Thresholding Bandit for Dose-ranging: The Impact of Monotonicity

Aurélien Garivier, Pierre Ménard, Laurent Rossi

Thresholding Bandit for Dose-ranging: The Impact of Monotonicity

Aurélien Garivier AURELIEN.GARIVIER@MATH.UNIV-TOULOUSE.FR

Pierre Ménard PIERRE.MENARD@MATH.UNIV-TOULOUSE.FR

Laurent Rossi LAURENT.ROSSI@MATH.UNIV-TOULOUSE.FR

Institut de Mathématiques de Toulouse; UMR5219
Université de Toulouse; CNRS
UPS IMT, F-31062 Toulouse Cedex 9, France

Abstract

We analyze the sample complexity of the thresholding bandit problem, with and without the assumption that the mean values of the arms are increasing. In each case, we provide a lower bound valid for any risk δ and any δ-correct algorithm; in addition, we propose an algorithm whose sample complexity is of the same order of magnitude for small risks. This work is motivated by phase 1 clinical trials, a practically important setting where the arm means are increasing by nature, and where no satisfactory solution is available so far.

Keywords: thresholding bandits, multi-armed bandits, best arm identification, unimodal regression.

1. Introduction

The phase 1 of clinical trials is devoted to the testing of a drug on healthy volunteers for *dose-ranging*. The first goal is to determine the maximum tolerable dose (MTD), that is the maximum amount of the drug that can be given to a person before adverse effects become intolerable or dangerous. A tolerance level is chosen, and the trials aim at identifying quickly which is the dose entailing the toxicity coming closest to this level. Classical approaches are based on dose escalation, and the most well-known is the "traditional 3+3 Design": see Le Tourneau and Siu (2009); Genovese et al. (2013) for and references therein for an introduction.

We propose in this article a complexity analysis for a simple model of phase 1 trials, which captures the essence of this problem. We assume that the possible doses are $x_1 < \ldots < x_K$, for some positive integer K. The patients are treated in sequential order, and identified by their rank. When the patient number t is assigned a dose x_k, we observe a measure of toxicity $X_{k,t}$ which is assumed to be an independent random variable. Its distribution ν_k characterizes the toxicity level of dose x_k. We treat here mostly the case of Gaussian laws with known variance and unknown mean, but some results can easily be extended to other one-parameter exponential families such as Bernoulli distributions. The goal of the experiment is to identify as soon as possible the dose x_k which has the toxicity level μ_k closest to the target admissibility level S, with a controlled risk δ to make an error.

Content. This setting is an instance of the *thresholding bandit problem*: we refer to Locatelli et al. (2016) for an important contribution and a nice introduction. In this work, we focus on identifying

the *exact sample complexity* of the problem: we want to understand precisely (with the correct multiplicative constant) how many samples are necessary to take a decision at risk δ. We prove a lower bound which holds for all possible algorithms, and we propose an algorithm which matches this bound asymptotically when the risk δ tends to 0.

But the classical thresholding bandit problem does not catch a key feature of phase 1 clinical trials: the fact that the toxicity is *known in hindsight* to be *increasing* with the assigned dose. In other words, we investigate how many samples can be spared by algorithms using the fact that $\mu_1 < \mu_2 < \ldots < \mu_K$. Under this assumption, we prove another lower bound on the sample complexity, and provide an algorithm matching it.

Organization. These lower bounds are presented in Section 2. We compare the complexities of the non-monotonous case versus the increasing case. This comparison is particularly simple and enlightening when $K = 2$, a setting often referred to as *A/B testing*. We discuss this case in Section 2.1, which furnishes a gentle introduction to the general case. We present in Section 3 an algorithm and show that it is asymptotically optimal when the risk δ goes to 0. The implementation of this algorithm requires, in the increasing case, an involved optimization which relies on constraint sub-gradient ascent and *unimodal regression*: this is detailed in Section 3.1. Section 3.2 shows the results of some numerical experiments for different strategies with high level of risk that complement the theoretical results. Section 4 summarizes further possible developments, and preceeds most of the technical proofs which are given in appendix.

1.1. Notation and Setting

For $K \geqslant 2$, we consider a Gaussian bandit model $\big(\mathcal{N}(\mu_1, 1), \ldots, \mathcal{N}(\mu_K, 1)\big)$, which we unambiguously refer to by the vector of means $\boldsymbol{\mu} = \big(\mu_1, \ldots, \mu_K\big)$. Let $\mathbb{P}_{\boldsymbol{\mu}}$ and $\mathbb{E}_{\boldsymbol{\mu}}$ be respectively the probability and the expectation under the Gaussian bandit model $\boldsymbol{\mu}$. A threshold $S \in \mathbb{R} \cup \{\pm\infty\}$ is given, and we denote by $a^*_{\boldsymbol{\mu}} \in \operatorname{argmin}_{1 \leqslant a \leqslant K} |\mu_a - S|$ any optimal arm.

Let $\mathcal{M}$ be the set of Gaussian bandit models with an unique optimal arm and $\mathcal{I} = \{\boldsymbol{\mu} \in \mathcal{M} : \mu_1 < ... < \mu_K\}$ be the subset of models with increasing means.

Definition of a δ-correct algorithm. A risk level $\delta \in (0, 1)$ is fixed. At each step $t \in \mathbb{N}^*$ an agent chooses an arm $A_t \in \{1, \ldots, K\}$ and receives an independent reward $Y_t \sim \mathcal{N}(\mu_{A_t}, 1)$. Let $\mathcal{F}_t = \sigma(A_1, Y_1, \ldots, A_t, Y_t)$ be the information available to the player at step t. Her goal is to identify the optimal arm $a^*_{\boldsymbol{\mu}}$ while minimizing the number of draws τ. To this aim, the agent needs:

- a **sampling rule** $(A_t)_{t \geqslant 1}$, where A_t is $\mathcal{F}_{t-1}$-measurable,
- a **stopping rule** τ_δ, which is a stopping time with respect to the filtration $(\mathcal{F}_t)_{t \geqslant 1}$,
- a $\mathcal{F}_{\tau_\delta}$-measurable **decision rule** $\widehat{a}_{\tau_\delta}$.

For any setting $\mathcal{S} \in \{\mathcal{M}, \mathcal{I}\}$ (the non-monotonous or the increasing case), an algorithm is said to be δ-correct on $\mathcal{S}$ if for all $\boldsymbol{\mu} \in \mathcal{S}$ it holds that $\mathbb{P}_{\boldsymbol{\mu}}(\tau_\delta < +\infty) = 1$ and $\mathbb{P}_{\boldsymbol{\mu}}(\widehat{a}_{\tau_\delta} \neq a^*_{\boldsymbol{\mu}}) \leqslant \delta$.

2. Lower Bounds

For $\mathcal{S} \in \{\mathcal{M}, \mathcal{I}\}$, we define the set of *alternative bandit problems* of the bandit problem $\boldsymbol{\mu} \in \mathcal{M}$ by

$$\mathcal{A}lt(\boldsymbol{\mu}, \mathcal{S}) := \{\boldsymbol{\lambda} \in \mathcal{S} : a^*_{\boldsymbol{\lambda}} \neq a^*_{\boldsymbol{\mu}}\}\,, \tag{1}$$

and the simplex of dimension $K-1$ by Σ_K. The first result of this paper is a lower bound on the sample complexity of the thresholding bandit problem, which we show in the sequel to be tight when δ is small enough.

Theorem 1 *Let $\mathcal{S} \in \{\mathcal{M}, \mathcal{I}\}$ and $\delta \in (0, 1/2]$. For all δ-correct algorithm on $\mathcal{S}$ and for all bandit models $\boldsymbol{\mu} \in \mathcal{S}$,*

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta] \geqslant T^*_{\mathcal{S}}(\boldsymbol{\mu}) \operatorname{kl}(\delta, 1-\delta)\,, \tag{2}$$

where the characteristic time $T^*_{\mathcal{S}}(\boldsymbol{\mu})$ is given by

$$T^*_{\mathcal{S}}(\boldsymbol{\mu})^{-1} = \sup_{\omega \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2}\,. \tag{3}$$

In particular, this implies that

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \geqslant T^*_{\mathcal{S}}(\boldsymbol{\mu})\,.$$

This result is an adaptation of Theorem 1 in Garivier and Kaufmann (Jun. 2016), and can be proved along the same lines. In fact, our result is a generalization: the classical Best Arm Identification problem is a particular case of our setting with the choices $\mathcal{S} = \mathcal{M}$ and $S = +\infty$. As in Garivier and Kaufmann (Jun. 2016), one can prove that the supremum and the infimum are reached at a unique value, and in the sequel we denote by $\omega^*(\boldsymbol{\mu})$ the optimal weights

$$\omega^*(\boldsymbol{\mu}) := \operatorname*{argmax}_{\omega \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2}\,. \tag{4}$$

2.1. The Two-armed Bandit Case

As a warm-up, we treat in the section the case $K = 2$. Here (only), one can find an explicit formula for the characteristic times.

