q,n} - 1) = n - \sum_{q|T_{q,n}<3} 2 - \sum_{q|T_{q,n}\geq 3} 1 \geq n - 2K$)

$$\frac{w_p \sigma_p}{T_{p,n}}(n - 2K) \leq \sum_{q=1}^{K} w_q\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,n}-1}}\right). \quad \text{(A.12)}$$

Step 3. Lower bound. Plugging Equation A.7 in Equation A.12,

$$\begin{aligned}\frac{w_p \sigma_p}{T_{p,n}}(n - 2K) &\leq \sum_q w_q\left(\sigma_q + 2A\sqrt{\frac{1}{T_{q,n}-1}}\right)\\ &\leq \sum_q w_q\left(\sigma_q + 2A\sqrt{\frac{2K^{2/3}}{Cn^{2/3}}}\right)\\ &\leq \Sigma_{\mathcal{N}_K} + \frac{2\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{1/3}},\end{aligned}$$

on ξ, since $T_{q,n} - 1 \geq \frac{T_{q,n}}{2}$ (as $T_{q,n} \geq 2$). Finally as $n \geq 4K$, and since $\frac{1}{1-x} \geq 1 + 2x$, we obtain on ξ the following bound

$$\begin{aligned}\frac{w_p \sigma_p}{T_{p,n}} &\leq \frac{\Sigma_{\mathcal{N}_K}}{n - 2K} + \frac{2\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{(n-2K)n^{1/3}}\\ &\leq \frac{\Sigma_{\mathcal{N}_K}}{n}\frac{1}{1 - 2K/n} + \frac{4\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{4/3}}\\ &\leq \frac{\Sigma_{\mathcal{N}_K}}{n}(1 + \frac{4K}{n}) + \frac{4\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{4/3}}\\ &\leq \frac{\Sigma_{\mathcal{N}_K}}{n} + \frac{4\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{4/3}} + \frac{4K\Sigma_{\mathcal{N}_K}}{n^2}.\end{aligned} \quad \text{(A.13)}$$

Step 4. Regret. By summing and using Equation (A.13) which holds for all p, we obtain on ξ (with probability $1-\delta$)

$$L_{n,\mathcal{N}_K} = \sum_p \frac{w_p^2 \sigma_p^2}{T_{p,n}} \leq \frac{\Sigma_{\mathcal{N}_K}^2}{n} + \frac{4\Sigma_{\mathcal{N}_K}\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{4/3}} + \frac{4K\Sigma_{\mathcal{N}_K}^2}{n^2}.$$

This implies since $\mathbb{E}L_n = \mathbb{E}[L_n \mathbb{I}\{\xi\}] + \mathbb{E}[L_n \mathbb{I}\{\xi^c\}]$ and since $\delta = n^{-2}$

$$\begin{aligned}
\mathbb{E}L_{n,\mathcal{N}_K} &\leq \frac{\Sigma_{\mathcal{N}_K}^2}{n} + \frac{4\Sigma_{\mathcal{N}_K}\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{4/3}} + \frac{4K\Sigma_{\mathcal{N}_K}^2}{n^2} + (\sum_p w_p^2\sigma_p^2)n^{-2} \\
&\leq \frac{\Sigma_{\mathcal{N}_K}^2}{n} + \frac{4\Sigma_{\mathcal{N}_K}\sqrt{2}A}{\sqrt{C}}\frac{K^{1/3}}{n^{4/3}} + \frac{5K\Sigma_{\mathcal{N}_K}^2}{n^2},
\end{aligned}$$

since $\sum_p w_p^2\sigma_p^2 \leq \Sigma_{\mathcal{N}_K}^2$.

Since $\delta = n^{-2}$ and $n \geq 4K \geq 8$, we have

$$A \leq 6(2f_{\max}+1)\sqrt{(2f_{\max}+3b+12f_{\max}^2)\log(nK)}.$$

Also, we have $A \geq 2\sqrt{\log(2^{11})} \geq 4$, which implies (since C is an increasing function of A)

$$C \geq \Big(\frac{2A}{f_{\max}+2A}\Big)^{2/3} \geq \Big(\frac{8}{f_{\max}+8}\Big)^{2/3}.$$

These last equations lead to

$$\mathbb{E}L_{n,\mathcal{N}_K} \leq \frac{\Sigma_{\mathcal{N}_K}^2}{n} + 24\sqrt{2}\Sigma_{\mathcal{N}_K}\sqrt{(2f_{\max}+3b+12f_{\max}^2)}\Big(2f_{\max}+1\Big)^{4/3}\frac{K^{1/3}}{n^{4/3}}\sqrt{\log(nK)} + \frac{14K\Sigma_{\mathcal{N}_K}^2}{n^2}.$$

□

Appendix B. Proof of Proposition 1

Step 1: Expression of the variance of the stratified estimate. Note that the samples $f(x)+s(x)\epsilon_t$ are such that $\epsilon_t \sim \nu_x$ and $\mathbb{E}_{\nu_x}[\epsilon_t]=0$, $\mathbb{V}_{\nu_x}[\epsilon_t]=1$, and the ϵ_t are independent. We have

$$\begin{aligned}
\sigma_k^2 &= \frac{1}{w_k}\int_{\Omega_k}\mathbb{E}_{\nu_x}[(X_x(t)-\mu_k)^2]dx \\
&= \frac{1}{w_k}\int_{\Omega_k}\mathbb{E}_{\nu_x}\Big[(f(x)+s(x)\epsilon_t-\frac{1}{w_k}\int_{\Omega_k}f(u)du)^2\Big]dx \\
&= \frac{1}{w_k}\int_{\Omega_k}\mathbb{E}_{\nu_x}\Big[(f(x)-\frac{1}{w_k}\int_{\Omega_k}f(u)du)^2\Big]dx + \frac{1}{w_k}\int_{\Omega_k}\mathbb{E}_{\nu_x}\Big[s(x)^2\epsilon_t^2\Big]dx \\
&= \frac{1}{w_k}\int_{\Omega_k}\big(f(x)-\frac{1}{w_k}\int_{\Omega_k}f(u)du\big)^2dx + \frac{1}{w_k}\int_{\Omega_k}s(x)^2dx
\end{aligned}$$

Step 2: Proof for the uniformly continuous functions. We first prove the result for a subset of $L_2([0,1]^d)$, namely the set of functions f and s that are uniformly continuous.

Proposition 5. *If the functions f and s are uniformly continuous and if the strata satisfy the Assumptions of Proposition 1, we have*

$$\sum_k w_{k,p}\sigma_{k,p} - \int_{[0,1]^d} s(x)dx \to_p 0.$$

Proof. Let $\upsilon > 0$. As s and f are uniformly continuous, we know that $\exists \eta$ such that $\forall x$, $\forall u \in \mathcal{B}_{2,d}(\eta)$ (where $B_{2,d}(\eta)$ is the ball of center 0 and radius η according to the $||.||_2$ norm) we have $|s(x+u) - s(x)| \leq \upsilon$ and $|f(x+u) - f(x)| \leq \upsilon$.

By Assumption AS1, we know that $w_{k,p} \leq \upsilon_p$. Note that the diameter of strata $\Omega_{k,p}$ is such that $D(w_{k,p}) \leq D(\upsilon_p)$. Let us choose p big enough, i.e. such that $D(\upsilon_p) \leq \eta$ and $\upsilon_p \leq \upsilon$.
We have

$$\begin{aligned}
\sigma_{k,p}^2 - (\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s)^2 &= \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s^2 - \Big(\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s\Big)^2 + \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} \Big(f - \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} f\Big)^2 \\
&= \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} \Big(s - \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s\Big)^2 + \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} \Big(f - \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} f\Big)^2 \\
&\leq \upsilon^2 + \upsilon^2 \leq 2\upsilon^2.
\end{aligned}$$

By concavity of the square-root function, we have

$$\sigma_{k,p} - (\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s) \leq \sqrt{2}\upsilon,$$

and by summing up we get

$$\sum_k w_{k,p}\sigma_{k,p} - \int_{[0,1]^d} s \leq \sqrt{2}\upsilon.$$

□

Step 3: Density of uniformly continuous functions in L_2. We first remind a property of the functions in $L_2([0,1]^d)$.

Proposition 6. *The uniformly continuous functions according to the $||.||_2$ norm are dense in $L_2([0,1]^d)$.*

Proof. The result follows directly from the facts that

- The continuous functions are dense in $L_2(\Omega)$ (Stone-Weierstrass Theorem).
- The uniformly continuous functions on a compact space Ω according to the $||.||_2$ norm are dense in the space of continuous functions.
- $[0,1]^d$ is a compact.

□

This means that we can approximate with arbitrary precision according to the $||.||_2$ measure on $L_2([0,1]^d)$ any function in $L_2([0,1]^d)$ by an uniformly continuous function. Using this proposition, we can prove the following Lemma.

Lemma 2. *For a given p and a given υ, there exist two uniformly continuous function m_υ and s_υ such that:*

$$\Big|\sum_{k=1}^{K_p} w_{k,p}\sigma_{k,p} - \sum_{k=1}^{K_p}\sqrt{w_{k,p}}\sqrt{\int_{\Omega_{k,p}}\Big(f_\upsilon(x) - \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} f_\upsilon(u)du\Big)^2 dx - \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s_\upsilon^2(x)dx}\Big| \le \upsilon.$$

Proof. Let us fix p and $\upsilon \le 1$.

