



ICMA

PROCEEDINGS-FULL PAPER

2020 International Conference on
Machine Learning, Artificial Intelligence
and Applications

Online Conference, August 18, 2020



WWW.GAICS.ORG

Contents

ICMA_0031	2
Weighted Empirical Risk Minimization: Sample Selection Bias Correction based on Importance Sampling	
ICMA_0034	22
Ternary Generative Adversarial Networks	
ICMA_0052	38
Thresholding Bandit for Dose-ranging: The Impact of Monotonicity	

Weighted Empirical Risk Minimization: Sample Selection Bias Correction based on Importance Sampling

Robin Vogel^{a*}, Mastane Achab^a, Stéphan Cléménçon^a and Charles Tillier^a

^aLTCI, Télécom Paris, France

*Corresponding Author: robin.vogel@telecom-paris.fr

Abstract

Prediction problems are of major importance in statistical learning. The main paradigm of predictive learning is Empirical Risk Minimization (ERM in abbreviated form). In the standard setup, training observations originate from the same distribution as the testing observations. We consider statistical learning problems, when the distribution P' of the training observations Z'_1, \dots, Z'_n differs from the distribution P involved in the risk one seeks to minimize (referred to as the test distribution) but is still defined on the same measurable space as P and dominates it. In the unrealistic case where the likelihood ratio $\Phi(z) = dP/dP'(z)$ is known, one may straightforwardly extend the Empirical Risk Minimization (ERM) approach to this specific transfer learning setup using the same idea as that behind Importance Sampling, by minimizing a weighted version of the empirical risk functional computed from the 'biased' training data Z'_i with weights $\Phi(Z'_i)$. Although the importance function $\Phi(z)$ is generally unknown in practice, we show that, in various situations frequently encountered in practice - such as learning with class imbalance, learning in a stratified population, solving binary classification with only positive and unlabeled data (PU learning) or learning under random censorship - it takes a simple form and can be directly estimated from the Z'_i 's and some auxiliary information on the statistical population P . By means of linearization techniques, we then prove that the generalization capacity of the approach aforementioned is preserved when plugging the resulting estimates of the $\Phi(Z'_i)$'s into the weighted empirical risk. Beyond these theoretical guarantees, numerical results provide strong empirical evidence of the relevance of the approach promoted in this article. Specifically, we show on ImageNet - an image dataset based on the hierarchical lexical database WordNet - that correcting bias on high-level categories can lead to significant performance improvements for the classification task.

Keyword: Statistical Learning Theory, Importance Sampling, Transfer Learning

1. Introduction

Prediction problems are of major importance in statistical learning. The main paradigm of predictive learning is *Empirical Risk Minimization* (ERM in abbreviated form), see *e.g.* Devroye et al. (1996). In the standard setup, Z is a random variable (r.v. in short) that takes its values in a feature space \mathcal{Z} with distribution P , Θ is a parameter space and $\ell: \Theta \times \mathcal{Z} \rightarrow \mathbb{R}_+$ is a (measurable) loss function. The risk is then defined by: $\forall \theta \in \Theta$,

$$\mathcal{R}_P(\theta) = \mathbb{E}_P[\ell(\theta, Z)], \tag{1}$$

and more generally for any measure Q on \mathcal{Z} : $\mathcal{R}_Q(\theta) = \int_{\mathcal{Z}} \ell(\theta, z) dQ(z)$. In most practical situations, the distribution P involved in the definition of the risk is unknown and learning is based on the sole observation of an independent and identically distributed (i.i.d.) sample Z_1, \dots, Z_n drawn from P and the risk (1) must be replaced by an empirical counterpart (or a possibly smoothed/penalized version of it), typically:

$$\hat{\mathcal{R}}_P(\theta) = \frac{1}{n} \sum_{i=1}^n \ell(\theta, Z_i) = \mathcal{R}_{\hat{P}_n}(\theta), \tag{2}$$

where $\hat{P}_n = (1/n) \sum_{i=1}^n \delta_{Z_i}$ is the empirical measure of P and δ_z denotes the Dirac measure at any point z . With the design of successful algorithms such as neural networks, support vector machines or boosting methods to perform ERM, the practice of predictive learning has recently received a significant attention and is now supported by a sound theory based on results in empirical process theory. The performance of minimizers of (2) can be indeed studied by means of concentration inequalities, quantifying the fluctuations of the maximal deviations $\sup_{\theta \in \Theta} |\hat{\mathcal{R}}_P(\theta) - \mathcal{R}_P(\theta)|$ under various complexity assumptions for the functional class $\mathcal{F} = \{\ell(\theta, \cdot): \theta \in \Theta\}$ (*e.g.* VC dimension, metric entropies, Rademacher averages), see Boucheron et al. (2013) for instance. Although, in the Big Data era, the availability of massive digitized information to train predictive rules is an undeniable opportunity for the widespread deployment of machine-learning solutions, the poor control of the data acquisition process one is confronted with in many applications puts practitioners at risk of jeopardizing the generalization ability of the rules produced by the algorithms implemented. Bias selection issues in machine-learning are now the subject of much attention in the literature, see Bolukbasi et al. (2016), Zhao et al. (2017), Burns et al. (2019), Liu et al. (2016) or Huang et al. (2007). In the context of face analysis, a research area including a broad range of applications such as face detection, face recognition or face attribute detection, machine learning algorithms trained with biased training data, *e.g.* in terms of gender or ethnicity, raise concerns about fairness in machine learning. Unfair algorithms may induce systemic undesired disadvantages for specific social groups, see Das et al. (2018) for further details. Several examples of bias in deep learning based face recognition systems are discussed in Nagpal et al. (2019).

Throughout the present article, we consider the case where the i.i.d. sample Z'_1, \dots, Z'_n available for training is not drawn from P but from another distribution P' , with respect to which P is absolutely continuous, and the goal pursued is to set theoretical grounds for the application of ideas behind Importance Sampling (IS in short) methodology to extend the ERM approach to this learning setup. We highlight that the problem under study is a very particular case of *Transfer Learning* (see

e.g. Pan and Yang (2010), Ben-David et al. (2010) and Storkey (2009)), a research area currently receiving much attention in the literature and encompassing general situations where the information/knowledge one would like to transfer may take a form in the *target* space very different from that in the *source* space (referred to as *domain adaptation*).

Weighted ERM (WERM). In this paper, we investigate conditions guaranteeing that values for the parameter θ that nearly minimize (1) can be obtained through minimization of a weighted version of the empirical risk based on the Z'_i 's, namely

$$\tilde{\mathcal{R}}_{w,n}(\theta) = \mathcal{R}_{\tilde{P}_{w,n}}(\theta), \quad (3)$$

where $\tilde{P}_{w,n} = (1/n) \sum_{i=1}^n w_i \delta_{Z'_i}$ and $w = (w_1, \dots, w_n) \in \mathbb{R}_+^n$ is a certain weight vector. Of course, ideal weights w^* are given by the likelihood function $\Phi(z) = (dP/dP')(z)$: $w_i^* = \Phi(Z'_i)$ for $i \in \{1, \dots, n\}$. In this case, the quantity (3) is obviously an unbiased estimate of the true risk (1):

$$\mathbb{E}_P, \left[\mathcal{R}_{\tilde{P}_{w^*,n}}(\theta) \right] = \mathcal{R}_P(\theta), \quad (4)$$

and generalization bounds for the \mathcal{R}_P -risk excess of minimizers of $\tilde{\mathcal{R}}_{w^*,n}$ can be directly established by studying the concentration properties of the empirical process related to the Z'_i 's and the class of functions $\{\Phi(\cdot)\ell(\theta, \cdot) : \theta \in \Theta\}$ (see section 2 below). However, the *importance function* Φ is unknown in general, just like distribution P . It is the major purpose of this article to show that, in far from uncommon situations, the (ideal) weights w_i^* can be estimated from the Z'_i 's combined with auxiliary information on the target population P . As shall be seen below, such favorable cases include in particular classification problems where class probabilities in the test stage differ from those in the training step, risk minimization in stratified populations (see Bekker and Davis (2018)), with strata statistically represented in a different manner in the test and training populations, positive-unlabeled learning (PU-learning, see e.g. du Plessis et al. (2014)). In each of these cases, we show that the stochastic process obtained by plugging the weight estimates in the weighted empirical risk functional (3) is much more complex than a simple empirical process (i.e. a collection of i.i.d. averages) but can be however studied by means of *linearization techniques*, in the spirit of the ERM extensions established in Cl  men  on et al. (2008) or Cl  men  on and Vayatis (2009). Learning rate bounds for minimizers of the corresponding risk estimate are proved and, beyond these theoretical guarantees, the performance of the weighted ERM approach is supported by convincing numerical results.

The article is structured as follows. In section 2, the ideal case where the importance function Φ is known is preliminarily considered and a first basic example where the optimal weights can be easily inferred and plugged into the risk without deteriorating the learning rate is discussed. The main results of the paper are stated in section 3, which shows that the methodology promoted can be applied to three important problems in practice, risk minimization in stratified populations, PU-learning and learning from censored data, with generalization guarantees. Illustrative numerical experiments are displayed in section 4, while some concluding remarks are collected in section 5. Proofs and additional results are deferred to the Supplementary Material.

2. Importance Sampling - Risk Minimization with Biased Data

Here and throughout, the indicator function of any event \mathcal{E} is denoted by $\mathbb{I}\{\mathcal{E}\}$, the sup norm of any bounded function $h: \mathcal{Z} \rightarrow \mathbb{R}$ by $\|h\|_\infty$. We place ourselves in the framework of statistical learning based on biased training data previously introduced. As a first go, we consider the unrealistic situation where the importance function Φ is known, insofar as we shall subsequently develop techniques aiming at mimicking the minimization of the ideally weighted empirical risk

$$\tilde{\mathcal{R}}_{w^*,n}(\theta) = \frac{1}{n} \sum_{i=1}^n w_i^* \ell(\theta, Z'_i), \quad (5)$$

namely the (unbiased) Importance Sampling estimator of (1) based on the instrumental data Z'_1, \dots, Z'_n . The following result describes the performance of minimizers $\tilde{\theta}_n^*$ of (5). Since the goal of this paper is to promote the main ideas of the approach rather than to state results with the highest level of generality due to space limitations, we assume throughout the article for simplicity that ℓ and Φ are both bounded functions. For $\sigma_1, \dots, \sigma_n$ independent Rademacher random variables (*i.e.* symmetric $\{-1,1\}$ -valued r.v.'s), independent from the Z'_i 's, we define the Rademacher average associated to the class of function \mathcal{F} as $R'_n(\mathcal{F}) := \mathbb{E}_\sigma \left[\sup_{\theta \in \Theta} \frac{1}{n} \left| \sum_{i=1}^n \sigma_i \ell(\theta, Z'_i) \right| \right]$. This quantity can be bounded by metric entropy methods under appropriate complexity assumptions on the class \mathcal{F} , it is for instance of order $O_{\mathbb{P}}(1/\sqrt{n})$ when \mathcal{F} is a VC major class with finite VC dimension, see *e.g.* Boucheron et al. (2005).

Lemma 1. With probability at least $1 - \delta$, we have: $\forall n \geq 1$,

$$\mathcal{R}_P(\tilde{\theta}_n^*) - \min_{\theta \in \Theta} \mathcal{R}_P(\theta) \leq 4\|\Phi\|_\infty \mathbb{E}[R'_n(\mathcal{F})] + 2\|\Phi\|_\infty \sup_{(\theta,z) \in \Theta \times \mathcal{Z}} \ell(\theta, z) \sqrt{\frac{2 \log(1/\delta)}{n}}.$$

Of course, when $P' = P$, we have $\Phi \equiv 1$ and the bound stated above simply describes the performance of standard empirical risk minimizers. The proof is based on the standard bound

$$\mathcal{R}_P(\tilde{\theta}_n^*) - \min_{\theta \in \Theta} \mathcal{R}_P(\theta) \leq 2 \sup_{\theta \in \Theta} |\tilde{\mathcal{R}}_{w^*,n}(\theta) - \mathbb{E}[\tilde{\mathcal{R}}_{w^*,n}(\theta)]|,$$

combined with basic concentration results for empirical processes, see the Supplementary Material for further details. Of course, the importance function Φ is generally unknown and must be estimated in practice. As illustrated by the elementary example below (related to binary classification, in the situation where the probability of occurrence of a positive instance significantly differs in the training and test stages), in certain statistical learning problems with biased training distribution, Φ takes a simplistic form and can be easily estimated from the Z'_i 's combined with auxiliary information on P .

Binary classification with varying class probabilities. The flagship problem in supervised learning corresponds to the simplest situation, where $Z = (X, Y)$, Y being a binary variable valued in $\{-1, +1\}$ say, and the r.v. X takes its values in a measurable space \mathcal{X} and models some information hopefully useful to predict Y . The parameter space Θ is a set \mathcal{G} of measurable mappings (*i.e.* classifiers) $g: \mathcal{X} \rightarrow \{-1, +1\}$ and the loss function is given by $\ell(g, (x, y)) = \mathbb{I}\{g(x) \neq y\}$ for all g in \mathcal{G} and any $(x, y) \in \mathcal{X} \times \{-1, +1\}$. The distribution P of the random pair (X, Y) can be either described by X 's marginal distribution $\mu(dx)$ and the posterior probability $\eta(x) = \mathbb{P}\{Y =$

$+1 | X = x\}$ or else by the triplet (p, F_+, F_-) where $p = \mathbb{P}\{Y = +1\}$ and $F_\sigma(dx)$ is X 's conditional distribution given $Y = \sigma 1$ with $\sigma \in \{-, +\}$. It is very common that the fraction of positive instances in the training dataset is significantly lower than the rate p expected in the test stage, supposed to be known here (see the Supplementary Material for the case where the rate p is only approximately known). We thus consider the case where the distribution P' of the training data $(X'_1, Y'_1), \dots, (X'_n, Y'_n)$ is described by the triplet (p', F_+, F_-) with $p' < p$. The likelihood function takes the simple following form

$$\Phi(x, y) = \mathbb{I}\{y = +1\} \frac{p}{p'} + \mathbb{I}\{y = -1\} \frac{1-p}{1-p'} \stackrel{\text{def}}{=} \phi(y),$$

which reveals that it depends on the label y solely, and the ideally weighted empirical risk process is

$$\tilde{\mathcal{R}}_{w^*, n}(g) = \frac{p}{p' n} \sum_{i: Y'_i=1} \mathbb{I}\{g(X'_i) = -1\} + \frac{1-p}{1-p' n} \sum_{i: Y'_i=-1} \mathbb{I}\{g(X'_i) = +1\}. \quad (6)$$

In general the theoretical rate p' is unknown and one replaces (6) with

$$\tilde{\mathcal{R}}_{\hat{w}^*, n}(g) = \frac{p}{n'_+} \sum_{i: Y'_i=1} \mathbb{I}\{g(X'_i) = -1\} + \frac{1-p}{n'_-} \sum_{i: Y'_i=-1} \mathbb{I}\{g(X'_i) = +1\}, \quad (7)$$

where $n'_+ = \sum_{i=1}^n \mathbb{I}\{Y'_i = +1\} = n - n'_-$, $\hat{w}_i^* = \hat{\phi}(Y'_i)$ and $\hat{\phi}(y) = \mathbb{I}\{y = +1\}np/n'_+ + \mathbb{I}\{y = -1\}n(1-p)/n'_-$. The stochastic process above is not a standard empirical process but a collection of sums of two ratios of basic averages. However, the following result provides a uniform control of the deviations between the ideally weighted empirical risk and that obtained by plugging the empirical weights into the latter.

Lemma 2. Let $\varepsilon \in (0, 1/2)$. Suppose that $p' \in (\varepsilon, 1 - \varepsilon)$. For any $\delta \in (0, 1)$, we have with probability larger than $1 - \delta$:

$$\sup_{g \in \mathcal{G}} |\tilde{\mathcal{R}}_{\hat{w}^*, n}(g) - \tilde{\mathcal{R}}_{w^*, n}(g)| \leq \frac{2}{\varepsilon^2} \sqrt{\frac{\log(2/\delta)}{2n}},$$

as soon as $n \geq 2\log(2/\delta)/\varepsilon^2$.

See the Appendix for the technical proof. Consequently, minimizing (7) nearly boils down to minimizing (6). Combining Lemmas [2](#) and [1](#), we immediately get the generalization bound stated in the result below.

Corollary 1. Suppose that the hypotheses of Lemma [2](#) are fulfilled. Let \tilde{g}_n be any minimizer of $\tilde{\mathcal{R}}_{\hat{w}^*, n}$ over class \mathcal{G} . We have with probability at least $1 - \delta$:

$$\mathcal{R}_P(\tilde{g}_n) - \inf_{g \in \mathcal{G}} \mathcal{R}_P(g) \leq \frac{2\max(p, 1-p)}{\varepsilon} \left(2\mathbb{E}[R'_n(\mathcal{G})] + \sqrt{\frac{2\log(2/\delta)}{n}} \right) + \frac{4}{\varepsilon^2} \sqrt{\frac{\log(4/\delta)}{2n}},$$

as soon as $n \geq 2\log(4/\delta)/\varepsilon^2$; where $R'_n(\mathcal{G}) = (1/n)\mathbb{E}_{\sigma}[\sup_{g \in \mathcal{G}} |\sum_{i=1}^n \sigma_i \mathbb{I}\{g(X'_i) \neq Y'_i\}|]$.

Hence, some side information (*i.e.* knowledge of parameter p) has permitted to weight the training data in order to build an empirical risk functional that approximates the target risk and to show that minimization of this risk estimate yields prediction rules with optimal (in the minimax sense) learning rates. The purpose of the subsequent analysis is to show that this remains true for more general problems. Observe in addition that the bound in Corollary [1](#) deteriorates as ε decays to zero: the method used here is not intended to solve the *few shot* learning problem, where almost no training data with positive labels is available (*i.e.* $p' \approx 0$). As shall be seen in subsection [3.2](#), alternative estimators of the importance function must be considered in this situation.

Remark 1. Although the quantity (7) can be viewed as a *cost-sensitive* version of the empirical classification risk based on the (X'_i, Y'_i) 's (see *e.g.* Bach et al. (2006)), we point out that the goal pursued here is not to achieve an appropriate trade-off between type I and type II errors in the P' classification problem as in biometric applications for instance (*i.e.* optimization of the (F_+, F_-) -ROC curve at a specific point) but to transfer knowledge gained in analyzing the biased data drawn from P' to the classification problem related to distribution P .

Related work. We point out that the natural idea of using weights in ERM problems that mimic those induced by the importance function has already been used in Sugiyama et al. (2008) for *covariate shift adaptation* problems (*i.e.* supervised situations, where the conditional distribution of the output given the input information is the same in the training and test domains), when, in contrast to the framework considered here, a test sample is additionally available (a method for estimating directly the importance function based on Kullback-Leibler divergence minimization is proposed, avoiding estimation of the test density). Importance sampling estimators have been also considered in Garcke and Vanck (2014) in the setup of *inductive transfer learning* (the tasks between source and target are different, regardless of the similarities between source and target domains), where the authors have proposed two methods to approximate the importance function, among which one is again based on minimizing the Kullback-Leibler divergence between the two distributions. In Cortes et al. (2008), the sample selection bias is assumed to be independent from the label, which is not true under our stratum-shift assumption or for the PU learning problem (see section [2](#)). Lemma [1](#) assumes that the exact importance function is known, as does Cortes et al. (2010). The next section introduces new results for more realistic settings where it has to be learned from the data.

3. Weighted Empirical Risk Minimization - Generalization Guarantees

Through two important and generic examples, relevant for many applications, we show that the approach sketched above can be applied to general situations, where appropriate auxiliary information on the target distribution is available, with generalization guarantees.

