Active Learning in Heteroscedastic Noise ☆

András Antos[a,*], Varun Grover[b], Csaba Szepesvári[a,b,*]

[a] *Computer and Automation Research Institute of the Hungarian Academy of Sciences, Kende u. 13-17, Budapest 1111, Hungary*
[b] *Department of Computing Science, University of Alberta, Edmonton T6G 2E8, Canada*

Abstract

We consider the problem of actively learning the mean values of distributions associated with a finite number of options. The decision maker can select which option to generate the next observation from, the goal being to produce estimates with equally good precision for all the options. If sample means are used to estimate the unknown values then the optimal solution, assuming that the distributions are known up to a shift, is to sample from each distribution proportional to its variance. No information other than the distributions' variances is needed to calculate the optimal solution. In this paper we propose an incremental algorithm that asymptotically achieves the same loss as an optimal rule. We prove that the excess loss suffered by this algorithm, apart from logarithmic factors, scales as $n^{-3/2}$, which we conjecture to be the optimal rate. The performance of the algorithm is illustrated on a simple problem.

Key words: active learning, heteroscedastic noise, regression, sequential analysis, sequential allocation

---

1. Introduction

Consider the problem of production quality assurance in a factory equipped with a number of machines that output products of different quality. The quality of production can be monitored by inspecting the products manufactured: An inspection of a product results in a (random) number which, without the loss of generality, can be assumed to lie between zero and one, one meaning the best, zero the poorest quality. It is assumed that the mean of these measurements characterizes the maintenance state of the machine. The goal is to estimate these mean values for the individual machines so as to minimize the maximum prediction error over the machines. Since the inspection results are random, multiple measurements are necessary for each machine. If the inspection of a

---

☆A previous version of this paper appeared at ALT-08 [1].
*Corresponding author.
Email addresses: antos@szit.bme.hu (András Antos), vgrover@cs.ualberta.ca (Varun Grover), szepesva@cs.ualberta.ca (Csaba Szepesvári)

product is expensive (as is the case when inspection requires the destruction of the products) then inspecting all machines equally often can be wasteful, since the precision of the estimate of the quality of any machine will be proportional to the variance of the inspection outcomes for that machine and hence, if there is a machine with high variance outcomes, one can inspect that machine more often at the price of inspecting machines with low variance outcomes less frequently, thus equalizing the quality of the estimates. Based on this sample, one suspects that a good sequential algorithm can result in significant cost-savings as compared to inspecting the products produced by each machine equally often.

This is an instance of active learning [6], which is also closely related to optimal experimental design of statistics [9, 5]. In particular, the problem can be cast as learning a regression function over a finite domain. The problem is also similar to *multi-armed bandit problems* [12, 3] in that only one option (arm) can be probed at any time. However, the performance criterion is different from that used in bandits where the observed values are treated as rewards and performance during learning is what matters. Nevertheless, we will see that the exploration-exploitation dilemma which characterizes classical bandit problems will still play a role here.

The formal description of this problem is as follows: We are interested in estimating the expected values (μ_k) of some distributions ($\mathcal{D}_k$), each associated with an option. If K is the number of options then $1 \leq k \leq K$. For any k, the decision maker can take independent observations $\{X_{kt}\}_t$ from $\mathcal{D}_k$. The value X_{kt} is observed when an observation is requested from option k the t^{th} time. (These observations correspond to the outcomes of inspections in the previous example). The observations are drawn sequentially: Given the information collected up to trial n the decision maker can decide which option to choose next.

The loss minimized by the decision maker is defined as follows: After trial n, let $\hat{\mu}_{kn}$ denote the estimate of μ_k as computed by the decision maker ($1 \leq k \leq K$). Let the error of predicting μ_k with $\hat{\mu}_{kn}$ be measured with the expected squared error,

$$L_{kn} = \mathbb{E}\left[(\hat{\mu}_{kn} - \mu_k)^2\right].$$

The overall loss is measured by the worst-case loss over the K options:

$$L_n = \max_{1 \leq k \leq K} L_{kn}.$$

The motivation for considering this loss function is as follows: Let M_K denote the set of probability distributions over $\{1, 2, \ldots, K\}$. Pick some $p \in M_K$. Imagine that after learning, an option will be randomly chosen from p. The task of the decision maker is to estimate μ_k if option k is selected. Assume that the decision maker uses $\hat{\mu}_{kn}$ to estimate μ_k. The associated least-squares loss then becomes $\mathbb{E}\left[\sum_{k=1}^{K} p_k(\hat{\mu}_{kn} - \mu_k)^2\right]$. Since during learning p is not known, taking a pessimistic approach, the loss is minimized for the worst distribution given the estimates, i.e., the goal is to minimize

$$L_n^{\mathcal{W}} = \max_{p \in M_K} \mathbb{E}\left[\sum_{k=1}^{K} p_k (\hat{\mu}_{kn} - \mu_k)^2\right].$$

It is not hard to see that $L_n^{\mathcal{W}} = L_n$, thus minimizing $L_n^{\mathcal{W}}$ and L_n are the same.

In this paper we will assume that the estimates $\hat{\mu}_{kn}$ are produced by computing the sample means of the respective options:

$$\hat{\mu}_{kn} = \frac{1}{T_{kn}} \sum_{t=1}^{T_{kn}} X_{kt}.$$

Here T_{kn} denotes the number of times an observation was requested from option k up to trial n.

Consider the non-sequential version of the problem, i.e., the problem of choosing $T_{1n}, \ldots, T_{Kn}$ such that $T_{1n} + \ldots + T_{Kn} = n$ so as to minimize the loss. Let us assume for a moment that we know the distributions up to an unknown shift. In particular, this means that we do not know (say) the mean of the distributions, but we know the variances of the distributions and all higher order moments. In this case there is no value in making the choice of $T_{1n}, \ldots, T_{Kn}$ data dependent. Due to the independence of observations

$$L_{kn} = \frac{\sigma_k^2}{T_{kn}},$$

where $\sigma_k^2 = \mathrm{Var}\left[X_{k1}\right]$. For simplicity assume that $\sigma_k^2 > 0$ holds for all k. It is not hard to see then that the minimizer of $L_n = \max_k L_{kn}$ is the allocation $\{T_{kn}^*\}_k$ that makes all the losses L_{kn} (approximately) equal, hence (apart from rounding issues)

$$T_{kn}^* = n \frac{\sigma_k^2}{\Sigma^2} = n\, \lambda_k.$$

Here $\Sigma^2 = \sum_{j=1}^{K} \sigma_j^2$ is the sum of the variances and

$$\lambda_k = \frac{\sigma_k^2}{\Sigma^2}$$

is the ratio of the k^{th} variance and the sum of the variances. The value of λ_k gives the *optimal allocation ratio* for option k. The corresponding loss is

$$L_n^* = \frac{\Sigma^2}{n}.$$

We conclude that to calculate the optimal allocations, all one needs to know about the distributions is their respective variances.

We note in passing that the optimal allocation is easy to extend to the case when some options have zero variance: in such a case it is both necessary and sufficient to make a single observation on such options. The case when all variances are zero (i.e., $\Sigma^2 = 0$) is uninteresting, hence we will assume from now on that $\Sigma^2 > 0$.

We expect a good sequential algorithm $\mathcal{A}$ to achieve a loss $L_n = L_n(\mathcal{A})$ close to the loss L_n^*. We will therefore look into the excess loss

$$\mathcal{E}_n(\mathcal{A}) = L_n(\mathcal{A}) - L_n^*.$$

Since L_{kn}, the loss of option k, can only decrease if we request a new observation from $\mathcal{D}_k$, one simple idea is to request the next observation from option k whose estimated loss, $\hat{\sigma}_{kn}^2/T_{kn}$, is the largest amongst all estimated losses. Here $\hat{\sigma}_{kn}^2$ is an estimate of the variance of the k^{th} option based on the history. The problem with this approach is that the variance might be underestimated in which case the option will not be selected for a long time, which prevents refining the estimated variance, ultimately resulting in a large excess loss. Thus we face a problem similar to the exploration-exploitation dilemma in bandit problems where a greedy policy might incur a large loss if the payoff of the optimal option is underestimated. One simple remedy is to make sure that the estimated variances converge to their true values. This can be ensured if the algorithm is forced to select all the options indefinitely in the limit, which is often called the method of forced sampling in the bandit literature [13]. One way to implement this idea is to introduce phases of increasing length. Then in each phase the algorithm could choose all options exactly once at the beginning, while in the rest of the phase it can sample all the options k proportionally to their respective variance estimates computed at the beginning of the phase. The problem then becomes to select the appropriate phase lengths to make sure that the proportion of forced selections diminishes at an appropriate rate with an increasing horizon n. (An algorithm along these lines have been described and analyzed by [8] in the context of stratified sampling. We shall discuss this algorithm further in Section 6.) While the introduction of phases allows a direct control of the proportion of forced selections, the algorithm is not incremental and is thus less appealing.

In this paper we propose and study an alternative algorithm that implements forced selections but remains completely incremental. The idea is to select the option with the largest estimated loss except if some of the options is seriously under-sampled, in which case the under-sampled option is selected. It turns out that a good definition for an option being under-sampled is $T_{kn} \leq \alpha\sqrt{n}$ with some constant $\alpha > 0$. (The algorithm will be formally stated in the next section.) We will show that the excess loss of this algorithm decreases with n as $\tilde{O}(n^{-3/2})$.[1]

[1] A nonnegative sequence $\{a_n\}$ is said to be $\tilde{O}(f_n)$, where $\{f_n\}$ is a positive valued sequence, if $a_n \leq C\, f_n \log^p(f_n)$ with suitable constants $C,p > 0$.

