Non-Asymptotic Sequential Tests for Overlapping Hypotheses and application to near optimal arm identification in bandit models

Aurélien Garivier, Emilie Kaufmann

Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

NON-ASYMPTOTIC SEQUENTIAL TESTS FOR OVERLAPPING HYPOTHESES AND APPLICATION TO NEAR OPTIMAL ARM IDENTIFICATION IN BANDIT MODELS

BY AURÉLIEN GARIVIER, AND EMILIE KAUFMANN

Abstract. In this paper, we study sequential testing problems with *overlapping* hypotheses. We first focus on the simple problem of assessing if the mean μ of a Gaussian distribution is $\geq -\epsilon$ or $\leq \epsilon$; if $\mu \in (-\epsilon, \epsilon)$, both answers are considered to be correct. Then, we consider PAC-best arm identification in a bandit model: given K probability distributions on $\mathbb{R}$ with means $\mu_1, \dots, \mu_K$, we derive the asymptotic complexity of identifying, with risk at most δ, an index $I \in \{1, \dots, K\}$ such that $\mu_I \geq \max_i \mu_i - \epsilon$. We provide non asymptotic bounds on the error of a parallel General Likelihood Ratio Test, which can also be used for more general testing problems. We further propose lower bound on the number of observation needed to identify a correct hypothesis. Those lower bounds rely on information-theoretic arguments, and specifically on two versions of a change of measure lemma (a high-level form, and a low-level form) whose relative merits are discussed.

1. Introduction. In order to extract information on a random source $\mathbb{P}$, the classical statistical framework relies on a fixed sample $X_1, \dots, X_n$ of a given size n. For example, a classical hypothesis test is a (possibly randomized) function $\hat{\imath}_n$ of the sample. In *sequential statistics*, on the contrary, the statistician is allowed to adaptively decide on the sample size. For example, the sequential hypothesis testing of $\mathcal{H}_0$ versus $\mathcal{H}_1$ consists of a stopping time τ with respect to the filtration $(\mathcal{F}_t)_t$, where $\mathcal{F}_t = \sigma(X_1, \dots, X_t)$, and a decision rule $\hat{\imath}_\tau$ which is $\mathcal{F}_\tau$-measurable.

Even more power is devoted to the statistician in *active testing*: the statistician not only chooses when to stop, but also sequentially influences the observations. For example, in the *multi-armed bandit problem*, a set $\mathcal{A}$ of unknown random sources is available on demand, and the statistician chooses one of them $A_t \in \mathcal{A}$ at every time step t, from which she gets a new sample. The choice of A_t is made based only upon past observations: A_t is $\mathcal{F}_{t-1}$-measurable. The goal of the statistician is typically *best-arm identification*, that is to infer which source has the highest expectation using as few samples as possible. Given a risk parameter δ, a strategy (consisting of a sampling, a stopping and a decision rule) is called δ-*correct* if the probability that the decision $\hat{a}_\tau$ is not the source with highest expectation is smaller than δ. The *sample complexity* is the smallest achievable expectation for the stopping time of a δ-correct strategy.

In the case of a finite set of sources in a one-dimensional canonical exponential

MSC 2010 subject classifications: Primary 62L05

model, the complexity of the best-arm identification problem is well understood since [8, 18]. A non-asymptotic lower bound has been given on the sample complexity: the expectation of the stopping of any δ-correct strategy is always larger than the product of a model-dependent *characteristic time*, which can be computed as the solution of a convex optimization problem depending on the sources, and a confidence term which scales like $\ln(1/\delta)$. This lower-bound is tight: a strategy asymptotically matching it when the risk δ goes to 0 has been described.

But the problem of δ-correct best-arm identification is not really well-posed: the characteristic time of a model tends to infinity when one of the sources has an expectation almost as high as the best one, and δ-correct strategies never stop (with high probability) if two sources are simultaneously optimal. In fact, a much more natural and useful objective is to search for a *Probably Approximately Correct* (PAC) decision: given a risk parameter δ and a tolerance parameter $\epsilon > 0$, a strategy is called (ϵ, δ)-PAC if for every bandit model of the considered class the probability that the chosen source $\hat{a}_\tau$ has a mean smaller that the highest one *minus* ϵ is smaller than δ. Note that a strategy is δ-correct if it is $(0, \delta)$-PAC: the latter notion is thus a generalization of the former.

The purpose of this paper is to show to what extend the results on the complexity of δ-correct best-arm identification can be generalized to the PAC setting. An interesting feature of the latter is that several decisions may simultaneously be correct if several sources are ϵ-optimal. In other words, this supposes to deal with *sequential testing of overlapping hypotheses*, a surprisingly little investigated problem of interest on its own.

The paper is organized as follows: Section 2 is devoted to the general problem of sequential testing of overlapping hypotheses. A generic sequential test based on generalized likelihood ratio tests (called the parallel GLRT) is proposed, together with a non-asymptotic analysis proving its correctness in exponential bandit models. Section 3 details the most simple example: testing if the mean of a Gaussian distribution is smaller than ϵ or larger that $-\epsilon$. Building on these ideas, Section 4 provides our main result on the complexity of PAC best-arm identification. In a nutshell, it appears that a complexity generalizing the results known for δ-correct procedures can be identified, but only asymptotically, under stronger hypothesis, and using less elegant arguments: high-level information-theoretic reasoning cannot be applied here. In Section 5 we then propose an (ϵ, δ)-PAC strategy that combines the parallel GLRT with a "tracking" sampling rule designed for converging to an optimal allocation. This strategy is matching our lower bound for a large family of regular bandit instances.