Proposition 2 *When $K = 2$,*

$$T^*_{\mathcal{I}}(\boldsymbol{\mu})^{-1} = \frac{(2S - \mu_1 - \mu_2)^2}{8} \quad \textit{and} \quad T^*_{\mathcal{M}}(\boldsymbol{\mu})^{-1} = \frac{\min\big((2S - \mu_1 - \mu_2)^2, (\mu_1 - \mu_2)^2\big)}{8}\,. \tag{5}$$

The proof of Proposition 2 is given in Appendix A.1.

Note that for both alternative sets the optimal weights defined in Equation (4) are uniform: $\omega^* = [1/2, 1/2]$. If the alternative set is $\mathcal{I}$, the optimal alternative, i.e. the element $\boldsymbol{\lambda}$ of $\overline{\mathcal{A}lt(\boldsymbol{\mu}, \mathcal{I})}$ (the closure of $\mathcal{A}lt(\boldsymbol{\mu}, \mathcal{I})$) which reaches the infimum in (3) for the optimal weights ω^*, is $\boldsymbol{\lambda} = [S - (\mu_2 - \mu_1)/2, S + (\mu_2 - \mu_1)/2]$. In words, in the optimal alternative the arms are translated in such a way that the mean of the two mean values is moved to the threshold S. If the alternative set is $\mathcal{M}$ and $\boldsymbol{\mu} \in \mathcal{I}$, the optimal alternatives can be of two different forms. If the threshold is between the two mean values, then the optimal alternative is the same as for the increasing case. Otherwise, the optimal alternative is identical to the one of Best Arm Identification (see Garivier and Kaufmann

(Jun. 2016)): $\boldsymbol{\lambda} = [(\mu_1 + \mu_2)/2, (\mu_1 + \mu_2)/2]$. Thus, if $\mu_1 \leqslant S \leqslant \mu_2$, the two characteristic times coincide, as can be seen in Figure 1 .

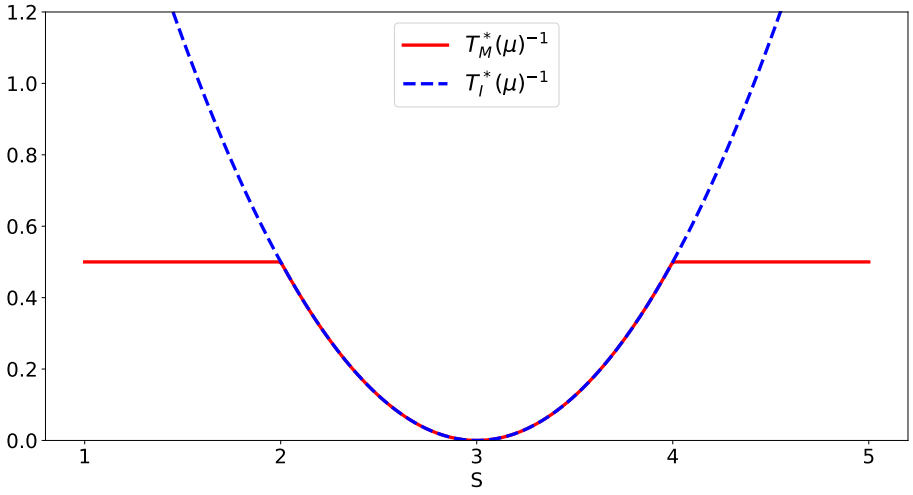

Figure 1: Inverse of the characteristic times as a function of the threshold S, for $\boldsymbol{\mu} = [2, 4]$. Solid red: general thresholding case ($\mathcal{S} = \mathcal{M}$). Dotted blue: increasing case ($\mathcal{S} = \mathcal{I}$).

2.2. On the Characteristic Time and the Optimal Proportions

We now illustrate, compare and comment the different complexities for a general bandit model $\boldsymbol{\mu} \in \mathcal{I}$ with $K \geqslant 2$. Since $\mathcal{I} \subset \mathcal{M}$, it holds trivially that $T^*_{\mathcal{I}}(\boldsymbol{\mu}) \leqslant T^*_{\mathcal{M}}(\boldsymbol{\mu})$. The difference $T^*_{\mathcal{M}}(\boldsymbol{\mu}) - T^*_{\mathcal{I}}(\boldsymbol{\mu})$ is almost everywhere positive, and can be very large. Both $T^*_{\mathcal{I}}(\boldsymbol{\mu})$ and $T^*_{\mathcal{M}}(\boldsymbol{\mu})$ tend to $+\infty$ as S tends to middle of two consecutive arms.

In the non-monotonous case $\mathcal{S} = \mathcal{M}$, there are two types of optimal alternatives (as in Section 2.1). Indeed, the proof of Lemma 5 in Appendix A shows that the best alternative takes one of the two following forms. Either the optimal arm μ^*_a and its challenger μ_b are moved to a pondered mean (by the optimal weights ω^*) of the two arms (just like in the Best Arm Identification problem), leading to a constant $(\mu_{a^*} - \mu_b)^2$ in Equation (19). Or, as in the increasing case $\mathcal{S} = \mathcal{I}$ (see the proof of Proposition 2), both arms μ_{a^*} and μ_b are translated in the same direction, leading to the constant $(2S - \mu_{a^*} - \mu_b)^2$. Figure 2 summarizes the different possibilities on a simple example with $K = 4$ arms, for different values of the threshold S. According to the value of S, the best alternative is shown in the second plot from the top.

On the structure of the optimal weights in the increasing case. In the increasing case $\mathcal{S} = \mathcal{I}$, it is particularly remarkable that *the optimal weights $\omega^*(\boldsymbol{\mu})$ put mass only on the optimal arm and its two closest arms*. This property is shown in the proof of Proposition 4, see Appendix A.2. This strongly contrasts with the non-monotonous case, as illustrated at the bottom of Figure 2.

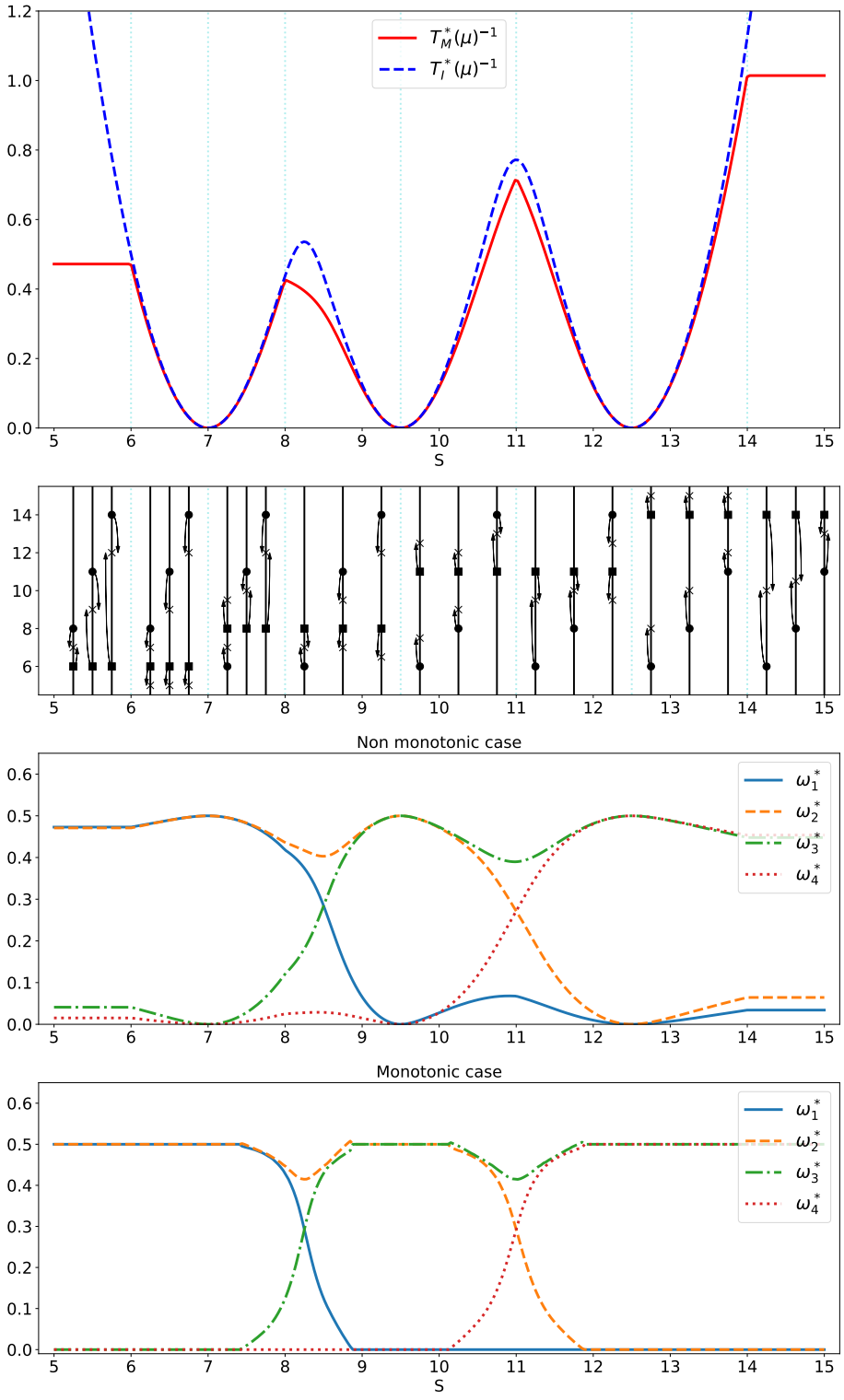

Figure 2: The complexity terms in the bandit model $\boldsymbol{\mu} = (6, 8, 11, 14)$. *Top:* inverse of the characteristic time as a function of the threshold S; red solid line: non-monotonous case $\mathcal{S} = \mathcal{M}$; blue dotted line: increasing case $\mathcal{S} = \mathcal{I}$. *Middle:* how to move the means to get from the initial bandit model to the optimal alternative in $\mathcal{M}$. *Bottom:* the optimal weights in function of the threshold S.