2. The Algorithm

The formal description of the algorithm, that we call GAFS-MAX (greedy allocation with forced selections for max-norm loss), is as follows:

Algorithm GAFS-MAX
In the first K trials choose each option once
Set $T_{k,K} = 1$ $(1 \le k \le K)$, $n = K$
At time $n+1$ do:

Predict $\hat{\mu}_{kn} = \frac{1}{T_{kn}} \sum_{t=1}^{T_{kn}} X_{kt}$

Compute $\hat{\sigma}^2_{kn} = \frac{1}{T_{kn}} \sum_{t=1}^{T_{kn}} X^2_{kt} - \hat{\mu}^2_{kn}$

Let

$$\hat{\lambda}_{kn} = \begin{cases} \hat{\sigma}^2_{kn}/(\sum_{j=1}^{K} \hat{\sigma}^2_{jn}), & \text{if } \sum_{j=1}^{K} \hat{\sigma}^2_{jn} \neq 0, \\ 1/K, & \text{otherwise} \end{cases}$$

Let $U_n = \operatorname{argmin}_{1 \le k \le K} T_{kn}$

Let

$$I_{n+1} = \begin{cases} U_n, & \text{if } T_{U_n,n} < \alpha\sqrt{n} + 1, \\ \operatorname{argmax}_{1 \le k \le K} \frac{\hat{\lambda}_{kn}}{T_{kn}}, & \text{otherwise} \end{cases}$$

Choose option I_{n+1} and let $T_{k,n+1} = T_{kn} + \mathbb{I}_{\{ I_{n+1} = k \}}$

Observe the feedback $X_{I_{n+1}, T_{I_{n+1},n+1}}$.

Of course, the variance estimates can be computed incrementally. Further, it is actually not necessary to compute the *estimated allocation ratios* $\hat{\lambda}_{kn}$ because in the computation of the option index I_{n+1}, $\hat{\lambda}_{kn}$ can be replaced by $\hat{\sigma}^2_{kn}$ without effecting the choices. The only parameter of the algorithm, α, determines the minimum amount of exploration. We normally set α to 1 (cf. Section 5.1).

3. Main Results

The main result of this paper is the following theorem:

Theorem 1. *Assume that the observations $\{X_{kt}\}$ are bounded with probability one.*[2] *Let L_n be the loss of GAFS-MAX after the n^{th} trial and let L^*_n be the optimal loss. Then*

$$L_n \le L^*_n + \tilde{O}(n^{-3/2}).$$

This result will be proved in the Section 4.3.We also prove high probability bounds on $T_{kn}/n - \lambda_k$ (Theorem 2). The proofs are somewhat involved, hence we start with an outline:

Clearly, the rate of growth of T_{kn} controls the rate of convergence of $\hat{\lambda}_{kn}$ to λ_k. In particular, we will show that given $T_{kn} \ge f_n$ it follows that $\hat{\lambda}_{kn}$ converges

[2] The results easily extend to the case when the tails of $\{X_{kt}\}$ are sub-Gaussian.

to λ_k at a rate of $\tilde{O}(1/f_n^{1/2})$ (Lemma 3). The second major tool is a result (cf. Lemma 4 and Corollary 1) that shows how a faster rate for $\hat{\lambda}_{kn}$ transforms into better bounds on T_{kn}. The actual proof is then started by observing that due to the forced selections $T_{kn} \geq \sqrt{n}$. Hence, by the first generic result the rate of convergence of $\hat{\lambda}_{kn}$ is at least $1/n^{1/4}$. The second device then enables us to show that T_{kn} grows at least as fast as $n\,\lambda_k/2$, i.e., linearly in n. Using again the first result we get that $\hat{\lambda}_{kn} - \lambda_k$ decays at least as fast as $1/n^{1/2}$, which, using the second result, allows us to conclude that $T_{kn}/n - \lambda_k$ converges to zero at the rate of $1/n^{1/2}$. Resorting to Wald's second identity then allows us to prove that the excess loss $L_{kn} - L_n^*$ decays at the rate of $1/(n^{3/2})$.

The convergence rate statements for $\hat{\lambda}_{kn}$ and T_{kn}/n hold with high probability. In particular, they all hold on the same event set A_δ.

4. Proof

The proof is presented in three sections. In Section 4.1, we introduce the necessary notation and some preliminary results that show the rate of convergence of the estimated allocation ratios $\hat{\lambda}_{kn}$ to the optimal allocation ratios λ_k. In Section 4.2, we show that these in turn give bounds on the actual allocation ratios, T_{kn}/n. Finally, in Section 4.3, we prove the main result.

4.1. Preliminaries and notation

First, we state Hoeffding's inequality in a form that suits the best our needs:

Lemma 1 (Hoeffding's inequality, [11]). *Let Z_t be a sequence of zero-mean, i.i.d. random variables, where $a \leq Z_t \leq b$, $a < b$ reals. Then, for any $0 < \delta \leq 1$,*

$$\mathbb{P}\left(\frac{1}{n}\sum_{t=1}^{n} Z_t > \sqrt{\frac{1}{2}\frac{(b-a)^2}{n}\,\log(1/\delta)}\right) \leq \delta.$$

Let us now introduce some notation. First, let

$$\Delta(R, n, \delta) = R\sqrt{\frac{\log(1/\delta)}{2n}}$$

denote the deviation bound that we can get from Hoeffding's equality for the confidence level δ after seeing n observations from a distribution whose support is included in an interval of length R. Further, let $\mu_k^{(2)} = \mathbb{E}\left[X_{kt}^2\right]$, R_k be the length of the (connected) range of the random variables $\{X_{kt}\}_t$ (i.e., $R_k = \operatorname{esssup} X_{kt} - \operatorname{essinf} X_{kt}$), S_k be the length of the (connected) range of the random variables $\{X_{kt}^2\}_t$, and B_k be the essential supremum of the random variables

$\{|X_{kt}|\}_t$. Note that $R_k \le 2B_k$ and $S_k \le B_k^2$. Let

$$\begin{aligned} A_\delta &= \bigcap_{1\le k\le K, n\ge 1} \left\{ \left| \frac{1}{n}\sum_{t=1}^{n} X_{kt}^2 - \mu_k^{(2)} \right| \le \Delta(S_k, n, \delta_n) \right\} \cap \\ & \bigcap_{1\le k\le K, n\ge 1} \left\{ \left| \frac{1}{n}\sum_{t=1}^{n} X_{kt} - \mu_k \right| \le \Delta(R_k, n, \delta_n) \right\}, \end{aligned}$$

where

$$\delta_n = \frac{\delta}{4Kn(n+1)}.$$

Note that δ_n is chosen so that $\sum_{k=1}^{K}\sum_{n=1}^{\infty}\delta_n = \delta/4$. Hence, we observe that by Hoeffding's inequality

$$\mathbb{P}(A_\delta) \ge 1-\delta.$$

The sets $\{A_\delta\}$ will play a key role in the proof: Many of the statements will be proved on these set.

Our first result connects a lower bound on T_{kn} to the rate of convergence of $\hat\lambda_{kn}$. Let $b_k = S_k + (|\mu_k| + B_k)R_k$ $(\le 5B_k^2)$, $a'_k = 2b_k/\sigma_k^2$, and $\ell_{K,\delta} = \log(4K/\delta)$. Note that, by $\sigma_k^2 \le (R_k/2)^2$,

$$a'_k \ge 8b_k/R_k^2 \ge 8B_k/R_k \ge 4 \tag{1}$$

and that

$$\log(\delta_n^{-1}) = \log(n(n+1)) + \ell_{K,\delta} \le 2\log n + 1 + \ell_{K,\delta}. \tag{2}$$

Lemma 2. *Fix $0<\delta\le 1$, $1\le k\le K$, and $n>0$, and assume that $T_{kn}\ge 1$ holds on A_δ. Then*

$$\left|\hat\sigma_{kn}^2 - \sigma_k^2\right| \le b_k\sqrt{\frac{\log(\delta_{T_{kn}}^{-1})}{2T_{kn}}}$$

also holds on A_δ.

Proof. Let $\hat\mu_{kn}^{(2)} = 1/T_{kn}\sum_{t=1}^{T_{kn}} X_{kt}^2$ and recall that $\hat\mu_{kn} = 1/T_{kn}\sum_{t=1}^{T_{kn}} X_{kt}$. Consider any element of A_δ. Then by the definition of A_δ,

$$\left| \frac{1}{m}\sum_{t=1}^{m} X_{kt}^2 - \mu_k^{(2)} \right| \le \Delta(S_k, m, \delta_m)$$

holds simultaneously for any $m\ge 1$. Hence, it also holds that

$$\left|\hat\mu_{kn}^{(2)} - \mu_k^{(2)}\right| = \left| \frac{1}{T_{kn}}\sum_{t=1}^{T_{kn}} X_{kt}^2 - \mu_k^{(2)} \right| \le \Delta(S_k, T_{kn}, \delta_{T_{kn}}).$$

Similarly, we get that

$$|\hat{\mu}_{kn} - \mu_k| = \left| \frac{1}{T_{kn}} \sum_{t=1}^{T_{kn}} X_{kt} - \mu_k \right| \leq \Delta(R_k, T_{kn}, \delta_{T_{kn}}).$$

Using $\hat{\sigma}^2_{kn} = \hat{\mu}^{(2)}_{kn} - \hat{\mu}^2_{kn}$ and $\sigma^2_k = \mathbb{E}\left[X^2_{kt}\right] - (\mathbb{E}\left[X_{kt}\right])^2 = \mu^{(2)}_k - \mu^2_k$, we get

$$\begin{aligned}
\left|\hat{\sigma}^2_{kn} - \sigma^2_k\right| &\leq \left|\hat{\mu}^{(2)}_{kn} - \mu^{(2)}_k\right| + \left|\hat{\mu}^2_{kn} - \mu^2_k\right| \\
&\leq \left|\hat{\mu}^{(2)}_{kn} - \mu^{(2)}_k\right| + |\hat{\mu}_{kn} - \mu_k|(|\hat{\mu}_{kn}| + |\mu_k|) \\
&\leq \Delta(S_k, T_{kn}, \delta_{T_{kn}}) + \Delta(R_k, T_{kn}, \delta_{T_{kn}})(|\mu_k| + B_k) \\
&= (S_k + R_k(|\mu_k| + B_k)) \sqrt{\frac{\log(\delta^{-1}_{T_{kn}})}{2T_{kn}}} = b_k \sqrt{\frac{\log(\delta^{-1}_{T_{kn}})}{2T_{kn}}}. \qquad \square
\end{aligned}$$