3.1 Statistical Learning from Biased Data in a Stratified Population

A natural extension of the simplistic problem considered in section 2 is multiclass classification in a stratified population. The random labels Y and Y' are supposed to take their values in $\{1, \dots, J\}$ say, with $J \geq 1$, and each labeled observation (X, Y) belongs to a certain random stratum S in $\{1, \dots, K\}$ with $K \geq 1$. Again, the distribution P of a random element $Z = (X, Y, S)$ may be described by the parameters $\{(p_{j,k}, F_{j,k}): 1 \leq j \leq J, 1 \leq k \leq K\}$ where $F_{j,k}$ is the conditional distribution of X given $(Y, S) = (j, k)$ and $p_{j,k} = \mathbb{P}_{(X,Y,S) \sim P}\{Y = j, S = k\}$. Then, we have

$$dP(x, y, s) = \sum_{j=1}^J \sum_{k=1}^K \mathbb{I}\{y = j, s = k\} p_{j,k} dF_{j,k}(x),$$

and considering a distribution P' with $F_{j,k} \equiv F'_{j,k}$ but possibly different class-stratum probabilities $p'_{j,k}$, the likelihood function becomes

$$\frac{dP}{dP'}(x, y, s) = \sum_{j=1}^J \sum_{k=1}^K \frac{p_{j,k}}{p'_{j,k}} \mathbb{I}\{y = j, s = k\} \stackrel{def}{=} \phi(y, s).$$

A more general framework can actually encompass this specific setup by defining 'meta-strata' in $\{1, \dots, J\} \times \{1, \dots, K\}$. Strata may often correspond to categorical input features in practice. The formalism introduced below is more general and includes the example considered in the preceding section, where strata are defined by labels.

Learning from biased stratified data. Consider a general mixture model, where distributions P and P' are stratified over $K \geq 1$ strata. Namely, $Z = (X, S)$ and $Z' = (X', S')$ with auxiliary random variables S and S' (the strata) valued in $\{1, \dots, K\}$. We place ourselves in a *stratum-shift* context, assuming that the conditional distribution of X given $S = k$ is the same as that of X' given $S' = k$, denoted by $F_k(dx)$, for any $k \in \{1, \dots, K\}$. However, stratum probabilities $p_k = \mathbb{P}(S = k)$ and $p'_k = \mathbb{P}(S' = k)$ may possibly be different. In this setup, the likelihood function depends only on the strata and can be expressed in a very simple form, as follows:

$$\frac{dP}{dP'}(x, s) = \sum_{k=1}^K \mathbb{I}\{s = k\} \frac{p_k}{p'_k} \stackrel{def}{=} \phi(s).$$

In this case, the ideally weighted empirical risk writes

$$\tilde{\mathcal{R}}_{w^*,n}(\theta) = \frac{1}{n} \sum_{i=1}^n \ell(\theta, Z'_i) \sum_{k=1}^K \mathbb{I}\{S'_i = k\} \frac{p_k}{p'_k}.$$

If the strata probabilities p_k 's for the test distribution are known, an empirical counterpart of the ideal empirical risk above is obtained by simply plugging estimates of the p'_k 's computed from the training data:

$$\tilde{\mathcal{R}}_{\hat{w}^*,n}(\theta) = \sum_{i=1}^n \ell(\theta, Z'_i) \sum_{k=1}^K \mathbb{I}\{S'_i = k\} \frac{p_k}{n'_k}, \quad (8)$$

with $n'_k = \sum_{i=1}^n \mathbb{I}\{S'_i = k\}$, $\hat{w}_i^* = \hat{\phi}(S'_i)$ and $\hat{\phi}(s) = \sum_{k=1}^K \mathbb{I}\{s = k\} n p_k / n'_k$.

A bound for the excess of risk is given in Theorem 1, that can be viewed as a generalization of Corollary 1.

Theorem 1. Let $\varepsilon \in (0, 1/2)$ and assume that $p'_k \in (\varepsilon, 1 - \varepsilon)$ for $k = 1, \dots, K$. Let $\tilde{\theta}_n^*$ be any minimizer of $\tilde{\mathcal{R}}_{\hat{w}^*,n}$ as defined in (8) over class θ . We have with probability at least $1 - \delta$:

$$\mathcal{R}_P(\tilde{\theta}_n^*) - \inf_{\theta \in \Theta} \mathcal{R}_P(\theta) \leq \frac{2 \max_k p_k}{\varepsilon} \left(2 \mathbb{E}[R'_n(\mathcal{F})] + L \sqrt{\frac{2 \log(2/\delta)}{n}} \right) + \frac{4L}{\varepsilon^2} \sqrt{\frac{\log(4K/\delta)}{2n}},$$

as soon as $n \geq 2 \log(4K/\delta) / \varepsilon^2$; where $R'_n(\mathcal{F}) = (1/n) \mathbb{E}_{\sigma}[\sup_{\theta \in \Theta} |\sum_{i=1}^n \sigma_i \ell(\theta, Z'_i)|]$, and the loss is bounded by $L = \sup_{(\theta, z) \in \Theta \times \mathcal{Z}} \ell(\theta, z)$.

Just like in Corollary 1, the bound in Theorem 1 explodes when ε vanishes, which corresponds to the situation where a stratum $k \in \{1, \dots, K\}$ is very poorly represented in the training data, *i.e.* when $p'_k \ll p_k$. Again, as highlighted by the experiments carried out, reweighting the losses in a frequentist (ERM) approach guarantees good generalization properties in a specific setup only, where the training information, though biased, is sufficiently informative.

3.2 Positive-Unlabeled Learning

Relaxing the *stratum-shift* assumption made in the previous subsection, the importance function becomes more complex and writes:

$$\phi(x, s) = \frac{dP}{dP'}(x, s) = \sum_{k=1}^K \mathbb{I}\{s = k\} \frac{p_k}{p'_k} \frac{dF_k}{dF'_k}(x),$$

where F_k and F'_k are respectively the conditional distributions of X given $S = k$ and of X' given $S' = k$. The Positive-Unlabeled (PU) learning problem, which has recently been the subject of much attention (see *e.g.* du Plessis et al. (2014), du Plessis et al. (2015), Kiryo et al. (2017)), provides a typical example of this situation. Re-using the notations introduced in section 2, in the PU problem, the testing and training distributions P and P' are respectively described by the triplets (p, F_+, F_-) and (q, F_+, F) , where $F = pF_+ + (1 - p)F_-$ is the marginal distribution of X . Hence, the objective pursued is to solve a binary classification task, based on the sole observation of a training sample pooling data with positive labels and unlabeled data, q denoting the theoretical fraction of positive data among the dataset. As noticed in du Plessis et al. (2014) (see also du Plessis et al. (2015), Kiryo et al. (2017)), the likelihood/importance function can be expressed in a simple manner, as follows, $\forall (x, y) \in \mathcal{X} \times \{-1, +1\}$:

$$\Phi(x, y) = \frac{p}{q} \mathbb{I}\{y = +1\} + \frac{1}{1-q} \mathbb{I}\{y = -1\} - \frac{p}{1-q} \frac{dF_+}{dF}(x) \mathbb{I}\{y = -1\}. \quad (9)$$

Based on an i.i.d. sample $(X'_1, Y'_1), \dots, (X'_n, Y'_n)$ drawn from P' combined with the knowledge of p (which can also be estimated from PU data, see e.g. du Plessis and Sugiyama (2014)) and using that $F_- = (1/(1-p))(F - pF_+)$, one may obtain estimators of q , F_+ and F by computing $n'_+/n = (1/n) \sum_{i=1}^n \mathbb{I}\{Y'_i = +1\}$, $\hat{F}_+ = (1/n'_+) \sum_{i=1}^n \mathbb{I}\{Y'_i = +1\} \delta_{X'_i}$ and $\hat{F} = (1/n'_-) \sum_{i=1}^n \mathbb{I}\{Y'_i = -1\} \delta_{X'_i}$. However, plugging these quantities into (9) do not permit to get a statistical version of the importance function, insofar as the probability measures \hat{F}_+ and \hat{F} are mutually singular with probability one, as soon as F_+ is continuous. Of course, as proposed in du Plessis et al. (2014), one may use statistical methods (e.g. kernel smoothing) to build distribution estimators, that ensures absolute continuity but are subject to the curse of dimensionality. However, WERM can still be applied in this case, by observing that: $\forall g \in \mathcal{G}$,

$$\mathcal{R}_P(g) = -p + \mathbb{E}_{P'} \left[\frac{2p}{q} \mathbb{I}\{g(X') = -1, Y' = +1\} + \frac{1}{1-q} \mathbb{I}\{g(X') = +1, Y' = -1\} \right], \quad (10)$$

which leads to the weighted empirical risk

$$\frac{2p}{n'_+} \sum_{i:Y'_i=+1} \mathbb{I}\{g(X'_i) = -1\} + \frac{1}{n'_-} \sum_{i:Y'_i=-1} \mathbb{I}\{g(X'_i) = +1\}. \quad (11)$$

Minimization of (11) yields rules \tilde{g}_n whose generalization ability regarding the binary problem related to (p, F_+, F_-) can be guaranteed, as shown by the following result, the form of the weighted empirical risk in this case being quite similar to (7).

Theorem 2. Let $\varepsilon \in (0, 1/2)$. Suppose that $q \in (\varepsilon, 1 - \varepsilon)$. Let \tilde{g}_n be any minimizer of the weighted empirical risk (11) over class \mathcal{G} . We have with probability at least $1 - \delta$:

$$\mathcal{R}_P(\tilde{g}_n) - \inf_{g \in \mathcal{G}} \mathcal{R}_P(g) \leq \frac{2 \max(2p, 1)}{\varepsilon} \left(2 \mathbb{E}[R'_n(\mathcal{G})] + \sqrt{\frac{2 \log(2/\delta)}{n}} \right) + \frac{4(2p+1)}{\varepsilon^2} \sqrt{\frac{\log(4/\delta)}{2n}},$$

as soon as $n \geq 2 \log(4/\delta)/\varepsilon^2$; where $R'_n(\mathcal{G}) = (1/n) \mathbb{E}_\sigma[\sup_{g \in \mathcal{G}} |\sum_{i=1}^n \sigma_i \mathbb{I}\{g(X'_i) \neq Y'_i\}|]$.

Remark 2. Let $\eta(x) = \mathbb{P}\{Y = +1 | X = x\}$ denote the posterior probability and recall that $(dF_+/dF_-)(x) = ((1-p)/p)(\eta(x)/(1-\eta(x)))$. Observing that

$$\Phi(x, y) = \frac{p}{q} \mathbb{I}\{y = +1\} + \frac{1 - \eta(x)}{1 - q} \mathbb{I}\{y = -1\}, \quad (12)$$

in the case when an estimate $\hat{\eta}(x)$ of $\eta(x)$ is available, one can perform WERM using the empirical weight function

$$\hat{\Phi}(x, y) = \frac{np}{n'_+} \mathbb{I}\{y = +1\} + \frac{1 - \hat{\eta}(x)}{1 - \frac{n'_+}{n}} \mathbb{I}\{y = -1\}. \quad (13)$$

A bound that describes how this approach generalizes, depending on the accuracy of estimate $\hat{\eta}$, can be easily established, for more details refer to Theorem 3 in the Supplementary Material, where it is also discussed how to exploit such formulas in order to design incremental WERM procedures.

3.3 Learning from Censored Data

Another important example of sample bias is the censorship setting where the learner has only access to (right) censored targets $\min(Y', C')$ instead of Y' . Intuitively, this situation occurs when Y' is a duration/date, e.g. the date of death of a patient modeled by covariates X' , and the study happens at a (random) date C' . Hence if $C' \leq Y'$, then we know that the patient is still alive at time C' but the target time Y' remains unknown. This problem has been extensively studied (see e.g. Ausset et al. (2019)): we show here that it is an instance of WERM. Formally, we respectively denote by P and P' the testing and training distributions of the r.v.'s $(X, \min(Y, C), \mathbb{I}\{Y \leq C\})$ and $(X', \min(Y', C'), \mathbb{I}\{Y' \leq C'\})$ both valued in $\mathbb{R}^d \times \mathbb{R}_+ \times \{0,1\}$ (with Y, Y', C, C' all nonnegative r.v.'s) and such that the pairs (X, Y) and (X', Y') share the same distribution Q . Moreover, $C > Y$ with probability 1 (i.e. the testing data are never censored) and Y' and C' are assumed to be conditionally independent given X' . Hence, for all $(x, y, \delta) \in \mathbb{R}^d \times \mathbb{R}_+ \times \{0,1\}$:

$$dP(x, y, \delta) = \delta dQ(x, y),$$

and

$$\delta dP'(x, y, \delta) = \delta \mathbb{P}(C' \geq y) d\mathbb{P}(X' = x, Y' = y | C' \geq y) = \delta S_{C'}(y|x) dQ(x, y),$$

where $S_{C'}(y|x) = \mathbb{P}(C' \geq y | X' = x)$ denotes the conditional survival function of C' given X' . Then, the importance function is:

$$\forall (x, y, \delta) \in \mathbb{R}^d \times \mathbb{R}_+ \times \{0,1\}, \quad \Phi(x, y, \delta) = \frac{dP}{dP'}(x, y, \delta) = \frac{\delta}{S_{C'}(y|x)}.$$

In survival analysis, the ratio $\delta/S_{C'}(y|x)$ is called IPCW (inverse of the probability of censoring weight) and $S_{C'}(y|x)$ can be estimated by using the Kaplan-Meier method, see Kaplan and Meier (1958).

4. Numerical Experiments

This section illustrates the impact of reweighting by the likelihood ratio on classification performances, as a special case of the general strategy presented in Section 3. Since the distribution shapes are unknown for real data, we infer that reweighting will have variable effectiveness, depending on the dataset. We detail here an experiment that uses the structure of ImageNet to illustrate reweighting with a stratified population and strata distribution bias or *strata bias*. The code of the experiments can be found at

<https://drive.google.com/drive/folders/1-tWJ4n4WyXuTza8dLPngyHSVprKUZFVJ?usp=sharing>.

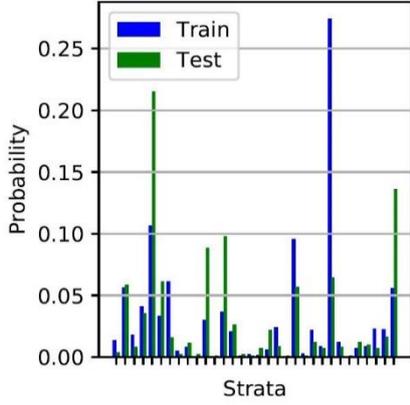
We focus on the *learning from biased stratified data* setting introduced in section 3.1 by leveraging the ImageNet Large Scale Visual Recognition Challenge (ILSVRC); a well-known benchmark for the image classification task, see Russakovsky et al. (2014) for more details.

The challenge consists in learning a classifier from 1.3 million training images spread out over 1,000 classes. Performance is evaluated using the validation dataset of 50,000 images of ILSVRC as our test dataset. ImageNet is an image database organized according to the WordNet hierarchy, which groups nouns in sets of related words called synsets. In that context, images are examples of very precise nouns, e.g. *flamingo*, which are contained in a larger synset, e.g. *bird*.

The impact of reweighting in presence of strata bias is illustrated on the ILSVRC classification problem with broad significance synsets for strata. To do this, we encode the data using deep neural networks. Specifically our encoding is the flattened output of the last convolutional layer of the network ResNet50 introduced in He et al. (2015). It was trained for classification on the training dataset of ILSVRC. The encodings X_1, \dots, X_n belong to a 2,048-dimensional space.

A total of 33 strata are derived from a list of high-level categories provided by ImageNet¹. The construction of the strata is postponed to the Appendix. By default, strata probabilities p_k and p_k' for $1 \leq k \leq K$ are equivalent between training and testing datasets, meaning that reweighting by Φ would have little to no effect. Since our testing data is the validation data of ILSVRC, we have around 25 times more training than testing data. Introducing a strata bias parameter $0 \leq \gamma \leq 1$, we set the strata train probabilities such that $p_k' = \gamma^{1-\lfloor K/2 \rfloor/k} p_k$ before renormalization and remove train instances so that the train set has the right distribution over strata; see the Appendix for more details on the generation of strata bias. When γ is close to one, there is little to no strata bias. In contrast, when γ approaches 0, strata bias is extreme.

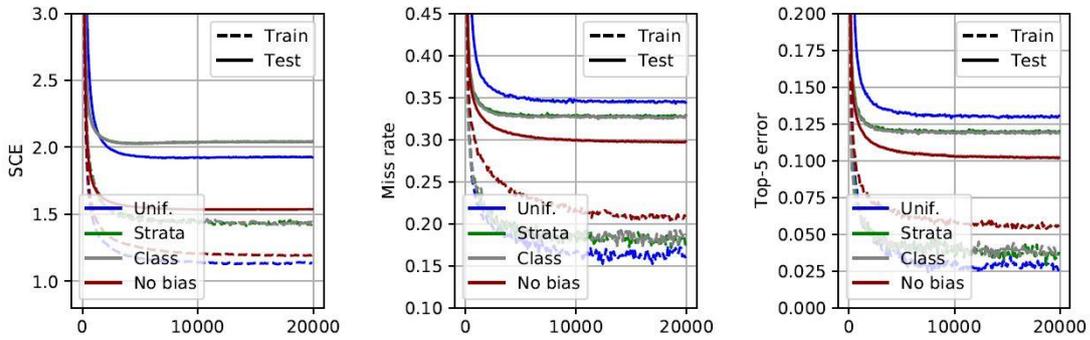
¹ <http://www.image-net.org/about-stats>



Comparison of p'_k 's and p_k 's

Model	Reweighting	Miss rate	Top-5 error
Linear	Unif. $\hat{\Phi} = 1$	0.344	0.130
	Strata $\hat{\Phi}$	0.329	0.120
	Class $\hat{\Phi}$	0.328	0.119
	No bias	0.297	0.102
MLP	Unif. $\hat{\Phi} = 1$	0.371	0.143
	Strata $\hat{\Phi}$	0.364	0.138
	Class $\hat{\Phi}$	0.363	0.138
	No bias	0.316	0.111

Figure 1: Results for the strata reweighting experiment with ImageNet.

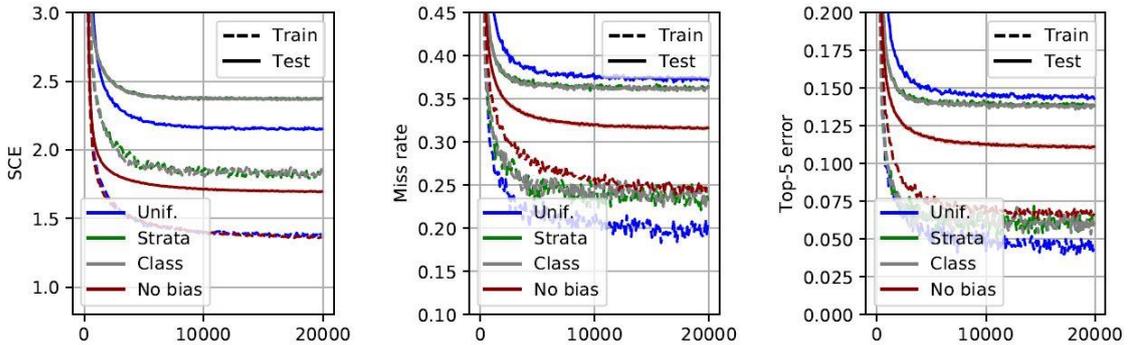


Dynamics for the SCE.

Dynamics for the miss rate.

Dynamics for the top-5 error.

Figure 2: Dynamics for the linear model for the strata reweighting experiment.