2. The Parallel GLRT for Testing Overlapping Hypotheses.

2.1. *From sequential to active testing.* In a classical parametric hypothesis testing framework, some data D is collected whose distribution depends on some parameter $\mu \in \mathcal{X}$ and is denoted by $\mathbb{P}_\mu$. Given two hypotheses

$$\mathcal{H}_0 : (\mu \in \mathcal{R}_0) \quad \text{and } \mathcal{H}_1 : (\mu \in \mathcal{R}_1) \ ,$$

where $\mathcal{R}_0$ and $\mathcal{R}_1$ are two *disjoint* subsets of $\mathcal{X}$, the goal is to build a *test* Φ mapping the data to a decision in $\{0,1\}$, where $\Phi(D) = 1$ means that we *reject* hypothesis $\mathcal{H}_0$ based on the observed data. Usually, the data collected consists in a fixed number n of i.i.d. samples from some distribution depending on μ, and the two hypotheses do not play symmetric roles: the risk to reject the null hypothesis H_0 when it is true, called the type I error $\alpha := \sup_{\mu \in \mathcal{R}_0} \mathbb{P}_\mu(\Phi(D) = 1)$, is controlled, and the decision $\Phi(D) = 0$ is not interpreted as an acceptation of $\mathcal{H}_0$, but as a absence of rejection. Under a controlled type I error, the procedure aims at minimizing the type II error $\beta := \sup_{\mu \in \mathcal{R}_1} \mathbb{P}_\mu(\Phi(D) = 0)$, that is at maximizing *power* $1 - \beta$ of the test.

In some cases, however, the hypotheses do play symmetric roles and both type I and type II errors have the same importance. When this happen, we really want to *decide for* $\mathcal{H}_0$ or $\mathcal{H}_1$. This distinction is of importance in what follows, and both cases will be illustrated in our approach.

We focus in this paper on *sequential testing* for *symmetric hypotheses*. In order to emphasize that there is no preferred (null) hypothesis, the hypotheses are denoted by $\mathcal{H}_1$ and $\mathcal{H}_2$. Given a data stream $X_1, X_2, \dots$ and two hypotheses $\mathcal{H}_1 : (\mu \in \mathcal{R}_1)$ and $\mathcal{H}_2 : (\mu \in \mathcal{R}_2)$, the goal is to stop the data collection process after some (random) number τ of observations, and based on this observation $D_\tau = (X_1, \dots, X_\tau)$ make a decision $\hat{\imath}_\tau = \Phi(D_\tau) \in \{1,2\}$, where $\hat{\imath}_\tau = i$ means that hypothesis $\mathcal{H}_i$ is selected. The duration τ of the test should be minimal, but needs to be long enough so that it is actually possible to discriminate between the two hypotheses. This area was pioneered by Wald [19], who propose the Sequential Probability Ratio Test (SPRT) for two simple hypotheses, e.g. in a parametric setting $\mathcal{H}_1 : (\mu = \mu_1)$ and $\mathcal{H}_2 : (\mu = \mu_2)$. Among the sequential tests with prescribed type I and type II errors, the SPRT is proved to have the smallest average duration $\mathbb{E}_{\mu_i}[\tau]$ for $i \in \{1,2\}$. Later, particular examples of sequential test of *composite* hypotheses (in which $\mathcal{R}_1$ and $\mathcal{R}_2$ are not reduced to a singleton) have also been studied, see e.g. [17, 13].

We introduce below in Section 2.2 a broader sequential testing framework, in which we allow for composite hypotheses that are *possibly overlapping*, allowing for $\mathcal{R}_1 \cap \mathcal{R}_2 \neq \varnothing$. Moreover, we allow to simultaneously test more than two hypotheses. In this setup, the goal is to stop the data collection as quickly as possible

in order to identify *one* correct hypothesis. We present a generic sequential test for this problem, which is an extension or the classical Generalized Likelihood Ratio test. Then, we illustrate its efficiency in the context of *active testing*, especially for active identification problems in a multi-armed bandit model.

Active testing can be traced back to the early work of Chernoff [4]: at each round, *an experiment is selected and performed* in order to gain information about which of a finite set of hypotheses is true. An active test is thus made of a *sampling rule* (that indicates which experiment A_t is selected at round t) together with the stopping rule and recommendation rules $(\tau, \hat{\imath})$ inherent to sequential testing. In a multi-armed bandit model, these experiments consists in sampling one of the K available *arms* (probability distributions) $\nu_1, \dots, \nu_K$ with respective means $\mu_1, \dots, \mu_K$. As explained in Section 2.4 below, we assume here that the arms belong to a one-parameter exponential family, and as such can be parameterized by their means.