Let f_υ be an uniformly continuous function such that

$$\int_\Omega (f(x) - f_\upsilon(x))^2 dx \le \min_k (w_{k,p})\frac{\upsilon^2}{16},$$

and s_υ be an uniformly continuous function such that

$$\int_\Omega (s(x) - s_\upsilon(x))^2 dx \le \min_k (w_{k,p})\frac{\upsilon^2}{16}.$$

A choice of such functions is possible from Proposition 6 and since $w_{k,p} > 0$. Note that we thus have

$$\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} (f(x) - f_\upsilon(x))^2 dx \le \frac{\upsilon^2}{16}, \tag{B.1}$$

and

$$\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} (s(x) - s_\upsilon(x))^2 dx \le \frac{\upsilon^2}{16}. \tag{B.2}$$

Since for a functional L_2 norm, $\|f-g\|_2 \ge |\|f\|_2 - \|g\|_2|$,

$$\sqrt{\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} (s(x) - s_\upsilon(x))^2 dx} \ge \Big|\sqrt{\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s(x)^2 dx} - \sqrt{\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s_\upsilon(x)^2 dx}\Big|,$$

by combining this with Equation (B.2), we get

$$\Big(\sqrt{\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s(x)^2 dx} - \frac{\upsilon}{4}\Big)^2 \le \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s_\upsilon(x)^2 dx \le \Big(\sqrt{\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s(x)^2 dx} + \frac{\upsilon}{4}\Big)^2,$$

which implies

$$\Big|\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s(x)^2 dx - \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s_\upsilon(x)^2 dx\Big| \le \sqrt{\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} s(x)^2 dx}\frac{\upsilon}{2} + \frac{\upsilon^2}{16}. \tag{B.3}$$

A bias variance decomposition implies

$$\begin{aligned}
&\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} (f(x) - f_\upsilon(x))^2 dx \\
&= \frac{1}{w_{k,p}}\int_{\Omega_{k,p}}\Big(f(x) - f_\upsilon(x) - \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} f(x)dx - \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} f_\upsilon(x)dx\Big)^2 dx \\
&+ \big(\frac{1}{w_{k,p}}\int_{\Omega_{k,p}} f(x)dx - \frac{1}{w_{k,p}}\int_{\Omega_{k,p}} f_\upsilon(x)dx\big)^2,
\end{aligned}$$

and together with Equation (B.1),

$$\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}\Big(f(x)-f_\upsilon(x)-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}f(x)dx-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}f_\upsilon(x)dx\Big)^2dx\leq\frac{\upsilon^2}{16}.$$

Applying a reasoning similar to the estimation of the integral in s, we have

$$\Big|\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}\Big(f(x)-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}f(x)dx\Big)^2dx-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}\Big(f_\upsilon(x)-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}f_\upsilon(x)dx\Big)^2dx\Big|$$
$$\leq\sqrt{\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}\Big(f(x)-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}f(x)dx\Big)^2dx}\frac{\upsilon}{2}+\frac{\upsilon^2}{16}.$$

These both steps (last equation and Equation (B.3)) imply (since $\sigma_{k,n}^2=\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}(f(x)-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}f(u)du)^2dx+\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}s(x)^2dx$)

$$\Big|\sigma_{k,p}^2-\Big(\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}\Big(f_\upsilon(x)-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}f_\upsilon(u)du\Big)^2dx+\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}s_\upsilon^2(x)dx\Big)\Big|\leq\sqrt{2}\sigma_{k,p}\upsilon+\frac{\upsilon^2}{8}.$$

By concavity of the square-root function, we have

$$\Big|\sigma_{k,p}-\sqrt{\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}\Big(f_\upsilon(x)-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}f_\upsilon(u)du\Big)^2dx+\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}s_\upsilon^2(x)dx}\Big|\leq\sqrt{\sqrt{2}\sigma_{k,p}\upsilon+\frac{\upsilon^2}{8}}.$$

And finally, by summing up

$$\Big|\sum_{k=1}^{K_p}w_{k,p}\sigma_{k,p}-\sum_{k=1}^{K_p}\sqrt{w_{k,p}}\sqrt{\int_{\Omega_{k,p}}\Big(f_\upsilon(x)-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}f_\upsilon(u)du\Big)^2dx+\int_{\Omega_{k,p}}s_\upsilon^2(x)dx}\Big|\leq 2\sqrt{\upsilon f_{\max}},$$

since $\upsilon\leq 1\leq f_{\max}$, $\sigma_{k,p}\leq f_{\max}$, and $\sqrt{\sqrt{2}+1/8}\leq 2$. This concludes the proof. □

Step 4: Combination of all the preliminary results to finish the proof. Finally, we finish the proof of Proposition 1.

Let $\upsilon>0$ and f_υ and s_υ be as in Lemma 2. We know that

$$\Big|\sum_{k=1}^{K_p}w_{k,p}\sigma_{k,p}-\sum_{k=1}^{K_p}\sqrt{w_{k,p}}\sqrt{\int_{\Omega_{k,p}}\Big(f_\upsilon(x)-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}f_\upsilon(u)du\Big)^2dx+\int_{\Omega_{k,p}}s_\upsilon^2(x)dx}\Big|\leq\upsilon,$$

and also that

$$\int_{[0,1]^d}(s(x)-s_\upsilon(x))^2dx\leq\frac{\upsilon}{2}.$$

Note that by Cauchy-Schwartz:

$$\int_{[0,1]^d}|s(x)-s_\upsilon(x)|dx\leq\sqrt{\int_{[0,1]^d}(s(x)-s_\upsilon(x))^2dx}\leq\sqrt{\frac{\upsilon}{2}}.$$

Note also that Proposition 5 tells us that for p large enough

$$\sum_{k=1}^{K_p}\sqrt{w_{k,p}}\sqrt{\int_{\Omega_{k,p}}\Big(f_\upsilon(x)-\frac{1}{w_{k,p}}\int_{\Omega_{k,p}}f_\upsilon(u)du\Big)^2dx+\int_{\Omega_{k,p}}s_\upsilon^2(x)dx}-\int_{[0,1]^d}s_\upsilon(x)dx\leq\upsilon.$$

When combining all those results, one gets the desired result.

Note finally that if we choose the strata as being small boxes of size $\frac{1}{K}$ and side $(\frac{1}{K})^{1/d}$, then the assumptions of Proposition 1 is verified.

Appendix C. Proof of Proposition 2

Note first that

$$\sigma_k^2 = \frac{1}{w_k}\int_{\Omega_k}\Big(f(x) - \frac{1}{w_k}\int_{\Omega_k} f(u)du\Big)^2 dx + \frac{1}{w_k}\int_{\Omega_k} s^2(x)dx.$$

The term in f. Since the function f is $(\alpha, M)-$ Hölder, we know that $\forall (x,y) \in \Omega, |f(x)-f(y)| \leq M||x-y||_2^{\alpha}$. Thus

$$\frac{1}{w_k}\int_{\Omega_k}\Big(f(x) - \frac{1}{w_k}\int_{\Omega_k} f(u)du\Big)^2 dx \leq M^2 D(\Omega_k)^{2\alpha}$$
$$\leq M^2 d(\frac{1}{K})^{2\alpha/d}.$$

The term in s. Since the function s is $(\alpha, M)-$ Hölder, we know that $\forall (x,y) \in \Omega, |s(x)-s(y)| \leq M||x-y||_2^{\alpha}$.

$$\frac{1}{w_k}\int_{\Omega_k} s^2(x)dx - (\frac{1}{w_k}\int_{\Omega_k} s(u)du\big)^2 = \frac{1}{w_k}\int_{\Omega_k}\big(s(x) - \frac{1}{w_k}\int_{\Omega_k} s(u)du\big)^2 dx \leq M^2 D(\Omega_k)^{2\alpha}$$
$$\leq M^2 d(\frac{1}{K})^{2\alpha/d}.$$

Finally... By combining those two results

$$w_k\sigma_k - \int_{\Omega_k} s(x)dx \leq w_k\sqrt{\sigma_k^2 - (\frac{1}{w_k}\int_{\Omega_k} s(x)dx\big)^2}$$
$$\leq w_k\sqrt{M^2 d(\frac{1}{K})^{2\alpha/d} + M^2 d(\frac{1}{K})^{2\alpha/d}}.$$

By summing over all the strata, one obtains

$$\Sigma_{\mathcal{N}_K} - \int_{[0,1]^d} s(x)dx \leq \sqrt{2d}M(\frac{1}{K})^{\alpha/d}.$$

Appendix D. Proof of Proposition 3

Let us consider a stratum $\Omega_k \in \mathcal{N}_K$ (which is an hypercube of side length $K^{-1/d}$). Let a_k be the center of this stratum. Let us define a function on $[0,1]^d$ such that

$$g_k(x) = \|x - a_k\|_2^{\alpha}(K^{-1/d}/4 - \|x - a_k\|_2)^{\alpha}\mathbf{1}\{\|x - a_k\|_2 \leq K^{-1/d}/4\}.$$

Consider the sphere $\mathcal{S}_k$ defined by the points at a distance $K^{-1/d}/4$ of a_k. Consider the function h_k defined on the ring defined as the difference between the ball of center a_k and diameter $K^{-1/d}/2$ and the ball of center a_k and diameter $K^{-1/d}/4$. For any point x in this ring, define u as the point of intersection between $\mathcal{S}_k$ and the segment $[x, a_k]$. Set $h_k(x) = -g_k(x - (x - u))$. In other words it is the symmetric through u of g_k restricted to the segment $[a_k, u - a_k]$. Let c be a constant, and consider the function

$$f_k^c(x) = g_k + ch_k.$$

Consider the c such that $\int_{[0,1]^d} f_k^c = \int_{\Omega_k} f_k^c = 0$ (since by definition, g_k, h_k are non zero only in Ω_k). By definition of h_k, g_k, we know that $c \in [-1, 0]$. We now refer to f_k as f_k^c with this specific c.

Le us define now for any $x \in [0,1]^d$

$$s(x) = 0 \quad \text{and} \quad f(x) = \sum_{k \le K} f_k.$$

These two functions are $(1, \alpha)$ Hölder (since each f_k is $(1, \alpha)$ Hölder and is non-zero only on stratum Ω_k). For such functions (s, f) the best partition on K convex strata is $\mathcal{N}_K$: indeed, there are K circular "bumps" of exactly the same shape and which are circular, so grouping some of them and separating some others is not going to provide a better partition in K strata than this one.