Dynamics for the SCE.

Dynamics for the miss rate.

Dynamics for the top-5 error.

Figure 3: Dynamics for the MLP model for the strata reweighting experiment.

The models used are a linear model and a multilayer perceptron (MLP) with one hidden layer; more details are given in the Appendix. We report better performance when reweighting using the strata information, compared to the case where the strata information is ignored, see [fig. 1](#). For comparison, we added two reference experiments: one which reweights the train instances by the class probabilities, which we do not know in a stratified population experiment, and one with more data and no strata bias

because it uses all of the ILSVRC train data. The dominance of the linear model over the MLP can be justified by the much higher number of parameters to estimate.

5. Conclusion

In this paper, we have considered specific transfer learning problems, where the distribution of the test data P differs from that of the training data, P' , and is absolutely continuous with respect to the latter. This setup encompasses many situations in practice, where the data acquisition process is not perfectly controlled. In this situation, a simple change of measure shows that the target risk may be viewed as the expectation of a weighted version of the basic empirical risk, with ideal weights given by the importance function $\Phi = dP/dP'$, unknown in practice. Throughout this article, we have shown that, in statistical learning problems corresponding to a wide variety of practical applications, these ideal weights can be replaced by statistical versions based solely on the training data combined with very simple information about the target distribution. The generalization capacity of rules learnt from biased training data by minimization of the weighted empirical risk has been established, with learning bounds. These theoretical results are also illustrated with several numerical experiments.

References

- Ausset, G., Cl  men  on, S., and Portier, F. (2019). Empirical risk minimization under random censorship: Theory and practice. *arXiv preprint arXiv:1906.01908*.
- Bach, F., Heckerman, D., and Horvitz, E. (2006). Considering cost asymmetry in learning classifiers.
- Bekker, J. and Davis, J. (2018). Beyond the selected completely at random assumption for learning from positive and unlabeled data. *CoRR*, abs/1809.03207.
- Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., and Wortman Vaughan, J. (2010). A theory of learning from different domains. *Machine Learning*, 79(1).
- Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., and Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *NIPS*, page 4349–4357.
- Boucheron, S., Bousquet, O., and Lugosi, G. (2005). Theory of classification : a survey of some recent advances. *ESAIM: Probability and Statistics*, 9:323–375.
- Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford.
- Burns, K., Hendricks, L. A., Saenko, K., Darrell, T., and Rohrbach, A. (2019). Women also snowboard: Overcoming bias in captioning models.

Chollet, F. et al. (2015). Keras. <https://keras.io>.

Cl  men  on, S. and Vayatis, N. (2009). Empirical performance maximization based on linear rank statistics. In *NIPS*, volume 3559 of *Lecture Notes in Computer Science*, pages 1–15. Springer.

Cl  men  on, S., Lugosi, G., and Vayatis, N. (2008). Ranking and Empirical Minimization of U-Statistics. *The Annals of Statistics*, 36(2):844–874.

Cortes, C., Mansour, Y., and Mohri, M. (2010). Learning bounds for importance weighting. In *Advances in neural information processing systems*, pages 442–450.

Cortes, C., Mohri, M., Riley, M., and Rostamizadeh, A. (2008). Sample selection bias correction theory. In *International conference on algorithmic learning theory*, pages 38–53. Springer.

Das, A., Dantcheva, A., and Bremond, F. (2018). Mitigating bias in gender, age and ethnicity classification: a multi-task convolution neural network approach. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Devroye, L., Gy  rfi, L., and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer.

du Plessis, M. C., Niu, G., and Sugiyama, M. (2014). Analysis of learning from positive and unlabeled data. In *NIPS*, pages 703–711.

du Plessis, M. C., Niu, G., and Sugiyama, M. (2015). Convex formulation for learning from positive and unlabeled data. In *ICML, ICML’15*, pages 1386–1394. JMLR.org.

Appendix - Technical Proofs

Here we detail the proofs of the results stated in the article and discuss their connection with related work.

Proof of Lemma 1

Let $\delta \in (0,1)$. Applying the classic maximal deviation bound stated in Theorem 3.2 of Boucheron et al. (2005) to the bounded class $\mathcal{K} = \{z \in \mathcal{Z} \mapsto \Phi(z)l(\theta, z): \theta \in \Theta\}$, we obtain that, with probability at least $1 - \delta$:

$$\sup_{\theta \in \Theta} |\widetilde{\mathcal{R}}_{w^*,n}(\theta) - \mathbb{E}[\widetilde{\mathcal{R}}_{w^*,n}(\theta)]| \leq 2\mathbb{E}[R'_n(\mathcal{K})] + \|\Phi\|_\infty \sup_{(\theta,z) \in \Theta \times \mathcal{Z}} |l(\theta, z)| \sqrt{\frac{2\log(1/\delta)}{n}}.$$

In addition, by virtue of the contraction principle, we have $R'_n(\mathcal{K}) \leq \|\Phi\|_\infty R'_n(\mathcal{F})$ almost-surely. The desired result can be thus deduced from the bound above combined with the classic bound

$$\mathcal{R}_P(\tilde{\theta}_n^*) - \min_{\theta \in \Theta} \mathcal{R}_P(\theta) \leq 2\sup_{\theta \in \Theta} |\widetilde{\mathcal{R}}_{w^*,n}(\theta) - \mathbb{E}[\widetilde{\mathcal{R}}_{w^*,n}(\theta)]|.$$

Proof of Lemma 2

Apply twice the Taylor expansion

$$\frac{1}{x} = \frac{1}{a} - \frac{x-a}{a^2} + \frac{(x-a)^2}{xa^2},$$

so as to get

$$\begin{aligned} \frac{1}{n_+/n} &= \frac{1}{p'} - \frac{n_+/n - p'}{p'^2} + \frac{(n_+/n - p')^2}{p'^2 n_+/n}, \\ \frac{1}{n_-/n} &= \frac{1}{1-p'} - \frac{n_-/n - 1 + p'}{(1-p')^2} + \frac{(n_-/n - 1 + p')^2}{(1-p')^2 n_-/n}. \end{aligned}$$

This yields the decomposition

$$\begin{aligned} &\frac{p}{p^2} \left(\frac{n_+}{n} - p' \right) \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{g(X'_i) = -1, Y'_i = +1\} \\ &- \frac{1-p}{(1-p')^2} \left(\frac{n_-}{n} - 1 + p' \right) \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{g(X'_i) = +1, Y'_i = -1\} \\ &+ \frac{p(n_+/n - p')(n_-/n - 1 + p')^2}{p'^2 n_+/n (1-p')^2 n_-/n} \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{g(X'_i) = +1, Y'_i = -1\}. \end{aligned}$$

We deduce that

$$|\widetilde{\mathcal{R}}_{\hat{w}^*,n}(g) - \widetilde{\mathcal{R}}_{w^*,n}(g)| \leq \frac{|n_+/n - p'|}{\varepsilon^2} \left(1 + |n_+/n - p'| \left(\frac{p}{n_+/n} + \frac{1-p}{1-n_+/n} \right) \right).$$

By virtue of Hoeffding inequality, we obtain that, for any $\delta \in (0,1)$, we have with probability larger than $1 - \delta$:

$$|n_+/n - p'| \leq \sqrt{\frac{\log(2/\delta)}{2n}},$$

so that, in particular, $\min\{n'_+/n, 1 - n'_+/n\} \geq \varepsilon - \sqrt{\log(2/\delta)/(2n)}$. This yields the desired result.

Proof of Corollary 1

Observe first that $\|\Phi\|_\infty \leq \max(p, 1 - p)/\varepsilon$ and

$$\mathcal{R}_P(\tilde{g}_n) - \inf_{g \in \mathcal{G}} \mathcal{R}_P(g) \leq 2 \sup_{g \in \mathcal{G}} |\tilde{\mathcal{R}}_{\hat{w}^*, n}(g) - \tilde{\mathcal{R}}_{w^*, n}(g)| + 2 \sup_{g \in \mathcal{G}} |\tilde{\mathcal{R}}_{w^*, n}(g) - \mathcal{R}_P(g)|.$$

The result then directly follows from the application of Lemmas 1-2 combined with the union bound.

Proof of Theorem 1

Observe first that $\|\Phi\|_\infty \leq \max_k p_k/\varepsilon$ and

$$\mathcal{R}_P(\tilde{\theta}_n^*) - \inf_{\theta \in \Theta} \mathcal{R}_P(\theta) \leq 2 \sup_{\theta \in \Theta} |\tilde{\mathcal{R}}_{\hat{w}^*, n}(\theta) - \tilde{\mathcal{R}}_{w^*, n}(\theta)| + 2 \sup_{\theta \in \Theta} |\tilde{\mathcal{R}}_{w^*, n}(\theta) - \mathcal{R}_P(\theta)|.$$

The result then directly follows from the application of Lemmas 1-3 combined with the union bound.

Lemma 3. Let $\varepsilon \in (0, 1/2)$. Suppose that $p'_k \in (\varepsilon, 1 - \varepsilon)$ for $k \in \{1, \dots, K\}$. For any $\delta \in (0, 1)$, we have with probability larger than $1 - \delta$:

$$\sup_{\theta \in \Theta} |\tilde{\mathcal{R}}_{\hat{w}^*, n}(\theta) - \tilde{\mathcal{R}}_{w^*, n}(\theta)| \leq \frac{2L}{\varepsilon^2} \sqrt{\frac{\log(2K/\delta)}{2n}},$$

as soon as $n \geq 2 \log(2K/\delta)/\varepsilon^2$, where $L = \sup_{(\theta, z) \in \Theta \times \mathcal{Z}} \ell(\theta, z)$.

Proof. Apply the Taylor expansion

$$\frac{1}{x} = \frac{1}{a} - \frac{x - a}{a^2} + \frac{(x - a)^2}{xa^2},$$

so as to get for all $k \in \{1, \dots, K\}$

$$\frac{1}{n'_k/n} = \frac{1}{p'_k} - \frac{n'_k/n - p'_k}{p'^2_k} + \frac{(n'_k/n - p'_k)^2}{p'^2_k n'_k/n}.$$

This yields the decomposition

$$\begin{aligned} & \tilde{\mathcal{R}}_{\hat{w}^*, n}(\theta) - \tilde{\mathcal{R}}_{w^*, n}(\theta) = \\ & \frac{1}{n} \sum_{i=1}^n \ell(\theta, Z'_i) \sum_{k=1}^K \mathbb{I}\{S'_i = k\} \left(-\frac{p_k}{p'^2_k} \left(\frac{n'_k}{n} - p'_k \right) + \frac{p_k (n'_k/n - p'_k)^2}{p'^2_k n'_k/n} \right). \end{aligned}$$

We deduce that

$$|\tilde{\mathcal{R}}_{\hat{w}^*, n}(\theta) - \tilde{\mathcal{R}}_{w^*, n}(\theta)| \leq \frac{L}{\varepsilon^2} \sum_{k=1}^K |n'_k/n - p'_k| p_k \left(1 + \frac{|n'_k/n - p'_k|}{n'_k/n} \right).$$

By virtue of Hoeffding inequality, we obtain that, for any $k \in \{1, \dots, K\}$ and $\delta \in (0, 1)$, we have with probability larger than $1 - \delta$:

$$|n'_k/n - p'_k| \leq \sqrt{\frac{\log(2/\delta)}{2n}},$$

so that, by a union bound, $\max_k \{n'_k/n\} \geq \varepsilon - \sqrt{\log(2K/\delta)/(2n)}$. This yields the desired result.

Proof of Theorem 2

Observe first that $\|\Phi\|_\infty \leq \max(2p, 1)/\varepsilon$ and

$$\mathcal{R}_P(\tilde{g}_n) - \inf_{g \in \mathcal{G}} \mathcal{R}_P(g) \leq 2 \sup_{g \in \mathcal{G}} |\tilde{\mathcal{R}}_{\tilde{w}^*,n}(g) - \tilde{\mathcal{R}}_{w^*,n}(g)| + 2 \sup_{g \in \mathcal{G}} |\tilde{\mathcal{R}}_{w^*,n}(g) - \mathcal{R}_P(g)|,$$

with weighted empirical risk $\tilde{\mathcal{R}}_{w^*,n}(g)$ defined in (11). The result then directly follows from the application of Lemmas 1-4 combined with the union bound.

Lemma 4. Let $\varepsilon \in (0, 1/2)$. Suppose that $q \in (\varepsilon, 1 - \varepsilon)$. For any $\delta \in (0,1)$, we have with probability larger than $1 - \delta$:

$$\sup_{g \in \mathcal{G}} |\tilde{\mathcal{R}}_{\tilde{w}^*,n}(g) - \tilde{\mathcal{R}}_{w^*,n}(g)| \leq \frac{2(2p+1)}{\varepsilon^2} \sqrt{\frac{\log(2/\delta)}{2n}},$$

as soon as $n \geq 2\log(2/\delta)/\varepsilon^2$.

Proof. Apply twice the Taylor expansion

$$\frac{1}{x} = \frac{1}{a} - \frac{x-a}{a^2} + \frac{(x-a)^2}{xa^2},$$

so as to get

$$\begin{aligned} \frac{1}{n_+'/n} &= \frac{1}{q} - \frac{n_+'/n - q}{q^2} + \frac{(n_+'/n - q)^2}{q^2 n_+'/n}, \\ \frac{1}{n_-'/n} &= \frac{1}{1-q} - \frac{n_-'/n - 1 + q}{(1-q)^2} + \frac{(n_-'/n - 1 + q)^2}{(1-q)^2 n_-'/n}. \end{aligned}$$

This yields the decomposition:

$$\begin{aligned} \tilde{\mathcal{R}}_{\tilde{w}^*,n}(g) - \tilde{\mathcal{R}}_{w^*,n}(g) &= -\frac{2p}{q^2} \left(\frac{n'_+}{n} - q \right) \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{g(X'_i) = -1, Y'_i = +1\} \\ &\quad - \frac{1}{(1-q)^2} \left(\frac{n'_-}{n} - 1 + q \right) \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{g(X'_i) = +1, Y'_i = -1\} \\ &\quad + \frac{2p(n'_+/n - q)^2}{q^2 n'_+} \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{g(X'_i) = -1, Y'_i = +1\} \\ &\quad + \frac{(n'_-/n - 1 + q)^2}{(1-q)^2 n'_-} \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{g(X'_i) = +1, Y'_i = -1\}. \end{aligned}$$

We deduce that

$$|\tilde{\mathcal{R}}_{\tilde{w}^*,n}(g) - \tilde{\mathcal{R}}_{w^*,n}(g)| \leq \frac{|n'_+/n - q|}{\varepsilon^2} \left(2p + 1 + |n'_+/n - q| \left(\frac{2p}{n'_+/n} + \frac{1}{1 - n'_+/n} \right) \right).$$

By virtue of Hoeffding inequality, we obtain that, for any $\delta \in (0,1)$, we have with probability larger than $1 - \delta$:

$$|n'_+/n - q| \leq \sqrt{\frac{\log(2/\delta)}{2n}},$$

so that, in particular, $\min\{n'_+/n, 1 - n'_+/n\} \geq \varepsilon - \sqrt{\log(2/\delta)/(2n)}$. This yields the desired result.

Appendix - Inaccurate Prior Information about the Test Distribution

As noticed in section 2, it may happen that the rate of positive instances in the target population is approximately known only. Suppose that our guess for p is \tilde{p} such that $|p - \tilde{p}| \leq \zeta$, with $\zeta \in (0,1)$. Denote by \tilde{P} the distribution over $\mathcal{X} \times \{-1, +1\}$ under which X is drawn from $\tilde{p}F_+ + (1 - \tilde{p})F_-$ and such that $\mathbb{P}_{(X,Y) \sim \tilde{P}}\{Y = 1 | X = x\} = \mathbb{P}_{(X,Y) \sim P}\{Y = 1 | X = x\} = \eta(x)$.

By a change of measure we have,

$$\mathbb{P}_{\tilde{P}}(Y \neq g(X)) = \mathbb{P}_P(Y \neq g(X)) + \mathbb{E}_P \left[\left(\frac{d\tilde{P}}{dP}(X, Y) - 1 \right) \mathbb{I}\{Y \neq g(X)\} \right],$$

which allows to bound the difference of the classification risks of g under P and \tilde{P} :

$$|\mathcal{R}_{\tilde{P}}(g) - \mathcal{R}_P(g)| \leq \mathbb{E}_P \left[\left| \frac{d\tilde{P}}{dP}(X, Y) - 1 \right| \right] = 2|\tilde{p} - p| \leq 2\zeta.$$

Appendix - Numerical Experiments

Strategy to induce bias in balanced datasets. In the real data experiment described in section 4, a strategy is used to induce class distribution bias or strata bias, since the data is uniformly distributed on strata for the train and test set. Since the experiment involves a small test dataset, it is kept intact, while we discard elements of the train dataset to induce bias between the train and test datasets. The bias is parameterized by a single parameter γ , such that when γ is close to one, there is little strata or class bias, while when γ approaches 0, bias is extreme.

The bias we induce is inspired by a power law, which is often used to model unequal distributions. The distribution on the strata of the train set is modified so that the generated train set follows a power law. Formally, the power law distribution $\{p_k'\}_{k=1}^K$ over $S \in \{1, \dots, K\}$, is defined for all $1 \leq k \leq K$ as $p_k' = \gamma \frac{1}{\sigma(k)^{\gamma/2}} p_k$ and then normalized to sum to 1, with σ is a random permutation in $\{1, \dots, K\}$.

To generate a train dataset with modality distribution $\{p_k'\}_{k=1}^K$, we sample instances from the original train data set $\mathcal{D}_n^\circ = \{(X_i', Y_i', S_i')\}_{i=1}^n$, where Y_i' is the class, S_i' is the strata. The generated train dataset is noted \mathcal{D}_n . First, we define candidates $\mathbb{I}_k = \{i \mid 1 \leq i \leq n, S_i' = k\}$ for each strata $k \in \{1, \dots, K\}$. Then we select one of the candidate sets \mathbb{I}_k with the probabilities p_k' 's, to remove one of its elements, selected at random, and place it in the train dataset \mathcal{D}_n . We repeat this operation until one of the candidate sets is empty. A more efficient implementation of this process was used in the provided code.

Models. We compare two models: a linear model and a multilayer perceptron (MLP) with one hidden layer of size 1,524. Given a classification problem of input \mathbf{x} of dimension \mathbf{d} with \mathbf{K} classes, precisely with $\mathbf{d} = 2048, \mathbf{K} = 1000$, a linear model simply learns the weights matrix $\mathbf{W} \in \mathbb{R}^{\mathbf{d} \times \mathbf{K}}$ and the bias vector $\mathbf{b} \in \mathbb{R}^{\mathbf{K}}$ and outputs logits $\mathbf{l} = \mathbf{W}^T \mathbf{x} + \mathbf{b}$. On the other hand, the MLP has a hidden layer of dimension $\mathbf{h} = \lfloor (\mathbf{d} + \mathbf{K})/2 \rfloor$ and learns the weights matrices $\mathbf{W}_1 \in \mathbb{R}^{\mathbf{d}, \mathbf{h}}, \mathbf{W}_2 \in \mathbb{R}^{\mathbf{h}, \mathbf{K}}$ and bias vectors $\mathbf{b}_1 \in \mathbb{R}^{\mathbf{h}}, \mathbf{b}_2 \in \mathbb{R}^{\mathbf{K}}$ and outputs logits $\mathbf{l} = \mathbf{W}_2^T \mathbf{h}(\mathbf{W}_1^T \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2$ where \mathbf{h} is the ReLU function, i.e. $\mathbf{h}: \mathbf{x} \mapsto \mathbf{max}(\mathbf{x}, \mathbf{0})$. The MLP model involves approximatively 5M (million) parameters, while the MLP model uses only 2M. The weight decay or l2 penalization for the linear model and MLP model are written $\mathcal{P} = \frac{1}{2} \|\mathbf{W}\|^2$ and $\mathcal{P} = \frac{1}{2} \|\mathbf{W}_1\|^2 + \frac{1}{2} \|\mathbf{W}_2\|^2$, respectively.