Many works on bandit models in the past decade have focus on the reinforcement learning framework, where the samples collected are viewed as rewards and the goal is to maximize cumulative rewards (see, e.g. [16, 14, 2]). A broader class of objectives is to *learn something* about the unknown parameter $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$ and several active testing problems have been studied, including the ϵ-best arm identification problem [6] where the goal is to find any arm whose expectation is within ϵ of the highest mean. We show in Section 2.4 that the Parallel GLRT test provides a valid stopping rule for very general tests in bandit models. We investigate in Section 5 how it can help reaching the optimal sample complexity for ϵ-best arm identification, when coupled with a good sampling rule.

2.2. *Sequential Test of Overlapping Hypotheses.* We formally define our framework in a rather general setting in which one sequentially collects samples $X_1, X_2, \dots$ and the joint distribution of these samples depends on some parameter $\mu \in \mathcal{R}$. Let $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_M$ be M regions that cover the parameter space (i.e. such that $\bigcup_{i=1}^{M} \mathcal{R}_i = \mathcal{R}$) but that do not necessarily form a partition of $\mathcal{R}$. We consider the corresponding M hypotheses

$$\mathcal{H}_1 : (\mu \in \mathcal{R}_1) \quad \mathcal{H}_2 : (\mu \in \mathcal{R}_2) \quad \dots \quad \mathcal{H}_M : (\mu \in \mathcal{R}_M) \, .$$

We denote by $(\mathcal{F}_t = \sigma(X_1, \dots, X_t))_{t \geq 1}$ the filtration generated by the observations. A sequential test of the hypotheses $\mathcal{H}_1, \dots, \mathcal{H}_M$ based on the stream of samples $(X_s)_{s \in \mathbb{N}}$ is a pair $(\tau, \hat{\imath}_\tau)$ where τ is a *stopping rule* indicating when to stop the test (hence a is a stopping time with respect to $(\mathcal{F}_t)_t$), and $\hat{\imath}_\tau \in \{1, \dots, M\}$ is a *recommendation rule* that proposes upon stopping a guess for a correct hypothesis based on the observed data (hence $\mathcal{F}_\tau$-measurable).

DEFINITION 1. *Let $\delta \in (0,1)$ be a risk parameter. A sequential test $(\tau_\delta, \hat{\imath}_{\tau_\delta})$ is* δ-correct *if for all $\mu \in \mathcal{R}$, $\mathbb{P}_\mu\left(\tau_\delta < \infty, \mu \notin \mathcal{R}_{\hat{\imath}_{\tau_\delta}}\right) \le \delta$.*

In words, the probability that the test stops and recommends a wrong hypothesis $\mathcal{H}_i$ (such that $\mu \notin \mathcal{R}_i$) should be under control. For example, in the presence of two overlapping hypotheses, the test should be such that $\mathbb{P}_\mu(\hat{\imath}_{\tau_\delta} = 2) \le \delta$ for all $\mu \in \mathcal{R}_1 \backslash \mathcal{R}_2$ and $\mathbb{P}_\mu(\hat{\imath}_{\tau_\delta} = 1) \le \delta$ for all $\mu \in \mathcal{R}_2 \backslash \mathcal{R}_1$; however, for all $\mu \in \mathcal{R}_1 \cap \mathcal{R}_2$ the test may decide for either hypothesis (at any time). Among all δ-correct sequential tests, the preferred one is the one that stops after using as few samples as possible: we want τ_δ to be small in expectation for every $\mu \in \mathcal{R}$.

2.3. *The Parallel Generalized Likelihood Ratio Test.* The likelihood of the observations collected up to round t under the distribution parameter λ is denoted by $\ell(X_1, \dots, X_t; \lambda)$. To decide between the M symmetric hypotheses, we run in parallel M sequential tests of the following two non-overlapping (and asymmetric) hypotheses

$$\tilde{\mathcal{H}}_0 : (\mu \in \mathcal{R} \backslash \mathcal{R}_i) \quad \text{against} \quad \tilde{\mathcal{H}}_1 : (\mu \in \mathcal{R}_i),$$

for each $i \in \{1, \dots, M\}$. We stop when one of these tests rejects hypothesis $\tilde{\mathcal{H}}_0$. The i-th test causing to stop means that μ is believed to belong to $\mathcal{R}_i$ (this does not exclude that it may also belong to other regions), in which case we set $\hat{\imath}_\tau = i$.

Our test of $\tilde{\mathcal{H}}_0$ against $\tilde{\mathcal{H}}_1$ is a Generalized Likelihood Ratio Test (GLRT). The Generalized Likelihood Ratio statistic based on t samples is defined as

$$\frac{\max_{\lambda \in \mathcal{R}} \ell(X_1, \dots, X_t; \lambda)}{\max_{\lambda \in \mathcal{R} \backslash \mathcal{R}_i} \ell(X_1, \dots, X_t; \lambda)} = \inf_{\lambda \in \mathcal{R} \backslash \mathcal{R}_i} \frac{\ell(X_1, \dots, X_t; \hat{\mu}(t))}{\ell(X_1, \dots, X_t; \lambda)} ,$$

where $\hat{\mu}(t)$ is the maximum likelihood estimator (in $\mathcal{R}$). Large values of this GLR statistic tend to reject $\tilde{\mathcal{H}}_0$. Calibrating the rejection threshold for a GLR based on a fixed sample size is often done by resorting to asymptotic arguments (like Wilks phenomenon [20]) describing the limit distribution of the GLR under the null hypothesis; this is however not useful for the finite-confidence bandit analysis that follows. We propose to use a threshold function $\beta(t, \delta)$ that depends on the current number of samples t and on the risk parameter δ. We provide in the cases considered here a valid choice for the threshold $\beta(t, \delta)$ in order to ensure δ-correctness of the corresponding test.