Now by definition we have

$$\begin{aligned} \sigma_{k,p}^2 &= \frac{1}{w_k} \int_{\Omega_k} \Big(f(x) - \frac{1}{w_k} \int_{\Omega_k} f(u) du \Big)^2 dx + \frac{1}{w_k} \int_{\Omega_k} s^2(x) dx \\ &= \frac{1}{w_k} \int_{\Omega_k} \Big(f_k(x) - \frac{1}{w_k} \int_{\Omega_k} f_k(u) du \Big)^2 dx \\ &= \frac{1}{w_k} \int_{\Omega_k} f_k(x)^2 dx. \end{aligned} \tag{D.1}$$

Now since the function f_k is increasing in any direction until the border of the sphere centered in a_k and with radius $K^{-1/d}/8$, we have

$$\begin{aligned} \frac{1}{w_k} \int_{\Omega_k} f_k(x)^2 dx &\ge \frac{1}{w_k} \int_{x: \|x - a_k\|_2 \le K^{-1/d}/8} f_k(x)^2 dx \\ &\ge \frac{1}{w_k} \int_{x: K^{-1/d}/16 \le \|x - a_k\|_2 \le K^{-1/d}/8} \big(K^{-\alpha/d}/32 \big)^2 dx \\ &\ge \frac{1}{w_k} \frac{\pi^{d/2} \big(K^{-1/d}/8\big)^d}{\Gamma(d/2+1)} \big(K^{-\alpha/d}/32 \big)^2 \big(1 - 1/2^d\big) \\ &\ge \frac{1}{w_k} \frac{\pi^{d/2}}{2 \times 32^2 8^d \Gamma(d/2+1)} K^{-1} K^{-2\alpha/d} \\ &= \frac{\pi^{d/2}}{2 \times 32^2 8^d \Gamma(d/2+1)} K^{-2\alpha/d}, \end{aligned}$$

where Γ is the classic Gamma function.

This implies, together with Equation (D.1), that

$$\sigma_{k,p} \ge c(d) K^{-\alpha/d},$$

where $c(d) = \sqrt{\frac{\pi^{d/2}}{2 \times 32^2 8^d \Gamma(d/2+1)}} > 0$ is some fixed constant that depends on d only.

Since $s = 0$, this implies that

$$\Sigma_{\mathcal{N}_K} - \int_{[0,1]^d} s(x) dx \ge c(d) K^{-\alpha/d},$$

which concludes the proof (the bound on the quality follows trivially from this).

Appendix E. Proof of Theorem 2

The definition of K_n implies that (since n is such that $n \geq 2^{d+3\alpha}$)

$$K_n \geq \left(n^{\frac{1}{d+3\alpha}} - 1\right)^d \geq \left(n^{\frac{1}{d+3\alpha}}/2\right)^d \geq n^{\frac{d}{d+3\alpha}}/2^d. \tag{E.1}$$

Also, by definition

$$K_n \leq n^{\frac{d}{d+3\alpha}}. \tag{E.2}$$

Proposition 2 implies together with Equation (E.1)

$$\Sigma_{\mathcal{N}_{K_n}} \leq \int_{[0,1]^d} s(x)dx + \sqrt{2d}Mn^{\frac{-\alpha}{d+3\alpha}}2^\alpha. \tag{E.3}$$

We thus have

$$\begin{aligned} Q_{n,\mathcal{N}_{K_n}} &\leq \frac{\Sigma^2_{\mathcal{N}_{K_n}}}{n} - \frac{\left(\int_{[0,1]^d} s(x)dx\right)^2}{n} \\ &\leq \frac{2\Sigma_{\mathcal{N}_{K_n}}\sqrt{2d}Mn^{\frac{-\alpha}{d+3\alpha}}2^\alpha}{n} = 2^{\alpha+3/2}\Sigma_{\mathcal{N}_{K_n}}\sqrt{d}Mn^{-\frac{d+4\alpha}{d+3\alpha}}. \end{aligned} \tag{E.4}$$

On this partition, using Equation(E.2), Proposition 4, and Equation (E.3), one gets

$$\mathbb{E}R_{n,\mathcal{N}_{K_n}} \leq 48\Sigma_{\mathcal{N}_{K_n}}\sqrt{(2f_{\max} + 3b + 12f^2_{\max})}\left(2f_{\max}+1\right)^{4/3}n^{-\frac{d+4\alpha}{d+3\alpha}}\sqrt{\log(n)} + 14\Sigma^2_{\mathcal{N}_{K_n}}n^{-\frac{d+6\alpha}{d+3\alpha}}.$$

By combining this last equation with Equation (E.4), we obtain

$$\begin{aligned} &\mathbb{E}[L_n(\mathcal{A}_{MC-UCB})] - \inf_{\mathcal{N}' measurable} L_{n,\mathcal{N}'}(\mathcal{A}^*) \\ &\leq 2^{\alpha+3/2}\Sigma_{\mathcal{N}_{K_n}}\sqrt{d}Mn^{-\frac{d+4\alpha}{d+3\alpha}} \\ &+ 48\Sigma_{\mathcal{N}_{K_n}}\sqrt{(2f_{\max} + 3b + 12f^2_{\max})}\left(2f_{\max}+1\right)^{4/3}n^{-\frac{d+4\alpha}{d+3\alpha}}\sqrt{\log(n)} + 14\Sigma^2_{\mathcal{N}_{K_n}}n^{-\frac{d+6\alpha}{d+3\alpha}} \\ &\leq \left(2^{\alpha+3/2}\sqrt{d}M + 48\sqrt{(2f_{\max} + 3b + 12f^2_{\max})}\left(2f_{\max}+1\right)^{4/3}\right)\Sigma_{\mathcal{N}_{K_n}}n^{-\frac{d+4\alpha}{d+3\alpha}}\sqrt{\log(n)} + 14\Sigma^2_{\mathcal{N}_{K_n}}n^{-\frac{d+6\alpha}{d+3\alpha}} \\ &\leq 112\sqrt{d}(M+1)\sqrt{(2f_{\max} + 3b + 12f^2_{\max})}\left(2f_{\max}+1\right)^{4/3}f_{\max}n^{-\frac{d+4\alpha}{d+3\alpha}}\sqrt{\log(n)} + 56f^2_{\max}n^{-\frac{d+6\alpha}{d+3\alpha}}. \end{aligned}$$

since $\Sigma_{\mathcal{N}_{K_n}} \leq \Sigma_{\mathcal{N}_1} \leq 2f_{\max}$. This concludes the proof.

Appendix F. Proof of Theorem 1

Let us write the proof of the lower bound using the terminology of multi-armed bandits. We first prove a lower bound of order $n^{-4/3}$ for a two-armed bandit, which will be the lower bound considered for a small K (for $K \leq 2 \times 288\log(2)$). We then prove a lower bound in $K^{1/3}n^{-4/3}$ for any bandit with more than $2 \times 288\log(2)$ arms. This implies in all the cases the desired lower bound, i.e. of order $K^{1/3}n^{-4/3}$ (since for small K, i.e. $K \leq 2 \times 288\log(2)$, then K is smaller than a constant).

Appendix F.1. Lower bound with two arms

Consider a bandit with 2 Bernoulli arms, with means respectively $\mu_1 > 0$ and $\mu_2 = 1/2$.

Let $\mu > 0$ and $\alpha = \mu/2$ such that

$$0 < \mu - \alpha < \mu < \mu + \alpha < 1/2.$$

Note that for arm 1, one has that $\sigma^2 = \mu(1-\mu)$, and thus: $\sqrt{\frac{1}{2}\mu} \leq \sigma \leq \sqrt{\mu}$. We define $\sigma_{-\alpha}$ and $\sigma_{+\alpha}$ the two other standard deviation, and notice that $\frac{1}{2}\sqrt{\mu} \leq \sigma_{-\alpha} \leq \sqrt{\mu}$, and that $\sqrt{\frac{1}{2}\mu} \leq \sigma_{+\alpha} \leq \sqrt{3\mu/2}$. We choose μ small enough so that $\sigma = \frac{1}{2^{2/3}n^{1/3}}$.

We consider 3 bandit environments $M(\sigma)$, $M(\sigma_{-\alpha})$, $M(\sigma_{+\alpha})$ (characterized by the standard deviation of the first arm), and we denote by $\mathbb{P}_\sigma$, $\mathbb{P}_{\sigma_{-\alpha}}$, $\mathbb{P}_{\sigma_{+\alpha}}$ the probability with respect to the corresponding environments.

The optimal static allocation for environment $M(\sigma')$ is to play arm 1, $t_1(\sigma') = \frac{\sigma'}{\sigma'+1/2}n$ times and arm 2, $t_2(\sigma') = \frac{1/2}{\sigma'+1/2}n$ times. The corresponding quadratic error of the resulting estimate is $l(\sigma') = \frac{(\sigma'+1/2)^2}{4n}$.

Consider deterministic algorithms first (extension to randomized algorithms will be discussed later). An algorithm is a set (for all $t = 1$ to $n-1$) of mappings from any sequence $(u_1, \dots, u_t) \in \{0,1\}$ of t observed samples (where $u_s \in \{0,1\}$ is the sample observed at the s-th round) to the choice of an arm $I_{t+1} \in \{1,2\}$. The algorithm then returns an estimate and we write $L(u_1, \dots, u_T)$ the (random variable) corresponding quadratic error.