Cost function. The cost function is the Softmax Cross-Entropy (SCE), which is the most used classification loss in deep learning. Specifically, given logits $\mathbf{l} = (l_1, \dots, l_K) \in \mathbb{R}^{\mathbf{K}}$, the softmax function is $\gamma: \mathbb{R}^{\mathbf{K}} \rightarrow [0, 1]^{\mathbf{K}}$ with $\gamma = (\gamma_1, \dots, \gamma_K)$ and for all $k \in \{1, \dots, \mathbf{K}\}, \gamma_k: \mathbf{l} \mapsto \exp(l_k) / \sum_{j=0}^{\mathbf{K}} \exp(l_j)$. Given an instance with logits \mathbf{l} and ground truth class value y , the expression of the softmax cross-entropy $c(\mathbf{l}, y)$ is $\mathbf{c}(\mathbf{l}, y) = \sum_{k=1}^{\mathbf{K}} \mathbb{I}\{y = k\} \cdot \mathbf{log}(\gamma_k(\mathbf{l}))$.

The loss that is reweighted depending on the cases as described in is this quantity $c(\mathbf{l}, y)$. The loss on the test set is never reweighted, since the test set is the target distribution. The weights and bias of the model that yield the logits are tuned using backpropagation on this loss averaged on random batches of B elements of the training data summed with the regularization term $\lambda \cdot \mathcal{P}$ where λ is a hyperparameter that controls the strength of the regularization.

Preprocessing, optimization, parameters. The images of ILSVRC were encoded using the implementation of ResNet50 provided by the library *keras*, see Chollet et al. (2015), by taking the flattened output of the last convolutional layer.

Optimization is performed using a momentum batch gradient descent algorithm on batches of size 1,000, with a learning rate parameter of 0.001 and a momentum of 0.9, see Ruder (2016) for more details. The weight decay parameters λ were cross-validated by trying values on the logarithmic scale $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$ and then we tried more fine-grained values between the two best results, in practice 10^{-3} was best and 10^{-2} was second best so we tried $\{0.002, 0.003, 0.004, 0.005\}$. The standard deviation initialization of the weights $\sigma_0 = 0.01$ was chosen by trial-and-error to avoid overflows. The learning rate was fixed after trying different values to have fast convergence while keeping good convergence properties.

Stratified information for ImageNet. In this section, we detail the data preprocessing necessary to assign strata to the ILSVRC data. These were constructed using a list of 27 high-level categories found on the ImageNet website².

Each ILSVRC image has a ground truth low level synset, either from the name of the training instance, or in the validation textfile for the validation dataset, that is provided by the ImageNet website. The ImageNet API³ provides the hierarchy of synsets in the form of *is-a* relationships, e.g. *a flamingo is a bird*. Using this information, for each synset in the validation and training database, we gathered all of its ancestors in the hierarchy that were high-level categories. Most of the synsets had only one ancestor, which then accounts for one stratum. Some of the synsets had no ancestors, or even several ancestors in the table, which creates extra strata, either a *no-category* stratum or a strata composed of the union of several ancestors. The definitions of the strata can be requested to the API.

² <http://www.image-net.org/about-stats>

³ <http://image-net.org/download-API>

Ternary Generative Adversarial Networks

Ching-Hsun Tseng^a and Shin-Jye Lee^{a*}

^aInstitute of Management of Technology, National Chiao Tung University, Taiwan

*Corresponding Author:camhero@gmail.com

Abstract

As variety learning methods introducing, using deep learning structures to fix present problems is a prevalent option, image tasks especially. Among image distinguishing, convolutional networks (CNNs) have been seen as a vital feat and overwhelmed a series of methods during competitions. Recently, the semi-supervised learning, such as GAN, has also spread a different spectrum on unsupervised image classifications. In this paper, in order to offering a more robust solution, we propose the ternary generative adversarial networks (TGAN), which we draw a lesson from DCGAN, WGAN-GP, ACGAN, and Triple GAN. Different from above novel GANs, TGAN owns three structures, the generator, discriminator, and supervisor, and thus TGAN not only can fulfill the original duty of distinguishing fake or real images and producing images but also classifies images' label properly and sends loss to three structures to update properly toward low resolution images. Among our experiments and model comparisons, TGAN's structure can efficiently converge and offer a decent accuracy on label classification, as TGAN has a readily trainable ability on label distinguishing, compared with ACGAN. Most importantly, this structure can help to output more reasonable generated images than rival's samples. Moreover, giving above a series of experiments, except of inception score performance, TGAN brings better performance than competitors. We can also see TGAN's generated images outperformance others' generated images and witness a right generated effect toward specific images' category because TGAN also bring the test generated images accuracy to further level. Therefore, TGAN can truly produce fake images based on the mechanism of label loss and TGAN structure. We believe that if TGAN was stacked up enough layers and added more parameters, the performance will become much decent. Surely, the result and the mechanism of TGAN is still far from perfect or robust. Considering the constraint computer setting and much more unused novel skills, modifying current TGAN structure so we can propose improved TGAN is a deduced path in near future.

Keyword: Deep Learning, CNNs, GANs

1. Introduction

Convolutional neural networks (CNNs) (Krizhevsky, Sutskever, & Hinton, 2012), to begin with, is a milestone in image classifications field by taking spatial information into account and delivering these values into following deep neural networks, such as multi-layer perceptron. By stacking deep enough layers, CNNs can provide a much accuracy result than traditional methods among image tasks. The development of generative adversarial networks (GAN) (Goodfellow et al., 2014) have sprayed the light on unsupervised and semi-supervised learning, especially proposing a combination of CNNs and GAN structure. With generator learning noise and labels, the discriminator will receive more challenged distractions, which are fake images or information. By the discriminator learning distract information from generators, this structure can boost the prediction accuracy and allows people to use the trained discriminators models to distinguish image tasks if the discriminator embedded with an ability of output prediction label. In this regard, GAN allow models to learn present data with minor labeled data and still can structure a robust prediction. This means GANs can learn from each other infinitely without human involving. Further, it must be mentioned that the method of bring CNNs into GAN and then introduced the deep convolutional generative adversarial networks (DCGAN) (Radford, Metz, & Chintala, 2015), which builds the bridge between image learning and unsupervised learning. DCGAN enables stacking convolutions among GAN so we can practice unsupervised learning on image tasks. Therefore, we can leverage unlimited unlabeled images to improve discriminators with multi-convolutions because real world images are notoriously minor compared with unknown data. General speaking, the unstable result of GAN is an admitted issue. Although there are already a series of varietal GAN-based networks to prevent this dilemma, these methods basically have to dedicate relatively large calculating cost or propose a complex loss function to lead a proper learning path. On the other hand, avoiding overfitting in deep structure is also a big obstacle during training deep neuron networks, especially training such complex structure with considering spatial information. Generally, it is normal to see overfitting problems in deep learning, so the general solution is using dropout, which is discarding partial parameter to prevent model decaying, but the effect might only go to minor. Even there are presented GAN-based structures, the accuracy will easily be dimmed by low resolution and then the accuracy remains in under 30% or even worse. This causes the original meaning of non-supervised GAN become limited because GANs can only distinguish fake/real images rather than images labels. Even worse, the poor ability of distinguishing low-definition images will readily output blurry images. Therefore, aforementioned issues are waiting for being fixed.

In this paper, in order to address above problems, we propose a new three-unit generative adversarial network, Ternary Generative Adversarial Networks (TGAN). By TGAN, we mainly structure a third networks, which is called supervisor, after the discriminator so we separate the job of label prediction from the discriminator. Therefore, the only duty of the supervisor is to classify the right images' label toward either to real images or fake images. In order to pick up easy-trainable model and early-converged result, we practice a series of novel skills in deep learning structures, including weight initializations and advanced activation functions among each unit.

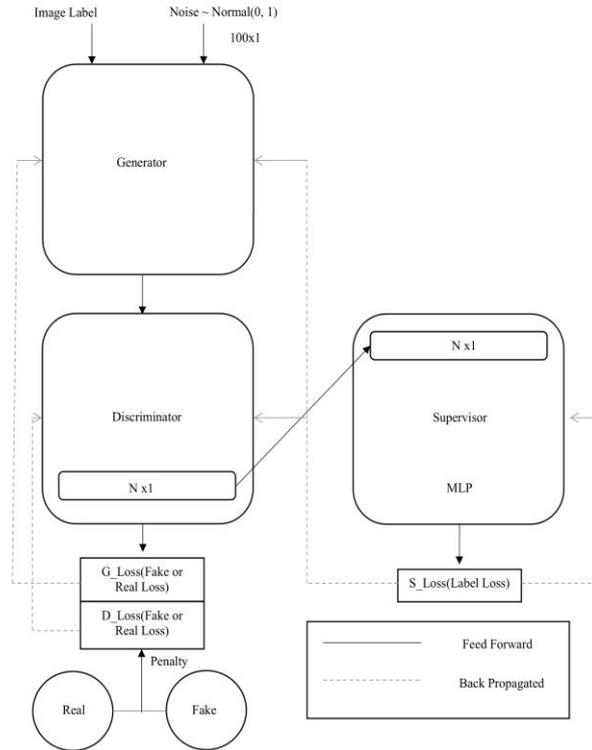


Fig. 1. A graphical abstract of the proposed structure

As the work flow illustrated, the noise is following the normal distribution $(0, 1)$ with 100 dimensions and then using this noise to combine the real label to produce fake images. Among producing fake images, the important convolutions will assign the weight to each convolution. Secondly, we assign the final hidden layer of the discriminator to our supervisor, whose duty only distinguish the images' label both from the generator and real images and thus the discriminator's duty is classifying true or fake. Therefore, the supervisor has to distinguish the images label toward to both fake and real images. Then, the supervisor's loss will be updated by backpropagation. Among backpropagation, we separate the loss of discriminator, which only contain fake/real loss, and supervisor, which only includes label loss. Also, we apply gradient penalty in discriminator by sampling images among real and fake images toward each label and then creating interpolates. The detail description of loss mechanism will be offered in the following sections. In the end, TGAN can update the loss toward fake/real and label properly so produce reasonable fake image to distract the discriminator in order to boost the prediction level. Among this paper, the introduction, motivation, purpose, and work flows are presented in this section. By section II, we will go through the related researches, which are motivated our work and help our work to better performance or even make our idea come true. In section III, the mechanism and crystal-clear work flows will be shown in this part. Then, the part IV will illustrate our experiment design and results, which contain the samples of our generator and prediction performance of our discriminator. Lastly, the conclusion will be suggested in the end, section V.

2. Related Work

2.1 Generative Adversarial Networks-based Structures

In order to implement an unsupervised learning, Generative Adversarial Networks (Goodfellow et al., 2014), introduce a novel way to train generative models. It constructs a two-player game in the model so two networks can compete each other and then spontaneously reach to better performance than a solo model. Later, the introducing of Conditional Generative Adversarial Nets (Mirza & Osindero, 2014) shows using label in the generator in MNIST data to help the generator output a decent and mimic image. On the other hand, another similar structure of Info-GAN (Chen et al., 2016) bring a GAN structure with the generator only being inputted with fake images. Stack-GAN (Zhang et al., 2017) bring an another advanced c-GAN with two c-GAN in this model to produce image from textures. AC-GAN (Odena, Olah, & Shlens, 2017) proposed a model, which can predict ten labels in the discriminators. We can also see a slightly same picture in Semi-Supervised-GAN (Odena, 2016). In this work (Mirza & Osindero, 2014), it shows how this model could be used to learn a multi-modal model, and provide preliminary examples of an application to image tagging. Therefore, we can apply semi-supervised structure in GAN because the generator literally uses the real data's label to produce fake information. Recently, some works have shown that GAN can offer reasonable image samples with low resolution (Denton, Chintala, & Fergus, 2015). Besides, the works (Sutskever, Vinyals, & Le, 2014) (Ramsundar et al., 2015) of feeding class into the generator also can ramp up the quality of picture, as these class lead the generator to produce human-readable picture instead of white noise, and performance of discriminator indirectly. On the other side of spectrum, in order to confront the inherit problem of GAN, WGAN (Arjovsky, Chintala, & Bottou, 2017) and WGAN-GP (Gulrajani, Ahmed, Arjovsky, Dumoulin, & Courville, 2017), proposed multiple methods, including replacing non-linear activation function with linear and getting rid of log to prevent model decaying on the flaws of shrinking distance between true and fake images. Conversely, although LSGAN (Mao et al., 2017) suggested using least square error to evaluate the performance can fix possible situations in GAN, the existing problems still remain.

2.2 Image Classification- Convolution Networks

Among image classification, before CNNs (Krizhevsky et al., 2012), it is hard to fix three-dimensional data, spatial data, into deep learning structure. With this big feature, we can see a lot of novel structure used extended-CNNs structure to leap the model's performance in ILVRC, including VGG (Simonyan & Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), and ResNet (He, Zhang, Ren, & Sun, 2016a) (He, Zhang, Ren, & Sun, 2016b). Most importantly, SE-Nets can boost above convolutional models with minor computational cost and ~25% improvement compared with time/s and accuracy in (Hu, Shen, & Sun, 2018). On the other hand, the deconvolutional networks, fractionally-strided convolutions, was firstly introduced in (Zeiler, Krishnan, Taylor, & Fergus, 2010). The structure is applied in unsupervised feature learning in (Zeiler, Taylor, & Fergus, 2011) (Zeiler &

Fergus, 2014) (Bengio, Courville, & Vincent, 2013) (Dosovitskiy & Brox, 2016), and are used in visualizing (Zeiler & Fergus, 2014), upsampling (Long, Shelhamer, & Darrell, 2015), semantic segmentation (Wei et al., 2017) (Long et al., 2015), learning high-level features (Wei et al., 2017), and generating fake images in DC-GAN' generator (Radford et al., 2015).

3. The Proposed Work – Ternary GAN

In our structure, we separate the function of prediction label from the discriminator and create a supervisor, which is a normal multilayer perceptron. By our proposing loss mechanism, the output can become easier trainable and relative robust GAN compared with current situations.

3.1 TGAN’s Generator

In this section, we want to illustrate proposed model’s structure with crystal-clear plots step-by-step. Firstly, we use the general structure. The only different part is using the recommend leaky-ReLU activation each layer activation function. The algorithm can be shown as:

$$x = G(LReLU(DeCNN(z, y))) \quad (1)$$

where the z follows the normal distribution $(0,1) \in \mathbb{R}^{100 \times 1}$, the y is the training data’s images’ label, x is the generated images, $DeCNN$ is deconvolution structure, $LReLU$ is Leaky ReLU.

3.2 TGAN’s Discriminator

Different from general structure of ACGAN or the general discriminators which distinguish label and Real images’ label, our discriminator only classify whether the input image is real or not by feeding generated images and real images into CNNs structure. Because we also draw a lesson from WGAN-GP, we use the recommended structure setting of WGAN-GP, which is using layer normalizer during the discriminator. Also, we apply Leaky ReLU as each layer’s activation function in CNNs. In the end, the last hidden layer will be sent to the supervisors as its input layer. So, the discriminator’s algorithm can be illustrated as:

$$\mathbb{E}_{x \sim p_r} [D(CNN(x))] \quad (2)$$

$$\mathbb{E}_{x \sim p_g} [D(CNN(x))] \quad (3)$$

where p_r means the data is from real images, p_g is from generated images, CNN means convolutional networks.

3.3 TGAN’s Supervisor

By our third unit, the supervisor, it is a typical multilayer perceptron neuron network. By picking up the last hidden layer from the discriminator and viewing it as the supervisor’s input layer, the label loss of supervisor’s output will influence the discriminator’s weight by backpropagation. The supervisor’s algorithm can be illustrated as below:

$$\mathbb{E}_{x \sim p_r} [S(MLP(x))] \quad (4)$$

$$\mathbb{E}_{x \sim p_g} [S(MLP(x))] \quad (5)$$

where p_r means the data is from real images, p_g is from generated images, MLP means multi-layer perceptron networks.

3.4 Loss Function of TGAN

Based on the above novel structure, basically every aspect of the inherit problems of deep structure and GAN have been taken into account in our proposed structure. However, the mode collapse is still a difficult challenge among cGAN training. Generally, the role of loss function is a vital factor. The original loss function of GAN is under below:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_r} [\log D(x|y)] + \mathbb{E}_{x \sim p_g} [\log(1 - D(G(x|y)))] \quad (6)$$

where D and G is the discriminator and the generator in the GAN, p_r is the probability from real images, p_g is the probability from generated images whose noise follow the uniform distribution $(0, 1)$ with $\mathbb{R}^{100 \times 1}$.

From above original loss function, the model only be evaluated the ability of distinguishing true and fake images but the reality usually needs the capability of predicting label from real and fraud information. Also, the well-known dilemma in GAN is hard-trained result, either in generator's output nor GAN's loss convergence. In WGAN, the authors clearly show that because of the GAN's non-reasonable KL divergence, GANs cannot update the generator weight to close to the discriminator. Also, as the gradient in sigmoid is too smooth in both sides and the images from real and fake images' low dimensional manifold too hard to overlap or even impossible by proof in WGAN's thesis. Aforementioned reasons bring a hard-trained GAN. In order to face the true problem toward loss function, which is also distance problem between both networks, we use gradient penalty in the discriminator's loss to limit the gradient.

Firstly, in the discriminator and generator, we simply use the loss function like below:

- Toward Discriminator:

$$L_{TGAN}(D) = -\mathbb{E}_{x \sim p_r} [D(x)] + \mathbb{E}_{x \sim p_g} [(1 - D(x))] \quad (7)$$

where D is the discriminator in TGAN, p_r is the probability from real images, p_g is the probability from generated images whose noise follow the normal distribution $(0, 1)$ with $\mathbb{R}^{100 \times 1}$. We want to minimize the difference between the real data and generated data as low as possible.

- Toward Generator:

$$L_{TGAN}(G) = -\mathbb{E}_{x \sim p_g} [D(x)] \quad (8)$$

where G is the generator in TGAN, p_g is the probability from generated images whose noise follow the normal distribution $(0, 1)$ with $\mathbb{R}^{100 \times 1}$. We want to minimize the $L_{TGAN}(G)$, and thus we hope the generated ability become better as possible.

- Toward Supervisor:

$$L_{TGAN}(S) = -\mathbb{E}_{x \sim p_r} [\log S(C = c|x)] - \mathbb{E}_{x \sim p_g} [\log S(C = c|x)] \quad (9)$$

where S is the supervisor in TGAN, p_r is the probability from real images, p_g is the probability from generated images whose noise follow the normal distribution $(0, 1)$ with $\mathbb{R}^{100 \times 1}$. We want to minimize the $L_{TGAN}(S)$, which means we hope the prediction can become close to real label toward to both real images and generated images.

3.5 Gradient Penalty in TGAN

As the method in WGAN-GP (Gulrajani et al., 2017), we simply import the gradient penalty into our discriminator's loss. Thus, our loss toward the discriminator goes to:

$$L_{TGAN}(D) = -\mathbb{E}_{x \sim p_r}[D(x)] + \mathbb{E}_{x \sim p_g}[(1 - D(x))] + \lambda \mathbb{E}_{x \sim p_{\hat{x}}}[\|\nabla_x D(x)\|_p - 1]^2 \quad (10)$$

where D is the discriminator in TGAN, p_r is the probability from real images, p_g is the probability from generated images whose noise follow the normal distribution $(0, 1)$ with $\mathbb{R}^{100 \times 1}$, \hat{x} is the sample among the real image and the fake image, $p_{\hat{x}}$ is the probability among the real image and the fake image. Similarly, we want $L_{TGAN}(D)$ as low as possible. On the other hands, λ is generally set 10. With higher λ , the loss and training history will be more stable but the output will become fuzzier in images and hard to lead the accuracy to right prediction. By lower λ , the fluctuation will become more severe and then we might easily see a mode collapse. Aforementioned situations are related to the λ which affects the impact of gradient penalty.