3. An Asymptotically Optimal Algorithm

We present in this section an asymptotically optimal algorithm inspired by the *Direct-tracking* procedure of Garivier and Kaufmann (Jun. 2016). At any time $t \geqslant 1$ let $h(t) = (\sqrt{t} - K/2)_+$ and $U_t = \{a : N_a(t) < h(t)\}$ be the set of "abnormally rarely sampled" arms.

Algorithm 1: Algorithm for the general case (Direct-tracking).

Sampling rule

$$A_{t+1} \in \begin{cases} \operatorname*{argmin}_{a \in U_t} N_a(t) \text{ if } U_t \neq \emptyset & (\textit{forced exploration}) \\ \operatorname*{argmax}_{1 \leqslant a \leqslant K} t\, w_a^*(\widehat{\boldsymbol{\mu}}(t)) - N_a(t) & (\textit{direct tracking}) \end{cases}$$

Stopping rule

$$\tau_\delta = \inf \left\{ t \in \mathbb{N}^* : \widehat{\boldsymbol{\mu}}(t) \in \mathcal{M} \text{ and } \inf_{\boldsymbol{\lambda} \in \mathcal{A}lt\left(\widehat{\boldsymbol{\mu}}(t), \mathcal{S}\right)} \sum_{a=1}^{K} N_a(t) \frac{(\widehat{\mu}_a(t) - \lambda_a)^2}{2} > \beta(t, \delta) \right\} . \quad (6)$$

Decision rule

$$\widehat{a}_{\tau_\delta} \in \operatorname*{argmin}_{1 \leqslant a \leqslant K} \left| \widehat{\mu}_a(\tau_\delta) - S \right| .$$

When $L := \mathrm{Card}\{\mathrm{argmin}_{1 \leqslant a \leqslant K} |\widehat{\mu}_a(t) - S|\} > 1$, we adopt the convention that $T^*_{\mathcal{S}}(\widehat{\boldsymbol{\mu}}(t)^{-1} = 0$ and

$$w_a^*(\widehat{\boldsymbol{\mu}}(t)) = \begin{cases} 1/L & \text{if } a \in \mathrm{argmin}_{1 \leqslant a \leqslant K} |\widehat{\mu}_a(t) - S| , \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 3 (Asymptotic optimality) *For $\mathcal{S} \in \{\mathcal{I}, \mathcal{M}\}$, for the constant C defined in Equation (20) of Appendix B and for $\beta(t, \delta) = \log(tC/\delta) + (3K + 2)\log\log(tC/\delta)$, Algorithm 1 is δ-correct on $\mathcal{S}$ and asymptotically optimal, i.e.*

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \leqslant T^*_{\mathcal{S}}(\boldsymbol{\mu}) . \quad (7)$$

The analysis of Algorithm 1 is the same in both the increasing case $\mathcal{S} = \mathcal{I}$ and the non-monotonous case $\mathcal{S} = \mathcal{M}$. However, the practical implementations are quite specific to each case, and we detail them in the next section.

3.1. On the Implementation of Algorithm 1

The implementation of Algorithm 1 requires to compute efficiently the optimal weights $w^*(\boldsymbol{\mu})$ given by Equation (4). For the non-monotonous case $\mathcal{S} = \mathcal{M}$, one can follow the lines of Garivier and Kaufmann (Jun. 2016), Section 2.2 and replace their Lemma 3 by Lemma 5 below.

In the increasing case $\mathcal{S} = \mathcal{I}$, however, implementing the algorithm is more involved. Let $\mathcal{I}_b := \{\boldsymbol{\lambda} \in \mathcal{I}, a^*_{\boldsymbol{\lambda}} = b\}$. Noting that the function

$$F : w \mapsto \inf_{\boldsymbol{\lambda} \in Alt(\boldsymbol{\mu}, \mathcal{I})} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2} = \min_{b \neq a^*_{\boldsymbol{\mu}}} \inf_{\boldsymbol{\lambda} \in \mathcal{I}_b} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \tag{8}$$

is concave (since it is the infimum of linear functions), one may access to its maximum by a sub-gradient ascent on the simplex Σ_K (see e.g. Boyd et al. (2003)). Let $\overline{\mathcal{I}}_b$ denote the closure of $\mathcal{I}_b$, and let

$$\boldsymbol{\lambda}^b := \operatorname*{argmin}_{\lambda \in \overline{\mathcal{I}}_b} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \tag{9}$$

be the argument of the second infimum in Equation (8). The sub-gradient of F at ω is

$$\partial F(\omega) = \operatorname*{Conv}_{b \in B_{Opt}} \left[\frac{(\mu_a - \lambda^b_a)^2}{2} \right]_{a \in \{1, \cdots, K\}},$$

where Conv denotes the convex hull operator and where B_{Opt} is the set of points that reach the minimum in (8). Thus, performing the sub-gradient ascent simply requires to solve efficiently the minimization program (9). It appears that this problem boils down to *unimodal regression* (a problem closely related to isotonic regression, see R. E. Barlow (1973) and T. Robertson (1988)). Indeed, we can write

$$\begin{aligned}\{\boldsymbol{\lambda} \in \mathcal{I} : a^*_{\boldsymbol{\lambda}} = b\} = \{\boldsymbol{\lambda} \in \mathcal{M} :& \\ \lambda_1 < \ldots < \lambda_{b-1} < \min(\lambda_b, 2S - \lambda_b) \leqslant \max(\lambda_b, 2S - \lambda_b) < \lambda_{b+1} < \ldots < \lambda_K\}&.\end{aligned}$$

Assume that $\mu_b \leqslant S$ (the other case is similar). Then $\lambda^b_b < S$, since λ_b and $2S - \lambda_b$ play a symmetric role in the constraints. Thus, in this case, one may only consider the set

$$\begin{aligned}\{\boldsymbol{\lambda} \in \mathcal{M} :& \lambda_1 < \ldots < \lambda_{b-1} < \lambda_b, \\ & 2S - \lambda_K < \ldots < 2S - \lambda_{b+1} < \lambda_b, \\ & \lambda_b \leqslant S\}.\end{aligned}$$

Let $\boldsymbol{\lambda}'$ be the new variables such that

$$\lambda'_a = \begin{cases} \lambda_a & \text{if } 1 \leqslant a \leqslant b, \\ 2S - \lambda_a & \text{else}. \end{cases} \tag{10}$$

Then $\boldsymbol{\lambda}^{b\prime}$ is the solution of

$$\boldsymbol{\lambda}^{b\prime} = \operatorname*{argmin}_{\substack{\lambda'_1 \leqslant \ldots \leqslant \lambda'_b \\ \lambda'_K \leqslant \ldots \leqslant \lambda'_b \\ \lambda'_b \leqslant S}} \sum_{a=1}^{K} \omega_a \frac{(\mu'_a - \lambda'_a)^2}{2}. \tag{11}$$

Thanks to Lemma 11 in Appendix C, it is simply related to

$$\widehat{\boldsymbol{\lambda}}^b := \operatorname*{argmin}_{\substack{\lambda'_1 \leqslant \ldots \leqslant \lambda'_b \\ \lambda'_K \leqslant \ldots \leqslant \lambda'_b}} \sum_{a=1}^{K} \omega_a \frac{(\mu'_a - \lambda'_a)^2}{2},$$

	BC-$\mathcal{M}$	R-$\mathcal{M}$	DT-$\mathcal{M}$	$T^*_{\mathcal{M}}(\boldsymbol{\mu})\log\frac{1}{\delta}$	BC-$\mathcal{I}$	R-$\mathcal{I}$	DT-$\mathcal{I}$	$T^*_{\mathcal{I}}(\boldsymbol{\mu})\log\frac{1}{\delta}$
$\boldsymbol{\mu}_1$, S_1	3913	3609	4119	2033	483	494	611	247
$\boldsymbol{\mu}_2$, S_2	3064	3164	3098	1861	2959	2906	3072	1842