Lemma 3. *Fix $0 < \delta \leq 1$, $n_0 > 0$, and assume that for $n \geq n_0$, $1 \leq k \leq K$, $T_{kn} \geq f_n \geq 2$ holds on A_δ, and that for $n \geq n_0$, for each $1 \leq k \leq K$ such that $\sigma_k \neq 0$*

$$f_n \geq \frac{{a'_k}^2}{2} \left(2 \log f_n + 1 + \ell_{K,\delta}\right). \tag{3}$$

Then there exists a constant $c > 0$ such that for any $n \geq n_0$, $1 \leq k \leq K$, on A_δ

$$\left|\hat{\lambda}_{kn} - \lambda_k\right| \leq c \sqrt{\frac{\log(\delta^{-1}_{f_n})}{f_n}} \tag{4}$$

holds. In particular, c can be chosen as

$$\frac{\sqrt{2}}{\Sigma^2} \max_{1 \leq k \leq K} \left(b_k + \lambda_k \sum_{j=1}^{K} b_j \right) = \frac{1}{\sqrt{2}} \max_{1 \leq k \leq K} \lambda_k \left(a'_k + \sum_{j=1}^{K} \lambda_j a'_j \right).$$

Remark 1. If $f_n = \beta n^p$ (p, $n > 0$) then (3) can be written as

$$\log n \leq \frac{\beta}{p{a'_k}^2} n^p - \frac{1 + \ell_{K,\delta} + 2 \log \beta}{2p}. \tag{5}$$

Remark 2. Note that, using (1) and $\lambda_k \leq 1$, the choice of c above can be sandwiched as

$$\begin{aligned}
4\sqrt{2} \max_{1 \leq k \leq K} \lambda_k &\leq \frac{1}{\sqrt{2}} \max_{1 \leq k \leq K} \lambda_k \left(a'_k + \sum_{j=1}^{K} \lambda_j a'_j \right) \\
&\leq \frac{1}{\sqrt{2}} \left(\max_{1 \leq k \leq K} \lambda_k a'_k + \sum_{j=1}^{K} \lambda_j a'_j \right) \leq \frac{5\sqrt{2}}{\Sigma^2} \left(\max_{1 \leq k \leq K} B^2_k + \sum_{j=1}^{K} B^2_j \right).
\end{aligned}$$

In what follows, for simplicity, we define c as

$$c = \frac{1}{\sqrt{2}} \left(\max_{1 \le k \le K} \lambda_k a'_k + \sum_{j=1}^{K} \lambda_j a'_j \right) \ge \sqrt{8}. \tag{6}$$

Proof. Using Lemma 2, for $n \ge n_0$, $1 \le k \le K$,

$$\left| \hat{\sigma}_{kn}^2 - \sigma_k^2 \right| \le b_k \sqrt{\frac{\log(\delta_{T_{kn}}^{-1})}{2T_{kn}}} \le b_k \sqrt{\frac{\log(\delta_{f_n}^{-1})}{2f_n}} \tag{7}$$

holds on A_δ, where we have used that $(\log(x(x+1)) + \ell_{K,\delta})/x$ is monotonically decreasing when $x \ge 2$, $\ell_{K,\delta} > 0$ and that $T_{kn} \ge f_n \ge 2$. Denote the right-hand side of (7) by $\Delta_{kn}(\delta)$.

Now, let us develop a lower bound on $\hat{\lambda}_{kn}$ in terms of λ_k. For $n \ge n_0$,

$$\begin{aligned}
\hat{\lambda}_{kn} &= \frac{\hat{\sigma}_{kn}^2}{\sum_{j=1}^{K} \hat{\sigma}_{jn}^2} \ge \frac{\sigma_k^2 - \Delta_{kn}(\delta)}{\Sigma^2 + \sum_{j=1}^{K} \Delta_{jn}(\delta)} \\
&= \frac{\sigma_k^2}{\Sigma^2} \left(1 + \frac{\sum_{j=1}^{K} \Delta_{jn}(\delta)}{\Sigma^2} \right)^{-1} - \frac{\Delta_{kn}(\delta)}{\Sigma^2 + \sum_{j=1}^{K} \Delta_{jn}(\delta)} \\
&\ge \lambda_k \left(1 - \frac{\sum_{j=1}^{K} \Delta_{jn}(\delta)}{\Sigma^2} \right) - \frac{\Delta_{kn}(\delta)}{\Sigma^2},
\end{aligned}$$

where we used $1/(1+x) \ge 1 - x$ that holds for $x > -1$.

An upper bound can be obtained analogously: For $n \ge n_0$, if

$$\Sigma^2 \ge 2 \sum_{j=1}^{K} \Delta_{jn}(\delta) \tag{8}$$

then

$$\begin{aligned}
\hat{\lambda}_{kn} &= \frac{\hat{\sigma}_{kn}^2}{\sum_{j=1}^{K} \hat{\sigma}_{jn}^2} \le \frac{\sigma_k^2 + \Delta_{kn}(\delta)}{\Sigma^2 - \sum_{j=1}^{K} \Delta_{jn}(\delta)} \\
&= \frac{\sigma_k^2}{\Sigma^2} \left(1 - \frac{\sum_{j=1}^{K} \Delta_{jn}(\delta)}{\Sigma^2} \right)^{-1} + \frac{\Delta_{kn}(\delta)}{\Sigma^2 - \sum_{j=1}^{K} \Delta_{jn}(\delta)} \\
&\le \lambda_k \left(1 + 2 \frac{\sum_{j=1}^{K} \Delta_{jn}(\delta)}{\Sigma^2} \right) + 2 \frac{\Delta_{kn}(\delta)}{\Sigma^2},
\end{aligned}$$

where we used $1/(1-x) = 1 + x/(1-x) \le 1 + 2x$ that holds for $0 \le x \le 1/2$. This constraint follows from (8), that is implied if n is big enough so that

$$\sigma_j^2 \ge 2\Delta_{jn}(\delta), \quad 1 \le j \le K. \tag{9}$$

The upper and lower bounds above, together with (7), give

$$\begin{aligned} |\hat{\lambda}_{kn} - \lambda_k| &\le \frac{2}{\Sigma^2}\left(\lambda_k \sum_{j=1}^K \Delta_{jn}(\delta) + \Delta_{kn}(\delta)\right) \\ &\le \frac{\sqrt{2}}{\Sigma^2}\left(\lambda_k \sum_{j=1}^K b_j + b_k\right)\sqrt{\frac{\log(\delta_{f_n}^{-1})}{f_n}} \end{aligned}$$

proving (4).

At last, to satisfy (9), by (7), it suffices if $\sigma_j^4 f_n \ge 2b_j^2 \log(\delta_{f_n}^{-1})$, $1 \le j \le K$. Note that if $\sigma_j = 0$ then $R_j = S_j = 0$, and so $b_j = 0$ and both sides above are 0. Otherwise we need

$$f_n \ge \frac{2b_j^2}{\sigma_j^4}\log(\delta_{f_n}^{-1}) = \frac{{a'_j}^2}{2}\log(\delta_{f_n}^{-1})$$

that is guaranteed by (2) and (3) provided that $n \ge n_0$. □

4.2. Bounds on the actual allocation ratios

Now we show how a rate of convergence result for $\hat{\lambda}_{kn}$ can be turned into bounds on the difference between the actual allocation ratios T_{kn}/n and λ_k. Note that this lemma holds pointwise, i.e., for any element ω of the probability space Ω underlying the random variables considered. For brevity, we write below $\hat{\lambda}_{kn}$ instead of $\hat{\lambda}_{kn}(\omega)$, T_{kn} instead of $T_{kn}(\omega)$, etc.

Let

$$\lambda_{\min} = \min_{1\le j\le K} \lambda_j \qquad \text{and} \qquad \rho = 1 + \frac{2}{\lambda_{\min}}.$$

In what follows, unless otherwise stated, we will assume that $\lambda_{\min} > 0$. For $K = 1$ the results are obvious, so without the loss of generality we can also assume that $K \ge 2$, in which case $\lambda_{\min} \le 1/K \le 1/2$ and $5 \le \rho \le 2.5/\lambda_{\min}$.

Lemma 4. *Fix $n_0 > 0$. Assume that g_n is such that for $n \ge n_0$, ng_n is monotone increasing in n, $5ng_n \ge \lceil\sqrt{n}\rceil$, and*

$$g_n \le \lambda_{\min}/2, \tag{10}$$

$$|\hat{\lambda}_{kn} - \lambda_k| \le g_n, \qquad 1 \le k \le K \tag{11}$$

hold. Then the following inequalities hold for $n \ge 1$ and $1 \le k \le K$:

$$-(K-1)\max\left(\frac{n_0}{n}, \frac{1}{n} + \rho g_n\right) \le \frac{T_{kn}}{n} - \lambda_k \le \max\left(\frac{n_0}{n}, \frac{1}{n} + \rho g_n\right).$$

Proof. By definition $T_{k,n+1} = T_{kn} + \mathbb{I}_{\{I_{n+1}=k\}}$. Let $E_{kn} = T_{kn} - n\lambda_k$ with $E_{k0} = 0$. Note that $E_{kn} \le n(1-\lambda_k)$ and

$$\sum_{k=1}^K E_{kn} = 0 \tag{12}$$

hold for any $n \geq 0$. Notice that the desired result can be stated as bounds on E_{kn}. Hence, our goal now is to study E_{kn}. If b_{jn} is an upper bound for E_{jn} ($1 \leq j \leq K$) then from (12) we get the lower bound $E_{kn} = -\sum_{j\neq k} E_{jn} \geq -\sum_{j\neq k} b_{jn} \geq -(K-1)\max_j b_{jn}$. Hence, we target upper bounds on $\{E_{kn}\}_k$.

Assume now that $n \geq n_0$. Note that (10) and (11) imply $\lambda_k - \hat{\lambda}_{kn} \leq |\hat{\lambda}_{kn} - \lambda_k| \leq \lambda_k/2$, and thus $\hat{\lambda}_{kn} \geq \lambda_k/2 > 0$ for each k.