From Pinsker's inequality, we have:

$$\mathbb{P}_{\sigma_{-\alpha}}[T_1 \leq t_1(\sigma)] \leq \mathbb{P}_\sigma[T_1 \leq t_1(\sigma)] + \sqrt{KL(\mathbb{P}_{\sigma_{-\alpha}}, \mathbb{P}_\sigma)/2},$$

and

$$\mathbb{P}_{\sigma_{+\alpha}}[T_1 > t_1(\sigma)] \leq \mathbb{P}_\sigma[T_1 \leq t_1(\sigma)] + \sqrt{KL(\mathbb{P}_{\sigma_{+\alpha}}, \mathbb{P}_\sigma)/2},$$

which implies

$$\frac{1}{2}\big(\mathbb{P}_{\sigma_{-\alpha}}[T_1 \leq t_1(\sigma)] + \mathbb{P}_{\sigma_{+\alpha}}[T_1 > t_1(\sigma)]\big) \leq \mathbb{P}_\sigma[T_1 \leq t_1(\sigma)] + \sqrt{\max\big(KL(\mathbb{P}_{\sigma_{-\alpha}}, \mathbb{P}_\sigma), KL(\mathbb{P}_{\sigma_{+\alpha}}, \mathbb{P}_\sigma)\big)/2} \tag{F.1}$$

Now by the "chain rule" for Kullback Leibler divergence, we have

$$KL(\mathbb{P}_{\sigma_{-\alpha}}, \mathbb{P}_\sigma) = E_{\sigma_{-\alpha}}[T_1] kl(\mu - \alpha, \mu),$$

where $kl(a,b) = a\log(\frac{a}{b}) + (1-a)\log(\frac{1-a}{1-b})$ denotes the KL for Bernoulli distributions with parameters a anb b. Using the property $kl(a,b) \leq \frac{(a-b)^2}{b(1-b)}$, we deduce

$$\begin{aligned} KL(\mathbb{P}_{\sigma_{-\alpha}}, \mathbb{P}_\sigma) &\leq E_{\sigma_{-\alpha}}[T_1]\frac{\alpha^2}{\mu(1-\mu)} \\ &\leq E_{\sigma_{-\alpha}}[T_1]\frac{\alpha^2}{\sigma^2}. \end{aligned}$$

Let us assume that the algorithm has access to μ and α, and knows which arm is which (but does not know the environment). An optimal strategy will pull arm 1 such that: $\left(\frac{\sigma_{-\alpha}}{\sigma_{-\alpha}+1/2}\right)n \leq T_1 \leq \left(\frac{\sigma_{+\alpha}}{\sigma_{+\alpha}+1/2}\right)n$. We thus have

$$KL(\mathbb{P}_{\sigma_{-\alpha}}, \mathbb{P}_\sigma) \leq \Big(\frac{\sigma_{+\alpha}}{\sigma_{+\alpha}+1/2}\Big)\frac{\alpha^2}{\sigma^2}n.$$

The same reasoning provides exactly the same bound for $KL(\mathbb{P}_{\sigma_{+\alpha}}, \mathbb{P}_\sigma)$. We thus deduce (since $\alpha \leq \sigma^2$ from the previous bound, and Equation (F.1))

$$\begin{aligned}\frac{1}{2}\big(\mathbb{P}_{\sigma_{-\alpha}}[T_1 \leq t_1(\sigma)] + \mathbb{P}_{\sigma_{+\alpha}}[T_1 > t_1(\sigma)]\big) &\leq \frac{1}{2} + \frac{\alpha}{\sigma}\sqrt{\frac{n}{2}}\sqrt{\frac{\sigma_{+\alpha}}{\sigma_{+\alpha}+1/2}} \\ &\leq \frac{1}{2} + \frac{2\alpha}{\sigma}\sqrt{n}\sqrt{\sigma} \\ &\leq \frac{1}{2} + \frac{2\alpha}{\sqrt{\sigma}}\sqrt{n}.\end{aligned}$$

Thus

$$\min\Big(\mathbb{P}_{\sigma_{-\alpha}}[T_1 \leq t_1(\sigma)], \mathbb{P}_{\sigma_{+\alpha}}[T_1 > t_1(\sigma)]\Big) \leq \frac{1}{2} + \frac{2\alpha}{\sqrt{\sigma}}\sqrt{n}.$$

Note that $\frac{\sigma^2}{8} \leq \alpha \leq \frac{\sigma^2}{4}$. This says that (at least) one of the two events:

1. $\{T_1 > t_1(\sigma)\}$ under environment $M(\sigma_{-\alpha})$
2. $\{T_1 \leq t_1(\sigma)\}$ under environment $M(\sigma_{+\alpha})$

holds with probability at least $\frac{1}{2} - \frac{2\alpha}{\sqrt{\sigma}}\sqrt{n} \geq \frac{1}{2} - \frac{\sigma^{3/2}}{2}\sqrt{n}$. Let us assume that it is the first, i.e. $\mathbb{P}(\{T_1 > t_1(\sigma)\}) \geq \frac{1}{2} - \frac{\sigma^{3/2}}{2}\sqrt{n}$. Under the environment $M(\mu - \alpha)$, the event $\{T_1 > t_1(\sigma)\}$ means that the algorithm pulls arm 1 too many times by at least this number of pulls:

$$\begin{aligned}\Delta T = t_1(\sigma) - t_1(\sigma_{-\alpha}) &= \frac{\sigma}{\sigma+1/2}n - \frac{\sigma_{-\alpha}}{\sigma_{-\alpha}+1/2}n \\ &\geq \frac{\sigma}{\sigma+1/2}n - \frac{\sigma_{-\alpha}}{\frac{\sigma}{2}+1/2}n \\ &\geq \frac{\sqrt{\mu(1-\mu)}}{\sigma+1/2}n - \frac{\sqrt{\frac{\mu}{2}(1-\frac{\mu}{2})}}{\frac{\sigma}{2}+1/2}n \\ &\geq C\sigma n.\end{aligned}$$

Note also that for a reasonable algorithm, $\left(\frac{\sigma_{-\alpha}}{\sigma_{-\alpha}+1/2}\right)n \leq T_1 \leq \left(\frac{\sigma_{+\alpha}}{\sigma_{+\alpha}+1/2}\right)n$, and this leads to

$$\Delta T \leq D\sigma n.$$

The regret is at least:

$$
\begin{aligned}
R_{n,\mu-\alpha} &\geq \mathbb{P}(\{T_1 > t_1(\sigma)\})\Bigg[\frac{1}{4}\Big(\frac{\sigma_{-\alpha}^2}{t_1(\sigma_{-\alpha}) + \Delta T} + \frac{1}{4(n - t_1(\sigma_{-\alpha}) - \Delta T)}\Big) - \frac{1}{4}\frac{\Big(\sigma_{-\alpha} + 1/2\Big)^2}{n}\Bigg] \\
&= \mathbb{P}(\{T_1 > t_1(\sigma)\})\Bigg[\frac{1}{4}\frac{\Big(\sigma_{-\alpha} + 1/2\Big)}{n}\Big(\sigma_{-\alpha}\Big(1 - \Delta T\frac{(\sigma_{-\alpha} + 1/2)}{1/2n}\Big) + \frac{1}{2}\Big(1 + \frac{(\sigma_{-\alpha} + 1/2)}{n\sigma_{-\alpha}}\Delta T\Big)\Big) \\
&\quad \times \frac{1}{(1 - \Delta T\frac{(\sigma_{-\alpha}+1/2)}{1/2n})(1 + \frac{(\sigma_{-\alpha}+1/2)}{n\sigma_{-\alpha}}\Delta T)} - \frac{1}{4}\frac{\Big(\sigma_{-\alpha} + 1/2\Big)^2}{n}\Bigg] \\
&= \mathbb{P}(\{T_1 > t_1(\sigma)\})\Bigg[\frac{1}{4}\frac{\Big(\sigma_{-\alpha} + 1/2\Big)^2}{n}\Big(1 - \sigma_{-\alpha}\Delta T\frac{2}{n} + \frac{1}{2}\frac{1}{n\sigma_{-\alpha}}\Delta T\Big) \\
&\quad \times \frac{1}{(1 - \Delta T\frac{(\sigma_{-\alpha}+1/2)}{1/2n})(1 + \frac{(\sigma_{-\alpha}+1/2)}{n\sigma_{-\alpha}}\Delta T)} - \frac{1}{4}\frac{\Big(\sigma_{-\alpha} + 1/2\Big)^2}{n}\Bigg] \\
&= \mathbb{P}(\{T_1 > t_1(\sigma)\})\frac{1}{4}\frac{\Big(\sigma_{-\alpha} + 1/2\Big)^2}{n}\frac{2(\Delta T)^2(\sigma_{-\alpha} + 1/2)^2}{n^2\sigma_{-\alpha}} \\
&\quad \times \frac{1}{(1 - \Delta T\frac{(\sigma_{-\alpha}+1/2)}{1/2n})(1 + \frac{(\sigma_{-\alpha}+1/2)}{n\sigma_{-\alpha}}\Delta T)}.
\end{aligned}
$$

Since $\sigma \leq \frac{1}{2^{2/3}n^{1/3}}$, then $\mathbb{P}(\{T_1 > t_1(\sigma)\}) \leq 1/4$, and also $\Delta T \geq Cn^{2/3}$, and we know that

$$
\begin{aligned}
R_{n,\mu-\alpha} &\geq \frac{1}{256n}\frac{2C^2n^{4/3}n^{1/3}}{n^2}\frac{1}{(1 - \frac{2Cn^{2/3}}{n})(1 + \frac{2}{n^{2/3}}Dn^{2/3})} \\
&\leq C_2\frac{1}{n^{4/3}}.
\end{aligned}
$$

The same holds for environment $M(\sigma_{+\alpha})$ if it is event $\{T_1 \leq t_1(\sigma)\}$ that has a large probability in this event. We have thus proved the lower bound.

Now, the extension to randomized algorithms is straightforward: any randomized algorithm can be seen as a static (i.e., does not depend on samples) mixture of deterministic algorithms (which can be defined before the game starts). Each deterministic algorithm satisfies the lower bound above in expectation, thus any static mixture does so too.

Appendix F.2. Lower bound when the number of arms is such that $K \geq 288\log(2)$

Assume now that the number of arms satisfies $K \geq 288\log(2)$. Each arm k represents a stratum and the distribution associated to this arm is defined as the distribution of a noisy sample of the function collected when sampling uniformly in the strata.