3.6 Demonstrating Structure of TGAN

Among our proposing structure, ternary generative adversarial networks, our goal is making each networks to learn from loss properly to help sample images to access closely to real image and fulfill the duty of predict label accurately. With the favor of gradient penalty in both discriminator and supervisor, we can further make sure the over-sloped gradient won't update the weight falsely. Finally, the proposing loss function is experimentally helpful for preventing model decaying. Our proposing structure's detail work flows is illustrated in Fig. 2.

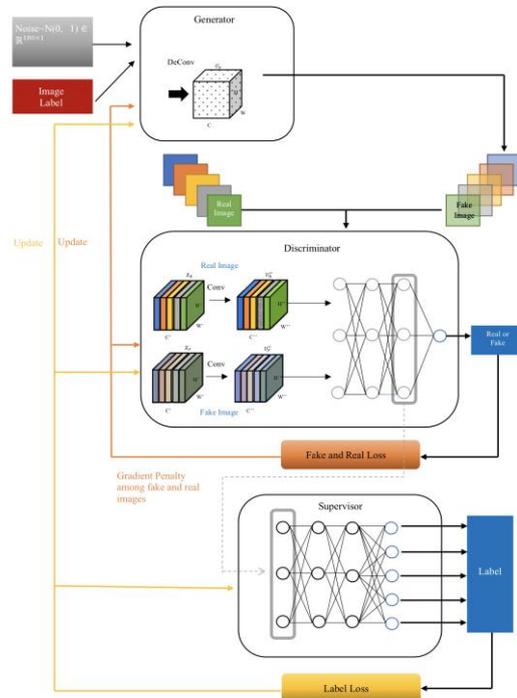


Fig. 2. The structure of proposed TGAN

4. Experiments and Results

4.1 Experiments' Description and Design

While our proposing model's, Ternary Generative Adversarial Networks (TGAN), mechanisms and work flows are shown in above sections, we still have to examine our model's performance by revealing each network's loss among TGAN and sampling generated images. To begin with, in order to make our model enclose reality situations, we use cifar-10 as our training and testing data, which is a dataset contains 10 kinds of transportation and animals, such as airplanes, horses, boats, flogs, etc., with 5000 in training data and 1000 in testing data. Secondly, we separate our comparison into two parts, (1) comparing TGAN's label classifying performance in accuracy with ACGAN and modified ACGANs, which is ACGAN with WGAN-GP, (2) sampling generated images from aforementioned GANs and evaluating these images. In our comparison, we reimplement aforementioned models and compare the performance with same structures as close as possible. Most importantly, as training GAN-based model is highly computing costing, we have to mention our advice setup is AMD Ryzen 7 2700X Eight-Core Processor 3.70 GHz, 32.0 GB RAM, and Nvidia RTX 2070 8GB.

4.2 The Comparison of Performance of Images' Label Classification

In this part, we compare the ability of distinguishing images' label in GANs. The following Table is the accuracy comparison toward real training images, real testing images, and fake(generated) testing images. The value in the bracket are the improving ratio compared with ACGAN. Every recorded value is the best figure in the last 100 epochs. When it comes to each index purpose, training images accuracy is evaluating the prediction ability in existing images. Testing images accuracy is for assessing the prediction ability in unknown images. Lastly, testing generated images accuracy is bound to evaluate whether the GAN can generate the proper fake images by learning the feature in testing images because producing images which cannot be recognized by discriminator also could be viewed as an unpersuaded learning result.

From the Table. 1., ACGAN has a decent performance in classifying real training images, 92.7%. However, we witnessed a severe overfitting problem because the testing images accuracy is apparently lower than training images accuracy with 23.4% difference. We can see a degradation when we apply gradient penalty and earth mover distance in ACGAN. The possible explanation for this result could be the loss of simply using linear activation function, earth mover distance, and gradient penalty unable update the weight toward the label distinguishing ability properly. In our proposing model, TGAN, while the training images accuracy is lower, 79% with -19% difference ration, TGAN put off the overfitting problem because there is not much difference between training images accuracy (79%) and testing images accuracy (69.3%), with only 5.5% difference. When we compared the former difference in training and testing, there is. 76.5% improvement in overfitting problem. Therefore, above situation hints the improvement in label prediction toward real images. Also, TGAN not only maintain the good performance in distinguish fake images' label but also escalated to 95.8%. From Table.2. and Fig. 4.,

applying novel skills in TGAN not only brings the much better accuracy performance, we can also see that the testing images accuracy already reach over 70% by 3% improvement compared with ACGAN. Therefore, the improving version TGAN is used as our final proposed model of TGAN. The comparison table and bar chart are under below:

TABEL. 1. Accuracy comparison toward each GANs in training images accuracy, testing images accuracy, and testing generated images accuracy.

Accuracy Comparison		Index		
		Training Images Accuracy	Testing Images Accuracy	Testing Generated Images Accuracy *
Models	ACGAN	92.7%	69.3%	94.5%
	ACGAN (with WGAN-GP)	37.8% (-59%)	35.5% (-49%)	13.3% (-86%)
	TGAN	74.8% (-19%)	69.3% (0%)	95.8% (1%)

*Fake Image Accuracy

ABEL. 2 Improved TGAN compares with each targeted GANs

Accuracy comparison		Index		
		Training Images Accuracy	Testing Images Accuracy	Testing Generated Images Accuracy*2
Models	ACGAN	92.7%	69.3%	94.5%
	ACGAN (with WGAN-GP)	37.8% (-59%)	35.5% (-49%)	13.3% (-8%)
	TGAN (Original)	74.8% (-19%)	69.3% (0%)	95.8% (1%)
	TGAN (Improving)*1	77% (-17%)	71.5% (3%)	96.7% (2%)

- *1 The TGAN (cross entropy-based supervisor) embedded with PReLU, He-value, and Xavier-value.
- *2 Fake Image Accuracy

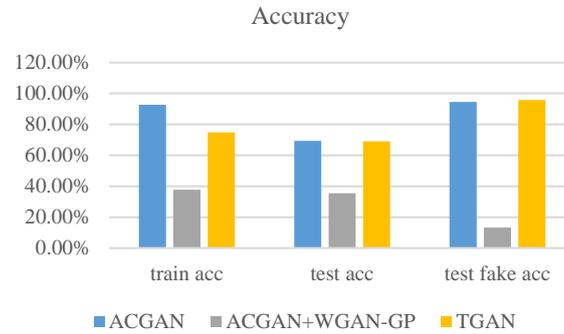


Fig. 3. Accuracy comparison in ACGAN, ACGAN+WGAN-GP, and TGAN

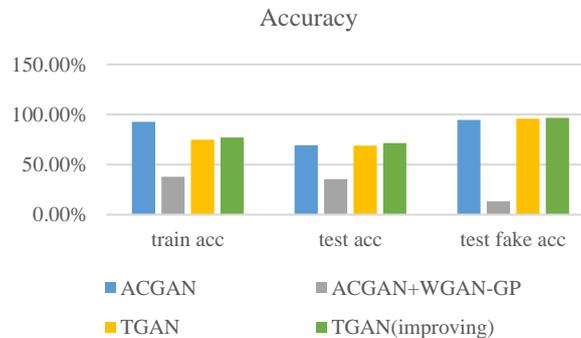


Fig. 4. Improved TGAN compares with each targeted GANs

In the following comparisons, the improving TGAN is regard as our proposed model.

4.3 The Comparison of Generated Images Performance

In order to evaluating the generated images, the general way is using artificial method, which is using eye and intuitive sense to evaluate, so we sample out the generated images from each comparing GANs. Nonetheless, only using intuition to assessment is not objective. As a result, we also apply popular assessing method, inception score, to evaluate our generated images performance. First of all, we sample the generated images from ACGAN as our standard. The sampling is as the Fig. 5.

In using WGAN-GP in ACGAN, Fig. 6., although we pick up the best inception score in this combination, we have to admit that the generated images might be really fuzzy and ambiguous. What have to mention is we sample the images which is examined the highest inception score among the last 100 epochs. Even with these unpromising results, we still tell a little bit that the 1st might be horses, 5th might be airplanes, 7th might hint to flogs. Others are really blurry compared with the results of proposed models in the following discussion. Lastly, in Fig. 7., sampling from our proposed TGAN

also see a decent result as the picture under below. Except the rows of 3 and 4, we can readily see what the object among each row of generated images is. In first row, we can see they are airplanes from long distance. The second row might hint to cars. Also, the 6th row are dogs, 7th are flogs, 8th are horses, 9th are bots, and 10th are trucks. Therefore, although the result of TGAN might not be clear or high resolution, the targeted object in each generated image is quite obvious.



Fig. 5. The generated images from ACGAN



Fig. 6. The generated images from ACGAN with WGAN-GP



Fig. 7. The generated images from TGAN

4.4 Inception Score toward Each GANs' Samples

It is true that using intuition to evaluate the generated images has its own merit. While following general method is easier, it is hard to prevent the standard goes to subjective and biased. Therefore, current novel method is using inception net, which is proposed by Google, to assessment the images' mean as an objective and unbiased evaluation. What has to be mentioned is the inception score only means an objective reference value instead of an absolute index. The whole recoded figure is under the comparison table:

From the comparison in Table. 3. and Fig. 8., we use ACGAN as our standard in this part experiment. Toward ACGAN+WGAN-GP, we receive highest inception score mean among our reimplement result, ~2.83. Obviously, using earth mover distance and proper gradient penalty can truly shrink the distance among real and generated data's weight in the discriminator and generator respectively. In the TGAN's inception score outcome, TGAN maybe not the highest value in this index, thought, TGAN still outshine ACGAN performance especially the images generated quality in the above samples' discussions. While TGAN's inception score doesn't transcend the ACGAN+WGAN-GP's, the difference between them is only 0.033, which is pretty minor and hints a close performance.

TABEL. 3. Inception score comparison toward each GANs with standard deviation.

Inception score comparison		Index
		Inception Score Mean
Models	ACGAN	2.6294839
	ACGAN+WGAN-GP	2.8292003
	TGAN	2.7964336

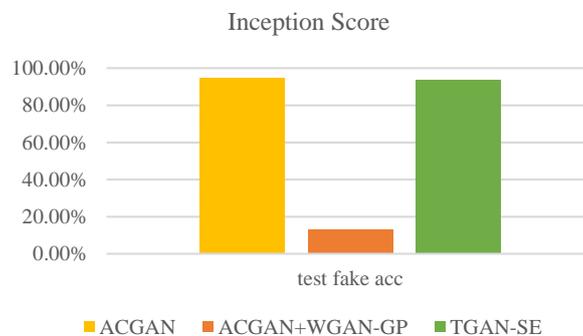


Fig. 8. Inception score comparison in each GANs. We can see that although the combination of ACGAN+WGAN-GP picks up the best performance, the TGAN's images' quality is admittedly close to this combination.

5. Conclusion and Future Works

Giving above a series of experiments, except of inception score performance, TGAN brings better performance than competitors. We can also see TGAN's generated images outperform others' generated images and witness a right generated effect toward specific images' category because TGAN also bring the test generated images accuracy to further level. Therefore, TGAN can truly produce fake images based on the mechanism of label loss and TGAN structure. We believe that if TGAN was stacked up enough layers and added more parameters, the performance will become much decent. Surely, the result and the mechanism of TGAN is still far from perfect or robust. Considering the constraint computer setting and much more unused novel skills, modifying current TGAN structure so we can propose improved TGAN is a deduced path in near future.

References

- Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein gan. *arXiv preprint arXiv:1701.07875*.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1798-1828.
- Chen, X., Duan, Y., Houthoofd, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). *Infogan: Interpretable representation learning by information maximizing generative adversarial nets*. Paper presented at the Advances in neural information processing systems.
- Denton, E. L., Chintala, S., & Fergus, R. (2015). *Deep generative image models using a laplacian pyramid of adversarial networks*. Paper presented at the Advances in neural information processing systems.
- Dosovitskiy, A., & Brox, T. (2016). *Inverting visual representations with convolutional networks*. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). *Generative adversarial nets*. Paper presented at the Advances in neural information processing systems.
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). *Improved training of wasserstein gans*. Paper presented at the Advances in Neural Information Processing Systems.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016a). *Deep residual learning for image recognition*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016b). *Identity mappings in deep residual networks*. Paper presented at the European conference on computer vision.

- Hu, J., Shen, L., & Sun, G. (2018). *Squeeze-and-excitation networks*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks*. Paper presented at the Advances in neural information processing systems.
- Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., & Paul Smolley, S. (2017). *Least squares generative adversarial networks*. Paper presented at the Proceedings of the IEEE International Conference on Computer Vision.
- Mirza, M., & Osindero, S. (2014). Conditional generative adversarial networks. *Manuscript: <https://arxiv.org/abs/1709.02023>*, 9, 24.
- Odena, A. (2016). Semi-supervised learning with generative adversarial networks. *arXiv preprint arXiv:1606.01583*.
- Odena, A., Olah, C., & Shlens, J. (2017). *Conditional image synthesis with auxiliary classifier gans*. Paper presented at the Proceedings of the 34th International Conference on Machine Learning-Volume 70.
- Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.
- Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., & Pande, V. (2015). Massively multitask networks for drug discovery. *arXiv preprint arXiv:1502.02072*.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). *Sequence to sequence learning with neural networks*. Paper presented at the Advances in neural information processing systems.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). *Going deeper with convolutions*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Wei, Y., Liang, X., Chen, Y., Shen, X., Cheng, M.-M., Feng, J., . . . Yan, S. (2017). Stc: A simple to complex framework for weakly-supervised semantic segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(11), 2314-2320.
- Zeiler, M. D., & Fergus, R. (2014). *Visualizing and understanding convolutional networks*. Paper presented at the European conference on computer vision.
- Zeiler, M. D., Krishnan, D., Taylor, G. W., & Fergus, R. (2010). *Deconvolutional networks*. Paper presented at the Cvpr.
- Zeiler, M. D., Taylor, G. W., & Fergus, R. (2011). *Adaptive deconvolutional networks for mid and high level feature learning*. Paper presented at the ICCV.

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. N. (2017). *Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks*. Paper presented at the Proceedings of the IEEE International Conference on Computer Vision.

Thresholding Bandit for Dose-ranging: The Impact of Monotonicity

Garivier^{a*}, Ménard^b and Rossi^c

^aUMPA, LIP, ENS de Lyon, France

^bSequel, INRIA Lille, France

^cIMT, Univ. Toulouse 3, France

*Corresponding Author: aurelien.garivier@ens-lyon.fr

Abstract

Purpose- We address an active learning framework where several distributions can be sequentially sampled, and where the distribution whose expectation is the closest to a given threshold must be identified. We study two frameworks: the general case, and the case where the expectations are given in increasing order. The later setting is notably motivated by phase 1 clinical trials, a setting where the maximal dose with acceptable toxicity needs to be found (toxicity is known to increase with the dose). Despite its practical importance, no effective method is known for this problem. Classical approaches are based on dose escalation, and the most well-known is the "traditional 3+3 Design". The problem is here considered as a "thresholding bandit problem", but with a particular structure that has never been studied before.

Results- This paper provides an information-theoretic answer by identifying the exact sample complexity of the problem, that is the number of samples required before any algorithm may find the right answer with a certain confidence. The complexity is computed with and without the monotonicity assumption. In addition, in each case we propose and analyse an algorithm whose sample complexity is of the same order of magnitude. The computational complexity of the resulting algorithm is discussed.

Methods- These results are obtained by building on isotonic regression and on the recent progress in optimal best-arm identification in multi-armed bandit problems. The complexity comes as a solution to an optimization problem, for which an original solution is provided and analyzed.

Conclusion/discussion- We thus provide a complete treatment of the problem in the case where the number of provided distributions is not too large, and when the distributions can be sampled without constraint. These promising results are a key step for further developments dealing with many distributions and/or constrained sampling strategies (like escalation methods in clinical trials).

Keyword: sequential learning, multi-armed bandits, thresholding bandits, best arm identification, unimodal regression, isotonic regression

1. Introduction

The motivation of this work originally comes from an attempt to contribute to the theory of clinical trials, which we use here as a running example to explain the model that we study and the solution that we propose.

An important step of phase 1 (and sometimes early phase 2) clinical trials is 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 target tolerance level is chosen (typically 33%), 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 et al. \(2009\)](#) and [Genovese et al. \(2013\)](#) for and references therein for an introduction.

We propose in this paper a complexity analysis for a simple model of phase 1 trials, which captures the essence of this problem and which can be identified in machine learning literature as a fixed-confidence thresholding bandit problem. We assume that the possible doses are $x(1) < \dots < x(K)$, 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 characterizes the toxicity level of dose $x(k)$. To avoid obfuscating technicalities, we treat here 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 $\mu(k)$ 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 the fixed budget setting. Contrary to previous work, we focus here 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 $\mu(1) < \mu(2) < \dots < \mu(K)$. Under this assumption, we prove another lower bound on the sample complexity, and provide an algorithm matching it. The sample complexity does not take a simple form (like a sum of inverse squares), but identifying it exactly is essential even in practice, since it is the only way known so far to construct an algorithm which reaches the lower bound. We are thus able to quantify, for each

problem, how many samples can be spared when means are sorted, at the cost of a slight increase in the computation cost of the algorithm.

Connections to the State of the Art.

Phase 1 clinical trials have been an intense field of research in the statistical community (see [Le Tourneau et al. \(2009\)](#) and references therein), but not considered as a sequential decision problem using the tools of the bandit literature. The important progress made in the recent years in the understanding of bandit models has made it possible to shed a new light on this issue, and to suggest very innovative solutions. Some technical connections are to be found with [Combes & Proutiere \(2014\)](#), which deals with regret minimization for unimodal bandits. The closest contribution are the works of [Locatelli et al. \(2016\)](#) and [Chen et al. \(2014\)](#), which provides a general framework for combinatorial pure exploration bandit problems.

The present work tackles the more specific issue of phase 1 trials. It aims at providing strong foundations for such solutions: it does not yet tackle all the ethical and practical constraints. Observe that it might also be relevant to look for the highest dose with toxicity below the target level: we discuss this variant in Section 4; however, it seems that practitioners do not consider this alternative goal in priority.

From a technical point of view, the approach followed here extends the theory of Best-Arm Identification initiated by [Kaufmann et al. \(2016\)](#) to a different setting. Building on the mathematical tools of that paper, we analyze the characteristic time of a thresholding bandit problem with and without the assumptions that the means are increasing. Computing the complexity with such a structural constraint on the means is a challenging task that had never been done before. It induces significant difficulties in the theory, but (by using isotonic regression) we are still able to provide a simple algorithm for computing the complexity term, which is of fundamental importance in the implementation of the algorithm. The computational complexity of the resulting algorithm is discussed in Section 3.1.

Organization. The lower bounds on complexity are presented in Section 2. We compare the complexities of the non-monotonic case versus the increasing case $K=2$. This comparison is particularly simple and enlightening when $\delta > 0$, 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 discusses the interesting, but simpler variant of the problem where the goal is to identify the

arm with mean closest, but also below the threshold.

Section 5 summarizes further possible developments, and precedes most of the technical proofs which are given in appendix.