From the definition of E_{kn} and T_{kn} we get

$$E_{k,n+1} = E_{kn} - \lambda_k + \mathbb{I}_{\{\, I_{n+1}=k\,\}}.$$

By the definition of the algorithm

$$\mathbb{I}_{\{\, I_{n+1}=k\,\}} \leq \mathbb{I}_{\left\{\, T_{kn}\leq\lceil\sqrt{n}\rceil \text{ or } k=\operatorname{argmin}_{1\leq j\leq K} \frac{T_{jn}}{\hat{\lambda}_{jn}}\,\right\}},$$

Assume now that k is an index where $\{\frac{T_{jn}}{\hat{\lambda}_{jn}}\}_j$ takes its minimum, that is,

$$\frac{T_{kn}}{\hat{\lambda}_{kn}} \leq \min_j \frac{T_{jn}}{\hat{\lambda}_{jn}}.$$

Using $T_{jn} = E_{jn} + n\lambda_j$ and reordering the terms gives

$$E_{kn} + n\lambda_k \quad \leq \quad \hat{\lambda}_{kn} \min_j \frac{E_{jn} + n\lambda_j}{\hat{\lambda}_{jn}} \leq \hat{\lambda}_{kn}\left(\min_j \frac{E_{jn}}{\hat{\lambda}_{jn}} + n\max_j \frac{\lambda_j}{\hat{\lambda}_{jn}}\right).$$

By (12), there exists an index j such that $E_{jn} \leq 0$. Since $\hat{\lambda}_{jn} > 0$ for any j, it holds that $\min_j \frac{E_{jn}}{\hat{\lambda}_{jn}} \leq 0$. Hence,

$$E_{kn} + n\lambda_k \leq n\hat{\lambda}_{kn} \max_j \frac{\lambda_j}{\hat{\lambda}_{jn}}. \tag{13}$$

Using (11) and (10), we get

$$\frac{\lambda_j}{\hat{\lambda}_{jn}} \leq \frac{\lambda_j}{\lambda_j - g_n} = \frac{1}{1 - g_n/\lambda_j}.$$

This is upper bounded by

$$1 + \frac{2g_n}{\lambda_j}$$

using $1/(1-x) \leq 1 + 2x$ for $0 \leq x \leq 1/2$, where the latest constraint follows from (10). Using (13), $\hat{\lambda}_{kn} \leq 1$,and (11) again,

$$\begin{aligned} E_{kn} &\leq n\hat{\lambda}_{kn} \max_j \frac{\lambda_j}{\hat{\lambda}_{jn}} - n\lambda_k \\ &\leq n(\hat{\lambda}_{kn} - \lambda_k) + \frac{2ng_n}{\lambda_{\min}} \\ &\leq \left(1 + \frac{2}{\lambda_{\min}}\right) ng_n = \rho n g_n. \end{aligned}$$

Denote the right-hand side by F_n. Hence,

$$\mathbb{I}_{\{ I_{n+1}=k \}} \leq \mathbb{I}_{\{ T_{kn} \leq \lceil \sqrt{n} \rceil \text{ or } E_{kn} \leq F_n \}}.$$

We show that $T_{kn} \leq \lceil \sqrt{n} \rceil$ implies $E_{kn} \leq F_n$. By the definition of E_{kn}, from $T_{kn} \leq \lceil \sqrt{n} \rceil$ it follows that $E_{kn} = T_{kn} - n\lambda_k \leq \lceil \sqrt{n} \rceil \leq 5ng_n$. The bound $\rho \geq 5$ implies $5ng_n \leq F_n$. Hence, $E_{kn} \leq F_n$ follows. Therefore

$$\mathbb{I}_{\{ I_{n+1}=k \}} \leq \mathbb{I}_{\{ E_{kn} \leq F_n \}}.$$

Now we need the following technical lemma:

Lemma 5. *Let $0 \leq \lambda \leq 1$. Consider the sequences $E_n, \tilde{E}_n, I_n, \tilde{I}_n$ ($n \geq 1$) where $I_n, \tilde{I}_n \in 0, 1$, $E_{n+1} = E_n + I_n - \lambda$, $\tilde{E}_{n+1} = \tilde{E}_n + \tilde{I}_n - \lambda$, $\tilde{E}_1 = E_1$ and assume that $I_n \leq \tilde{I}_n$ holds whenever $E_n = \tilde{E}_n$. Then $E_n \leq \tilde{E}_n$ holds for $n \geq 1$.*

Proof. Consider the difference sequence $P_n = \tilde{E}_n - E_n$. The goal is to show that $P_n \geq 0$ holds for any n. It holds that $P_1 = 0$. Since

$$P_{n+1} - P_n = (\tilde{E}_{n+1} - \tilde{E}_n) - (E_{n+1} - E_n) = \tilde{I}_n - I_n \in \{ -1, 0, +1 \},$$

P_n is always an integer. Hence, it suffices to show that $P_{n+1} \geq 0$ if $P_n = 0$. However, this holds because if $P_n = 0$ then $I_n \leq \tilde{I}_n$. □

Now, returning to the proof of Lemma 4, define $\{\tilde{E}_{kn}\}_{n \geq n_0}$ by

$$\begin{aligned} \tilde{E}_{k,n_0} &= E_{k,n_0}, \\ \tilde{E}_{k,n+1} &= \tilde{E}_{kn} - \lambda_k + \mathbb{I}_{\{ \tilde{E}_{kn} \leq F_n \}}, \quad n \geq n_0. \end{aligned}$$

The conditions of Lemma 5 are clearly satisfied from index n_0. Consequently $E_{kn} \leq \tilde{E}_{kn}$ holds for any $n \geq n_0$. Further, since F_n is monotone increasing in n,

$$\tilde{E}_{kn} \leq \max(E_{k,n_0}, 1 + F_n) \leq \max(n_0(1 - \lambda_k), 1 + F_n), \qquad n \geq n_0,$$

and so $E_{kn} \leq \max(n_0(1 - \lambda_k), 1 + F_n) \leq \max(n_0, 1 + F_n)$ for $n \geq 0$, finishing the upper-bound. □

Corollary 1. *Fix $0 < \delta \leq 1$, $c \geq 1/5$, and $n_0 \geq 1$. Assume that $f_n > 0$ is such that for $n \geq n_0$, f_n is monotone increasing, but f_n/n^2 is monotone decreasing, $1 \leq f_n \leq n$,*

$$f_n \geq \frac{4c^2}{\lambda_{\min}^2}(2\log f_n + 1 + \ell_{K,\delta}), \qquad \textit{and} \tag{14}$$

$$|\hat{\lambda}_{kn} - \lambda_k| \leq c\sqrt{\frac{\log(\delta_{f_n}^{-1})}{f_n}}, \qquad 1 \leq k \leq K \tag{15}$$

hold. Let $F_n(\delta) = \rho n g_n(\delta)$, *where*

$$g_n(\delta) = c\sqrt{\frac{\log(\delta_{f_n}^{-1})}{f_n}}.$$

Then the following inequalities hold for $n \geq 0$ *and* $1 \leq k \leq K$*:*

$$-(K-1)\max(n_0, 1 + F_n(\delta)) \leq T_{kn} - n\lambda_k \leq \max(n_0, 1 + F_n(\delta)).$$

Further, these inequalities remain valid if δ_{f_n} *is replaced by* δ_n *in* $F_n(\delta)$.

Remark 3. If $f_n = \beta n^p$ (p,$n > 0$) then (14) can be written as

$$\log n \leq \frac{\beta\lambda_{\min}^2}{8pc^2}\, n^p - \frac{1 + \ell_{K,\delta} + 2\log\beta}{2p}. \tag{16}$$

Proof. Assume that $n \geq n_0$. Then $n g_n(\delta)$ is monotone increasing, (2) and (14) imply (10), and (15) implies (11). The bounds on f_n, K, and δ imply

$$5 n g_n(\delta) = 5nc\sqrt{\frac{\log(4K f_n(f_n+1)/\delta)}{f_n}} \geq 5c\sqrt{n\log(8 f_{n_0}(f_{n_0}+1))},$$

that is at least $\sqrt{n\log(16)} > \sqrt{2n} \geq \lceil\sqrt{n}\rceil$ by the bounds on c and f_{n_0}. Thus Lemma 4 gives the result. The last statement follows obviously from $\delta_{f_n}^{-1} \leq \delta_n^{-1}$ (since $f_n \leq n$). $\square$

Using the previous results we are now in the position to prove a linear lower bound on T_{kn}:

Lemma 6. *Let* $0 < \delta \leq 1$ *arbitrary. Then there exists an integer* N_1 *such that for any* $n \geq N_1$, $1 \leq k \leq K$,

$$T_{kn} \geq n\lambda_k/2$$

holds on A_δ. *In particular,*

$$N_1 = \max\left(\frac{2(K-1)}{\lambda_{\min}} n_0', D_2^4\left[\log D_2^4 + \frac{1}{2}\left(\ell_{K,\delta} + 1 + 7\cdot 10^{-9}\right)\right]^2\right), \tag{17}$$

where $D_2 = 4c(2K-1)/\lambda_{\min}^2$, *c is defined by* (6), *and*

$$n_0' = \max(K(K+1), n_1, n_2),$$

$$n_1 = \left(\frac{\lambda_{\min} n_1'}{2}\right)^2 \quad , \quad n_2 = \left(\frac{\lambda_{\min} n_2'}{2}\right)^2,$$

$$n_1' = \max_{1\leq k\leq K} \frac{2{a_k'}^2}{\lambda_{\min}}\left[4\log a_k' + 1 + \ell_{K,\delta}\right] \quad , \quad n_2' = \frac{(4c)^2}{\lambda_{\min}^3}\left[4\log\frac{\sqrt{8}c}{\lambda_{\min}} + 1 + \ell_{K,\delta}\right].$$

For the proof we need the following technical lemma that gives a bound on the point when for $a > 0$ the function $at^{1/2} + b$ overtakes $\log t$.

Lemma 7. *Let $a > 0$. For any $t \geq (2/a)^2 \left[\log((2/a)^2) - b\right]^2$, $at^{1/2} + b > \log t$.*

The proof of this lemma can be found in Appendix B (Proposition 6).