Table 1: Monte-Carlo estimation (with 10000 repetitions) of the expected number of draws $\mathbb{E}[\tau_\delta]$ for Algorithm 1 and Best Challenger Algorithm in the increasing and non-monotonous cases. Two thresholding bandit problems are considered: $\boldsymbol{\mu}_1 = [0.5,\ 1.1,\ 1.2,\ 1.3,\ 1.4,\ 5]$ with $S_1 = 1$, and $\boldsymbol{\mu}_2 = [1,\ 2,\ 2.5]$ with $S_2 = 1.55$. The target risk is $\delta = 0.1$ (it is approximately reached in the first scenario, while in the second the frequency of errors is of order 1%).

the unimodal regression of $\boldsymbol{\mu}'$ with weights ω and with a mode located at b. And it is efficiently computed via isotonic regressions (e.g. Frisén (1986), Geng and Shi (1990), Mureika et al. (1992)) with a computational complexity proportional to the number of arms K. From $\widehat{\boldsymbol{\lambda}}^b$, one can go back to $\boldsymbol{\lambda}^b$ by reversing Equation (10). Since we need to compute $\boldsymbol{\lambda}^b$ for each $b \neq a^*$, the overall cost of an evaluation of the sub-gradient is proportional to K^2.

3.2. Numerical Experiments

Table 1 presents the results of a numerical experiment of an increasing thresholding bandit. In addition to Algorithm 1 (DT), we tried the Best Challenger (BC) algorithm with the finely tuned stopping rule given by (6). We also tried the Racing algorithm (R), with the elimination criterion of (6). For a description of all those algorithms, see Garivier and Kaufmann (Jun. 2016) and references therein. It appears that the exploration function β prescribed in Theorem 3 is overly pessimistic. On the basis of our experiments, we recommend the use of $\beta(t,\delta) = \log\big(\big(\log(t)+1\big)/\delta\big)$ instead. It does, experimentally, satisfy the δ-correctness property. For each algorithm, the final letter in Table 1 indicates whether the algorithm is aware ($\mathcal{I}$) or not ($\mathcal{M}$) that the means are increasing.

We consider two frameworks: in the first one, knowing that the means are increasing provides much information and gives a substantial edge: it permits to spare a large portion of the trials for the same level of risk. In the second, the complexities of the non-monotonous setting is very close to that of the increasing setting. We chose a value of the risk δ which is relatively high (10%), in order to illustrate that in this regime, the most important feature for efficiency is a finely tuned stopping rule. This shows that, even without an optimal sampling strategy, the stopping rule of (6) is a key feature of an efficient procedure. When the risk goes down to 0, however, optimality really requires a sampling rule which respects the proportions of Equation (4), as shown by Theorem 3.

4. Conclusion

We provided a tight complexity analysis of the *dose-ranging* problem considered as a thresholding bandit problem with, and without, the assumption that the means of the arms are increasing. We proposed a lower bound on the expected number of draws for any δ-correct algorithm and adapted the *Direct-Tracking* algorithm to asymptotically reach this lower bound. We also compared the complexities of the non-monotonous and the increasing cases, both in theory and on an illustrative example. We showed in Section 3.1 how to compute the optimal weights thanks to a sub-gradient ascent in the increasing case, a new and non-trivial task relying on unimodal isotonic regression.

In order to complement the theoretical results, we presented some numerical experiments involving different strategies in a regime of high risk.

As a possibility of improvement, we mention the possible use of the unimodal regression algorithm of Stout (2000) in order to compute directly (8) with a complexity of order $O\big(K\big)$. We treated here mostly the case of Gaussian distributions with known variance. While the general form of the lower bound may easily be extended to other settings (including Bernoulli observations), the computation of the complexity terms is more involved and requires further investigations (in particular due to heteroscedasticity effects). The asymptotic optimality of Algorithm 1, however, can be extended directly. It remains important but very challenging tasks to make a tight analysis for moderate values of δ, to measure precisely the sub-optimality of Racing and Best Challenger strategies, and to develop a more simple and yet asymptotically optimal algorithm.

Acknowledgments

The authors thank Wouter Koolen for suggesting a key ingredient in the proof of Proposition 4. The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grants ANR-13-BS01-0005 (project SPADRO) and ANR-13-CORD-0020 (project ALICIA).

Appendix A. Proofs for the Lower Bounds

A.1. Proof of Proposition 2: Expression of the Complexities in the Two-armed Case

We treat here only the first equality: the expression of the characteristic time $T^*_{\mathcal{M}}(\boldsymbol{\mu})$, if the alternative set is $\mathcal{M}$, is a consequence of Lemma 5. Let $\boldsymbol{\mu} \in \mathcal{I}$ and suppose, without loss of generality, that arm 2 is optimal. Let $m = (\mu_1 + \mu_2)/2$ be the mean of two arms and $\Delta = \mu_2 - \mu_1$ be the gap. Noting that

$$\{\text{arm 1 is optimal}\} \Leftrightarrow m > S \quad \text{and} \quad \{\text{arm 2 optimal}\} \Leftrightarrow m < S\,,$$

we obtain

$$\begin{aligned} T^*_{\mathcal{I}}(\mu)^{-1} &= \sup_{\omega\in[0,1]} \inf_{\left\{\mu'_1<\mu'_2,\, |S-\mu'_1|<|S-\mu'_2|\right\}} \frac{\omega}{2}(\mu_1-\mu'_1)^2 + \frac{1-\omega}{2}(\mu_2-\mu'_2)^2 \\ &= \sup_{\omega\in[0,1]} \underbrace{\inf_{\{\Delta'>0,\, m'>S\}} \frac{\omega}{2}\big(m-m'-(\Delta-\Delta')/2\big)^2 + \frac{1-\omega}{2}\big(m-m'+(\Delta-\Delta')/2\big)^2}_{:=A(\omega)}\,, \end{aligned}$$

where $m' = (\mu_1 + \mu_2)/2$ and $\Delta' = \mu'_2 - \mu'_1$. Writing $\chi = S - m$, easy computations lead to

$$A(\omega) = \begin{cases} 2\omega(1-\omega)\chi^2 & \text{if } \Delta + 2(2\omega-1)\chi > 0\,, \\ \big(\chi^2 + (\Delta/2)^2 + (2\omega-1)\chi\Delta\big)/2 & \text{else.} \end{cases}$$

Thus, since the maximum of A is attained at $\omega = 1/2$, we just proved that $T^*_{\mathcal{I}}(\mu)^{-1} = \chi^2/2$.

A.2. Expression of the complexity in the increasing case

Fix $\boldsymbol{\mu} \in \mathcal{I}$ and let a^* be the optimal arm $a^* := a^*_{\boldsymbol{\mu}}$. Let $D^+(\theta, \widetilde{\omega})$ and $D^-(\theta, \widetilde{\omega})$ be two functions defined over $\mathbb{R} \times \Sigma_3$ by

$$D^+(\theta, \widetilde{\omega}) = \widetilde{\omega}_{-1} \frac{\big(\mu_{a^*-1} - \min(\mu_{a^*-1}, \theta)\big)^2}{2} + \widetilde{\omega}_0 \frac{(\mu_{a^*} - \theta)^2}{2} + \widetilde{\omega}_1 \frac{\big(\mu_{a^*+1} - (2S - \theta)\big)^2}{2} \quad (12)$$

$$D^-(\theta, \widetilde{\omega}) = \widetilde{\omega}_{-1} \frac{\big(\mu_{a^*-1} - (2S - \theta)\big)^2}{2} + \widetilde{\omega}_0 \frac{(\mu_{a^*} - \theta)^2}{2} + \widetilde{\omega}_1 \frac{\big(\mu_{a^*+1} - \max(\mu_{a^*+1}, \theta)\big)^2}{2} \quad (13)$$

with the convention $D^-(\cdot, \widetilde{\omega}) = +\infty$ if $a^* = 1$ and $D^+(\cdot, \widetilde{\omega}) = +\infty$ if $a^* = K$.