The Parallel GLRT using the threshold function $\beta(t, \delta)$ is formally defined in the following way. Given $\delta \in (0,1)$, the stopping rule is

$$\tau_\delta = \inf \left\{ t \in \mathbb{N} : \max_{i=1,\dots,M} \inf_{\lambda \in \mathcal{R} \backslash \mathcal{R}_i} \ln \frac{\ell(X_1, \dots, X_t; \hat{\mu}(t))}{\ell(X_1, \dots, X_t; \lambda)} > \beta(t, \delta) \right\}$$

and the decision rule is

$$\hat{\imath}_{\tau_\delta} \in \operatorname*{argmax}_{i=1,\dots,M} \inf_{\lambda \in \mathcal{R}\setminus\mathcal{R}_i} \ln \frac{\ell(X_1,\dots,X_\tau;\hat{\mu}(\tau_\delta))}{\ell(X_1,\dots,X_t;\lambda)}$$

(ties can be resolved arbitrarily). Note that the maximum over $i \in \{1,\dots,M\}$ in the definition of τ_δ and $\hat{\imath}_{\tau_\delta}$ can be reduced to the set of hypotheses to which $\hat{\mu}(t)$ belongs.

2.4. *A δ-Correct Parallel GLRT in a Bandit Model.* We consider bandit models where all arms belong to a canonical one-parameter exponential family

$$\mathcal{F} = \left\{ \nu_\theta \text{ of density } f_\theta(x) = \exp(\theta x - b(\theta)) \text{ w.r.t. } \xi;\ \ \theta \in \Theta \right\},$$

where ξ is some reference measure, $\Theta \subseteq \mathbb{R}$ is an interval of natural parameters and $b : \Theta \to \mathbb{R}$ is convex and twice differentiable. For appropriate choices of b and ξ, the class $\mathcal{F}$ can be the set of Bernoulli distributions, Gaussian distribution with a known variance, Poisson, Exponential distributions, etc. We refer the reader to [3] for more details about the use of these exponential families for bandit problems. We just recall that distributions in a one-parameter exponential family can be alternatively parameterized by their means $\mu = \dot{b}(\theta)$. Hence, if $\mathcal{I} = \dot{b}(\Theta)$, an exponential bandit model can be parameterized by $\boldsymbol{\mu} = (\mu_1,\dots,\mu_K) \in \mathcal{I}^K$. Letting ν^μ be the unique distribution that has mean $\mu \in \mathcal{I}$, the Kullback-Leibler divergence between the distribution of mean μ and than of mean μ' is

$$d(\mu,\mu') := \mathrm{KL}\left(\nu^\mu, \nu^{\mu'}\right) = \mu\left(\dot{b}^{-1}(\mu) - \dot{b}^{-1}(\mu')\right) - b\left(\dot{b}^{-1}(\mu)\right) + b\left(\dot{b}^{-1}(\mu')\right) .$$

Remarkable special cases are the family of Gaussian distributions with variance σ^2, with $d(\mu,\mu') = (\mu-\mu')^2/(2\sigma^2)$, and the family of Bernoulli distributions associated to the binary relative entropy $d(\mu,\mu') = \mu \ln \frac{\mu}{\mu'} + (1-\mu)\ln\frac{1-\mu}{1-\mu'}$.

The accumulation of observations in a bandit model is governed by a *sampling rule* which specifies, at each round t, the arm A_t that is selected; a sample X_t from the distribution $\nu^{\mu_{A_t}}$ is subsequently collected. Of course, A_t is only allowed to depend on the past observations $A_1, X_1, \dots, A_{t-1}, X_{t-1}$ and possibly on some exogenous (independent) source of randomness. We consider M regions $\mathcal{R}_1,\dots,\mathcal{R}_M$ covering $\mathcal{I}^K$: $\bigcup_{i=1}^M \mathcal{R}_i = \mathcal{I}^K$. For any given sampling rule, we are interested in building a sequential test $(\tau_\delta, \hat{\imath}_{\tau_\delta})$ that δ-correctly tests the M possibly overlapping hypotheses

$$\mathcal{H}_1 : (\boldsymbol{\mu} \in \mathcal{R}_1)\ \ \mathcal{H}_2 : (\boldsymbol{\mu} \in \mathcal{R}_2)\ \ \dots\ \ \mathcal{H}_M : (\boldsymbol{\mu} \in \mathcal{R}_M) \tag{2.1}$$

based on the stream of observations $A_1, X_1, A_2, X_2, \dots$. We denote by $N_a(t) = \sum_{s=1}^t \mathbb{1}_{(A_s=a)}$ the number of selections of arm a up to round t and use the notation $\hat{\mu}_a(t) = \frac{1}{N_a(t)} \sum_{s=1}^t X_s \mathbb{1}_{(A_s=a)}$ for the empirical mean of those observations.