Proof (Lemma 6). Due to the forced selection of the options built into the algorithm, $T_{kn} \geq \sqrt{n}$ holds for $n \geq K(K+1)$. The proof of this statement is somewhat technical and is moved into the Appendix A (Lemma 11). By Lemma 7, for

$$n \geq \max_{1 \leq k \leq K} {a'_k}^4 \left[4 \log a'_k + 1 + \ell_{K,\delta}\right]^2 = n_1,$$

(5) holds with $p = 1/2$, $\beta = 1$ for each k. Hence, we can apply Lemma 3 and Remark 1 following it with $n_0 = \max(K(K+1), n_1)$ and $f_n = n^{1/2}\,(\geq 2)$, and get that

$$\left|\hat{\lambda}_{kn} - \lambda_k\right| \leq c \sqrt{\frac{\log(\delta^{-1}_{n^{1/2}})}{n^{1/2}}} \tag{18}$$

on A_δ for $n \geq n_0$, $1 \leq k \leq K$, and $c \geq \sqrt{8}$ as defined by (6). By Lemma 7 again, for

$$n \geq 4 \left(\frac{2c}{\lambda_{\min}}\right)^4 \left[4 \log \frac{\sqrt{8}c}{\lambda_{\min}} + 1 + \ell_{K,\delta}\right]^2 = n_2,$$

(16) holds with $p = 1/2$, $\beta = 1$. Now, we can apply Corollary 1 and Remark 3 following it on A_δ with $n'_0 = \max(n_0, n_2) = \max(K(K+1), n_1, n_2)$ and $f_n = n^{1/2}\,(\geq 1)$, and get that on A_δ for $n \geq 0$, $1 \leq k \leq K$,

$$T_{kn} \geq n\lambda_k - (K-1)\max(n'_0, 1 + H_n(\delta)),$$

where

$$H_n(\delta) = D_1 n^{3/4} \sqrt{\log(\delta_n^{-1})}$$

and $D_1 = c\rho$. Hence, $T_{kn} \geq n\lambda_k/2$ by the time when $n \geq 2n'_0(K-1)/\lambda_{\min}$ and $n \geq 2(K-1)(1 + H_n(\delta))/\lambda_{\min}$. These two constrains are satisfied when $n \geq N_1$, where N_1 is defined as in equation (17); the first one is obvious, the second one follows from Proposition 7 in Appendix C. □

With the help of this result we can get better bounds on T_{kn}, resulting in our first main result:

Theorem 2. *Let $0 < \delta \leq 1$ be arbitrary. Then there exists a positive real number D_3 such that for any $n \geq 0$, $1 \leq k \leq K$,*

$$-(K-1)\max(N_1, 1 + G_n(\delta)) \leq T_{kn} - n\lambda_k \leq \max(N_1, 1 + G_n(\delta))$$

holds on A_δ, where

$$G_n(\delta) = D_3 \sqrt{n \log(\delta_n^{-1})}.$$

In particular, $D_3 = c\rho\sqrt{2/\lambda_{\min}}$*,* c *is defined by* (6)*, and* N_1 *is defined in Lemma 6.*

The theorem shows that asymptotically the GAFS-MAX algorithm behaves the same way as an optimal allocation rule that knows the variances. It also shows that the deviation of the proportion of choices of any option from the optimal value decays as $\tilde{O}(1/\sqrt{n})$.

For the proof we need the counterpart of Lemma 7 for linear functions. The proof is in Appendix B (Proposition 4). For a real number a, let a^+ denote its positive part: $a^+ = \max(a, 0)$.

Lemma 8. *Let* $a > 0$*. For any* $t \geq (2/a)[\log(1/a) - b]^+$*,* $at + b > \log t$*.*

Proof (Theorem 2). The proof is almost identical to that of Lemma 6. The difference is that now we start with a better lower bound on T_{kn}. In particular, by Lemma 6, $T_{kn} \geq n\lambda_k/2 \geq n\lambda_{\min}/2$ holds on A_δ for $n \geq N_1$. Note that, using the bounds on K, a'_k, c, $\lambda_{\min}$, and $\ell_{K,\delta}$, we have that

$$N_1 \geq \frac{2n'_0}{\lambda_{\min}} \geq \frac{\lambda_{\min}}{2} \max{}^2(n'_1, n'_2) \geq \max(n'_1, n'_2).$$

By Lemma 8, for

$$n \geq \max_{1\leq k\leq K} \frac{2{a'_k}^2}{\lambda_{\min}} \left[4 \log a'_k + 1 + \ell_{K,\delta}\right] = n'_1,$$

(5) holds with $p = 1$, $\beta = \lambda_{\min}/2$ for each k. Hence, we can apply Lemma 3 and Remark 1 following it with $(n_0 =) \max(N_1, n'_1) = N_1$ and $f_n = n\lambda_{\min}/2\,(\geq 2)$, and get that

$$\left|\hat{\lambda}_{kn} - \lambda_k\right| \leq c\sqrt{\frac{2\log(\delta^{-1}_{n\lambda_{\min}/2})}{n\lambda_{\min}}} \tag{19}$$

on A_δ for $n \geq N_1$, $1 \leq k \leq K$, and $c \geq \sqrt{8}$ as defined by (6). By Lemma 8 again, for

$$n \geq \frac{(4c)^2}{\lambda^3_{\min}} \left[4 \log \frac{\sqrt{8}c}{\lambda_{\min}} + 1 + \ell_{K,\delta}\right] = n'_2,$$

(16) holds with $p = 1$, $\beta = \lambda_{\min}/2$. Now, we can apply Corollary 1 and Remark 3 following it on A_δ with $(n_0 =) \max(N_1, n'_2) = N_1$ and $f_n = n\lambda_{\min}/2\,(\geq 1)$, and get that on A_δ for $n \geq 0$, $1 \leq k \leq K$,

$$-(K-1)\max(N_1, 1 + G_n(\delta)) \leq T_{kn} - n\lambda_k \leq \max(N_1, 1 + G_n(\delta)),$$

where

$$G_n(\delta) = D_3\sqrt{n \log(\delta_n^{-1})}$$

and $D_3 = c\rho\sqrt{2/\lambda_{\min}}$. □

This result yields a bound on the expected value of $\mathbb{E}\left[T_{kn}\right]$:

Theorem 3. *Let $N_2 = \sup_{0<\delta\le 1} N_1/\ell_{K,\delta}^2$, where N_1 is defined in Lemma 6. Then, $N_2 < \infty$ and there exists an index N_3 that depends only on N_2, $1/D_3$, and $\log K$ polynomially, such that for any k and $n \ge N_3$,*

$$\mathbb{E}\left[T_{kn}\right] \le n\lambda_k + D_3\sqrt{n(1+\log(4Kn(n+1)))} + 2. \tag{20}$$

PROOF. Recalling the definition of N_1 and that $\ell_{K,\delta} \ge \log 8$, we can easily see that $N_2 < \infty$. Note that N_2 does not depend on δ, and $N_1 \le N_2\ell_{K,\delta}^2 \le N_2\log^2(\delta_n^{-1})$ holds for any n and $0 < \delta \le 1$. Fix $0 < \delta \le 1$. If $n \ge N_1^2/(D_3^2\log(\delta_n^{-1}))$, then $1 + G_n(\delta) \ge N_1$, thus it follows from Theorem 2 that for such n,

$$\mathbb{P}\left(\frac{T_{kn} - n\lambda_k - 1}{D_3 n^{1/2}} > \sqrt{\log(\delta_n^{-1})}\right) \le \delta,$$

where we used $\mathbb{P}\left(A_\delta\right) \ge 1-\delta$. Let $Z = (T_{kn} - n\lambda_k - 1)/(D_3 n^{1/2})$ and $t = \sqrt{\log(\delta_n^{-1})}$. The above inequality is equivalent to

$$\mathbb{P}\left(Z > t\right) \le 4Kn(n+1)\, e^{-t^2}.$$

By the constraint that connects n and δ, this inequality holds for any pair (n,t) that satisfy

$$n \ge N_2^2\log^3(\delta_n^{-1})/D_3^2 = N_2^2 t^6/D_3^2,$$

that is, for any (n,t) such that

$$t \le (nD_3^2/N_2^2)^{1/6}.$$

Also, since $Z \le n^{1/2}/D_3$ is always true, $\mathbb{P}\left(Z > t\right) = 0$ holds for $t \ge n^{1/2}/D_3$. We need the following technical lemma, a variant of which can be found, e.g., as Exercise 12.1 in [7]:

Lemma 9. *Let $C > 1$, $c > 0$, $0 < a \le b$. Assume that the random variable Z satisfies $\mathbb{P}\left(Z > t\right) \le C\exp(-ct^2)$ for any $t \le a$ and $\mathbb{P}\left(Z > t\right) = 0$ for any $t \ge b$. Then*

$$\mathbb{E}\left[Z\right] \le \sqrt{(1+\log C)/c + Cb^2e^{-ca^2}}. \tag{21}$$

PROOF. By the monotonicity of $\mathbb{P}\left(Z > t\right) \le 1$, for any $u > 0$,

$$\begin{aligned}
\mathbb{E}\left[Z^2\right] &= \int_0^\infty \mathbb{P}\left(Z^2 > t\right) dt = \int_0^u + \int_u^{a^2} + \int_{a^2}^{b^2} + \int_{b^2}^\infty \\
&\le u + \left(\int_u^{a^2} Ce^{-ct}\, dt\right)^+ + \int_{a^2}^{b^2} \mathbb{P}\left(Z > a\right) dt + 0 \\
&\le u + \frac{C}{c}\left(e^{-cu} - e^{-ca^2}\right)^+ + (b^2 - a^2)Ce^{-ca^2}.
\end{aligned}$$

This gives

$$\mathbb{E}\left[Z^2\right] \le \frac{\log C + (1 - Ce^{-ca^2})^+}{c} + (b^2 - a^2)Ce^{-ca^2} \le \frac{1+\log C}{c} + Cb^2e^{-ca^2}$$

with the choice $u = (\log C)/c$. Now,

$$\mathbb{E}[Z] \leq \sqrt{\mathbb{E}[Z^2]} \leq \sqrt{\frac{1+\log C}{c} + Cb^2 e^{-ca^2}}. \qquad \square$$

Applying Lemma 9 with $a = (nD_3^2/N_2^2)^{1/6}$, $b = n^{1/2}/D_3$, $C = 4Kn(n+1)$, and $c = 1$,

$$\mathbb{E}[Z] \leq \sqrt{1 + \log(4Kn(n+1)) + 4Kn^2(n+1)e^{-(nD_3^2/N_2^2)^{1/3}}/D_3^2}.$$

Thus

$$\begin{aligned}\mathbb{E}[T_{kn}] &\leq 1 + n\lambda_k \\ &\quad + \sqrt{D_3^2 n(1+\log(4Kn(n+1))) + 4Kn^3(n+1)e^{-(nD_3^2/N_2^2)^{1/3}}}.\end{aligned}$$

Equation (20) then follows by straightforward algebra. $\square$

4.3. Bounding the loss: proof of Theorem 1

In order to develop a bound on the loss L_{kn} we need Wald's (second) identity:

Lemma 10 (Wald's Identity, Theorem 13.2.14 of [2]). *Let $\{\mathcal{F}_t\}$ be a filtration and let Y_t be an $\mathcal{F}_t$-adapted sequence of i.i.d. random variables. Assume that $\mathcal{F}_t$ and $\sigma(\{\, Y_s : s \geq t+1\,\})$ are independent and T is a stopping time w.r.t. $\mathcal{F}_t$ with a finite expected value: $\mathbb{E}[T] < +\infty$. Consider the partial sums $S_n = Y_1 + \ldots + Y_n$, $n \geq 1$. If $\mathbb{E}\left[Y_1^2\right] < +\infty$ then*

$$\mathbb{E}\left[(S_T - T\mathbb{E}[Y_1])^2\right] = \mathrm{Var}[Y_1]\,\mathbb{E}[T]. \tag{22}$$

Now, we can prove Theorem 1.