Proposition 4 *For all $\boldsymbol{\mu} \in \mathcal{I}$*

$$T^*_{\mathcal{I}}(\boldsymbol{\mu})^{-1} = \sup_{\widetilde{\omega} \in \Sigma_3} \min\left(\min_{\{2S - \mu_{a^*+1} \leqslant \theta \leqslant S\}} D^+(\theta, \widetilde{\omega}), \min_{\{S \leqslant \theta \leqslant 2S - \mu_{a^*-1}\}} D^-(\theta, \widetilde{\omega}) \right). \quad (14)$$

Proof We begin by proving that for all $\omega \in \Sigma_K$

$$\inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} = \min_{b \in \{a^*-1, a^*+1\}} \inf_{\{\boldsymbol{\lambda} \in \mathcal{I} \,:\, a^*_{\boldsymbol{\lambda}} = b\}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}. \quad (15)$$

Indeed, let $\boldsymbol{\lambda} \in \mathcal{I}$ such that $a^*_{\boldsymbol{\lambda}} \notin \{a^* - 1, a^* + 1\}$. Suppose for example that $a^*_{\boldsymbol{\lambda}} < a^* - 1$. Let $\boldsymbol{\lambda}^\alpha$ be the family of bandit problems defined for $\alpha \in [0, 1]$ by

$$\boldsymbol{\lambda}^\alpha = \alpha \boldsymbol{\lambda} + (1 - \alpha) \boldsymbol{\mu}\,.$$

For all $\alpha \in [0, 1]$, we have $\boldsymbol{\lambda} \in \mathcal{I}$. For $\boldsymbol{\nu} \in \mathcal{I}$ and $a \in \{0, .., K\}$, let $m_a(\boldsymbol{\nu}) = (\nu_a + \nu_{a+1})/2$ be the average of two consecutive means with the convention $m_0(\boldsymbol{\nu}) = -\infty$ and $m_K(\boldsymbol{\nu}) = +\infty$. As in the case of two arms we have that $a^*_{\boldsymbol{\nu}} = a$ is equivalent to $m_a(\boldsymbol{\nu}) > S$ and $m_a(\boldsymbol{\nu}) < S$. Therefore we have the following inequalities

$$\begin{aligned} &m_{a^*_{\boldsymbol{\lambda}}-1}(\boldsymbol{\mu}) < m_{a^*_{\boldsymbol{\lambda}}}(\boldsymbol{\mu}) \leqslant m_{a^*-2}(\boldsymbol{\mu}) < m_{a^*-1}(\boldsymbol{\mu}) < S < m_{a^*}(\boldsymbol{\mu}) \text{ and} \\ &m_{a^*_{\boldsymbol{\lambda}}-1}(\boldsymbol{\lambda}) < S < m_{a^*_{\boldsymbol{\lambda}}}(\boldsymbol{\lambda}) \leqslant m_{a^*-2}(\boldsymbol{\lambda}) < m_{a^*-1}(\boldsymbol{\lambda}) < m_{a^*}(\boldsymbol{\lambda})\,. \end{aligned}$$

Thus, by continuity of the applications $\alpha \mapsto m_a(\boldsymbol{\lambda}^\alpha)$ there exits $\alpha_0 \in (0, 1)$ such that

$$m_{a^*_{\boldsymbol{\lambda}}-1}(\boldsymbol{\lambda}^{\alpha_0}) < m_{a^*_{\boldsymbol{\lambda}}}(\boldsymbol{\lambda}^{\alpha_0}) \leqslant m_{a^*-2}(\boldsymbol{\lambda}^{\alpha_0}) < S < m_{a^*-1}(\boldsymbol{\lambda}^{\alpha_0}) < m_{a^*}(\boldsymbol{\lambda}^{\alpha_0})\,,$$

i.e. $a^*_{\boldsymbol{\lambda}^{\alpha_0}} = a^* - 1$. But $\alpha \mapsto \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda^\alpha_a)^2}{2}$ is an increasing function, and thus

$$\sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a^{\alpha_0})^2}{2} < \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}\,.$$

This holds for all λ, therefore

$$\inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \geqslant \min_{b \in \{a^*-1, a^*+1\}} \inf_{\{\boldsymbol{\lambda} \in \mathcal{I} \,:\, a^*_{\boldsymbol{\lambda}} = b\}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}\,.$$

The reverse inequality follows simply from the inclusion

$$\bigcup_{b\in\{a^*-1,a^*+1\}} \{\boldsymbol{\lambda}\in\mathcal{I} \,:\, a^*_{\boldsymbol{\lambda}}=b\} \subset \mathcal{A}lt(\boldsymbol{\mu},\mathcal{I})\,.$$

Fix $\omega\in\Sigma_K$ and let $\boldsymbol{\lambda}\in\mathcal{I}$ be such that, say, $a^*_{\boldsymbol{\lambda}}=a^*+1$ (the other case is similar). Then $\lambda_{a^*}\leqslant S$ and we can suppose, without loss of generality, that $\lambda_{a^*}\geqslant 2S-\mu_{a^*+1}$ since it holds

$$\inf_{\{\boldsymbol{\lambda}\in\mathcal{I}\,:\,a^*_{\boldsymbol{\lambda}}=a^*+1\}}\sum_{a=1}^K\omega_a\frac{(\mu_a-\lambda_a)^2}{2} = \inf_{\{\boldsymbol{\lambda}\in\mathcal{I}\,:\,a^*_{\boldsymbol{\lambda}}=a^*+1,\lambda_{a^*}\geqslant 2S-\mu_{a^*+1}\}}\sum_{a=1}^K\omega_a\frac{(\mu_a-\lambda_a)^2}{2}\,.$$

Let $\widetilde{\boldsymbol{\lambda}}$ be such that

$$\widetilde{\lambda}_a=\begin{cases}\mu_a & \text{if } a>a^*+1\,,\\ 2S-\lambda_{a^*} & \text{if } a=a^*+1\,,\\ \lambda_{a^*} & \text{if } a=a^*\,,\\ \min(\lambda_{a^*},\mu_a) & \text{if } a\leqslant a^*-1\,.\end{cases}$$

By construction we have $\widetilde{\lambda}\in\overline{\{\lambda\in\mathcal{I}:a^*_{\boldsymbol{\lambda}}=a^*+1\}}$. As $\lambda_{a^*+1}\leqslant 2S-\lambda_{a^*}$ and $\mu_{a^*+1}\geqslant 2S-\lambda_{a^*}$, we have

$$(\widetilde{\lambda}_{a^*+1}-\mu_{a^*+1})^2\leqslant(\lambda_{a^*+1}-\mu_{a^*+1})^2\,.$$

Similarly, for $a\leqslant a^*-1$ we have thanks to the fact that $\lambda_a\leqslant\lambda_{a^*}$ the inequality

$$(\widetilde{\lambda}_{a^*+1}-\mu_{a^*+1})^2\leqslant(\lambda_{a^*+1}-\mu_{a^*+1})^2\,.$$

Therefore

$$\sum_{a=1}^K\omega_a\frac{(\mu_a-\lambda_a)^2}{2}\geqslant\sum_{a=1}^K\omega_a\frac{(\mu_a-\widetilde{\lambda}_a)^2}{2}\,,$$

and we can rewrite the infimum as follows:

$$\begin{aligned}\inf_{\{\boldsymbol{\lambda}\in\mathcal{I}\,:\,a^*_{\boldsymbol{\lambda}}=a^*+1\}}\sum_{a=1}^K\omega_a\frac{(\mu_a-\lambda_a)^2}{2} &= \min_{2S-\mu_{a^*+1}\leqslant\theta\leqslant S}\sum_{a\leqslant a^*-1}\omega_a\frac{\left(\mu_a-\min\left(\theta,\mu_a\right)\right)^2}{2}\\ &\quad+\omega_{a^*}\frac{(\mu_{a^*}-\theta)^2}{2}+\omega_{a^*+1}\frac{(\mu_{a^*+1}-2S+\theta)^2}{2}\\ &= \min_{2S-\mu_{a^*+1}\leqslant\theta\leqslant S}\sum_{a<a^*-1}\omega_a\frac{\left(\mu_a-\min\left(\theta,\mu_a\right)\right)^2}{2}\\ &\quad+D^+(\theta,[\omega_{a^*-1},\omega_{a^*},\omega_{a^*+1}])\,.\end{aligned}\tag{16}$$

Similarly

$$\begin{aligned}\inf_{\{\boldsymbol{\lambda}\in\mathcal{I}\,:\,a^*_{\boldsymbol{\lambda}}=a^*-1\}}\sum_{a=1}^K\omega_a\frac{(\mu_a-\lambda_a)^2}{2} &= \min_{S\leqslant\theta\leqslant 2S-\mu_{a^*-1}}\sum_{a>a^*+1}\omega_a\frac{\left(\mu_a-\max\left(\theta,\mu_a\right)\right)^2}{2}\\ &\quad+D^-(\theta,[\omega_{a^*-1},\omega_{a^*},\omega_{a^*+1}])\,.\end{aligned}\tag{17}$$