Let us choose $3\mu/2 \leq 1/2$ and $\alpha = \frac{\mu}{2}$. Consider $2K$ Bernoulli bandits (i.e., $2K$ strata where the samples follow Bernoulli distributions) where the K first bandits have parameter $(\mu_k)_{1\leq k\leq K}$, where each $\mu_k \in \{\mu - \alpha, \mu, \mu + \alpha\}$, and the K last ones have parameter $1/2$.

Define $\sigma^2 = \mu(1-\mu)$ the variance of a Bernoulli of parameter μ, which is such that $\sqrt{\frac{1}{2}\mu} \leq \sigma \leq \sqrt{\mu}$ (since $\mu \leq 1/2$). We write $\sigma_{-\alpha}$ and $\sigma_{+\alpha}$ the two other standard deviations, and notice that $\frac{1}{2}\sqrt{\mu} \leq \sigma_{-\alpha} \leq \sqrt{\mu}$, and $\sqrt{\frac{1}{2}\mu} \leq \sigma_{+\alpha} \leq \sqrt{\frac{3}{2}\mu}$. We will in the sequel consider μ such that $3\mu/2 \leq 1/2$ and such that the associated standard deviation is such that $\sigma = \min\left(\frac{1}{20}(\frac{K}{n})^{1/3}, 1/2\right)$.

We consider the 2^K bandit environments $M(\upsilon)$ (characterized by $\upsilon = (\upsilon_k)_{1\leq k\leq K} \in \{-1,+1\}^K$) defined by $(\mu_k = \mu + \upsilon_k\alpha)_{1\leq k\leq K}$. We write $\mathbb{P}_\upsilon$ the probability with respect to the environment $M(\upsilon)$ at time n. We also write $M(\sigma)$ the environment defined by all K first arms having a parameter σ, and write $\mathbb{P}_\sigma$ the associated probability at time n.

The optimal oracle allocation for environment $M(\upsilon)$ consists in playing arm $k \leq K$, $\tau_k(\upsilon) = \frac{\sigma_{\upsilon_k\alpha}}{\sum_{i=1}^K \sigma_{\upsilon_i\alpha} + K/2} n$ times and arm $k > K$, $\tau_k(\upsilon) = \frac{1/2}{\sum_{i=1}^K \sigma_{\upsilon_i\alpha} + K/2} n$ times. The corresponding quadratic error of the resulting estimate is $l(\upsilon) = \frac{(\sum_{i=1}^K \sigma_{\upsilon_i\alpha} + K/2)^2}{n}$. For the environment $M(\sigma)$, the optimal oracle allocation consists in playing arm $k \leq K$, $t(\sigma) = \frac{\sigma}{K\sigma + K/2} n$ times (and arm $k > K$, $t_2(\sigma) = \frac{1/2}{K\sigma + K/2} n$ times).

Consider deterministic algorithms first (extension to randomized algorithms will be discussed later). An algorithm is a set (for all $t = 1$ to $n-1$) of mappings from any sequence $(r_1, \ldots, r_t) \in \{0,1\}$ of t observed samples (where $r_s \in \{0,1\}$ is the sample observed at the s-th round) to the choice of an arm $I_{t+1} \in \{1, \ldots, 2K\}$. Write $T_k(r_1, \ldots, r_n)$ the (random variable) corresponding to the number of pulls of arm k up to time n. We thus have $n = \sum_{k=1}^{2K} T_k$.

Now, consider the set of algorithms that know that the K first arms have parameter $\mu_k \in \{\mu-\alpha, \mu, \mu+\alpha\}$, and that also know that the K last arms have parameters $1/2$. Given this knowledge, an optimal algorithm will not pull any arm $k \leq K$ more than $\left(\frac{\sigma_{+\alpha}}{K\sigma_{-\alpha} + K/2}\right)n$ times. Indeed, the optimal oracle allocation in *all* such environments allocates less than $\left(\frac{\sigma_{+\alpha}}{K\sigma_{-\alpha} + K/2}\right)n$ samples to each arm $k \leq K$. In addition, since the samples of all arms are independent, a sample collected from arm k does not provide any information about the relative allocations for the other arms. Thus, once an arm has been pulled as many times as recommended by the optimal oracle strategy, there is no need to allocate more samples to that arm. Writing $\mathbb{A}$ the class of all algorithms that do not know the set of possible environments, $\mathbb{A}_\upsilon$ the class of algorithms that know the set of possible environments $M(\upsilon)$ and $\mathbb{A}_{opt}$ the subclass of $\mathbb{A}_\upsilon$ that pull all arms $k \leq K$ less than $\left(\frac{\sigma_{+\alpha}}{K\sigma_{-\alpha} + K/2}\right)n$ times, we have

$$\inf_{\mathbb{A}} \sup_{M(\upsilon)} \mathbb{E}R_n \geq \inf_{\mathbb{A}_\upsilon} \sup_{M(\upsilon)} \mathbb{E}R_n = \inf_{\mathbb{A}_{opt}} \sup_{M(\upsilon)} \mathbb{E}R_n,$$

where the first inequality comes from the fact that algorithms in $\mathbb{A}_\upsilon$ possess more information than those in $\mathbb{A}$, which they can use or not. Thus $\mathbb{A} \subset \mathbb{A}_\upsilon$.

Now for any $\upsilon = (\upsilon_1, \ldots, \upsilon_K)$, define the events

$$\Omega_\upsilon = \{\omega : \exists \mathcal{U} \subset \{1, \ldots, K\} : |\mathcal{U}| \leq \frac{K}{3} \text{ and } \forall k \in \mathcal{U}^c, \upsilon_k T_k \geq \upsilon_k t(\sigma)\}.$$

Note that by definition

$$\Omega_\upsilon = \bigcup_{p=1}^{\frac{K}{3}} \bigcup_{\mathcal{U}\subset\{1,\dots,K\}:|\mathcal{U}|=p} \left\{ \Big\{ \bigcap_{k\in\mathcal{U}} \{\upsilon_k T_k < \upsilon_k t(\sigma)\} \Big\} \bigcap \Big\{ \bigcap_{k\in\mathcal{U}^C} \{\upsilon_k T_k \geq \upsilon_k t(\sigma)\} \Big\} \right\}.$$

By the sub-additivity of the probabilities, we have

$$\mathbb{P}_\sigma(\Omega_\upsilon) \leq \sum_{p=1}^{\frac{K}{3}} \sum_{\mathcal{U}\subset\{1,\dots,K\}:|\mathcal{U}|=p} \mathbb{P}\left[\left\{ \Big\{ \bigcap_{k\in\mathcal{U}} \{\upsilon_k T_k < \upsilon_k t(\sigma)\} \Big\} \bigcap \Big\{ \bigcap_{k\in\mathcal{U}^C} \{\upsilon_k T_k \geq \upsilon_k t(\sigma)\} \Big\} \right\} \right].$$

Let us consider a fixed $\mathcal{U}$, and let us write for any υ, the corresponding $\upsilon^{\mathcal{U}} = (\mathbf{1}\{k \in \mathcal{U}\}\mathbf{1}\{\upsilon_k = -1\} + \mathbf{1}\{k \in \mathcal{U}^C\}\mathbf{1}\{\upsilon_k = 1\})_k$. This transformation is a bijection from $\{-1,1\}^K$ to $\{0,1\}^K$. This implies that the events $\left\{ \Big\{ \bigcap_{k\in\mathcal{U}} \{\upsilon_k T_k < \upsilon_k t(\sigma)\} \Big\} \bigcap \Big\{ \bigcap_{k\in\mathcal{U}^C} \{\upsilon_k T_k \geq \upsilon_k t(\sigma)\} \Big\} \right\}$ are disjoint for different υ, and cover all the space when all υ are considered. They thus form a partition of the space and

$$\sum_\upsilon \mathbb{P}_\sigma \left[\left\{ \Big\{ \bigcap_{k\in\mathcal{U}} \{\upsilon_k T_k < \upsilon_k t(\sigma)\} \Big\} \bigcap \Big\{ \bigcap_{k\in\mathcal{U}^C} \{\upsilon_k T_k \geq \upsilon_k t(\sigma)\} \Big\} \right\} \right] = 1.$$

We deduce that

$$\begin{aligned}
\sum_\upsilon \mathbb{P}_\sigma(\Omega_\upsilon) &\leq \sum_\upsilon \sum_{p=1}^{\frac{K}{3}} \sum_{\mathcal{U}\subset\{1,\dots,K\}:|\mathcal{U}|=p} \mathbb{P}_\sigma \left[\left\{ \Big\{ \bigcap_{k\in\mathcal{U}} \{\upsilon_k T_k < \upsilon_k t(\sigma)\} \Big\} \bigcap \Big\{ \bigcap_{k\in\mathcal{U}^C} \{\upsilon_k T_k \geq \upsilon_k t(\sigma)\} \Big\} \right\} \right] \\
&= \sum_{p=1}^{\frac{K}{3}} \sum_{\mathcal{U}\subset\{1,\dots,K\}:|\mathcal{U}|=p} \sum_\upsilon \left[\left\{ \Big\{ \bigcap_{k\in\mathcal{U}} \{\upsilon_k T_k < \upsilon_k t(\sigma)\} \Big\} \bigcap \Big\{ \bigcap_{k\in\mathcal{U}^C} \{\upsilon_k T_k \geq \upsilon_k t(\sigma)\} \Big\} \right\} \right] \\
&= \sum_{p=1}^{\frac{K}{3}} \sum_{\mathcal{U}\subset\{1,\dots,K\}:|\mathcal{U}|=p} 1 \\
&= \sum_{p=1}^{\frac{K}{3}} \begin{pmatrix} K \\ p \end{pmatrix}.
\end{aligned}$$

Since there are 2^K environments υ, we have

$$\min_\upsilon \mathbb{P}_\sigma(\Omega_\upsilon) \leq \frac{1}{2^K} \sum_\upsilon \mathbb{P}_\sigma(\Omega_\upsilon) \leq \frac{1}{2^K} \sum_{p=1}^{\frac{K}{3}} \begin{pmatrix} K \\ p \end{pmatrix}.$$

Note that $\frac{1}{2^K} \sum_{p=1}^{\frac{K}{3}} \begin{pmatrix} K \\ p \end{pmatrix} = \mathbb{P}(\sum_{k=1}^K X_k \leq \frac{K}{3})$ where $(X_1, \dots, X_K)$ are K independent Bernoulli random variables of parameter 1/2. By Chernoff-Hoeffding's inequality, we have $\mathbb{P}(\sum_{k=1}^K X_k \leq \frac{K}{3}) = \mathbb{P}(1/2 - \frac{1}{K}\sum_{k=1}^K X_k \geq \frac{K}{6}) \leq \exp(-K/72)$. Thus there exists $\upsilon_{\min}$ such that $\mathbb{P}_\sigma(\Omega_{\upsilon_{\min}}) \leq \exp(-K/72)$.