Proof (Theorem 1). Let $S_{kn} = \sum_{t=1}^n X_{kt}$,

$$\hat{L}_{kn} = \frac{S_{k,T_{kn}} - T_{kn}\mu_k}{T_{kn}},$$

$G'_n(\delta) = (K-1)\max(N_1, 1 + G_n(\delta))$ and

$$G''_n = D_3\sqrt{n(1+\log(4Kn(n+1)))} + 2.$$

Note that by Theorem 2,

$$\mathbb{P}(T_{kn} < n\lambda_k - G'_n(\delta)) \leq \delta \tag{23}$$

holds for any $n \geq 0$ and $0 < \delta \leq 1$. Then, for any $0 < \delta \leq 1$,

$$\begin{aligned} L_{kn} &= \mathbb{E}\left[\hat{L}_{kn}^2\right] = \mathbb{E}\left[\hat{L}_{kn}^2 \mathbb{I}_{\{\, T_{kn} \geq n\lambda_k - G'_n(\delta)\,\}}\right] + \mathbb{E}\left[\hat{L}_{kn}^2 \mathbb{I}_{\{\, T_{kn} < n\lambda_k - G'_n(\delta)\,\}}\right] \\ &\leq \frac{\mathbb{E}\left[(S_{k,T_{kn}} - T_{kn}\mu_k)^2\right]}{(n\lambda_k - G'_n(\delta))^2} + R_k^2\,\mathbb{P}(T_{kn} < n\lambda_k - G'_n(\delta)). \end{aligned}$$

Using Lemma 10 and then (20) of Theorem 3 for the first term, for $n \geq N_3$,

$$\mathbb{E}\left[(S_{k,T_{kn}} - T_{kn}\mu_k)^2\right] = \sigma_k^2 \mathbb{E}\left[T_{kn}\right] \leq \sigma_k^2(n\lambda_k + G''_n),$$

and thus

$$\begin{aligned} \frac{\mathbb{E}\left[(S_{k,T_{kn}} - T_{kn}\mu_k)^2\right]}{(n\lambda_k - G'_n(\delta))^2} &\leq \frac{\sigma_k^2(n\lambda_k + G''_n)}{(n\lambda_k - G'_n(\delta))^2} \\ &= \frac{\sigma_k^2}{n\lambda_k}\frac{1}{(1 - G'_n(\delta)/(n\lambda_k))^2} + \frac{\sigma_k^2 G''_n}{(n\lambda_k - G'_n(\delta))^2}, \end{aligned}$$

while, by (23), the second term is bounded above by $R_k^2\delta$.

Now choose $\delta = n^{-3/2}$. Then, recalling the definition of $G'_n(\delta)$, $G_n(\delta)$, δ_n, $\ell_{K,\delta}$, and that $N_1 \leq N_2\ell_{K,\delta}^2$, we have $G'_n(n^{-3/2}) = O(\sqrt{n \log n})$, thus for n sufficiently large, $G'_n(n^{-3/2})/(n\lambda_k) \leq 1/2$. Therefore, for such large n, using $1/(1-x) \leq 1 + 2x$ for $0 \leq x \leq 1/2$, we get,

$$L_{kn} \leq \frac{\sigma_k^2}{n\lambda_k}\left(1 + 2\frac{G'_n(n^{-3/2})}{n\lambda_k}\right)^2 + \frac{\sigma_k^2 G''_n}{(n\lambda_k - G'_n(n^{-3/2}))^2} + R_k^2 n^{-3/2},$$

which gives

$$L_{kn} \leq \frac{\sigma_k^2}{n\lambda_k} + \tilde{O}(n^{-3/2}) = \frac{\Sigma^2}{n} + \tilde{O}(n^{-3/2}) = L_n^* + \tilde{O}(n^{-3/2}).$$

Taking the maximum with respect to k yields the desired result. □

Let us now comment on the case when for some options $\lambda_k = 0$. Such options are chosen in the optimal allocation exactly once. Algorithm GAFS-MAX will select such options $\sqrt{n}$-times in n-steps since the estimated variance will be zero. Hence, we will have $T_{kn} \leq T_{kn}^* + O(\sqrt{n})$. Clearly, the loss for such an option will be zero. Further, since options with $\sigma_k^2 = 0$ are pulled only $O(\sqrt{n})$-times, they can not significantly influence the number of times the other options are chosen. Hence, the results go through if we replace $\min_k \lambda_k$ with $\min_{k:\lambda_k \neq 0} \lambda_k$.

5. Illustration

The purpose of this section is to illustrate the theory by means of some computer experiments. One particular goal of the experiments was to verify the excess loss rate obtained in the previous section. Another goal was to compare the adaptive strategy with a non-adaptive strategy.

5.1. Experimental Setup

Here we illustrate the behavior of the algorithm in a simple problem with $K = 2$, when the random responses are modeled as Bernoulli random variables for each of the options. In order to estimate the expected squared loss between the true mean and the estimated mean we repeat the experiment 100,000 times, then take the average. The error bars shown on the graphs show the standard deviations of these averages. The algorithms compared are GAFS-MAX (the algorithm studied here), GFSP-MAX (the algorithm described in the introduction that works in phases), and "UNIF", the uniform allocation rule.

The exploration parameter α of GAFS-MAX was set to 1. We have run experiments to test the sensitivity of GAFS-MAX to the choice of the value of α. The experiments showed that GAFS-MAX is largely insensitive to this choice unless a too small value is selected for α in which case if the algorithm underestimates the variance of some of the options then it will take a large amount of time for it to recover. In the limit, when $\alpha = 0$ (no forcing), as discussed before, the allocation ratios of the algorithm may fail to converge to the optimal ratios. For example, if initially the variance estimate for one of the options is zero (which happens with positive probability when the responses have Bernoulli distributions), that option will never be selected any more, in which case the loss L_n will fail to converge to zero, i.e., such an algorithm will suffer $\Omega(1)$ excess loss. Hence, for simplicity we sticked to $\alpha = 1$ which was proved to be an acceptable value for the problems tested (for details, see [10]).

The algorithm UNIF works in a round-robin fashion (i.e., tests the options systematically). In the case of GFSP-MAX, after the initialization phase where each option is observed twice, the phase length of the k^{th} phase is set to $K + k$. This ensures that at the end of the k^{th} phase, every options is explored at least k times, while the total number of observations is $2K+Kk+(1+2+\ldots+k) \approx k^2/2$. Thus, by time t each option is explored at least approximately $\sqrt{2t}$ times, which makes the comparison with GAFS-MAX running with $\alpha = 1$ fair, given our experience that normally the difference between the performance of GAFS-MAX running with $\alpha = \sqrt{2}$ and $\alpha = 1$ is small.

5.2. Results

In order for an adaptive algorithm to have any advantage the two options have to have different variances. For this purpose we chose $p_1 = 0.8$, $p_2 = 0.9$ so that $\lambda_1 = 0.64$ and $\lambda_2 = 0.36$.

Figure 1 shows the rescaled excess loss, $n^{3/2}(L_n - L_n^*)$, for the three algorithms. We see that the rescaled excess losses of the adaptive algorithms stay bounded, while the rescaled loss of the uniform sampling strategy grows as $\sqrt{n}$. It is remarkable that the limit of the rescaled loss seems to be a small number, showing the efficiency of the algorithm.[3] Incidentally, in this case the incremental method (GAFS-MAX) performs better than the algorithm that works in phases (GFSP-MAX), although their performance is quite similar and this does not need to hold generally.

Note that this example shows that the uniform allocation initially performs better than the adaptive rules. This is because the adaptive algorithms need to

[3] As far as we could measure it, the graph flattens out when considering larger sample sizes (see also Figure 3). However, it might still be the case that the rescaled loss goes to zero at a rate of (e.g.) $1/\log n$.

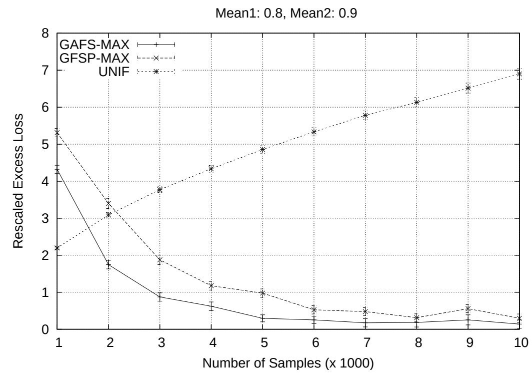

Figure 1: The rescaled excess loss, $n^{3/2}(L_n - L_n^*)$, against the number of observations. The losses were measured when the sample size is an integer multiple of 1000.

get a good estimate of the statistics before they can start exploiting. The cross-over point happens at ca. 1,700 for GAFS-MAX, while it happens just after 2,000 observations for GFSP-MAX. By selecting a larger exploration parameter α the cross-over point could be moved to the left.