Then, by noting that

$$\begin{aligned}
\left(\mu_a - \max\left(\theta, \mu_a\right)\right)^2 &\leqslant \left(\mu_{a^*+1} - \max\left(\theta, \mu_{a^*+1}\right)\right)^2 \qquad && \forall a \leqslant a^*+1 \\
\left(\mu_a - \min\left(\theta, \mu_a\right)\right)^2 &\leqslant \left(\mu_{a^*-1} - \min\left(\theta, \mu_{a^*-1}\right)\right)^2 \qquad && \forall a \leqslant a^*-1
\end{aligned}$$

and by using the new weights $\widetilde{\omega}$ defined by

$$\widetilde{\omega}_a = \begin{cases} \sum_{b \leqslant a^*-1} w_b & \text{if } a = a^*-1 \\ \omega_a & \text{if } a = a^* \\ \sum_{b \geqslant a^*+1} w_b & \text{if } a = a^*+1 \\ 0 & \text{else}\,, \end{cases}$$

we obtain thanks to Equation (15) and to the fact that $\widetilde{\omega}$ depends only on ω:

$$\inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \leqslant \min\left(\min_{\{2S - \mu_{a^*+1} \leqslant \theta \leqslant S\}} D^+(\theta, \widetilde{\omega}), \min_{\{S \leqslant \theta \leqslant 2S - \mu_{a^*-1}\}} D^-(\theta, \widetilde{\omega}) \right), \tag{18}$$

where we identified $\widetilde{\omega}$ to an element of Σ_3. Taking the supremum on each side of (18), one obtains:

$$T^*_{\mathcal{I}}(\boldsymbol{\mu})^{-1} \leqslant \sup_{\widetilde{\omega} \in \Sigma_3} \min\left(\min_{\{2S - \mu_{a^*+1} \leqslant \theta \leqslant S\}} D^+(\theta, \widetilde{\omega}), \min_{\{S \leqslant \theta \leqslant 2S - \mu_{a^*-1}\}} D^-(\theta, \widetilde{\omega}) \right).$$

In order to prove the reverse inequality and thus (14), we just need to use (17), (16) and restrict the weight ω to have a support included in $\{a^*-1, a^*, a^*+1\}$. ■

A.3. Expression of the Complexity in the Non-monotonous Case

Lemma 5 *For all $\boldsymbol{\mu} \in \mathcal{M}$ with the notation $a^*_{\boldsymbol{\mu}} = a^*$, all $\omega \in \Sigma_K$*

$$\inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{M})} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2} = \min_{b \neq a^*} \frac{\omega_{a^*}\omega_b}{2(\omega_{a^*} + \omega_b)} \min\left((\mu_{a^*} - \mu_b)^2, (2S - \mu_{a^*} - \mu_b)^2\right). \tag{19}$$

Proof Using that

$$\mathrm{Alt}(\boldsymbol{\mu}, \mathcal{M}) = \bigcup_{b \neq a^*} \left\{\boldsymbol{\lambda} \in \mathcal{M} : |\lambda_b - S| < |\lambda_{a^*} - S|\right\},$$

one has

$$\begin{aligned}
\inf_{\boldsymbol{\lambda} \in \mathcal{A}lt(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2} &= \min_{b \neq a^*} \inf_{|\lambda_b - S| < |\lambda_{a^*} - S|} \sum_{a=1}^{K} \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \\
&= \min_{b \neq a^*} \inf_{|\lambda_b - S| < |\lambda_{a^*} - S|} \omega_{a^*} \frac{(\mu_{a^*} - \lambda_{a^*})^2}{2} + \omega_b \frac{(\mu_b - \lambda_b)^2}{2}.
\end{aligned}$$

Since at the infimum $|\lambda_b - S| = |\lambda_{a^*} - S|$, denoting $x = \lambda_b - S$, we have $\lambda_{a^*} - S = x$ or $-x$. Therefore, one obtains

$$\inf_{|\lambda_b - S| < |\lambda_{a^*} - S|} \omega_{a^*} \frac{(\mu_{a^*} - \lambda_{a^*})^2}{2} + \omega_b \frac{(\mu_b - \lambda_b)^2}{2} = \min\Bigg(\inf_x \omega_a^* \frac{(\mu_{a^*} - S - x)^2}{2} + \omega_b \frac{(\mu_b - S - x)^2}{2}, \inf_x \omega_{a^*} \frac{(\mu_{a^*} - S + x)^2}{2} + \omega_b \frac{(\mu_b - S - x)^2}{2} \Bigg).$$

Noting that

$$\inf_x \omega_{a^*} \frac{(\mu_{a^*} - S - x)^2}{2} + \omega_b \frac{(\mu_b - S - x)^2}{2} = \frac{\omega_{a^*}\omega_b}{2(\omega_{a^*} + \omega_b)} (\mu_{a^*} - \mu_b)^2$$
$$\inf_x \omega_{a^*} \frac{(\mu_{a^*} - S - x)^2}{2} + \omega_b \frac{(\mu_b - S + x)^2}{2} = \frac{\omega_{a^*}\omega_b}{2(\omega_{a^*} + \omega_b)} (2S - \mu_{a^*} - \mu_b)^2$$

permits to conclude. ■

Appendix B. Correctness and Asymptotic Optimality of Algorithm 1

Proof (of Proposition 3) We follow and slightly adapt the proof of Theorem 14 of Garivier and Kaufmann (Jun. 2016). We fix a bandit problem $\boldsymbol{\mu} \in \mathcal{S}$ and the constant

$$C := e^{K+1} \left(\frac{2}{K}\right)^K \left(2(3K+2)\right)^{3K} \frac{4}{\log(3)}. \tag{20}$$

δ-correctness on $\mathcal{S}$

We will prove in the second part of proof that τ is almost surely finite, confer (23). Thus by definition of τ, we have

$$\mathbb{P}(\widehat{a}_\tau \neq a^*_{\boldsymbol{\mu}}) \leqslant \mathbb{P}\left(\exists t \in \mathbb{N}^*, \sum_{a=1}^K N_a(t) \frac{(\widehat{\mu}_a(t) - \mu_a)^2}{2} > \beta(t, \delta) \right), \tag{21}$$

where we used that $\boldsymbol{\mu} \in \mathcal{A}lt\big(\widehat{\boldsymbol{\mu}}(t), \mathcal{S}\big)$ since $\widehat{a}_\tau \neq a^*_{\boldsymbol{\mu}}$. Using the union bound then Theorem 10 (note that $\beta(t,\delta) \geqslant K+1$ thanks to the choice of C) we have

$$\begin{aligned}
\mathbb{P}(\widehat{a}_\tau \neq a^*_{\boldsymbol{\mu}}) &\leqslant \sum_{t=1}^{+\infty} \mathbb{P}\left(\sum_{a=1}^{K} N_a(t) \frac{(\widehat{\mu}_a(t) - \mu_a)^2}{2} > \beta(t,\delta) \right) \\
&\leqslant \sum_{t=1}^{+\infty} e^{K+1} \left(\frac{2}{K}\right)^K \Big(\beta(t,\delta)\big(\ln(t)\beta(t,\delta)+1\big)\Big)^K e^{-(t,\delta)} \\
&\leqslant e^{K+1} \left(\frac{2}{K}\right)^K \sum_{t=1}^{+\infty} \frac{\big(2(3K+2)\big)^{3K}}{\log(tC/\delta)^2} \frac{\delta}{tC} \\
&\leqslant e^{K+1} \left(\frac{2}{K}\right)^K \big(2(3K+2)\big)^{3K} \sum_{t=1}^{+\infty} \frac{1}{t\log(3t)^2} \frac{\delta}{C} \\
&\leqslant e^{K+1} \left(\frac{2}{K}\right)^K \big(2(3K+2)\big)^{3K} \frac{2}{\log(3)} \frac{\delta}{C} \leqslant \delta\,,
\end{aligned}$$

where in the third inequality we replaced $\beta(t,\delta)$ by its value and used in the fourth inequality ($C \geqslant 3$)

$$\sum_{t=1}^{+\infty} \frac{1}{t\log(t)^2} \leqslant \frac{1}{\log(3)^2} + \int_{t=1}^{+\infty} \frac{1}{t\log(3t)^2} \mathrm{d}t \leqslant \frac{2}{\log(3)}\,.$$

Asymptotic Optimality

Let $\epsilon \in (0,1)$. From the continuity of w^* in $\boldsymbol{\mu}$, there exists $\alpha = \alpha(\epsilon)$ such that

$$I_\epsilon := [\mu_1 - \alpha, \mu_1 + \alpha] \times \cdots \times [\mu_K - \alpha, \mu_K + \alpha]$$

is such that for all $\boldsymbol{\mu}' \in I_\epsilon$,

$$\boldsymbol{\mu}' \in \mathcal{S} \quad a^*_{\boldsymbol{\mu}} = a^*_{\boldsymbol{\mu}'} \quad \max_a |w^*_a(\boldsymbol{\mu}') - w^*_a(\boldsymbol{\mu})| \leqslant \epsilon\,.$$

Let $T \in \mathbb{N}^*$ and define the event

$$\mathcal{E}_T(\epsilon) = \bigcap_{t=T^{1/4}}^{T} (\widehat{\boldsymbol{\mu}}(t) \in I_\epsilon)\,.$$

The two following Lemmas are extracted from Garivier and Kaufmann (Jun. 2016).