Letting $\hat{\boldsymbol{\mu}}(t) = (\hat{\mu}_1(t), \dots, \hat{\mu}_K(t))$ be the vector of empirical means of the arms, the log-likelihood ratio can be written as

$$\begin{aligned}
&\ln \frac{\ell(X_1, \dots, X_t; \hat{\boldsymbol{\mu}}(t))}{\ell(X_1, \dots, X_t; \boldsymbol{\lambda})} \\
&= \sum_{a=1}^{K} \left[\hat{\mu}_a(t) \left(\dot{b}^{-1}(\hat{\mu}_a(t)) - \dot{b}^{-1}(\lambda_a) \right) - b\left(\dot{b}^{-1}(\hat{\mu}_a(t)) \right) + b\left(\dot{b}^{-1}(\lambda_a) \right) \right] \\
&= \sum_{a=1}^{K} N_a(t) d\left(\hat{\mu}_a(t), \lambda_a\right) \ . \qquad (2.2)
\end{aligned}$$

Hence, the parallel GLRT for testing (2.1) is defined by

$$\begin{aligned}
\tau_\delta &= \inf \left\{ t \in \mathbb{N} : \max_{a=1,\dots,M} \inf_{\lambda \in \mathcal{R} \setminus \mathcal{R}_a} \sum_{j=1}^{K} N_j(t) d\left(\hat{\mu}_j(t), \lambda_j\right) > \beta(t, \delta) \right\} \ , \qquad (2.3) \\
\hat{\imath}_{\tau_\delta} &= \operatorname*{argmax}_{a=1,\dots,M} \inf_{\lambda \in \mathcal{R} \setminus \mathcal{R}_a} \sum_{j=1}^{K} N_j(\tau_\delta) d\left(\hat{\mu}_j(\tau_\delta), \lambda_j\right) \ .
\end{aligned}$$

In Lemma 2 below, we provide a choice of the threshold function $\beta(t, \delta)$ for which the parallel GLRT is δ-correct, whatever the sampling rule. The definition of β requires to introduce the function

$$\mathcal{T}(x) = 2\tilde{h}\left(\frac{h^{-1}(1+x) + \ln(2\zeta(2))}{2} \right) \qquad (2.4)$$

where, for $u \geq 1$, $h(u) = u - \ln u$ and for any $x \geq 0$

$$\tilde{h}(x) = \begin{cases} e^{1/h^{-1}(x)} h^{-1}(x) & \text{if } x \geq h^{-1}(1/\ln(3/2)), \\ (3/2)(x - \ln\ln(3/2)) & \text{otherwise.} \end{cases}$$

The function $\mathcal{T}$ is easy to compute numerically. Its use for the construction of anytime, self-normalized confidence intervals is detailed in [11], where the following approximations are derived: $\mathcal{T}(x) \simeq x + 4\ln\left(1 + x + \sqrt{2x}\right)$ for $x \geq 5$ and $\mathcal{T}(x) \sim x$ when x is large.

LEMMA 2. *For any sampling rule, the parallel GLRT test $(\tau_\delta, \hat{\imath}_{\tau_\delta})$ using the threshold function*

$$\beta(t, \delta) = 3K \ln(1 + \ln t) + K\mathcal{T}\left(\frac{\ln(1/\delta)}{K} \right)$$

is δ-correct: for all $\boldsymbol{\mu} \in \mathcal{R}$, $\mathbb{P}_{\boldsymbol{\mu}}\left(\tau_\delta < \infty, \mu \notin \mathcal{R}_{\hat{\imath}_{\tau_\delta}}\right) \leq \delta$.

In can be noted that in the particular case of Gaussian bandits Lemma 2 holds for a slightly smaller function $\mathcal{T}$ that satisfies $\mathcal{T}(x) \simeq x + \ln(x)$.

PROOF. Let $\boldsymbol{\mu} \in \mathcal{R}$. The probability of error $\mathbb{P}_{\boldsymbol{\mu}}\left(\tau_\delta < \infty, \boldsymbol{\mu} \notin \mathcal{R}_{\hat{\imath}_{\tau_\delta}}\right)$ can be upper-bounded by

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\mu}}\left(\exists t, \exists i : \boldsymbol{\mu} \notin \mathcal{R}_i, \inf_{\boldsymbol{\lambda} \in \mathcal{R}\setminus\mathcal{R}_i} \ln \frac{\ell(X_1,\dots,X_t;\hat{\boldsymbol{\mu}}(t))}{\ell(X_1,\dots,X_t;\boldsymbol{\lambda})} > \beta(t,\delta)\right) \\
&\le \mathbb{P}_{\boldsymbol{\mu}}\left(\exists t, \exists i : \boldsymbol{\mu} \in \mathcal{R}\setminus\mathcal{R}_i, \ln \frac{\ell(X_1,\dots,X_t;\hat{\boldsymbol{\mu}}(t))}{\ell(X_1,\dots,X_t;\boldsymbol{\mu})} > \beta(t,\delta)\right) \\
&\le \mathbb{P}_{\boldsymbol{\mu}}\left(\exists t, \ln \frac{\ell(X_1,\dots,X_t;\hat{\boldsymbol{\mu}}(t))}{\ell(X_1,\dots,X_t;\boldsymbol{\mu})} > \beta(t,\delta)\right) .
\end{aligned}$$