From the point of view of an adaptive algorithm the most difficult case is when all variances are small (cf. Lemma 3), but (λ_k) is significantly different from the uniform distribution. This is explored further in Figure 2, which plots the cross-over point for a series of single-parameter problems. The parameter, κ, determines the means: $p_1(\kappa) = \kappa$, $p_2(\kappa) = \kappa/2$. This makes the allocation proportions non-uniform, but roughly constant (for small κ these proportions are $4/5$ and $1/5$, respectively for the first and the second option). This way we can measure the influence of the variance on the difficulty of competing with the uniform allocation. The figure also shows the curve a/κ^2 for an appropriate positive constant a. Based on the graph, we may conclude that the difficulty of catching up with the uniform allocation rule increases roughly proportionally to $\sigma_{\max}^{-2} = \max(\sigma_1, \ldots, \sigma_K)^{-2}$. This is very well expected: Indeed, as both variances become small, it becomes increasingly harder to figure out their relative sizes. Note, however, that as the variances become smaller the overall precision improves for the same sample size (independently of what algorithm is used).

Figure 3 shows the rescaled allocation ratio deviations, $\sqrt{n}|T_{kn}/n - \lambda_k|$, for $k = 1$. If we disregard logarithmic terms, our theory predicts that these rescaled deviations should stay bounded for the adaptive algorithms. The figure indeed supports this. The behavior of the curve for the uniform sampling method is markedly different: due to the mismatch of the allocation ratios, this curve grows as $\sqrt{n}$. Note that the variance of the algorithm that uses phases is much larger than the variance of the incremental algorithm. This is because the incremental algorithm is faster to update its statistics.

In conclusions, the experiments show that our method indeed performs better than a non-adaptive solution. In fact, depending on the problem parameters

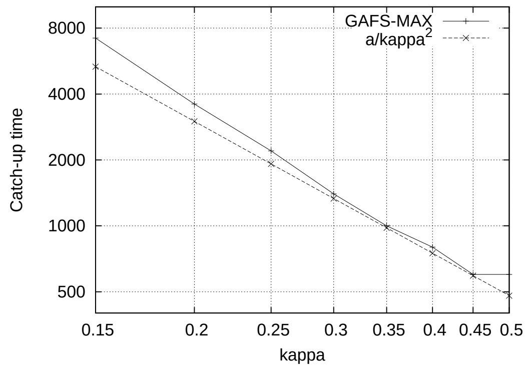

Figure 2: The number of observations required to perform better than uniform sampling for a range of problems parameterized with a single parameter $0 < \kappa < 1$. The solid line shows the data measured for GAFS-MAX, while the dashed curve shows a/κ^2 for an appropriate value of a. Note the log-log scale. For more information see the text.

the performance difference between the adaptive and non-adaptive algorithms can be large. Further, our experiments verified that the allocation strategy found by our algorithm converges to the optimal allocation strategy at the rate predicted by the theory (apart from logarithmic factors).

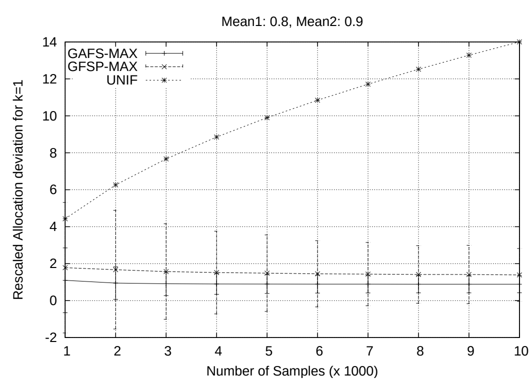

Figure 3: The rescaled allocation deviations, $\sqrt{n}|T_{kn}/n - \lambda_k|$, for $k = 1$ against the number of observations.

6. Related Work

As mentioned earlier, this work is closely related to active learning in a regression setting (e.g., [4]) and to optimal experimental design (OED) [9]. The connection is that the model studied here can be viewed as a linear regression problem over the finite domain $\mathcal{X} = \{e_1, \ldots, e_K\}$, where e_i is the i^{th} unit vector of the K-dimensional Euclidean space: $e_i \in \mathbb{R}^K$, $e_i = (0, \ldots, 1, \ldots, 0)$, i.e., all components of e_i are zero except its i^{th} component whose value is one. Indeed, with this definition of $\mathcal{X}$, the response to the choice of option k can be written as the linear regression model $Y = \theta^T e_k + W(e_k)$, where θ is the unknown parameter and $W(e_k)$ is a zero mean random variable.

Interestingly, in the rather extensive active learning and OED literature, to the best of our knowledge, no one looked into the problem of learning in a situation where the noise in the dependent variable varies in space, i.e., when the noise is *heteroscedastic*. Although the rate of convergence of an adaptive method that pays attention to heteroscedasticity will not be better than that of the one that does not, an adaptive algorithm's finite-time performance may be significantly better than that of underlying a non-adaptive algorithm. This has been demonstrated convincingly in a forthcoming related paper where Etore and Jourdain studied the utility of adapting the sampling proportions in stratified sampling [8].[4] Interestingly, this application is very closely related to the problem studied here. The only difference is that the loss is measured by taking the weighted sum of the losses of the individual prediction errors with some fix set of weights that sum to one. With obvious changes, the algorithm presented here can be modified to work in this setting and the analysis carries through with almost no changes (for details, see [10]). The algorithm studied in [8] is the phase-based algorithm. The results are weak consistency results, i.e., no bounds are given on the excess loss. In fact, the only condition the authors pose on the proportion of forced selections is that this proportion should go to zero such that the total number of forced selections for any option goes to infinity.

7. Conclusions and Future Work

When finite-sample performance is important, one may exploit heteroscedasticity to allocate more observations to parts of the input space where the variance is larger. In this paper we designed an algorithm for such a situation and showed that the excess loss of this algorithm compared with that of an optimal rule, that knows the variances, decays as $\tilde{O}(n^{-3/2})$. It remains an open question if this is the optimal rate for the class of problems studied here. Although currently we do not have a proof, the following heuristic argument provides some support for this conjecture: Take any algorithm $\mathcal{A}$ and let λ'_{kn} be the allocation ratios achieved when running $\mathcal{A}$. The loss of $\mathcal{A}$ is roughly $L_n(\mathcal{A}) \approx \max_k(\sigma_k^2/\lambda'_{kn})/n$, while the optimal loss is $L_n^* = \max_k(\sigma_k^2/\lambda_k)/n$. Let $\varepsilon_{kn} = \lambda'_{kn} - \lambda_k$. Assuming that the maximum above is taken at the same $k = k^*$ in both losses (which is reasonable if $\varepsilon_{kn} \to 0$ as $n \to \infty$), $\mathcal{E}_n(\mathcal{A}) = L_n(\mathcal{A}) - L_n^* \approx \sigma_{k^*}^2 |\varepsilon_{k^*n}|/(\lambda_{k^*}(\lambda_{k^*} + \varepsilon_{k^*n})n)$.

[4] In fact, we have learned about this paper just at the time when we submitted the first version of this paper. An earlier paper studying the same problem and achieving somewhat weaker results is due to Peierls and Yahav [14].

Since one expects that independently of the algorithm chosen $\varepsilon_{k^*n} = \Omega(1/\sqrt{n})$ (i.e., all λ'_{kn} converge at the parametric rate), we get $\mathcal{E}_n(\mathcal{A}) = \Omega(n^{-3/2})$.

Our analysis can probably be improved, e.g. in terms of the dependence of our constants in our bounds on $\lambda_{\min}^{-1}$. However, we think that the proof technique developed here might be useful to analyze similar algorithms in related contexts, such as active learning with heteroscedastic noise in related parametric and non-parametric models. We are currently investigating such models.

An interesting question is whether the algorithm and the results can be extended to other losses. An important class of losses can be expressed in terms of the expectation of a convex function. In this case, the natural algorithm is to minimize the empirical estimate of the loss based on the sample average (in fact, our algorithm is a special case of this general scheme). We believe that if the convex function is strictly convex in a small neighborhood of zero then a second order approximation to it can be used to prove results entirely similar to the ones obtained here. The analysis of the case when strict convexity does not hold looks more challenging and will probably require ideas that go beyond those presented in this paper.

Acknowledgements

This research was funded in part by the National Science and Engineering Research Council (NSERC) of Canada, iCore, the Alberta Ingenuity Fund, the Mobile Innovation Center of Hungary, the Hungarian Scientific Research Fund and the Hungarian National Office for Research and Technology (OTKA-NKTH CNK 77782), the PASCAL2 Network of Excellence (EC grant no. 216886), and by the Hungarian Academy of Sciences (Bolyai Fellowship for András Antos).

A. Forced selection lemma

Lemma 11. *For $1 \le k \le K$, $n \ge K(K+1)$*

$$T_{kn} \ge \sqrt{n} \tag{24}$$

holds.

Proof. For a positive integer l, let $C_l = \{ (l-1)^2+1, (l-1)^2+2, \dots, l^2 \}$, a partition of $\{1, 2, \dots\}$. Observe that if (24) holds for some $n = n' \in C_l$, then it holds also for any $n = n'' \in C_l$, $n'' > n'$, since $T_{k,n''} \ge T_{k,n'} \ge \sqrt{n'} > l-1$ which implies $T_{k,n''} \ge l \ge \sqrt{n''}$. Thus, it is enough to prove (24) for $n = K(K+1)$ and then for $n = l^2+1$, $l = K+1, K+2, \dots$.

By a careful analysis of the algorithm, we see that only forced selection steps happen till $n = K(K+2)$ in a uniform manner, during which each option is selected $K+2$ times. This implies that $T_{k,K(K+1)} = K+1 > \sqrt{K(K+1)}$ and that $T_{k,(K+1)^2+1} \ge T_{k,K(K+2)} = K+2 > \sqrt{(K+1)^2+1}$, i.e., (24) holds for $n = K(K+1)$ and $(K+1)^2+1$, for all k. Now we use induction for l. Assume that (24) holds for all k, for some $n = (l-1)^2+1$ ($l \ge K+2$), i.e.,

$T_{k,(l-1)^2+1} \geq \sqrt{(l-1)^2+1} > l-1$ implying $T_{k,(l-1)^2+1} \geq l$. Now at times $(l-1)^2+2, \ldots, l^2+1$ (which total up to $|C_l| = 2l-1 (\geq 2K-3)$ steps), one of those arms for which $T_{k,(l-1)^2+1} = l$ holds is forced to be selected exactly once. Hence each such arm is selected at least once in this phase, assuring $T_{k,l^2+1} \geq l+1 > \sqrt{l^2+1}$ for all k, i.e., (24) holds for $n = l^2+1$. □

B. Some elementary comparison lemmata

The purpose of this section is to provide upper bounds on the solutions of equations of the form

$$\log(t) = at^p + b, \tag{25}$$

where $a, p, t > 0$.