Lemma 6 *There exists two constants B, C (that depend on $\boldsymbol{\mu}$ and ϵ) such that*

$$\mathbb{P}_{\boldsymbol{\mu}}(\mathcal{E}_T^c) \leqslant BT\exp(-CT^{1/8})\,.$$

Lemma 7 *There exists a constant T_ϵ such that for $T \geqslant T_\epsilon$, it holds that on $\mathcal{E}_T$,*

$$\forall t \geqslant \sqrt{T},\ \max_a \left| \frac{N_a(t)}{t} - w^*_a(\mu) \right| \leqslant 2(K-1)\epsilon$$

Using Lemma 7, for $T \geqslant T_\epsilon$, introducing

$$C^*_\epsilon(\boldsymbol{\mu}) = \inf_{\substack{\boldsymbol{\mu}':||\boldsymbol{\mu}'-\boldsymbol{\mu}||\leqslant\alpha(\epsilon)\\ \boldsymbol{w}':||\boldsymbol{w}'-\boldsymbol{w}^*(\boldsymbol{\mu})||\leqslant 2(K-1)\epsilon}} \inf_{\lambda\in\mathcal{A}lt\left(\boldsymbol{\mu}',\mathcal{S}\right)} \sum_{a=1}^{K} w_a \frac{(\mu'_a(t)-\lambda_a)^2}{2},$$

on the event $\mathcal{E}_T$ it holds that for every $t \geqslant \sqrt{T}$,

$$t \inf_{\lambda\in\mathcal{A}lt\left(\widehat{\boldsymbol{\mu}}(t),\mathcal{S}\right)} \sum_{a=1}^{K} \frac{N_a(t)}{t} \frac{(\widehat{\mu}_a(t)-\lambda_a)^2}{2} \geqslant tC^*_\epsilon(\boldsymbol{\mu}). \tag{22}$$

Let $T \geqslant T_\epsilon$, thanks to (22), on $\mathcal{E}_T$, we have

$$\begin{aligned}\max(\tau_\delta, T) &\leqslant \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{I}_{(\tau_\delta > t)} \\ &\leqslant \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{I}_{(tC^*_\epsilon(\boldsymbol{\mu})\leqslant\beta(T,\delta))} \leqslant \sqrt{T} + \frac{\beta(T,\delta)}{C^*_\epsilon(\boldsymbol{\mu})}.\end{aligned}$$

Introducing

$$T_0(\delta) = \inf\left\{T\in\mathbb{N} : \sqrt{T} + \frac{\beta(T,\delta)}{C^*_\epsilon(\boldsymbol{\mu})} \leqslant T\right\},$$

for every $T \geqslant \max(T_0(\delta), T_\epsilon)$, one has $\mathcal{E}_T \subseteq (\tau_\delta \leqslant T)$, therefore

$$\mathbb{P}_{\boldsymbol{\mu}}\left(\tau_\delta > T\right) \leqslant \mathbb{P}(\mathcal{E}_T^c) \leqslant BT\exp(-CT^{1/8})$$

and

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta] \leqslant T_0(\delta) + T_\epsilon + \sum_{T=1}^{\infty} BT\exp(-CT^{1/8}). \tag{23}$$

We now provide an upper bound on $T_0(\delta)$. Introducing the constant

$$H(\epsilon) = \inf\{T\in\mathbb{N} : T - \sqrt{T} \geqslant T/(1+\epsilon)\}$$

one has

$$\begin{aligned}T_0(\delta) &\leqslant H(\epsilon) + \inf\left\{T\in\mathbb{N} : \beta(T,\delta) \leqslant \frac{C^*_\epsilon(\boldsymbol{\mu})T}{1+\epsilon}\right\} \\ &\leqslant H(\epsilon) + \inf\left\{T\in\mathbb{N} : \log(TC/\delta) + (3K+2)\,\mathrm{loglog}(TC/\delta) \leqslant \frac{C^*_\epsilon(\boldsymbol{\mu})T}{1+\epsilon}\right\}.\end{aligned}$$

Using technical Lemma 9, for δ small enough to have $\left(C^*_\epsilon(\boldsymbol{\mu})\delta\right)/\left((1+\epsilon)^2C\right) \leqslant e$, one obtains

$$\begin{aligned}T_0(\delta) &\leqslant C(\epsilon) + \frac{\delta}{C}\max\left(g\left(\frac{C^*_\epsilon(\boldsymbol{\mu})\delta}{(1+\epsilon)^2C}\right), \exp\left(g\left(\frac{\epsilon}{3K+2}\right)\right)\right) \\ &\leqslant C(\epsilon) + \max\left(\frac{(1+\epsilon)^2}{C^*_\epsilon(\boldsymbol{\mu})}\log\left(\frac{e(1+\epsilon)^2C}{C^*_\epsilon(\boldsymbol{\mu})\delta}\log\left(\frac{(1+\epsilon)^2C}{C^*_\epsilon(\boldsymbol{\mu})\delta}\right)\right), \frac{\delta}{C}\exp\left(g\left(\frac{\epsilon}{3K+2}\right)\right)\right).\end{aligned}$$

This last upper bound yields, for every $\epsilon > 0$,

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \leqslant \frac{(1+\epsilon)^2}{C^*_\epsilon(\boldsymbol{\mu})}.$$

Letting ϵ tend to zero and by definition of w^*,

$$\lim_{\epsilon \to 0} C^*_\epsilon(\boldsymbol{\mu}) = T^*_{\mathcal{S}}(\boldsymbol{\mu})^{-1}\,,$$

yields

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \leqslant T^*_{\mathcal{S}}(\boldsymbol{\mu})\,.$$

■

Appendix C. Some Technical Lemmas

C.1. A Useful Inequality

For $0 < y \leqslant 1/e$ let g be the function

$$g(y) = \frac{1}{y} \ln\!\left(\frac{e}{y} \ln\!\left(\frac{1}{y}\right)\right)\,. \tag{24}$$

Lemma 8 *Let $A > 0$ such that $1/A > e$, then for all $x \geqslant g(A)$*

$$\log(x) \leqslant Ax\,. \tag{25}$$

Proof Since $g(A) \geqslant 1/A$, the function $x \mapsto A - 1/x$ is non-decreasing, we just need to prove (25) for $x = g(A)$. It remains to remark that

$$\begin{aligned}\log\big(g(A)\big) &\leqslant \log\!\left(\frac{2}{A}\log\!\left(\frac{1}{A}\right)\right)\\ &\leqslant \log\!\left(\frac{e}{A}\log\!\left(\frac{1}{A}\right)\right) = A\,g(A)\,,\end{aligned}$$

as $\log(x) \leqslant x/e$. ■

Lemma 9 *Let A, $B > 0$, then for all $\epsilon \in (0,1)$ such that $(1+\epsilon)/A < e$ and $B/\epsilon > e$, for all $x \geqslant \max\Big(g\big(A/(1+\epsilon)\big), \exp\big(g(\epsilon/B)\big)\Big)$*

$$\log(x) + B\,\mathrm{loglog}(x) \leqslant Ax\,. \tag{26}$$

Proof Since $\log(x) \geqslant g(\epsilon/B)$ thanks to Lemma 8 we have $B\,\mathrm{loglog}(x) \leqslant \epsilon \log(x)$. Therefore , still using Lemma 8 with $x \geqslant g\big(A/(1+\epsilon)\big)$,

$$\begin{aligned}\log(x) + B\,\mathrm{loglog}(x) &\leqslant (1+\epsilon)\log(x)\\ &\leqslant Ax\,.\end{aligned}$$

■

C.2. A Deviation Bound

We recall here for self-containment the Theorem 2 of Magureanu et al. (2014).

Theorem 10 *For all $\delta \geqslant (K+1)$ and $t \in \mathbb{N}^*$ we have*

$$\mathbb{P}\left(\sum_{a=1}^{K} N_a(t)\frac{(\widehat{\mu}_a(t)-\mu_a)^2}{2} \geqslant \delta\right) \leqslant e^{K+1}\left(\frac{2\delta(\delta\log(t)+1)}{K}\right)^K e^{-\delta}\,. \tag{27}$$

The factor 2 that differs from Theorem 2 of Magureanu et al. (2014) comes from the fact that we consider deviation at the right and left of the mean.