1.1 Notation and Setting

For $K \geq 2$, we consider a Gaussian bandit model $(\mathcal{N}(\mu_1, 1), \dots, \mathcal{N}(\mu_K, 1))$, which we unambiguously refer to by the vector of means $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$. 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 $S \in \mathbb{R}$ is given, and we denote by $a_{\boldsymbol{\mu}}^* \in \arg \min_{1 \leq a \leq K} |\mu_a - S|$ any optimal arm.

Let \mathcal{M} be the set of Gaussian bandit models with a unique optimal arm and $\mathcal{I} = \{\boldsymbol{\mu} \in \mathcal{M} : \mu_1 < \dots < \mu_K\}$ 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 $A_t \in \{1, \dots, K\}$ and receives a conditionally independent reward $Y_t \sim \mathcal{N}(\mu_{A_t}, 1)$. Let $\mathcal{F}_t = \sigma(A_1, Y_1, \dots, A_t, Y_t)$ 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** $(A_t)_{t \geq 1}$, where A_t is \mathcal{F}_{t-1} -measurable,
- a **stopping rule** τ_{δ} , which is a stopping time with respect to the filtration $(\mathcal{F}_t)_{t \geq 1}$,
- a $\mathcal{F}_{\tau_{\delta}}$ -measurable **decision rule** $\hat{a}_{\tau_{\delta}}$.

For any setting $\mathcal{S} \in \{\mathcal{M}, \mathcal{I}\}$ (the non-monotonic 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 $\mathbb{P}_{\boldsymbol{\mu}}(\hat{a}_{\tau_{\delta}} \neq a_{\boldsymbol{\mu}}^*) \leq \delta$.

2. Lower Bounds

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

$$\text{Alt}(\boldsymbol{\mu}, \mathcal{S}) := \{\boldsymbol{\lambda} \in \mathcal{S} : a_{\boldsymbol{\lambda}}^* \neq a_{\boldsymbol{\mu}}^*\}, \quad (1)$$

and we denote the probability 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 $\boldsymbol{\mu} \in \mathcal{S}$,*

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau_{\delta}] \geq T_{\mathcal{S}}^*(\boldsymbol{\mu}) \text{kl}(\delta, 1 - \delta), \quad (2)$$

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

$$T_{\mathcal{S}}^*(\boldsymbol{\mu})^{-1} = \sup_{\omega \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}. \quad (3)$$

In particular, this implies that $\liminf_{\delta \rightarrow 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_{\delta}]}{\log(1/\delta)} \geq T_{\mathcal{S}}^*(\boldsymbol{\mu})$.

This result is a generalization of Theorem 1 of [Garivier & Kaufmann \(2016\)](#): the classical Best Arm Identification problem is a particular case of our non-monotonic setting $\mathcal{S} = \mathcal{M}$ with an infinite threshold $S = +\infty$. It is proved along the same lines. As [Garivier & Kaufmann \(2016\)](#), one proves 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

$$\omega^*(\boldsymbol{\mu}) := \arg \max_{\omega \in \Sigma_K} \inf_{\lambda \in \text{Alt}(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}. \quad (4)$$

This result is a generalization of Theorem 1 of [Garivier & Kaufmann \(2016\)](#): the classical Best Arm Identification problem is a particular case of our non-monotonic setting $\mathcal{S} = \mathcal{M}$ with an infinite threshold $S = +\infty$. It is proved along the same lines. As [Garivier & Kaufmann \(2016\)](#), one proves 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

$$\omega^*(\boldsymbol{\mu}) := \arg \max_{\omega \in \Sigma_K} \inf_{\lambda \in \text{Alt}(\boldsymbol{\mu}, \mathcal{S})} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}. \quad (4)$$

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 1. When $K = 2$,

$$T_{\mathcal{I}}^*(\boldsymbol{\mu})^{-1} = \frac{(2S - \mu_1 - \mu_2)^2}{8}, \quad (5)$$

$$T_{\mathcal{M}}^*(\boldsymbol{\mu})^{-1} = \frac{\min((2S - \mu_1 - \mu_2)^2, (\mu_1 - \mu_2)^2)}{8}. \quad (6)$$

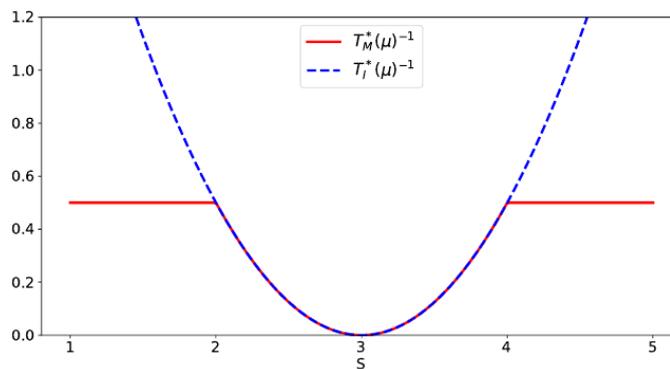


FIGURE 1. Inverse of the characteristic times as a function of the threshold S , for $\boldsymbol{\mu} = [2, 4]$. Solid red: general thresholding case ($\mathcal{S} = \mathcal{M}$). Dotted blue: increasing case ($\mathcal{S} = \mathcal{I}$).

Proof. The Equality (6) is a simple consequence of Lemma 1 proved in Section A.2. It remains to treat the first Equality (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 {arm 1 is optimal} $\Leftrightarrow m > S$ and that {arm 2 optimal} $\Leftrightarrow m < S$, we obtain:

$$T_{\mathcal{I}}^*(\boldsymbol{\mu})^{-1} = \sup_{\omega \in [0,1]} \inf_{\{\mu'_1 < \mu'_2, |S - \mu'_1| < |S - \mu'_2|\}} \frac{\omega}{2}(\mu_1 - \mu'_1)^2 + \frac{1-\omega}{2}(\mu_2 - \mu'_2)^2 = \sup_{\omega \in [0,1]} A(\omega),$$

where $m' = (\mu'_1 + \mu'_2)/2$, $\Delta' = \mu'_2 - \mu'_1$ and we denote by $A(\omega)$ the function

$$A(\omega) := \inf_{\{\Delta' > 0, m' > S\}} \frac{\omega}{2}(m - m' - (\Delta - \Delta')/2)^2 + \frac{1-\omega}{2}(m - m' + (\Delta - \Delta')/2)^2.$$

Writing $\chi = S - m$, easy computations lead to

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

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

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 $\overline{\text{Alt}(\boldsymbol{\mu}, \mathcal{I})}$ (the closure of $\text{Alt}(\boldsymbol{\mu}, \mathcal{I})$) 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 & Kaufmann (2016)): $\boldsymbol{\lambda} = [(\mu_1 + \mu_2)/2, (\mu_1 + \mu_2)/2]$. Thus, if $\mu_1 \leq S \leq \mu_2$, the two characteristic times coincide, as can be seen on Figure 1.

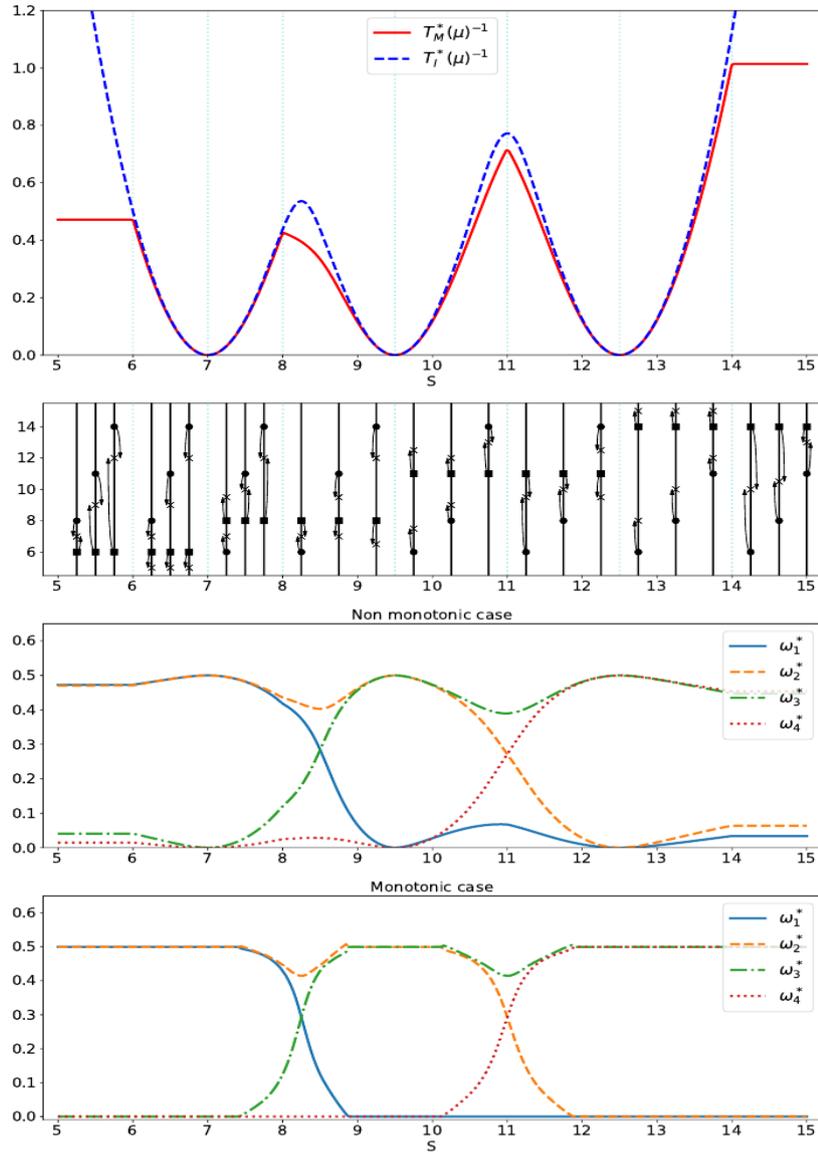


FIGURE 2. The complexity terms in the bandit model $\mu = (6, 8, 11, 14)$. Top: inverse of the characteristic time c as a function of the threshold S ; red solid line: non-monotonic case $S = \mathcal{M}$; blue dotted line: increasing case $S = \mathcal{I}$. Middle: how to move the means to get from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

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 $K \geq 2$ (see Figure 2). Since $\mathcal{I} \subset \mathcal{M}$, it is obvious that $T_{\mathcal{I}}^*(\boldsymbol{\mu}) \leq T_{\mathcal{M}}^*(\boldsymbol{\mu})$. 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.

On the structure of the optimal weights in the non-monotonic case. In the non-monotonic case $\mathcal{S} = \mathcal{M}$, there are two types of optimal alternatives (as in Section 2.1). Indeed, the proof of Lemma 1 in Appendix A.2 shows that the best alternative takes one of the two following forms. Either the optimal arm $\mu_{a_{\boldsymbol{\mu}}^*}$ 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 $(\mu_{a_{\boldsymbol{\mu}}^*} - \mu_b)^2$ in Equation (1). Or, as in the increasing case $\mathcal{S} = \mathcal{I}$ (see the proof of Proposition 1), both arms $\mu_{a_{\boldsymbol{\mu}}^*}$ and μ_b are translated in the same direction, leading to the constant $(2S - \mu_{a_{\boldsymbol{\mu}}^*} - \mu_b)^2$. 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.

Lemma 1. For all $\boldsymbol{\mu} \in \mathcal{M}$, $T_{\mathcal{M}}^*(\boldsymbol{\mu})^{-1} = \max_{\omega \in \Sigma_K} \min_{b \neq a^*} \frac{\omega_{a_{\boldsymbol{\mu}}^*} \omega_b}{2(\omega_{a_{\boldsymbol{\mu}}^*} + \omega_b)} \min((\mu_{a_{\boldsymbol{\mu}}^*} - \mu_b)^2, (2S - \mu_{a_{\boldsymbol{\mu}}^*} - \mu_b)^2)$.

On the structure of the optimal weights in the increasing case. In the increasing case $\mathcal{S} = \mathcal{I}$, one can show the remarkable property that *the optimal weights $\omega^*(\boldsymbol{\mu})$ put mass only on the optimal arm and its two closest arms*. This strongly contrasts with the non-monotonic case, as illustrated at the bottom of Figure 2. For simplicity we assume that $1 < a_{\boldsymbol{\mu}}^* < K$. Let $\tilde{\omega}$ be some weights in Σ_3 . Let $D^+(\theta, \tilde{\omega})$ be the cost, with weights $\tilde{\omega}$, for moving from the initial bandit problem $\boldsymbol{\mu}$ to a bandit problem $\tilde{\boldsymbol{\lambda}}^+$ where arm $a_{\boldsymbol{\mu}}^*$ has mean $\theta \leq S$ and S is halfway between $\mu_{a_{\boldsymbol{\mu}}^*}$ and $\mu_{a_{\boldsymbol{\mu}}^*+1}$,

$$\tilde{\lambda}_a^+ = \begin{cases} \mu_a & \text{if } a > a_{\boldsymbol{\mu}}^* + 1, \\ 2S - \theta & \text{if } a = a_{\boldsymbol{\mu}}^* + 1, \\ \theta & \text{if } a = a_{\boldsymbol{\mu}}^*, \\ \min(\theta, \mu_a) & \text{if } a \leq a_{\boldsymbol{\mu}}^* - 1. \end{cases}$$

The explicit formula for $D^+(\theta, \tilde{\omega})$ is

$$D^+(\theta, \tilde{\omega}) = \tilde{\omega}_{-1} \frac{(\mu_{a_{\boldsymbol{\mu}}^*-1} - \min(\mu_{a_{\boldsymbol{\mu}}^*-1}, \theta))^2}{2} + \tilde{\omega}_0 \frac{(\mu_{a_{\boldsymbol{\mu}}^*} - \theta)^2}{2} + \tilde{\omega}_1 \frac{(\mu_{a_{\boldsymbol{\mu}}^*+1} - (2S - \theta))^2}{2}.$$

Similarly we can do the same with arm $a_{\boldsymbol{\mu}}^* - 1$: moving from $\boldsymbol{\mu}$ to a bandit problem $\tilde{\boldsymbol{\lambda}}^-$, defined for $\theta \geq S$ by

$$\tilde{\lambda}_a^- = \begin{cases} \mu_a & \text{if } a < a_{\boldsymbol{\mu}}^* - 1, \\ 2S - \theta & \text{if } a = a_{\boldsymbol{\mu}}^* - 1, \\ \theta & \text{if } a = a_{\boldsymbol{\mu}}^*, \\ \max(\theta, \mu_a) & \text{if } a \geq a_{\boldsymbol{\mu}}^* + 1, \end{cases}$$

where both arms $a_{\boldsymbol{\mu}}^* - 1$ and $a_{\boldsymbol{\mu}}^*$ are optimal. For this alternative the cost is

$$D^-(\theta, \tilde{\omega}) = \tilde{\omega}_{-1} \frac{(\mu_{a_{\boldsymbol{\mu}}^*-1} - (2S - \theta))^2}{2} + \tilde{\omega}_0 \frac{(\mu_{a_{\boldsymbol{\mu}}^*} - \theta)^2}{2} + \tilde{\omega}_1 \frac{(\mu_{a_{\boldsymbol{\mu}}^*+1} - \max(\mu_{a_{\boldsymbol{\mu}}^*+1}, \theta))^2}{2}.$$

It appears, see the proof of Proposition 2 in Appendix A.1, that these two types of alternative $\tilde{\boldsymbol{\lambda}}^+$ and $\tilde{\boldsymbol{\lambda}}^-$ are the optimal one. Note that they are also in $\overline{\text{Alt}(\boldsymbol{\mu}, \mathcal{I})}$, the closure of the set of alternatives of $\boldsymbol{\mu}$.

Proposition 2. For all $\boldsymbol{\mu} \in \mathcal{I}$,

$$T_{\mathcal{I}}^*(\boldsymbol{\mu})^{-1} = \sup_{\tilde{\omega} \in \Sigma_3} \min \left(\min_{\{2S - \mu_{a_{\boldsymbol{\mu}}^* + 1} \leq \theta \leq S\}} D^+(\theta, \tilde{\omega}), \min_{\{S \leq \theta \leq 2S - \mu_{a_{\boldsymbol{\mu}}^* - 1}\}} D^-(\theta, \tilde{\omega}) \right). \quad (7)$$

The intuition behind this proposition is that if we try to transform $\boldsymbol{\mu}$ into an alternative $\boldsymbol{\lambda}$ with $b > a_{\boldsymbol{\mu}}^* + 1$ as optimal arm we have to pass by an alternative with optimal arm $a_{\boldsymbol{\mu}}^* + 1$ since we impose to the means to be increasing. It remains to see that this intermediate alternative has always a smaller cost. The cases with $a_{\boldsymbol{\mu}}^* = 1$ or K are similar considering only the the alternatives $\tilde{\boldsymbol{\lambda}}^+$ if $a_{\boldsymbol{\mu}}^* = 1$ and $\tilde{\boldsymbol{\lambda}}^-$ if $a_{\boldsymbol{\mu}}^* = K$. We can also derive bounds on the characteristic time to see that the dependence in K disappear. It is important to note that this property is really asymptotic when δ goes to zero and it is not clear at all that the dependence of the complexity in K would also disappear for moderate value of δ , we think it is not the case.

Proposition 3. For all $\boldsymbol{\mu} \in \mathcal{I}$ such that $1 < a_{\boldsymbol{\mu}}^* < K$, considering the gaps: $\Delta_{-1}^2 = (2S - \mu_{a_{\boldsymbol{\mu}}^* - 1} - \mu_{a_{\boldsymbol{\mu}}^*})^2/8$, $\Delta_1^2 = (2S - \mu_{a_{\boldsymbol{\mu}}^* + 1} - \mu_{a_{\boldsymbol{\mu}}^*})^2/8$ and $\Delta_0^2 = \min(\Delta_{-1}^2, \Delta_1^2)$,

$$\frac{1}{\Delta_0^2} \leq T_{\mathcal{I}}^*(\boldsymbol{\mu}) \leq \sum_{k=-1}^1 \frac{1}{\Delta_k^2} \leq \frac{3}{\Delta_0^2}. \quad (8)$$

3. An asymptotically Optimal Algorithm

We present in this section an asymptotically optimal algorithm inspired by the *Direct-tracking* procedure of [Garivier & Kaufmann \(2016\)](#) (which borrows the idea of tracking from GAFS-MAX algorithm of [Antos et al. \(2008\)](#)). At any time $t \geq 1$ let $h(t) = (\sqrt{t} - K/2)_+$ (where $(x)_+$ stands for the positive part of x) and $U_t = \{a : N_a(t) < h(t)\}$ be the set of "abnormally rarely sampled"

arms. After t rounds the empirical mean of arm a is $\hat{\mu}_a(t) = \hat{\mu}_{a, N_a(t)} = \frac{1}{N_a(t)} \sum_{s=1}^t Y_s \mathbb{1}_{\{A_s=a\}}$, where $N_a(t) = \sum_{s=1}^t \mathbb{1}_{\{A_s=a\}}$ denotes the number of draws of arm a up to and including time t .

Theorem 2 (Asymptotic optimality).

For $\mathcal{S} \in \{\mathcal{I}, \mathcal{M}\}$, for the constant $C = e^{K+1} \left(\frac{2}{K}\right)^K (2(3K+2))^{3K} \frac{4}{\log(3)}$ 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, in the sense that

$$\limsup_{\delta \rightarrow 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \leq T_{\mathcal{S}}^*(\boldsymbol{\mu}). \quad (9)$$

Algorithm 1: Algorithm (Direct-tracking).