Hence, to prove the δ-correctness of the parallel GLRT, it is sufficient to control the deviations of the quantity $\ln \frac{\ell(X_1,\dots,X_t;\hat{\boldsymbol{\mu}}(t))}{\ell(X_1,\dots,X_t;\boldsymbol{\mu})}$ uniformly over time. For this purpose, in the case of an exponential family bandit model, one can use the deviations bounds recently obtained by [11]. More precisely, thanks to Equation (2.2),

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\mu}}\left(\tau_\delta < \infty, \boldsymbol{\mu} \notin \mathcal{R}_{\hat{\imath}_{\tau_\delta}}\right) \le \mathbb{P}_{\boldsymbol{\mu}}\left(\exists t \in \mathbb{N} : \sum_{a=1}^{K} N_a(t) d(\hat{\mu}_a(t), \mu_a) > \beta(t,\delta)\right) \\
&\le \mathbb{P}_{\boldsymbol{\mu}}\left(\exists t \in \mathbb{N} : \sum_{a=1}^{K} N_a(t) d(\hat{\mu}_a(t), \mu_a) > 3\sum_{a=1}^{K} \ln(1 + \ln N_a(t)) + K\mathcal{T}\left(\frac{\ln(1/\delta)}{K}\right)\right),
\end{aligned}$$

which is upper bounded by δ according to Theorem 14 in [11]. In the Gaussian case, Corollary 10 in the same paper justifies the use of a slightly smaller function $\mathcal{T}$.

□

Lemma 2 provides a universal choice of threshold for which the parallel GLRT is δ-correct, whatever the sampling rule. Obviously, however, in order to minimize the number of samples τ_δ needed before stopping the test, the sampling rule also plays a crucial role. For example, if for $K = 3$ the question is whether $\mu_2 > \mu_3$, there is no point in sampling from ν_1. We show below how a *lower-bound* analysis of the sample complexity indirectly leads to an efficient sampling rule.

In the remainder of the paper we investigate the expected number of samples used by a parallel GLRT in two different settings. We first consider the simplest possible data-generating process (a one-armed Gaussian bandit): based on i.i.d. Gaussian samples, the goal is to decide whether the mean is "rather positive" or "rather negative". Then, with the particular example of PAC identification in a bandit model, we study *active testing* based on parallel GLRT, in which we further *optimize the sampling rule* in order to minimize the number of samples used.

3. Sequential Test of Overlapping Hypotheses: an Archetypal Example. We focus here in this section on a stream of independent samples $X_1, X_2, \dots$ of the distribution $\mathcal{N}\left(\mu, \sigma^2\right)$, where the variance σ^2 is known. The statistical problem is to determine whether μ is positive or negative, but with and indifference region of width 2ϵ: we consider the two hypotheses

$$\mathcal{H}_1 : (\mu < \epsilon) \quad \text{and} \quad \mathcal{H}_2 : (\mu > -\epsilon) \, .$$

The hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$ are not mutually exclusive. In this particular example, one aims at building a stopping rule τ_δ and a recommendation rule $\hat{\imath}_{\tau_\delta}$ such that

$$\forall \mu \le -\epsilon, \ \mathbb{P}_\mu(\hat{\imath}_{\tau_\delta} = 2) \le \delta \quad \text{and} \quad \forall \mu \ge \epsilon, \ \mathbb{P}_\mu(\hat{\imath}_{\tau_\delta} = 1) \le \delta \, , \tag{3.1}$$

but any answer $\hat{\imath}_{\tau_\delta}$ is considered correct when $\mu \in \mathcal{R}_1 \cap \mathcal{R}_2 = (-\epsilon, \epsilon)$.

3.1. *Properties of the Parallel GLRT Test.* The expression of the parallel GLRT for this test can be obtained by particularizing the general expression obtained in exponential family bandits (2.3) to $K = 1$, Gaussian distribution for which $d(x, y) = (x-y)^2/(2\sigma^2)$ and the two above hypotheses for which $\mathcal{R}\backslash\mathcal{R}_1 = [\epsilon, +\infty)$ and $\mathcal{R}\backslash\mathcal{R}_2 = (-\infty, -\epsilon]$. Letting $\hat{\mu}_s = \frac{1}{s}\sum_{i=1}^s X_i$ denote the empirical mean of the observations, one can easily show that

$$\tau_\delta = \inf\left\{ t \in \mathbb{N} : \frac{t(|\hat{\mu}_t| + \epsilon)^2}{2\sigma^2} > \beta(t, \delta) \right\} \, , \tag{3.2}$$

and $(\hat{\imath}_{\tau_\delta} = 2)$ if and only if $(\hat{\mu}_{\tau_\delta} > 0)$.

From Lemma 2, we know that the parallel GLRT is δ-correct for the choice of threshold

$$\beta(t, \delta) = 3\ln(\ln(t) + 1) + \mathcal{T}\left(\ln(1/\delta)\right) , \tag{3.3}$$

where $\mathcal{T}$ is such that $\mathcal{T}(x) \simeq x + \ln(x)$. Under the simple sampling rule considered in this section (that always selects arm 1), it is possible to give an upper bound on the expected number of samples used by the parallel GLRT that employs the above threshold.