Let us write $p = \mathbb{P}_{\upsilon_{\min}}(\Omega_{\upsilon_{\min}})$ and $p_\sigma = \mathbb{P}_\sigma(\Omega_{\upsilon_{\min}})$. Let $kl(a,b) = a\log(\frac{a}{b}) + (1-a)\log(\frac{1-a}{1-b})$ denote the KL for Bernoulli distributions with parameters a and b. Note that because $\forall\Omega$, $KL(\mathbb{P}_{\upsilon_{\min}}(.|\Omega), \mathbb{P}_\sigma(.|\Omega)) \geq 0$, we have

$$kl(p, p_\sigma) \le KL(\mathbb{P}_{\upsilon_{\min}}, \mathbb{P}_\sigma).$$

From that we deduce that $p(\log(p) - \log(p_\sigma)) + (1-p)(\log(1-p) - \log(1-p_\sigma)) \le KL(\mathbb{P}_{\upsilon_{\min}}, \mathbb{P}_\sigma)$, which implies, using $p_\sigma \le \exp(-K/72)$, that

$$\begin{aligned} KL(\mathbb{P}_{\upsilon_{\min}}, \mathbb{P}_\sigma) &\ge \frac{p}{72}K + p\log(p) + (1-p)\log(1-p) \\ &\ge \frac{p}{72}K - \log(2), \end{aligned}$$

since the entropy $-p\log(p) - (1-p)\log(1-p)$ of a Bernoulli random variable is upper bounded by $\log(2)$ for any p. This leads to

$$p \le \frac{72}{K}\Big(KL(\mathbb{P}_{\upsilon_{\min}}, \mathbb{P}_\sigma) + \log(2)\Big). \tag{F.2}$$

Let us now consider any environment (υ). Let $R_t = (r_1, \dots, r_t)$ be the sequence of observations, and let $\mathbb{P}^t_\upsilon$ be the law of R_t for environment $M(\upsilon)$. Note first that $\mathbb{P}_\upsilon = \mathbb{P}^n_\upsilon$. Adapting the chain rule for Kullback-Leibler divergence, we get

$$\begin{aligned} &KL(\mathbb{P}^n_\upsilon, \mathbb{P}^n_\sigma) \\ &= KL(\mathbb{P}^1_\upsilon, \mathbb{P}^1_\sigma) + \sum_{t=2}^{n}\sum_{R_{t-1}} \mathbb{P}^{t-1}_\upsilon(R_{t-1}) KL(\mathbb{P}^t_\upsilon(.|R_{t-1}), \mathbb{P}^t_\sigma(.|R_t)) \\ &= KL(\mathbb{P}^1_\sigma, \mathbb{P}^1_\upsilon) + \sum_{t=2}^{n}\Big[\sum_{R_{t-1}|\upsilon_{I_t}=+1} \mathbb{P}^{t-1}_\sigma(R_{t-1}) kl(\mu+\alpha, \mu) + \sum_{R_{t-1}|\upsilon_{I_t}=-1} \mathbb{P}^{t-1}_\sigma(R_{t-1}) kl(\mu-\alpha, \mu)\Big] \\ &= kl(\mu-\alpha, \mu)\mathbb{E}_\upsilon[\sum_{k:\upsilon_k=-1} T_k] + kl(\mu+\alpha, \mu)\mathbb{E}_\upsilon[\sum_{k:\upsilon_k=+1} T_k]. \end{aligned}$$

We thus have, using the property that $kl(a,b) \le \frac{(a-b)^2}{b(1-b)}$,

$$\begin{aligned} KL(\mathbb{P}_\upsilon, \mathbb{P}_\sigma) &= kl(\mu-\alpha, \mu)\mathbb{E}_\upsilon[\sum_{k:\upsilon_k=-1} T_k] + kl(\mu+\alpha, \mu)\mathbb{E}_\upsilon[\sum_{k:\upsilon_k=+1} T_k] \\ &\le \mathbb{E}_\sigma[\sum_{k\le K} T_k]\frac{\alpha^2}{\mu(1-\mu)} \\ &= E_\sigma[\sum_{k\le K} T_k]\frac{\alpha^2}{\sigma^2}. \end{aligned}$$

Note that for an algorithm in $\mathbb{A}_{opt}$, we have $\sum_{k=1}^{K} T_k \le K\Big(\frac{\sigma_{+\alpha}}{K\sigma_{-\alpha}+K/2}\Big)n$. Since $\alpha = \frac{\mu}{2}$ and $0 < \mu \le \frac{1}{2}$ we have

$$\begin{aligned} KL(\mathbb{P}_\upsilon, \mathbb{P}_\sigma) &\le \Big(K\frac{\sigma_{+\alpha}}{K\sigma_{-\alpha}+K/2}\Big)\frac{\alpha^2}{\sigma^2}n \\ &\le 4\sigma_{+\alpha}\frac{\alpha^2}{\sigma^2}n \\ &\le 8\frac{\alpha^2}{\sigma}n, \end{aligned}$$

We thus deduce using Equation F.2

$$\begin{aligned}
\mathbb{P}_{\upsilon_{\min}}(\Omega_{\upsilon_{\min}}) = p &\leq \frac{72}{K}\Big(KL(\mathbb{P}_{\upsilon_{\min}}, \mathbb{P}_\sigma) + \log(2)\Big) \\
&\leq \frac{576}{K}\frac{\alpha^2}{\sigma}n + \frac{72\log(2)}{K} \\
&\leq 1/4 + \frac{72\log(2)}{K} \\
&\leq \frac{1}{2},
\end{aligned}$$

since $\sigma \leq \frac{1}{20}(\frac{K}{n})^{1/3}$ and $\alpha = \frac{\mu}{2}$ where $\frac{\mu}{2} \leq \sigma^2 \leq \mu$, and $K \geq 288\log(2)$.

Let $\omega \in \Omega^c_{\upsilon_{\min}}$. We know that for any ω, there are at least $\frac{K}{3}$ arms among the K first ones, that are not pulled correctly: either $\frac{K}{6}$ arms among the arms with parameter $\mu - \alpha$ or among the arms with parameter $\mu + \alpha$ are not pulled correctly. Assume that for this fixed ω, there are $\frac{K}{6}$ arms among the arms with parameter $\mu - \alpha$ which are not pulled correctly. Let $\mathcal{U}(\omega)$ be this subset of arms.

We write $\Delta T = \sum_{k\in\mathcal{U}} T_k - \sum_{k\in\mathcal{U}} \tau_k(\upsilon)$ the number of times those arms are over pulled. Note that on ω we have $\Delta T \geq \frac{K}{6}\big(t(\sigma) - \frac{\sigma_{-\alpha}}{\sum_{i=1}^K \sigma_{\upsilon_i\alpha} + K/2}n\big)$. We have

$$\begin{aligned}
\Delta T = \frac{K}{6}t(\sigma) - \frac{K}{6}t(\sigma_{-\alpha}) &= \frac{1}{6}\frac{K\sigma}{K\sigma + K/2}n - \frac{1}{6}\frac{K\sigma_{-\alpha}}{\sum_{i=1}^K \sigma_{\upsilon_i\alpha} + K/2}n \\
&\geq \frac{1}{6}\frac{K\sigma}{K\sigma + K/2}n - \frac{1}{6}\frac{K\sigma_{-\alpha}}{K\sigma_{-\alpha} + K/2}n \\
&\geq \frac{n}{6}\Big(\frac{\sigma}{\sigma + 1/2} - \frac{\sigma_{-\alpha}}{\sigma_{-\alpha} + 1/2}\Big) \\
&\geq \frac{n}{12}\frac{\sigma - \sigma_{-\alpha}}{(\sigma + 1/2)(\sigma_{-\alpha} + 1/2)} \\
&\geq \frac{n}{12}(\sigma - \sigma_{-\alpha}) \\
&\geq \frac{n}{120}\sigma \\
&\geq \min\big(\frac{n}{240}, \frac{1}{2400}K^{1/3}n^{2/3}\big),
\end{aligned}$$

since $\sigma_{-\alpha} \leq \frac{\sqrt{3}}{2}\sigma \leq 0.9\sigma$, by definition of these quantities and since $\sigma = \min(\frac{1}{20}\big(\frac{K}{n}\big)^{1/3}, 1/2)$.