Let

$$\begin{aligned} \ell(t) &= \log t, \\ q(t) &= at^p + b, \text{ and} \\ t_0 &= (pa)^{-1/p}. \end{aligned}$$

Here t_0 is the point where ℓ and q have the same growth rate, i.e., where $\ell'(t_0) = q'(t_0)$. Note that $q'(t)/l'(t) = apt^p$ is strictly monotone increasing in t, hence for $t > t_0$, $q'(t) > \ell'(t)$. Hence, if $q(t_0) > \ell(t_0)$ then (25) has no solutions on $[t_0, \infty)$. Now observe that it also holds that $q'(t) < \ell'(t)$ when $t < t_0$. Hence, if $q(t_0) > \ell(t_0)$ then (25) has no solutions on $(0, t_0]$ since ℓ decreases faster than q as we move from t_0 towards zero. Now, consider the case when $q(t_0) \leq \ell(t_0)$. Since for $t > t_0$, $q'(t) > \ell'(t)$ and $q(t)/\ell(t) \stackrel{t\to\infty}{\to} \infty$, (25) will have exactly one solution in $[t_0, \infty)$.

These findings are summarized in the next proposition:

Proposition 1. *Consider $t_0 = (pa)^{-1/p}$, $q(t) = at^p + b$, and $\ell(t) = \log t$, where $a, p, t > 0$. Then $q(t_0) \leq \ell(t_0)$ is a sufficient and necessary condition for the existence of a solution to $q(t) = \ell(t)$. Further, when $q(t_0) \leq \ell(t_0)$ then there is exactly one solution on $[t_0, \infty)$.*

Remark 4. Note that $q(t_0) \leq \ell(t_0)$ is equivalent to $1 + bp \leq -\log(pa)$, which is thus a sufficient and necessary condition for the existence of a solution to $q(t) = \ell(t)$.

In the sequel we will derive upper bounds on the solutions of (25) by picking some t^* such that $q(t^*) \geq \ell(t^*)$ and $q'(t^*) \geq \ell'(t^*)$. In doing so we will first consider the homogeneous version of (25),

$$\log u = a'u^p. \tag{26}$$

The following proposition gives the link between the solutions of the homogeneous and inhomogeneous equations.

Proposition 2. *Any solution of (25) can be obtained by solving (26) with $a' = ae^{pb}$ and then using $t = e^b u$ and vice versa. Further, if u^* is an upper bound on the solutions of (26) then $t^* = e^b u^*$ is an upper bound on the solutions of (25).*

Now, let us consider the linear case, i.e., when $p = 1$.

Proposition 3. *Let $q(t) = at$, $\ell(t) = \log t$, where $a > 0$. Let $t^* = (2/a)\log(1/a)$. Then for any positive t such that $t \geq t^*$, $q(t) > \ell(t)$ holds.*

Proof. We may assume that $\log(1/a) \geq 1$, otherwise by Remark 4, $q(t) = \ell(t)$ does not have a solution and the statement follows trivially. It suffices to show that $\ell(t^*) < q(t^*)$ and $\ell'(t) < q'(t)$ holds for $t \geq t^*$. The second inequality follows from $\log(1/a) \geq 1$ and the monotonicity of $q'(t)/l'(t)$, while the first follows from the inequality $\log(z^2) < z$ $(z > 0)$. □

Proposition 4. *Let $q(t) = at + b$, $\ell(t) = \log t$, where $a > 0$. Let $t^* = (2/a)[\log(1/a) - b]$. Then for any positive t such that $t \geq t^*$, $q(t) > \ell(t)$ holds.*

Proof. The statement follows immediately from Propositions 2 and 3. □

Now, let us turn to the case when $p = 1/2$.

Proposition 5. *Let $q(t) = at^{1/2}$, $\ell(t) = \log t$, where $a > 0$. Let $t^* = (2/a)^2 \log^2((2/a)^2)$. Then for any positive t such that $t \geq t^*$, $q(t) > \ell(t)$ holds.*

Proof. We may assume that $\log(2/a) \geq 1$, otherwise by Remark 4, $q(t) = \ell(t)$ does not have a solution and the statement follows trivially. It suffices to show that $\ell(t^*) < q(t^*)$ and $\ell'(t) < q'(t)$ holds for $t \geq t^*$. The second inequality follows from $\log(2/a) \geq 1$ and the monotonicity of $q'(t)/l'(t)$, while the first follows from the inequality $\log(z^2) < z$ $(z > 0)$. □

Proposition 6. *Let $q(t) = at^{1/2} + b$, $\ell(t) = \log t$, where $a > 0$. Let $t^* = (2/a)^2 \left[\log((2/a)^2) - b\right]^2$. Then for any positive t such that $t \geq t^*$, $q(t) > \ell(t)$ holds.*

Proof. The statement follows immediately from Propositions 2 and 5. □

C. Technical calculation for Lemma 6

Proposition 7. *$n \geq N_1$ implies $n \geq 2(K-1)(1 + H_n(\delta))/\lambda_{\min}$.*

Proof. Recalling that $H_n(\delta) = D_1 n^{3/4}\sqrt{\log(\delta_n^{-1})}$, $D_1 = c\rho$, $\rho = (1 + 2/\lambda_{\min})$, we would like to have

$$n \geq 2(K-1)\left(1 + D_1 n^{3/4}\sqrt{\log(\delta_n^{-1})}\right)/\lambda_{\min},$$

or equivalently, if both $n \geq 2(K-1)/\lambda_{\min}$ and

$$\left(\frac{\lambda_{\min} n^{1/4}}{2D_1(K-1)} - \frac{1}{D_1 n^{3/4}}\right)^2 \geq \log(\delta_n^{-1}). \tag{27}$$

The first inequality follows immediately from $n \geq N_1$. Introducing $D_2' = 4D_1(K-1)/\lambda_{\min} = 4c\rho(K-1)/\lambda_{\min}$, we have

$$D_2' = 4c(2K - 2 + K\lambda_{\min} - \lambda_{\min})/\lambda_{\min}^2 \leq 4c(2K-1)/\lambda_{\min}^2 = D_2.$$

Using (2), (27) follows from

$$\frac{4\sqrt{n}}{D_2'^2} + \frac{1}{D_1^2 n^{3/2}} - \frac{4}{D_1 D_2' \sqrt{n}} = \left(\frac{2n^{1/4}}{D_2'} - \frac{1}{D_1 n^{3/4}}\right)^2 \geq 2\log n + 1 + \ell_{K,\delta},$$

that follows from

$$\frac{2\sqrt{n}}{D_2'^2} - \frac{2}{D_1 D_2' \sqrt{n}} - \frac{1}{2}(1 + \ell_{K,\delta}) \geq \log n.$$

Whenever $n \geq N_1 > D_2^4[\log D_2^4 + (\ell_{K,\delta}+1)/2]^2$, then

$$\frac{2}{D_1 D_2' \sqrt{n}} < \frac{2}{D_1 D_2' D_2^2 [\log D_2^4 + (\ell_{K,\delta}+1)/2]},$$

which is, after substituting D_1, D_2, D_2' and using $1/\lambda_{\min} \geq K \geq 2$, $\delta \leq 1$, $c \geq \sqrt{8}$, bounded above by

$$\frac{1}{450(128)^2[8\log 6 + 13\log 8 + 1]} < 10^{-8}/3.$$

Thus, it is enough to have

$$\frac{2\sqrt{n}}{D_2'^2} - \frac{1}{2}(1 + 7\cdot 10^{-9} + \ell_{K,\delta}) \geq \log n.$$

This is implied by Lemma 7 and $n \geq D_2'^4[\log(D_2'^4) + (\ell_{K,\delta} + 1 + 7\cdot 10^{-9})/2]^2$, which follows from $n \geq N_1$ and $D_2 \geq D_2'$. □

References

[1] A. Antos, V. Grover, and Cs. Szepesvári. Active learning in multi-armed bandits. In *Proc. of the 19th International Conference on Algorithmic Learning Theory*, volume LNCS/LNAI 5254, pages 287–302. Springer-Verlag, 2008.

[2] K.B. Athreya and S.N. Lahiri. *Measure Theory and Probability Theory*. Springer, 2006.

[3] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.

[4] R. Castro, R. Willett, and R.D. Nowak. Faster rates in regression via active learning. In *Advances in Neural Information Processing Systems 18 (NIPS-05)*, pages 179–186. MIT Press, 2006.

[5] P. Chaudhuri and P. Mykland. On efficient designing of nonlinear experiments. *Statistica Sinica*, 5:421–440, 1995.

[6] D. Cohn, Z. Ghahramani, and M. Jordan. Active learning with statistical models. *Journal of Artificial Intelligence Research*, 4:129–145, 1996.

[7] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Applications of Mathematics: Stochastic Modelling and Applied Probability. Springer-Verlag New York, 1996.

[8] P. Etore and B. Jourdain. Adaptive optimal allocation in stratified sampling methods. *Methodology and Computing in Applied Probability*, 2008. http://www.citebase.org/abstract?id=oai:arXiv.org:0711.4514.

[9] V. V. Fedorov. *Theory of Optimal Experiments*. Academic Press, 1972.

[10] V. Grover. Active learning and its application to heteroscedastic problems. Master's thesis, Department of Computing Science, University of Alberta, 2009.

[11] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[12] T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

[13] T.L. Lai and S. Yakowitz. Machine learning and nonparametric bandit theory. *IEEE Transactions on Automatic Control*, 40:1199–1209, 1995.

[14] R.F. Peierls and J.A. Yahav. Adaptive allocation for importance sampling. In J. van Ryzin, editor, *Adaptive Statistical Procedures and Related Topics*, volume 8, pages 204–218. Institute of Mathematical Statistics, 1986.