We first give a crude, asymptotic analysis. Fix $\mu \in \mathbb{R}$ and let $\alpha \in [0, \epsilon)$.

$$\begin{aligned}
\mathbb{E}[\tau_\delta] &\le \sum_{t=1}^{\infty} \mathbb{P}\left(t(|\hat{\mu}_t| + \epsilon)^2 \le 2\sigma^2\beta(t, \delta)\right) \\
&\le \sum_{t=1}^{\infty} \mathbb{P}\left(|\hat{\mu}_t - \mu| > \alpha\right) + \sum_{t=1}^{\infty} \mathbb{P}\left(t(|\hat{\mu}_t| + \epsilon)^2 \le 2\sigma^2\beta(t, \delta), |\hat{\mu}_t - \mu| \le \alpha\right) \\
&\le \sum_{t=1}^{\infty} \mathbb{P}\left(|\hat{\mu}_t - \mu| > \alpha\right) + \sum_{t=1}^{\infty} \mathbb{P}\left(t(|\mu| - \alpha + \epsilon)^2 \le 2\sigma^2\beta(t, \delta), |\hat{\mu}_t - \mu| \le \alpha\right) \, .
\end{aligned}$$

The first term is upper bounded by a constant (independent of δ), while the second is upper bounded by

$$T_0(\delta) = \inf\left\{t \in \mathbb{N}^* : \forall t' \geq t, t(|\mu| - \alpha + \epsilon)^2 \leq 2\sigma^2\beta(t,\delta)\right\} .$$

For $\beta(t,\delta)$ as given in (3.3), Lemma 17 in Appendix A shows that

$$T_0(\delta) = \frac{2\sigma^2}{(|\mu| - \alpha + \epsilon)^2} \ln\frac{1}{\delta} + o_{\delta\to 0}\left(\ln\frac{1}{\delta}\right) .$$

Letting α go to zero, one obtains, for all $\mu \in \mathbb{R}$,

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_\mu[\tau_\delta]}{\ln(1/\delta)} \leq \frac{2\sigma^2}{(|\mu| + \epsilon)^2} .$$

The following non-asymptotic statement is proved in Appendix A.

THEOREM 3.1. *Introducing the notation* $\ell(\delta) = \mathcal{T}(\ln(1/\delta)) + \frac{3}{e}$, *the parallel GLRT with threshold function* $\beta(t,\delta)$ *given in* (3.3) *satisfies*

$$\mathbb{E}[\tau_\delta] \leq \frac{2\sigma^2}{(|\mu| + \epsilon)^2}\left[\ell(\delta) + \frac{6}{e}\ln\left(\frac{2\sigma^2}{(|\mu| + \epsilon)^2}\ell(\delta)\right) + 8\sqrt{\ell(\delta) + \frac{6}{e}\ln\left(\frac{2\sigma^2}{(|\mu| + \epsilon)^2}\ell(\delta)\right)} + 32\delta^{1/8}\right] + 1 .$$

To summarize, the parallel GLRT with threshold function (3.3) is δ-correct and the expected number of samples uses satisfies $\mathbb{E}_\mu[\tau_\delta] \lesssim \frac{2\sigma^2}{(|\mu|+\epsilon)^2}\ln\left(\frac{1}{\delta}\right)$ for small values of the risk parameter δ, for every $\mu \in \mathbb{R}$, even in the indifference zone $(-\epsilon, \epsilon)$. We prove in the next section that this is the smallest possible number of samples for a δ-correct test.

3.2. *Asymptotic Optimality of the Parallel GLRT.* In order to prove the optimality of the parallel GLRT, we need to provide a *lower bound* on the sample complexity of any δ-correct algorithm. To provide lower bound it is quite common to rely on a *change of distribution* argument.

A change of distribution consists in relating the probability of an event C under two different probability distributions. It is well-known that for any event $C \in \mathcal{F}_t$,

$$\begin{aligned}
\mathbb{P}_\lambda(C) = \mathbb{E}_\lambda[\mathbb{1}_C] &= \int \mathbb{1}_C(x_1,\dots,x_t)\, d\mathbb{P}_\lambda^{X_1,\dots,X_n}(x_1,\dots,x_n) \\
&= \int \mathbb{1}_C(x_1,\dots,x_t) \frac{d\mathbb{P}_\lambda^{X_1,\dots,X_t}(x_1,\dots,x_t)}{d\mathbb{P}_\mu^{X_1,\dots,X_t}(x_1,\dots,x_t)}\, d\mathbb{P}_\mu^{X_1,\dots,X_t}(x_1,\dots,x_t) \\
&= \int \mathbb{1}_C(x_1,\dots,x_t) \frac{\ell(X_1,\dots,X_t;\lambda)}{\ell(X_1,\dots,X_t;\mu)}\, d\mathbb{P}_\mu^{X_1,\dots,X_t}(x_1,\dots,x_t) \\
&= \mathbb{E}_\mu\Big[\mathbb{1}_C \exp\big(-L_t(\mu,\lambda)\big)\Big],
\end{aligned}$$

where $L_t(\mu,\lambda)$ denotes the log-likelihood ratio of the observations:

$$L_t(\mu,\lambda) := \ln \frac{\ell(X_1,\dots,X_t;\mu)}{\ell(X_1,\dots,X_t;\lambda)}.$$

This elementary change of distribution has been a key ingredient for the derivation of lower bounds in the bandit literature. The most notable example is the famous regret lower bound by Lai and Robbins in the seminal paper [14], and further examples include notably [1].