Thus on ω, the regret is such that

$$
\begin{aligned}
R_{n,\upsilon_{\min}}(\omega) &\geq \sum_{k=1}^{2K} \frac{w_k^2\sigma_k^2}{T_k(\omega)} - \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha} + K/2\right)^2}{n}\\
&\geq \sum_{k\in\mathcal{U}(\omega)} \frac{w_k^2\sigma_k^2}{T_k(\omega)} + \sum_{k\in\mathcal{U}(\omega)^C} \frac{w_k^2\sigma_k^2}{T_k(\omega)} - \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha} + K/2\right)^2}{n}\\
&\geq \frac{1}{K^2}\frac{K}{6}\frac{\sigma_{-\alpha}^2}{t_k(\sigma_{-\alpha}) + 6\Delta T/K} + \frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha} - K\sigma_{-\alpha}/6 + K/2\right)^2}{(2K-K/6)^2(n-\Delta T)} - \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha} + K/2\right)^2}{n}\\
&\geq \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha} + K/2\right)^2}{n}\frac{1+\left(\frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}+K/2\right)\Delta T}{\left(K\sigma_{-\alpha}/6\right)n} - \frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}+K/2\right)\Delta T}{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}-K\sigma_{-\alpha}/6+K/2\right)n}\right)}{\left(1+\frac{6\Delta T\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}+K/2\right)}{K\sigma_{-\alpha}n}\right)\left(1-\frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}+K/2\right)\Delta T}{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}-K\sigma_{-\alpha}/6+K/2\right)n}\right)}\\
&\quad - \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha} + K/2\right)^2}{n}\\
&\geq \frac{1}{(2K)^2}\frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha} + K/2\right)^2}{n}\frac{\left(\frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}+K/2\right)\Delta T}{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}-K\sigma_{-\alpha}/6+K/2\right)n}\right)\left(\frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}+K/2\right)\Delta T}{\left(K\sigma_{-\alpha}/6\right)n}\right)}{\left(1+\frac{6\Delta T\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}+K/2\right)}{K\sigma_{-\alpha}n}\right)\left(1-\frac{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}+K/2\right)\Delta T}{\left(\sum_{i=1}^K \sigma_{\upsilon_i\alpha}-K\sigma_{-\alpha}/6+K/2\right)n}\right)}\\
&\geq C'\frac{(\Delta T)^2}{n^3\sigma}\\
&\geq C\frac{K^{1/3}}{n^{4/3}},
\end{aligned}
$$

where C', C are numerical constants. Note that for events ω where there are $\frac{K}{6}$ arms among the arms with parameter $\mu + \alpha$ which are not pulled correctly, the same result holds.

Note finally that $\mathbb{P}(\Omega^c_{\upsilon_{\min}}) \geq 1/2$. We thus have that the regret is lower-bounded as

$$
\begin{aligned}
\mathbb{E}R_{n,\upsilon_{\min}} &\geq \sum_{\omega\in\Omega^c_{\upsilon_{\min}}} R_{n,\upsilon_{\min}}(\omega)\mathbb{P}_{\upsilon_{\min}}(\omega)\\
&\geq \sum_{\omega\in\Omega^c_{\upsilon_{\min}}} C\frac{K^{1/3}}{n^{4/3}}\mathbb{P}_{\upsilon_{\min}}(\omega)\\
&\geq \frac{1}{2}C\frac{K^{1/3}}{n^{4/3}} \geq C_1\frac{K^{1/3}}{n^{4/3}},
\end{aligned}
$$

which proves the lower bound for deterministic algorithms. Again, the extension to randomized algorithms is straightforward: any randomized algorithm can be seen as a static (i.e., does not depend on samples) mixture of deterministic algorithms (which can be defined before the game starts). Each deterministic algorithm satisfies the lower bound above in expectation, thus any static mixture does so too.

Appendix F.3. Conclusion

Consider a multi-armed bandit with K arms. If $K \leq 2\times 288\log(2)$, we know from the lower bound with 2 arms that there exists a bandit with K arms (with 2 arms defined as in the lower bound with two arms, and all other arms that always output 0) such that

$$\mathbb{E}R_{n,\upsilon_{\min}} \geq C_2 n^{-4/3} = \frac{C_2}{K^{1/3}}\frac{K^{1/3}}{n^{4/3}} \geq \frac{C_2}{24}\frac{K^{1/3}}{n^{4/3}},$$

since $K \leq 2 \times 288\log(2)$. Now if $K \geq 2 \times 288\log(2)$, we know from the lower bound with K arms that there exists a bandit with K arms such that

$$\mathbb{E}R_{n,\upsilon_{\min}} \geq C_1\frac{K^{1/3}}{n^{4/3}}.$$

This concludes the proof.

Appendix G. Proof of Theorem 3

Set $K = n^{\frac{d}{d+3\alpha}}$, and D be some large constant.

Let $\nu \in \{0,1\}^K$. We define the functions $(f,s) = (f_\nu, s_\nu)$ as follows.

- On the first fifth of the domain (i.e. on $[0,1/5] \times [0,1]^{d-1}$), set f and s such that $s = 0$ and $f = \sum_{k \in \mathcal{N}_{5DK}, \Omega_k \subset [0,1/5]\times[0,1]^{d-1}} f_k$ (where f_k as in the proof of Proposition 3).
- On the third fifth of the domain (i.e. on $[2/5,3/5] \times [0,1]^{d-1}$) set f, s and the distribution of the noise such that the distribution when one samples in x is a Bernoulli of parameter $1/2$ (i.e. $f = 1/2$ and $s = 1/2$).
- On the fifth part of the domain (i.e. on $[4/5,1] \times [0,1]^{d-1}$), the construction is more involved. Consider first a stratification of the fifth part in K strata of same size. For each stratum, consider the ball that is in the middle of the stratum, and whose radius is a quarter of the side length of the stratum. Let g_ν be the function whose value is $\sigma_{\nu_k} \in \{\sigma_-, \sigma_+\}$ on the ball (where we write σ_- and σ_+ for the $\sigma_{-\alpha}$ and the $\sigma_{+\alpha}$ in the proof of Theorem 1, in order to avoid confusion with the Hölder exponent). Outside the ball, let the function go linearly to 0 from σ_{ν_k} (it reaches 0 on the ball that is in the middle of the stratum, and whose radius is half of the side length of the stratum. Then the function remains 0 in the rest of the stratum. We define $f = f_\nu$ and $s = s_\nu$ such that the distribution in x is a Bernoulli of standard deviation g_ν and associated mean smaller than $1/2$ (i.e. $f = g_\nu$ and $s = \sqrt{g_\nu(1-g_\nu)}$).
- Finally, connect linearly the three parts of the function f, s on the second and fourth fifth, so that they are continuous.

This construction defines the functions $(f,s) = (f_\nu, s_\nu)$ (which differ according to ν on the fifth part).

The functions (f_ν, s_ν) are clearly $(5,\alpha)$-Hölder on parts 1 to 4 of the domain, and continuous on the whole domain. We remind that in the proof of Theorem 1, the quantity σ is of order $(K/n)^{1/3} = K^{-\alpha/d}$, and σ_-, σ_+ are of the same order. Let us write c for the constant such that $\sigma_- \leq \sigma_+ = cK^{-\alpha/d}$. The linear part linking σ_{ν_k} with 0 is of length $K^{1/d}/4$. So for any two points $(x, x+u)$ in the linear junction part, we have $\frac{\|s(x+u)-s(x)\|_2}{\|u\|_2^\alpha} \leq \sup_{0\leq v\leq K^{1/d}/4} \frac{vcK^{-\alpha/d}}{v^\alpha} c4^{\alpha-1\leq c}$. So s is $(\max(5,c),\alpha)$-Hölder on the whole domain. The same holds for f since f is the Bernoulli mean associated with f, and since f is very small on the fifth domain, then $f \approx s^2$ and has a regularity higher than f on this domain. So it is also $(\max(5,c),\alpha)$-Hölder on the whole domain.

As proved in Proposition 3, for any partition of DK or less than DK strata, we have that the difference in quality is larger than the difference in quality in the first part of the domain (since the quality on the rest of the domain is positive by definition), and

$$\inf_{\mathcal{N}} Q_{n,\mathcal{N}} \geq c\frac{K^{-\alpha/d}}{n} = cn^{-\frac{d+4\alpha}{d+3\alpha}},$$

where $c > 0$ depends on d only. This concludes the proof for any algorithm that considers partitions in less than DK strata (since the pseudo-regret is positive by definition).

Now consider partitions in more than DK strata, and consider a partition $\mathcal{N}_{K'}$, $K' \geq DK$. For such a partition, we have more than $K'/5$ arms which are Bernoulli of variance $1/2$ (third part of the domain). For $K' \geq DK$ (where we choose D large enough for this), the number of arms which are Bernoulli of standard deviation σ_- is larger than uK' where $u > 0$ is some constant that depends on d only. The same holds for the number of arms which are Bernoulli of standard deviation σ_+. We can thus apply Theorem 1 since we are in the same configuration when ν is varying (up to some constants) than in this theorem, and we get that there exists some constant c such that

$$\inf \sup_{\nu} R_{n,\mathcal{N}_{K'}} \geq c\frac{(uK')^{1/3}}{n^{4/3}} \geq c'n^{-\frac{d+4\alpha}{d+3\alpha}}.$$

This concludes the proof.

Appendix H. Large deviation inequalities for independent sub-Gaussian random variables

We first state Bernstein inequality for large deviations of independent random variables around their mean.

Lemma 3. *Let $(X_1,\ldots,X_n)$ be n independent random variables of mean $(\mu_1,\ldots,\mu_n)$ and of variance $(\sigma_1^2,\ldots,\sigma_n^2)$. Assume that there exists $b > 0$ such that for any $\lambda < \frac{1}{b}$, for any $i \leq n$, it holds that $\mathbb{E}\Big[\exp(\lambda(X_i-\mu_i))\Big] \leq \exp\Big(\frac{\lambda^2\sigma_i^2}{2(1-\lambda b)}\Big)$. Then with probability larger than $1-\delta$*

$$|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i| \leq \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n}.$$

Proof. If the assumptions of Lemma 3 are verified, then

$$\begin{aligned}
\mathbb{P}\Big(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i \geq n\upsilon\Big) &= \mathbb{P}\Big[\exp\Big(\lambda(\textstyle\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i)\Big) \geq \exp(n\lambda\upsilon)\Big] \\
&\leq \mathbb{E}\Bigg[\frac{\exp\Big(\lambda(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i)\Big)}{\exp(n\lambda\upsilon)}\Bigg] \\
&\leq \textstyle\prod_{i=1}^n \mathbb{E}\Bigg[\frac{\exp\Big(\lambda(X_i-\mu_i)\Big)}{\exp(\lambda\upsilon)}\Bigg] \\
&\leq \exp(\tfrac{\lambda^2}{2}\textstyle\sum_{i=1}^n \frac{\sigma_i^2}{2(1-\lambda b)} - n\lambda\upsilon).
\end{aligned}$$

By setting $\lambda = \frac{n\upsilon}{\sum_{i=1}^n \sigma_i^2 + bn\upsilon}$ we obtain

$$\mathbb{P}\Big(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i \geq n\upsilon\Big) \leq \exp(-\frac{n^2\upsilon^2}{2(\sum_{i=1}^n \sigma_i^2 + bn\upsilon)}).$$

By an union bound we obtain

$$\mathbb{P}\Big(|\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i| \geq n\upsilon\Big) \leq 2\exp(-\frac{n^2\upsilon^2}{2(\sum_{i=1}^n \sigma_i^2 + bn\upsilon)}).$$

This means that with probability $1-\delta$,

$$|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i| \leq \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n}.$$

□

We also state the following Lemma on large deviations for the variance of independent random variables.