C.3. Unimodal Regression under Bound Restriction

For $\boldsymbol{\mu} \in \mathcal{M}$, $\omega \in \mathring{\Sigma}_K$ (where $\mathring{\Sigma}_K$ stands for the interior of Σ_K) and $b \in \{1,\cdots,K\}$, let $\mathcal{U}$ be the set of unimodal vector with maximum localized at b

$$\mathcal{U} = \{\boldsymbol{\lambda} : \lambda_1 \leqslant \cdots \leqslant \lambda_b \geqslant \lambda_{b+1} \geqslant \cdots \lambda_K\}\,, \tag{28}$$

and $\mathcal{U}_S$ be the same set with an additional bound restriction on λ_b

$$\mathcal{U}_S = \{\boldsymbol{\lambda} : \lambda_1 \leqslant \cdots \leqslant \lambda_b \geqslant \lambda_{b+1} \geqslant \cdots \lambda_K,\ \lambda_b \leqslant S\}\,. \tag{29}$$

Let $\widehat{\boldsymbol{\lambda}}$ be the unimodal regression of $\boldsymbol{\mu}$

$$\widehat{\boldsymbol{\lambda}} := \operatorname*{argmin}_{\boldsymbol{\lambda}\in\mathcal{U}} \sum_{a=1}^{K} \omega_a \frac{(\mu_a-\lambda_a)^2}{2}\,, \tag{30}$$

and $\boldsymbol{\lambda}^*$ be the projection of $\boldsymbol{\mu}$ on $\mathcal{U}_S$

$$\boldsymbol{\lambda}^* := \operatorname*{argmin}_{\boldsymbol{\lambda}\in\mathcal{U}_S} \sum_{a=1}^{K} \omega_a \frac{(\mu_a-\lambda_a)^2}{2}\,. \tag{31}$$

We have, as in the case of isotonic regression (see Hu (1997)), the following simple relation between $\boldsymbol{\lambda}^*$ and $\widehat{\boldsymbol{\lambda}}$

Lemma 11 *It holds that*

$$\lambda^*_a = \min(\widehat{\lambda}_a, S) \text{ for all } a \in \{1,\ldots,K\}\,.$$

To prove Lemma 11 we need the following properties on $\widehat{\boldsymbol{\lambda}}$.

Lemma 12 *Let $c_{-k} < \ldots < c_0 > \ldots > c_l$ be real numbers and $(A_{-k}, \ldots, A_0, \ldots, A_k)$ be integer intervals forming a partition of $\{1, \ldots, K\}$ be such that $\widehat{\boldsymbol{\lambda}}$ is constant on the sets A_i equals to c_i for all $-k \leqslant i \leqslant l$ and $b \in A_0$. Then, for all $-k \leqslant i \leqslant l$ and $\boldsymbol{\lambda} \in \mathcal{U}$*

$$\sum_{a \in A_i} (\mu_a - \widehat{\lambda}_a)\omega_a = 0 \tag{32}$$

$$\sum_{a \in A_i} (\mu_a - \widehat{\lambda}_a)\omega_a \lambda_a \leqslant 0\,. \tag{33}$$

Proof Since $\widehat{\boldsymbol{\lambda}}$ is the projection of $\boldsymbol{\mu}$ on the closed convex $\mathcal{U}$ we know that for all $\boldsymbol{\lambda}$ in $\mathcal{U}$

$$\sum_{A \in \{1,\ldots,K\}} (\mu_a - \widehat{\lambda}_a)(\widehat{\lambda}_a - \lambda_a)\omega_a \geqslant 0\,. \tag{34}$$

Fix $\boldsymbol{\lambda} \in \mathcal{U}$ and $-k \leqslant i \leqslant l$ and suppose, for example, that $i < 0$. The other cases $i = 0$ and $i > 0$ are similar. Introduce, for $|\epsilon| < \min\big(|c_i - c_{i-1}|, |c_{i+1} - c_i|\big)$, the vector $\boldsymbol{\lambda}^\epsilon$ such that

$$\lambda_a^\epsilon = \begin{cases} c_i - \epsilon & \text{if } a \in A_i\,, \\ \widehat{\lambda}_a & \text{else.} \end{cases}$$

By construction $\boldsymbol{\lambda}^\epsilon \in \mathcal{U}$ and thanks to (34) we have

$$\sum_{A \in \{1,\ldots,K\}} (\mu_a - \widehat{\lambda}_a)(\widehat{\lambda}_a - \lambda_a^\epsilon)\omega_a = \epsilon \sum_{a \in A_i} (\mu_a - \widehat{\lambda}_a)\omega_a \geqslant 0\,.$$

Taking ϵ positive or negative proves (32). Let $x, y \in \{1, \ldots, K\}$ be such that $A_i = \{x, x+1, \ldots, y-1, y\}$ and $\boldsymbol{\lambda}'$ be such that

$$\lambda'_a = \begin{cases} \lambda_a & \text{if } a \in A_i\,, \\ \lambda_x & \text{if } a < x\,, \\ \lambda_y & \text{if } a > y\,. \end{cases}$$

By construction $\boldsymbol{\lambda}' \in \mathcal{U}$ and thanks to (34) we have

$$\begin{aligned} \sum_{A \in \{1,\ldots,K\}} (\mu_a - \widehat{\lambda}_a)(\widehat{\lambda}_a - \lambda_a)\omega_a &= \sum_{a \in A_i} (\mu_a - \widehat{\lambda}_a)(c_i - \lambda'_a)\omega_a \\ &\quad + \lambda_x \sum_{j<i} \sum_{a \in A_j} (\mu_a - \widehat{\lambda}_a)\omega_a + \lambda_y \sum_{j>i} \sum_{a \in A_j} (\mu_a - \widehat{\lambda}_a)\omega_a \\ &= -\sum_{a \in A_i} (\mu_a - \widehat{\lambda}_a)\lambda'_a \omega_a\,, \end{aligned}$$

where we used (32). Equation (34) allows us to prove (33). ■

We now adapt the proof of Hu (1997) to the case of unimodal regression.

Proof [of Lemma 11]

Since $\mathcal{U}_S$ is a closed convex we just need to check that for all $\lambda \in \mathcal{U}_S$

$$\sum_{a\in\{1,\dots,K\}} \big(\mu_a - \min(\widehat{\lambda}_a, S)\big)\big(\min(\widehat{\lambda}_a, S) - \lambda_a\big)\omega_a \geqslant 0\,.$$

We have, using the same notation of Lemma 12,

$$\begin{aligned}\sum_{a\in\{1,\dots,K\}} \big(\mu_a - \min(\widehat{\lambda}_a, S)\big)\big(\min(\widehat{\lambda}_a, S) - \lambda_a\big)\omega_a &= \sum_{i:c_i\leqslant S}\sum_{a\in A_i}(\mu_a - \widehat{\lambda}_a)(\widehat{\lambda}_a - \lambda_a)\omega_a \\ &\quad + \sum_{i:c_i>S}\sum_{a\in A_i}(\mu_a - S)(S - \lambda_a)\omega_a \\ &= \sum_{i:c_i\leqslant S}\sum_{a\in A_i}(\mu_a - \widehat{\lambda}_a)(\widehat{\lambda}_a - \lambda_a)\omega_a \\ &\quad + \sum_{i:c_i>S}\sum_{a\in A_i}(\mu_a - \widehat{\lambda}_a)(S - \lambda_a)\omega_a + \sum_{i:c_i>S}\sum_{a\in A_i}(c_i - S)(\lambda_a - S)\omega_a \geqslant 0\,,\end{aligned}$$

where we used the Lemma 12 for the two first sums and the fact that $\lambda_a < S$ for the last sum. ■

References

Stephen Boyd, Lin Xiao, and Almir Mutapcic. Subgradient methods. *Lecture notes of EE392o, Stanford University, Autumn Quarter*, 2004, 2003.

M Frisén. Unimodal regression. *The Statistician*, pages 479–485, 1986.

Aurlien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference On Learning Theory*, pages 998–1027, Jun. 2016.

Zhi Geng and Ning-Zhong Shi. Algorithm as 257: isotonic regression for umbrella orderings. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39(3):397–402, 1990.

Mark C Genovese, Patrick Durez, Hanno B Richards, Jerzy Supronik, Eva Dokoupilova, Vadim Mazurov, Jacob A Aelion, Sang-Heon Lee, Christine E Codding, Herbert Kellner, Takashi Ikawa, Sophie Hugot, and Shephard Mpofu. Efficacy and safety of secukinumab in patients with rheumatoid arthritis: a phase ii, dose-finding, double-blind, randomised, placebo controlled study. *Annals of the Rheumatic Diseases*, 72(6):863–869, 2013. ISSN 0003-4967. doi: 10.1136/annrheumdis-2012-201601.

Xiaomi Hu. Maximum-likelihood estimation under bound restriction and order and uniform bound restrictions. *Statistics & probability letters*, 35(2):165–171, 1997.

J. Jack Lee Le Tourneau, Christophe and Lillian L. Siu. Dose escalation methods in phase i cancer clinical trials. *JNCI Journal of the National Cancer Institute*, 101.10:708720, 2009.

Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pages 1690–1698, 2016.

Stefan Magureanu, Richard Combes, and Alexandre Proutière. Lipschitz bandits: Regret lower bound and optimal algorithms. In *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, pages 975–999, 2014.

RA Mureika, TR Turner, and PC Wollan. An algorithm for unimodal isotonic regression, with application to locating a maximum, univ. new brunswichk dept. math. Technical report, and Stat. Tech. Report 92–4, 1992.

J. M. Bremner H. D. Brunk R. E. Barlow, D. J. Bartholomew. *Statistical inference under order restrictions*. John Wiley and Sons, 1973.

Quentin F Stout. Optimal algorithms for unimodal regression. *Ann Arbor*, 1001:48109–2122, 2000.

R. L. Dykstra T. Robertson, F. T. Wright. *Order restricted statistical inference*. John Wiley and Sons, 1988.