Sampling rule

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

Stopping rule

$$\tau_\delta = \inf \left\{ t \in \mathbb{N}^* : \hat{\boldsymbol{\mu}}(t) \in \mathcal{M} \text{ and } \inf_{\lambda \in \mathcal{A}t(\hat{\boldsymbol{\mu}}(t), \mathcal{S})} \sum_{a=1}^K N_a(t) \frac{(\hat{\mu}_a(t) - \lambda_a)^2}{2} > \beta(t, \delta) \right\}. \quad (10)$$

Decision rule

$$\hat{a}_{\tau_\delta} \in \operatorname{argmin}_{1 \leq a \leq K} |\hat{\mu}_a(\tau_\delta) - S|.$$

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-monotonic case $\mathcal{S} = \mathcal{M}$, one can follow the lines of [Garivier & Kaufmann \(2016\)](#), Section 2.2 and replace their [Lemma 3](#) by [Lemma 1](#) above.

In the increasing case $\mathcal{S} = \mathcal{I}$, however, implementing the algorithm is more involved. It is not sufficient to simply use Proposition 2, since $\hat{\mu}(t)$ is not necessarily in \mathcal{I} . Let $\mathcal{I}_b := \{\lambda \in \mathcal{I}, a_\lambda^* = b\}$ be the set of alternatives with b as optimal arm. Noting that the function

$$F : w \mapsto \inf_{\lambda \in \text{Alt}(\mu, \mathcal{I})} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} = \min_{b \neq a_\mu^*} \inf_{\lambda \in \mathcal{I}_b} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \quad (11)$$

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

$$\lambda^b := \arg \min_{\lambda \in \bar{\mathcal{I}}_b} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \quad (12)$$

be the argument of the second infimum in Equation (11). The sub-gradient of F at ω is

$$\partial F(\omega) = \text{Conv}_{b \in B_{Opt}} \left[\frac{(\mu_a - \lambda_a^b)^2}{2} \right]_{a \in \{1, \dots, K\}},$$

where Conv denotes the convex hull operator and where B_{Opt} is the set of arms that reach the minimum in (11). Thus, performing the sub-gradient ascent simply requires to solve efficiently the minimization program (12). It appears that this problem boils down to *unimodal regression* (a problem closely related to isotonic regression, see for example Barlow et al. (1973) and Robertson et al. (1988)). Indeed, we can rewrite the set

$$\{\lambda \in \mathcal{I} : a_\lambda^* = b\} = \{\lambda \in \mathcal{M} : \lambda_1 < \dots < \lambda_{b-1} < \min(\lambda_b, 2S - \lambda_b) \leq \max(\lambda_b, 2S - \lambda_b) < \lambda_{b+1} < \dots < \lambda_K\}.$$

Assume that $\mu_b \leq S$ (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

$$\left\{ \lambda \in \mathcal{M} : \lambda_1 < \dots < \lambda_{b-1} < \lambda_b, 2S - \lambda_K < \dots < 2S - \lambda_{b+1} < \lambda_b, \lambda_b \leq S \right\}.$$

Let λ' be the new variables such that

$$\lambda'_a = \begin{cases} \lambda_a & \text{if } 1 \leq a \leq b, \\ 2S - \lambda_a & \text{else.} \end{cases} \quad (13)$$

Then $\lambda^{b'}$ is the solution of the following minimization program

$$\lambda^{b'} = \arg \min_{\substack{\lambda'_1 \leq \dots \leq \lambda'_b \\ \lambda'_K \leq \dots \leq \lambda'_b \\ \lambda'_b \leq S}} \sum_{a=1}^K \omega_a \frac{(\mu'_a - \lambda'_a)^2}{2}. \quad (14)$$

Lemma 6 in Appendix C shows that

$$\lambda_a^{b'} = \min(\widehat{\lambda}_a^b, S) \text{ for all } a \in \{1, \dots, K\}, \quad \text{where } \widehat{\lambda}^b := \arg \min_{\substack{\lambda'_1 \leq \dots \leq \lambda'_b \\ \lambda'_K \leq \dots \leq \lambda'_b}} \sum_{a=1}^K \omega_a \frac{(\mu'_a - \lambda'_a)^2}{2},$$

is the unimodal regression of $\boldsymbol{\mu}'$ with weights ω and with a mode located at b . It is efficiently computed via isotonic regressions (e.g. Frisén (1986), Geng & Shi (1990), Mureika et al. (1992)) with a computational complexity proportional to the number of arms K . From $\widehat{\lambda}^b$, one can go back to $\boldsymbol{\lambda}^b$ by reversing Equation (13). Since we need to compute $\boldsymbol{\lambda}^b$ for each $b \neq a_\mu^*$, 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 (10). We also tried the Racing algorithm (R), with the elimination criterion of (10). For a description of all those algorithms, see Garivier & Kaufmann (2016) and references therein. Finally, in order to allow comparison with the state of the art, we added the sampling rule of algorithm APT (Anytime Parameter-free Thresholding algorithm) from Locatelli et al. (2016) in combination with the stopping rule (10). We chose to set the parameter ε of APT to be roughly equal to a tenth of the gap. It appears that the exploration function β prescribed in Theorem 2 is overly

	BC- \mathcal{M}	R- \mathcal{M}	DT- \mathcal{M}	APT- \mathcal{M}	$T_{\mathcal{M}}^*(\boldsymbol{\mu}) \log \frac{1}{\delta}$
1	3913	3609	4119	5960	2033
2	3064	3164	3098	3672	1861
	BC- \mathcal{I}	R- \mathcal{I}	DT- \mathcal{I}	APT- \mathcal{I}	$T_{\mathcal{I}}^*(\boldsymbol{\mu}) \log \frac{1}{\delta}$
1	483	494	611	1127	247
2	2959	2906	3072	3531	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-monotonic cases. Two thresholding bandit problems are considered: bandit problem 1, $\boldsymbol{\mu}_1 = [0.5, 1.1, 1.2, 1.3, 1.4, 5]$ with $S_1 = 1$, and bandit problem 2, $\boldsymbol{\mu}_2 = [1, 2, 2.5]$ 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%).

pessimistic. On the basis of our experiments, we recommend the use of $\beta(t, \delta) = \log((\log(t) + 1)/\delta)$ 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-monotonic 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 (10) 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 2. The poor performances of APT can be explained by the crude adaptation of this algorithm to the fixed confidence setting. This possibly comes from the fact that it was originally designed for the fixed budget setting and it appears that these two frameworks are fundamentally different, as argued by Carpentier & Locatelli (2016).

4. Closest Mean Below the Threshold

In this section we briefly discuss a variant of the previous problem: finding the arm with the closest mean *below* the threshold, i.e. $a_{\mu}^* \in \arg \min_{1 \leq a \leq K : \mu_a \leq S} |\mu_a - S|$, still under the assumption that the means are increasing. Surprisingly this new problem is simpler than the previous one and it is possible to compute exactly, in this case, the optimal weights and the characteristic time.

Let $\mu \in \mathcal{I}$ be a bandit problem such that the optimal arm a_{μ}^* for this new setting is unique (with $1 < a_{\mu}^* < K$ for the sake of clarity). As in Section 2.2, we only need to consider alternative bandit problems such that arm $a_{\mu}^* - 1$ or $a_{\mu}^* + 1$ is optimal. But only one arm needs to be moved: the optimal alternative $\tilde{\lambda}^-$ (resp. $\tilde{\lambda}^+$) where $a_{\mu}^* - 1$ (resp. $a_{\mu}^* + 1$) is optimal are defined by

$$\tilde{\lambda}_a^- = \begin{cases} S & \text{if } a = a_{\mu}^*, \\ \mu_a & \text{else,} \end{cases} \quad \text{and} \quad \tilde{\lambda}_a^+ = \begin{cases} S & \text{if } a = a_{\mu}^* + 1, \\ \mu_a & \text{else.} \end{cases} \quad (15)$$

With the same arguments used to prove Proposition 2, one obtains that

$$T_{\mathcal{I}}^*(\mu)^{-1} = \sup_{\tilde{\omega} \in \Sigma_2} \min \left(\tilde{\omega}^- \frac{(S - \mu_{a_{\mu}^*})^2}{2}, \tilde{\omega}^+ \frac{(\mu_{a_{\mu}^*+1} - S)^2}{2} \right) = \frac{1}{\frac{2}{(S - \mu_{a_{\mu}^*})^2} + \frac{2}{(\mu_{a_{\mu}^*+1} - S)^2}}, \quad (16)$$

and that the associated optimal weights $\omega^*(\mu)$ are defined by

$$\omega^*(\mu)_a = \begin{cases} \frac{2}{(S - \mu_{a_{\mu}^*})^2} \left(\frac{2}{(S - \mu_{a_{\mu}^*})^2} + \frac{2}{(\mu_{a_{\mu}^*+1} - S)^2} \right)^{-1} & \text{if } a = a_{\mu}^*, \\ \frac{2}{(\mu_{a_{\mu}^*+1} - S)^2} \left(\frac{2}{(S - \mu_{a_{\mu}^*})^2} + \frac{2}{(\mu_{a_{\mu}^*+1} - S)^2} \right)^{-1} & \text{if } a = a_{\mu}^* + 1, \\ 0 & \text{else.} \end{cases}$$

In some way, this problem is closer to the classical threshold bandit problem (Locatelli et al., 2016) with two arms since we are testing if the mean of arms a_{μ}^* and $a_{\mu}^* + 1$ is below or above the threshold.

For an optimal strategy, Algorithm 1 can be adapted to this new setting, as well as the Theorem 2 and its asymptotic optimality proof. The only point that needs to be discussed is how to implement this algorithm in practice. One may follow the procedure described in Section 3.1: perform an gradient ascent on the simplex to compute the maximum of F defined in (11). The main difficulty is to compute the sub-gradient of F at $\omega \in \Sigma_K$ and in particular the projection given by (12), that rewrites in this setting

$$\lambda^b = \arg \min_{\substack{\lambda'_1 \leq \dots \leq \lambda'_b \leq S \\ S \leq \lambda'_{b+1} \leq \dots \leq \lambda'_K}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda'_a)^2}{2}. \quad (17)$$

But this projection can also be easily computed using two isotonic regressions under bound restrictions, see for example Hu (1997).

5. 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 proved that, surprisingly, the complexity terms can be computed almost as easily as in the best-arm identification case, despite the important constraints of our setting. 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-monotonic 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. In fact, despite the asymptotic nature of the results presented here, the procedure proposed here appears to be the most efficient in practice *even when the number of trials implied is rather low* (which is often the case in clinical trials).

In the case where several arms are simultaneously closest to the threshold, the complexity of the problem is infinite. This suggests to extend the results presented here to the PAC setting, where the goal is to find *any ε -closest arm* with probability at least $1 - \delta$. This extension, and extensions to the non-Gaussian case, are left for future investigation since they induce significant technical difficulties.

As a possibility of improvement, we can also mention the possible use of the unimodal regression algorithm of Stout (2000) in order to compute directly (11) with a complexity of order $O(K)$. 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.

References

- Antos, A., Grover, V., & Szepesvári, C. (2008). Active learning in multi-armed bandits. In International conference on algorithmic learning theory (pp. 287–302).
- Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. (1973). Statistical inference under order restrictions. John Wiley and Sons.
- Berge, C. (1963). Topological spaces: including a treatment of multi-valued functions, vector spaces, and convexity. Courier Corporation.
- Boyd, S., Xiao, L., & Mutapcic, A. (2003). Subgradient methods. Lecture notes of EE392o, Stanford University, Autumn Quarter, 2004 .
- Carpentier, A., & Locatelli, A. (2016). Tight (lower) bounds for the fixed budget best arm identification bandit problem. In Conference on learning theory (pp. 590–604).
- Chen, S., Lin, T., King, I., Lyu, M. R., & Chen, W. (2014). Combinatorial pure exploration of multi-armed bandits. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), Advances in neural information processing systems 27 (pp. 379–387). Curran Associates, Inc.
- Combes, R., & Proutiere, A. (2014). Unimodal bandits: Regret lower bounds and optimal algorithms. In Proceedings of the 31st international conference on international conference on machine learning - volume 32 (pp. I-521–I-529).
- Frisén, M. (1986). Unimodal regression. *The Statistician*, 479–485.
- Garivier, A., & Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In Conference on learning theory (pp. 998–1027).
- Geng, Z., & Shi, N.-Z. (1990). Algorithm as 257: isotonic regression for umbrella orderings. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39 (3), 397–402.
- Genovese, M C., Durez, P., Richards, H. B., Supronik, J., Dokoupilova, E., Mazurov, V., Mpofu, S. (2013). 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.

- Hu, X. (1997). Maximum-likelihood estimation under bound restriction and order and uniform bound restrictions. *Statistics & probability letters*, 35 (2), 165–171.
- Kaufmann, E., Cappé, O., & Garivier, A. (2016). On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17 (1), 1–42.
- Le Tourneau, C., Lee, J. J., & Siu, L. L. (2009). Dose escalation methods in phase i cancer clinical trials. *JNCI: Journal of the National Cancer Institute*, 101 (10), 708–720.
- Locatelli, A., Gutzeit, M., & Carpentier, A. (2016). An optimal algorithm for the thresholding bandit problem. In *Proceedings of the 33rd international conference on machine learning, ICML 2016, New York City, June 19-24, 2016* (pp. 1690–1698).
- Magureanu, S., Combes, R., & Proutière, A. (2014). 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* (pp. 975–999).
- Mureika, R., Turner, T., & Wollan, P. (1992). An algorithm for unimodal isotonic regression, with application to locating a maximum, univ. new brunswick dept. math (Tech. Rep.). and Stat. Tech. Report 92–4.
- Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference*. John Wiley and Sons.
- Stout, Q. F. (2000). Optimal algorithms for unimodal regression. *Ann Arbor*, 1001 , 48109–2122.

Appendix A: Proofs of the Lower Bounds

A.1 Expression of the Complexity in the Increasing Case

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

$$D^+(\theta, \tilde{\omega}) = \tilde{\omega}_{-1} \frac{(\mu_{a^*-1} - \min(\mu_{a^*-1}, \theta))^2}{2} + \tilde{\omega}_0 \frac{(\mu_{a^*} - \theta)^2}{2} + \tilde{\omega}_1 \frac{(\mu_{a^*+1} - (2S - \theta))^2}{2} \quad (18)$$

$$D^-(\theta, \tilde{\omega}) = \tilde{\omega}_{-1} \frac{(\mu_{a^*-1} - (2S - \theta))^2}{2} + \tilde{\omega}_0 \frac{(\mu_{a^*} - \theta)^2}{2} + \tilde{\omega}_1 \frac{(\mu_{a^*+1} - \max(\mu_{a^*+1}, \theta))^2}{2}, \quad (19)$$

if $1 < a^* < K$. Else, if $a^* = 1$ we define

$$D^+(\theta, \tilde{\omega}) = \tilde{\omega}_0 \frac{(\mu_{a^*} - \theta)^2}{2} + \tilde{\omega}_1 \frac{(\mu_{a^*+1} - (2S - \theta))^2}{2}, \quad D^-(\theta, \tilde{\omega}) = +\infty,$$

and if $a^* = K$ we define $D^+(\theta, \tilde{\omega}) = +\infty$, $D^-(\theta, \tilde{\omega}) = \tilde{\omega}_{-1} \frac{(\mu_{a^*-1} - (2S - \theta))^2}{2} + \tilde{\omega}_0 \frac{(\mu_{a^*} - \theta)^2}{2}$.

Proof of Proposition 2. We just treat here the case $1 < a^* < K$, the two other limit cases are very similar. We begin by proving that for all $\omega \in \Sigma_K$

$$\inf_{\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\mu}, S)} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} = \min_{b \in \{a^*-1, a^*+1\}} \inf_{\{\boldsymbol{\lambda} \in \mathcal{I} : a_{\boldsymbol{\lambda}}^* = b\}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}. \quad (20)$$

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 $\boldsymbol{\lambda}^\alpha \in \mathcal{I}$. For $\boldsymbol{\nu} \in \mathcal{I}$ and $a \in \{0, \dots, K\}$, 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

$$m_{a_{\boldsymbol{\lambda}}^*-1}(\boldsymbol{\mu}) < m_{a_{\boldsymbol{\lambda}}^*}(\boldsymbol{\mu}) \leq m_{a^*-2}(\boldsymbol{\mu}) < m_{a^*-1}(\boldsymbol{\mu}) < S < m_{a^*}(\boldsymbol{\mu}) \text{ and} \\ m_{a_{\boldsymbol{\lambda}}^*-1}(\boldsymbol{\lambda}) < S < m_{a_{\boldsymbol{\lambda}}^*}(\boldsymbol{\lambda}) \leq m_{a^*-2}(\boldsymbol{\lambda}) < m_{a^*-1}(\boldsymbol{\lambda}) < m_{a^*}(\boldsymbol{\lambda}).$$

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

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

i.e. $a_{\boldsymbol{\lambda}^{\alpha_0}}^* = a^* - 1$. But $\alpha \mapsto \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a^{\alpha_0})^2}{2}$ 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 $\boldsymbol{\lambda}$, therefore

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

The reverse inequality follows from the inclusion $\bigcup_{b \in \{a^*-1, a^*+1\}} \{\boldsymbol{\lambda} \in \mathcal{I} : a_{\boldsymbol{\lambda}}^* = b\} \subset \text{Alt}(\boldsymbol{\mu}, \mathcal{I})$.