Lemma 4. *Let $(X_1, \ldots, X_n)$ be n independent random variables of mean $(\mu_1, \ldots, \mu_n)$ and of variance $(\sigma_1^2, \ldots, \sigma_n^2)$. Assume that there exists $b > 0$ such that for any $\lambda < \frac{1}{b}$, for any $i \leq n$, it holds that $\mathbb{E}\Big[\exp(\lambda(X_i - \mu_i))\Big] \leq \exp\Big(\frac{\lambda^2\sigma_i^2}{2(1-\lambda b)}\Big)$ and also $\mathbb{E}\Big[\exp(\lambda(X_i - \mu_i)^2 - \lambda\sigma_i^2)\Big] \leq \exp\Big(\frac{\lambda^2\sigma_i^2}{2(1-\lambda b)}\Big)$. Assume also that there exists $f_{\max}$ such that $\max_i(\max(\sigma_i, |\mu_i|)) \leq f_{\max}$, with $f_{\max} \geq 1$.*

Let $V = \frac{1}{n}\sum_i(\mu_i - \frac{1}{n}\sum_i \mu_i)^2 + \frac{1}{n}\sum_n \sigma_i^2$ be the variance of a sample chosen uniformly at random among the n distributions, and $\hat{V} = \frac{1}{n}\sum_{i=1}^n \big(X_i - \frac{1}{n}\sum_{j=1}^n X_j\big)^2$ the corresponding empirical variance. Then if $n \geq b\log(2/\delta)$, with probability larger than $1-\delta$,

$$\begin{aligned}
|\sqrt{\hat{V}} - \sqrt{V}| &\leq 2(2f_{\max}+1)\sqrt{\frac{(2f_{\max} + 3b + 4V)\log(6/\delta)}{n}} \\
&\leq 2(2f_{\max}+1)\sqrt{\frac{(2f_{\max} + 3b + 12f_{\max}^2)\log(6/\delta)}{n}}.
\end{aligned}$$

Proof. By decomposing the estimate of the empirical variance in bias and variance, we obtain with probability $1-\delta$

$$\begin{aligned}\hat{V} =& \frac{1}{n}\sum_i (X_i - \frac{1}{n}\sum_j \mu_j)^2 - (\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mu_i)^2 \\ =& \frac{1}{n}\sum_i (X_i-\mu_i)^2 + 2\frac{1}{n}\sum_i (X_i-\mu_i)(\mu_i - \frac{1}{n}\sum_j \mu_j) \\ &+ \frac{1}{n}\sum_i(\mu_i - \frac{1}{n}\sum_j \mu_j)^2 - (\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mu_i)^2.\end{aligned}$$

We then have by the definition of V,

$$\hat{V} - V = \frac{1}{n}\sum_{i=1}^n (X_i-\mu_i)^2 - \frac{1}{n}\sum_{i=1}^n \sigma_i^2 + 2\frac{1}{n}\sum_i (X_i-\mu_i)(\mu_i - \frac{1}{n}\sum_j \mu_j) - (\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mu_i)^2. \quad \text{(H.1)}$$

Similar to Lemma 3 (since for any $i \le n$, $|\mu_i - \frac{1}{n}\sum_j \mu_j| \le 2f_{\max}$), it holds that with probability larger than $1-\delta$

$$\Big|\frac{1}{n}\sum_i (X_i-\mu_i)(\mu_i - \frac{1}{n}\sum_j \mu_j)\Big| \le 2f_{\max}\sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + 2f_{\max}\frac{b\log(2/\delta)}{n}. \quad \text{(H.2)}$$

If the assumptions of Lemma 4 are verified, we have

$$\begin{aligned}\mathbb{P}\Big(\sum_{i=1}^n (X_i-\mu_i)^2 - \sum_{i=1}^n \sigma_i^2 \ge n\upsilon\Big) &= \mathbb{P}\Bigg[\exp\Big(\lambda(\sum_{i=1}^n |X_i-\mu_i|^2 - \sum_{i=1}^n \sigma_i^2)\Big) \ge \exp(n\lambda\upsilon)\Bigg] \\ &\le \mathbb{E}\Bigg[\frac{\exp\Big(\lambda(\sum_{i=1}^n |X_i-\mu_i|^2 - \sum_{i=1}^n \sigma_i^2)\Big)}{\exp(n\lambda\upsilon)}\Bigg] \\ &\le \prod_{i=1}^n \mathbb{E}\Bigg[\frac{\exp\Big(\lambda(|X_i-\mu_i|^2 - \sigma_i^2)\Big)}{\exp(\lambda\upsilon)}\Bigg] \\ &\le 2\exp(\lambda^2\sum_{i=1}^n \frac{\sigma_i^2}{2(1-\lambda b)} - n\lambda\upsilon).\end{aligned}$$

By choosing $\lambda = \frac{n\upsilon}{\sum_{i=1}^n \sigma_i^2 + nb\upsilon}$ we obtain

$$\mathbb{P}\Big(\sum_{i=1}^n (X_i-\mu_i)^2 - \sum_{i=1}^n \sigma_i^2 \ge n\upsilon^2\Big) \le \exp(-\frac{n^2\upsilon^2}{2(\sum_{i=1}^n \sigma_i^2 + bn\upsilon)}). \quad \text{(H.3)}$$

By a union bound we get

$$\mathbb{P}\Big(|\sum_{i=1}^n (X_i-\mu_i)^2 - \sum_{i=1}^n \sigma_i^2| \ge n\upsilon\Big) \le 2\exp(-\frac{n^2\upsilon^2}{2(\sum_{i=1}^n \sigma_i^2 + bn\upsilon)}).$$

This means that with probability larger than $1-\delta$,

$$|\frac{1}{n}\sum_{i=1}^n (X_i-\mu_i)^2 - \frac{1}{n}\sum_{i=1}^n \sigma_i^2| \le \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n}. \quad \text{(H.4)}$$

Finally, by combining Equations (H.1), (H.2) and (H.4) with Lemma 3, we obtain with probability larger than $1-3\delta$

$$\begin{aligned}
|\hat{V}-V| &\le \frac{4(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n} + \frac{2b^2\log(2/\delta)^2}{n^2} + \sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{b\log(2/\delta)}{n} \\
&\quad + 2f_{\max}\sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + 2f_{\max}\frac{b\log(2/\delta)}{n} \\
&\le (2f_{\max}+1)\sqrt{\frac{2(\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n}} + \frac{(2f_{\max}+3b+4\frac{1}{n}\sum_{i=1}^n \sigma_i^2)\log(2/\delta)}{n} \\
&\le (2f_{\max}+1)\sqrt{\frac{2V\log(2/\delta)}{n}} + \frac{(3b+4V+2f_{\max})\log(2/\delta)}{n},
\end{aligned}$$

when $n \ge b\log(2/\delta)$, since $V \ge \frac{1}{n}\sum_{i=1}^n \sigma_i^2$.

This implies with probability larger than $1-3\delta$ that

$$\begin{aligned}
&V - (2f_{\max}+1)\sqrt{\frac{2V\log(2/\delta)}{n}} + (2f_{\max}+1)^2\frac{\log(2/\delta)}{2n} \le \hat{V} + \frac{(2f_{\max}+3b+4V)\log(2/\delta)}{n} + (2f_{\max}+1)^2\frac{\log(2/\delta)}{2n} \\
&\Leftrightarrow \sqrt{V} - (2f_{\max}+1)\sqrt{\frac{\log(2/\delta)}{2n}} \le \sqrt{\hat{V} + \frac{(2f_{\max}+3b+4V+(2f_{\max}+1)^2)\log(2/\delta)}{n}} \\
&\Rightarrow \sqrt{V} - (2f_{\max}+1)\sqrt{\frac{\log(2/\delta)}{2n}} \le \sqrt{\hat{V}} + (2f_{\max}+1)\sqrt{\frac{(2f_{\max}+3b+4V)\log(2/\delta)}{n}} \\
&\Rightarrow \sqrt{V} \le \sqrt{\hat{V}} + 2(2f_{\max}+1)\sqrt{\frac{(2f_{\max}+3b+4V)\log(2/\delta)}{n}},
\end{aligned}$$

since $f_{\max} \ge 1$.

On the other hand, we have also with probability larger than $1-3\delta$ (on the same event as before)

$$\begin{aligned}
&\hat{V} \le V + (2f_{\max}+1)\sqrt{\frac{2V\log(2/\delta)}{n}} + \frac{(3b+4V+2f_{\max})\log(2/\delta)}{n} \\
&\Rightarrow \sqrt{\hat{V}} \le \sqrt{V} + 2(2f_{\max}+1)\sqrt{\frac{(2f_{\max}+3b+4V)\log(2/\delta)}{n}}.
\end{aligned}$$

Finally, we have with probability larger than $1-3\delta$

$$|\sqrt{\hat{V}} - \sqrt{V}| \le 2(2f_{\max}+1)\sqrt{\frac{(2f_{\max}+3b+4V)\log(2/\delta)}{n}}. \tag{H.5}$$

□