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

$$\inf_{\{\lambda \in \mathcal{I} : a_\lambda^* = a^* + 1\}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} = \inf_{\{\lambda \in \mathcal{I} : a_\lambda^* = a^* + 1, \lambda_{a^*} \geq 2S - \mu_{a^*+1}\}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}. \quad (21)$$

Let $\tilde{\lambda}$ be such that

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

By construction we have $\tilde{\lambda} \in \overline{\{\lambda \in \mathcal{I} : a_\lambda^* = a^* + 1\}}$. As $\lambda_{a^*+1} \leq 2S - \lambda_{a^*}$ and $\mu_{a^*+1} \geq 2S - \lambda_{a^*}$ hold, we get $(\tilde{\lambda}_{a^*+1} - \mu_{a^*+1})^2 \leq (\lambda_{a^*+1} - \mu_{a^*+1})^2$. Similarly, for $a \leq a^* - 1$ we have thanks to the fact that $\lambda_a \leq \lambda_{a^*}$ the inequality $(\tilde{\lambda}_{a^*+1} - \mu_{a^*+1})^2 \leq (\lambda_{a^*+1} - \mu_{a^*+1})^2$. Therefore, combining these two inequalities and using the definition of $\tilde{\lambda}$, one obtains

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

and we can rewrite the infimum in Equation 21, indexing the alternative $\tilde{\lambda}$ by θ the mean of arm a , as follows:

$$\begin{aligned} \inf_{\{\lambda \in \mathcal{I} : a_\lambda^* = a^* + 1\}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} &= \min_{2S - \mu_{a^*+1} \leq \theta \leq S} \sum_{a \leq a^* - 1} \omega_a \frac{(\mu_a - \min(\theta, \mu_a))^2}{2} \\ &\quad + \omega_{a^*} \frac{(\mu_{a^*} - \theta)^2}{2} + \omega_{a^*+1} \frac{(\mu_{a^*+1} - 2S + \theta)^2}{2} \\ &= \min_{2S - \mu_{a^*+1} \leq \theta \leq S} \sum_{a < a^* - 1} \omega_a \frac{(\mu_a - \min(\theta, \mu_a))^2}{2} \\ &\quad + D^+(\theta, [\omega_{a^*-1}, \omega_{a^*}, \omega_{a^*+1}]). \end{aligned} \quad (22)$$

Similarly, if the optimal arm of the alternative is $a^* - 1$, we get

$$\begin{aligned} \inf_{\{\lambda \in \mathcal{I} : a_\lambda^* = a^* - 1\}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} &= \min_{S \leq \theta \leq 2S - \mu_{a^*-1}} \sum_{a > a^* + 1} \omega_a \frac{(\mu_a - \max(\theta, \mu_a))^2}{2} \\ &\quad + D^-(\theta, [\omega_{a^*-1}, \omega_{a^*}, \omega_{a^*+1}]). \end{aligned} \quad (23)$$

Then, by noting that

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

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

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

we obtain thanks to Equation (20) and to the fact that $\tilde{\omega}$ depends only on ω :

$$\inf_{\lambda \in \text{Alt}(\mu, S)} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2} \leq \min \left(\min_{\{2S - \mu_{a^*+1} \leq \theta \leq S\}} D^+(\theta, \tilde{\omega}), \min_{\{S \leq \theta \leq 2S - \mu_{a^*-1}\}} D^-(\theta, \tilde{\omega}) \right), \quad (24)$$

where we identified $\tilde{\omega}$ to an element of Σ_3 . Taking the supremum on each side of (24), one obtains:

$$T_{\mathcal{I}}^*(\mu)^{-1} \leq \sup_{\tilde{\omega} \in \Sigma_3} \min \left(\min_{\{2S - \mu_{a^*+1} \leq \theta \leq S\}} D^+(\theta, \tilde{\omega}), \min_{\{S \leq \theta \leq 2S - \mu_{a^*-1}\}} D^-(\theta, \tilde{\omega}) \right).$$

In order to prove the reverse inequality and thus (7), we just need to use (23), (22) and restrict the weight ω to have a support included in $\{a^* - 1, a^*, a^* + 1\}$. \square

Proof of Proposition 3. We recall the definitions of the gaps: $\Delta_{-1}^2 = (2S - \mu_{a^*-1} - \mu_{a^*})^2/8$, $\Delta_1^2 = (2S - \mu_{a^*+1} - \mu_{a^*})^2/8$ and $\Delta_0^2 = \min(\Delta_{-1}^2, \Delta_1^2)$. For the lower bound we consider the particular weights $\bar{\omega} \in \Sigma_3$ defined by

$$\bar{\omega}_i = \frac{1/\Delta_i^2}{\sum_{k=-1}^1 1/\Delta_k^2} \quad \text{for } -1 \leq i \leq 1.$$

Thanks to the Proposition 2, we know that

$$T_{\mathcal{I}}^*(\mu)^{-1} \geq \min \left(\min_{\{2S - \mu_{a^*+1} \leq \theta \leq S\}} D^+(\theta, \bar{\omega}), \min_{\{S \leq \theta \leq 2S - \mu_{a^*-1}\}} D^-(\theta, \bar{\omega}) \right).$$

Then we can lower bound the two terms that appear in the minimum. Indeed, we have, denoting

the mean $\bar{\theta} = \bar{\omega}_0 \mu_{a^*} + \bar{\omega}_1 (2S - \mu_{a^*+1})$,

$$\begin{aligned} \min_{\{2S - \mu_{a^*+1} \leq \theta \leq S\}} D^+(\theta, \bar{\omega}) &\geq \min_{\{2S - \mu_{a^*+1} \leq \theta \leq S\}} \bar{\omega}_0 \frac{(\mu_{a^*} - \theta)^2}{2} + \bar{\omega}_1 \frac{((2S - \mu_{a^*+1}) - \theta)^2}{2} \\ &= \bar{\omega}_0 \frac{(\mu_{a^*} - \bar{\theta})^2}{2} + \bar{\omega}_1 \frac{((2S - \mu_{a^*+1}) - \bar{\theta})^2}{2} \\ &\geq \min(\bar{\omega}_1, \bar{\omega}_0) \Delta_1^2 \geq \frac{1}{\sum_{k=-1}^1 1/\Delta_k^2}, \end{aligned}$$

where we used the definition of the weights $\bar{\omega}$ for the last inequality and the fact that either $(\mu_{a^*} - \bar{\theta})^2/2 \geq \Delta_1^2$ or $((2S - \mu_{a^*+1}) - \bar{\theta})^2/2 \geq \Delta_1^2$ since by definition $\bar{\theta}$ belongs to the interval with bounds $2S - \mu_{a^*+1}$ and μ_{a^*} for the one before. Similarly one can prove the same inequality with the second term:

$$\min_{\{S \leq \theta \leq 2S - \mu_{a^*-1}\}} D^-(\theta, \bar{\omega}) \geq \frac{1}{\sum_{k=-1}^1 1/\Delta_k^2},$$

therefore we obtain the lower bound $T_{\mathcal{I}}^*(\mu)^{-1} \geq \frac{1}{\sum_{k=-1}^1 1/\Delta_k^2}$. For the upper bound we just need to choose a particular θ in order to bound one of the two terms that appears in the expression of $T_{\mathcal{I}}^*(\mu)^{-1}$. Thus, with the choice $\theta_1 = (2S - \mu_{a^*+1})/2 + \mu_{a^*}/2$, we get

$$\min_{\{2S - \mu_{a^*+1} \leq \theta \leq S\}} D^+(\theta, \tilde{\omega}) \leq D^+(\theta_1, \tilde{\omega}) \leq (\tilde{\omega}_{-1} + \tilde{\omega}_0 + \tilde{\omega}_1) \frac{(2S - \mu_{a^*+1} - \mu_{a^*})^2}{8} = \Delta_1^2,$$

where we used that $(\mu_{a^*-1} - \min(\mu_{a^*-1}, \theta_1))^2/2 \leq (2S - \mu_{a^*+1} - \mu_{a^*})^2/8$ since θ_1 is at the middle between $2S - \mu_{a^*+1}$ and μ_{a^*} . In the same way with $\theta_{-1} = (2S - \mu_{a^*-1})/2 + \mu_{a^*}/2$ one can obtain the inequality

$$\min_{\{2S - \mu_{a^*+1} \leq \theta \leq S\}} \leq \Delta_{-1}^2.$$

Combining these two inequalities with Proposition 2 leads to $T_{\mathcal{T}}^*(\boldsymbol{\mu})^{-1} \leq \Delta_0^2$. \square

A.2. Expression of the Complexity in the Non-monotonic Case

Proof of Lemma 1. To simplify the notations we note $a_{\boldsymbol{\mu}}^* = a^*$. Thanks to the definition of the characteristic time, we just have to prove that

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

Using that $\text{Alt}(\boldsymbol{\mu}, \mathcal{M}) = \bigcup_{b \neq a^*} \{\boldsymbol{\lambda} \in \mathcal{M} : |\lambda_b - S| < |\lambda_{a^*} - S|\}$, one has

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

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

$$\begin{aligned} \inf_{|\lambda_b - S| < |\lambda_{a^*} - S|} \omega_{a^*} \frac{(\mu_{a^*} - \lambda_{a^*})^2}{2} + \omega_b \frac{(\mu_b - \lambda_b)^2}{2} &= \min \left(\inf_x \omega_{a^*} \frac{(\mu_{a^*} - S - x)^2}{2} + \omega_b \frac{(\mu_b - S - x)^2}{2}, \right. \\ &\quad \left. \inf_x \omega_{a^*} \frac{(\mu_{a^*} - S + x)^2}{2} + \omega_b \frac{(\mu_b - S + x)^2}{2} \right). \end{aligned}$$

The conclusion follows by noting that

$$\begin{aligned} \inf_x \omega_{a^*} \frac{(\mu_{a^*} - S - x)^2}{2} + \omega_b \frac{(\mu_b - S - x)^2}{2} &= \frac{\omega_{a^*} \omega_b}{2(\omega_{a^*} + \omega_b)} (\mu_{a^*} - \mu_b)^2, \\ \inf_x \omega_{a^*} \frac{(\mu_{a^*} - S + x)^2}{2} + \omega_b \frac{(\mu_b - S + x)^2}{2} &= \frac{\omega_{a^*} \omega_b}{2(\omega_{a^*} + \omega_b)} (2S - \mu_{a^*} - \mu_b)^2. \end{aligned}$$

\square

Appendix B: Correctness and Asymptotic Optimality of Algorithm 1

Proof of Proposition 2. We follow and slightly adapt the proof of Theorem 14 of Kaufmann et al. (2016). We fix a bandit problem $\mu \in \mathcal{S}$ and the constant

$$C := e^{K+1} \left(\frac{2}{K}\right)^K (2(3K+2))^{3K} \frac{4}{\log(3)}. \quad (25)$$

We prove Algorithm 1 is δ -correct on \mathcal{S} , before showing that it is asymptotically optimal.

δ -correctness on \mathcal{S} .

We will prove in the second part of proof that τ is almost surely finite, confer (28). By definition of τ , the probability that the predicted arm is the wrong one is upper-bounded by

$$\mathbb{P}_{\mu}(\hat{a}_{\tau} \neq a_{\mu}^*) \leq \mathbb{P}_{\mu}\left(\exists t \in \mathbb{N}^*, \sum_{a=1}^K N_a(t) \frac{(\hat{\mu}_a(t) - \mu_a)^2}{2} > \beta(t, \delta)\right), \quad (26)$$

where we used that $\mu \in \text{Alt}(\hat{\mu}(t), \mathcal{S})$ since $\hat{a}_{\tau} \neq a_{\mu}^*$. Using the union bound then Theorem 3 (note that $\beta(t, \delta) \geq K + 1$ thanks to the choice of C) we have

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

where in the third inequality we replaced $\beta(t, \delta)$ by its value and used in the fourth inequality, for $C \geq 3$, the following upper-bound

$$\sum_{t=1}^{+\infty} \frac{1}{t \log(3t)^2} \leq \frac{1}{\log(3)^2} + \int_{t=1}^{+\infty} \frac{1}{t \log(3t)^2} dt \leq \frac{2}{\log(3)}.$$

Asymptotic Optimality

We begin by remarking that the function $\mu \rightarrow \omega^*(\mu)$ is continuous on the sets $\mathcal{S}_b = \{\mu \in \mathcal{S} : a_{\mu}^* = b\}$ for $b \in \{1, \dots, K\}$. Indeed it is a consequence of Lemma 1 if $\mathcal{S} = \mathcal{M}$ and Proposition 2 if $\mathcal{S} = \mathcal{I}$ and the Maximum theorem from Berge (1963). Let ε be a real in $(0, 1)$. From the continuity of w^* in μ , there exists $\alpha = \alpha(\varepsilon)$ such that the neighbourhood of μ :

$$I_{\varepsilon} := [\mu_1 - \alpha, \mu_1 + \alpha] \times \dots \times [\mu_K - \alpha, \mu_K + \alpha]$$

is such that for all $\boldsymbol{\mu}' \in I_\varepsilon$, $\boldsymbol{\mu}' \in \mathcal{S}$, $a_\mu^* = a_{\boldsymbol{\mu}'}^*$ and $\max_a |w_a^*(\boldsymbol{\mu}') - w_a^*(\boldsymbol{\mu})| \leq \varepsilon$. Let $T \in \mathbb{N}^*$ and denote by $\mathcal{E}_T(\varepsilon) = \bigcap_{t=T^{1/4}}^T (\widehat{\boldsymbol{\mu}}(t) \in I_\varepsilon)$ the typical event where $\widehat{\boldsymbol{\mu}}(t)$ is not too far from $\boldsymbol{\mu}$. The two following Lemmas are extracted from [Kaufmann et al. \(2016\)](#).

Lemma 2. *There exist B and C (depending on $\boldsymbol{\mu}$ and ε) such that $\mathbb{P}_\mu(\mathcal{E}_T^c) \leq BT \exp(-CT^{1/8})$.*

Lemma 3. *There exists a constant T_ε such that for $T \geq T_\varepsilon$, it holds that on \mathcal{E}_T ,*

$$\forall t \geq \sqrt{T}, \max_a \left| \frac{N_a(t)}{t} - w_a^*(\boldsymbol{\mu}) \right| \leq 2(K-1)\varepsilon$$

We now assume that $T \geq T_\varepsilon$. Introducing the constant

$$C_\varepsilon^*(\boldsymbol{\mu}) = \inf_{\substack{\boldsymbol{\mu}': \|\boldsymbol{\mu}' - \boldsymbol{\mu}\| \leq \alpha(\varepsilon) \\ \boldsymbol{w}': \|\boldsymbol{w}' - \boldsymbol{w}^*(\boldsymbol{\mu})\| \leq 2(K-1)\varepsilon}} \inf_{\lambda \in \text{Alt}(\boldsymbol{\mu}', \mathcal{S})} \sum_{a=1}^K w_a \frac{(\mu'_a(t) - \lambda_a)^2}{2},$$

thanks to Lemma 3, on the event \mathcal{E}_T it holds that for every $t \geq \sqrt{T}$,

$$t \inf_{\lambda \in \text{Alt}(\widehat{\boldsymbol{\mu}}(t), \mathcal{S})} \sum_{a=1}^K \frac{N_a(t)}{t} \frac{(\widehat{\mu}_a(t) - \lambda_a)^2}{2} \geq t C_\varepsilon^*(\boldsymbol{\mu}). \quad (27)$$

Thus, combining (27) and the definition of the stopping rule (10), we have on the event \mathcal{E}_T

$$\min(\tau_\delta, T) \leq \sqrt{T} + \sum_{t=\sqrt{T}}^T \mathbb{1}_{\{\tau_\delta > t\}} \leq \sqrt{T} + \sum_{t=\sqrt{T}}^T \mathbb{1}_{\{t C_\varepsilon^*(\boldsymbol{\mu}) \leq \beta(T, \delta)\}} \leq \sqrt{T} + \frac{\beta(T, \delta)}{C_\varepsilon^*(\boldsymbol{\mu})}.$$

Introducing $T_0(\delta) = \inf \left\{ T \in \mathbb{N} : \sqrt{T} + \frac{\beta(T, \delta)}{C_\varepsilon^*(\boldsymbol{\mu})} \leq T/(1 + \varepsilon/2) \right\}$, for every $T \geq \max(T_0(\delta), T_\varepsilon)$, one has $\mathcal{E}_T \subseteq \{\tau_\delta \leq T\}$, therefore thanks to Lemma 2 $\mathbb{P}_\mu(\tau_\delta > T) \leq \mathbb{P}(\mathcal{E}_T^c) \leq BT \exp(-CT^{1/8})$ and

$$\mathbb{E}_\mu[\tau_\delta] \leq T_0(\delta) + T_\varepsilon + \sum_{T=1}^{\infty} BT \exp(-CT^{1/8}). \quad (28)$$

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

$$H(\varepsilon) = \inf \left\{ T \in \mathbb{N} : T/(1 + \varepsilon/2) - \sqrt{T} \geq T/(1 + \varepsilon) \right\}, \text{ one has}$$

$$\begin{aligned} T_0(\delta) &\leq H(\varepsilon) + \inf \left\{ T \in \mathbb{N} : \beta(T, \delta) \leq \frac{\cup_\varepsilon(\boldsymbol{\mu})T}{1 + \varepsilon} \right\} \\ &\leq H(\varepsilon) + \inf \left\{ T \in \mathbb{N} : \log(TC/\delta) + (3K + 2) \log \log(TC/\delta) \leq \frac{C_\varepsilon^*(\boldsymbol{\mu})T}{1 + \varepsilon} \right\}. \end{aligned}$$

Using technical Lemma 5, for δ small enough to have $(C_\varepsilon^*(\boldsymbol{\mu})\delta)/((1 + \varepsilon)^2 C) \leq e$, we get

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

This last upper bound yields, for every $\varepsilon > 0$, $\limsup_{\delta \rightarrow 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_{\delta}]}{\log(1/\delta)} \leq \frac{(1 + \varepsilon)^2}{C_{\varepsilon}^*(\boldsymbol{\mu})}$. Letting ε tend to zero and by definition of w^* , $\lim_{\varepsilon \rightarrow 0} C_{\varepsilon}^*(\boldsymbol{\mu}) = T_{\mathcal{S}}^*(\boldsymbol{\mu})^{-1}$, which allows to conclude that $\limsup_{\delta \rightarrow 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_{\delta}]}{\log(1/\delta)} \leq T_{\mathcal{S}}^*(\boldsymbol{\mu})$. \square

Appendix C: Some technical Lemma for the Analysis of Algorithm 1

C.1. An Inequality

For $0 < y \leq 1/e$ let g be the function $g(y) = \frac{1}{y} \log\left(\frac{e}{y} \log\left(\frac{1}{y}\right)\right)$.

Lemma 4. *Let $A > 0$ such that $1/A > e$, then for all $x \geq g(A)$, $\log(x) \leq Ax$.*

Proof. Since $g(A) \geq 1/A$, the function $x \mapsto A - 1/x$ is non-decreasing, we just need to prove the inequality for $x = g(A)$. It remains to remark that since $\log(x) \leq x/e$,

$$\log(g(A)) \leq \log\left(\frac{2}{A} \log\left(\frac{1}{A}\right)\right) \leq \log\left(\frac{e}{A} \log\left(\frac{1}{A}\right)\right) = Ag(A).$$

□

Lemma 5. *Let $A, B > 0$, then for all $\varepsilon \in (0, 1)$ such that $(1 + \varepsilon)/A < e$ and $B/\varepsilon > e$, for all $x \geq \max\left(g(A/(1 + \varepsilon)), \exp(g(\varepsilon/B))\right)$, $\log(x) + B \log\log(x) \leq Ax$.*

Proof. Since $\log(x) \geq g(\varepsilon/B)$ thanks to Lemma 4 we have $B \log\log(x) \leq \varepsilon \log(x)$. Therefore, still using Lemma 4 with $x \geq g(A/(1 + \varepsilon))$, $\log(x) + B \log\log(x) \leq (1 + \varepsilon) \log(x) \leq Ax$. □

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

Theorem 3. *For all $\delta \geq (K + 1)$ and $t \in \mathbb{N}^*$ we have*

$$\mathbb{P}\left(\sum_{a=1}^K N_a(t) \frac{(\hat{\mu}_a(t) - \mu_a)^2}{2} \geq \delta\right) \leq e^{K+1} \left(\frac{2\delta(\delta \log(t) + 1)}{K}\right)^K e^{-\delta}. \quad (29)$$

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.2. Unimodal Regression under Bound Restriction

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

$$\mathcal{U} = \{\boldsymbol{\lambda} : \lambda_1 \leq \dots \leq \lambda_b \geq \lambda_{b+1} \geq \dots \lambda_K\}, \quad (30)$$

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

$$\mathcal{U}_S = \{\boldsymbol{\lambda} : \lambda_1 \leq \dots \leq \lambda_b \geq \lambda_{b+1} \geq \dots \lambda_K, \lambda_b \leq S\}. \quad (31)$$

Let $\hat{\boldsymbol{\lambda}}$ be the unimodal regression of $\boldsymbol{\mu}$: $\hat{\boldsymbol{\lambda}} := \arg \min_{\boldsymbol{\lambda} \in \mathcal{U}} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^-}{2}$, and $\boldsymbol{\lambda}^*$ be the projection

of $\boldsymbol{\mu}$ on \mathcal{U}_S : $\boldsymbol{\lambda}^* := \arg \min_{\boldsymbol{\lambda} \in \mathcal{U}_S} \sum_{a=1}^K \omega_a \frac{(\mu_a - \lambda_a)^2}{2}$. We have, as in the case of isotonic regression (see [Hu \(1997\)](#)), the following simple relation between $\boldsymbol{\lambda}^*$ and $\hat{\boldsymbol{\lambda}}$.

Lemma 6. *It holds that $\lambda_a^* = \min(\hat{\lambda}_a, S)$ for all $a \in \{1, \dots, K\}$.*

ORGANIZER:

GLOBAL ACADEMIC-INDUSTRIAL COOPERATION SOCIETY (GAICS)



ISBN 978-986-98189-2-6