But this "low-level" ingredient can also serve as a basis for more elaborate arguments, which lead to more elegant and stronger results as for example illustrated in [8]. We propose in Lemma 3 below two forms of change of distribution that are useful: the low-level form (3.4), and the high-level form (3.5). The high-level form, of information-theoretic flavor, involves the Kullback-Leibler divergence $\mathrm{KL}(\cdot,\cdot)$ from one distribution to another. For any random variable Z defined on a probabilistic space $(\Omega,\mathcal{F},\mathbb{P})$, we denote by $\mathbb{P}^Z$ the law of Z.

LEMMA 3. *Let μ and λ be two parameters.*

1. ***High-level form:*** *for any stopping time τ and any event $C \in \mathcal{F}_\tau$,*

$$\mathbb{E}_\mu[L_\tau(\mu,\lambda)] = \mathrm{KL}\left(\mathbb{P}_\mu^{X_1,\dots,X_\tau}, \mathbb{P}_\lambda^{X_1,\dots,X_\tau}\right) \geq \mathrm{kl}\big(\mathbb{P}_\mu(C), \mathbb{P}_\lambda(C)\big)\,, \tag{3.4}$$

   where $\mathrm{kl}(x,y) = x\ln(x/y) + (1-x)\ln((1-x)/(1-y))$ is the binary relative entropy.
2. ***Low-level form:*** *for all $x \in \mathbb{R}$, $n \in \mathbb{N}^*$, for all event $C \in \mathcal{F}_n$,*

$$\mathbb{P}_\lambda(C) \geq e^{-x}\left[\mathbb{P}_\mu(C) - \mathbb{P}_\mu\left(L_n(\mu,\lambda) \geq x\right)\right]. \tag{3.5}$$

PROOF. The high-level form is a consequence of the data-processing inequality:

$$\mathbb{E}_\mu[L_\tau(\mu,\lambda)] = \mathrm{KL}\left(\mathbb{P}_\mu^{X_1,\dots,X_\tau}, \mathbb{P}_\lambda^{X_1,\dots,X_\tau}\right) \geq \mathrm{KL}\left(\mathbb{P}_\mu^{\mathbb{1}_C}, \mathbb{P}_\lambda^{\mathbb{1}_C}\right) = \mathrm{kl}\big(\mathbb{P}_\mu(C), \mathbb{P}_\lambda(C)\big)\,,$$

see [9] for details. The low-level form is even more elementary:

$$\begin{aligned}\mathbb{P}_\lambda(C) &= \mathbb{E}_\mu\Big[\mathbb{1}_C \exp\big(-L_t(\mu,\lambda)\big)\Big] \geq \mathbb{E}_\mu\Big[\mathbb{1}_C \mathbb{1}_{(L_t(\mu,\lambda)<x)} \exp\big(-L_t(\mu,\lambda)\big)\Big] \\ &\geq e^{-x}\,\mathbb{P}_\mu\Big(C \cap \big(L_t(\mu,\lambda) < x\big)\Big) \\ &\geq e^{-x}\left[\mathbb{P}_\mu(C) - \mathbb{P}_\mu\left(L_n(\mu,\lambda) \geq x\right)\right].\end{aligned}$$

□

The sample complexity lower bounds provided below illustrate the strengths and weaknesses of each form. In particular, Lemma 3 permits to identify the sample complexity of the testing problem with overlapping hypotheses studied in this section. Indeed, Theorem 4 below shows that any δ-correct sequential procedure satisfies $\mathbb{E}_\mu[\tau_\delta] \geq \frac{2\sigma^2}{(|\mu|+\epsilon)^2}\ln\left(\frac{1}{\delta}\right)$ for every μ, in a regime of small values of δ. In Section 3.1 we exhibited a δ-correct parallel GLRT that uses not more than this number of samples for small values of δ. This proves that the sample complexity of the testing problem, that is the minimal number of samples needed by a δ-correct test, is close to $\frac{2\sigma^2}{(|\mu|+\epsilon)^2}\ln\left(\frac{1}{\delta}\right)$ for small values of the risk δ, and justifies that the parallel GLRT, which attains this sample complexity, is asymptotically optimal.

It is notable that the lower bound of Theorem 4 is non-asymptotic for all μ that belong to a single hypothesis, which is obtained using the high-level form in Lemma 3, but only asymptotic when μ is in the intersection $(-\epsilon,\epsilon)$, which requires the low-level form, as can be seen in the proof below.

THEOREM 4. *Any δ-correct sequential test satisfies*

- $\forall \mu \notin (-\epsilon,\epsilon), \quad \mathbb{E}_\mu[\tau_\delta] \geq \dfrac{2\sigma^2}{\big(|\mu|+\epsilon\big)^2}\,\mathrm{kl}(\delta, 1-\delta)$,
- $\forall \mu \in \mathbb{R}, \quad \liminf\limits_{\delta\to 0} \dfrac{\mathbb{E}_\mu[\tau_\delta]}{\ln(1/\delta)} \geq \dfrac{2\sigma^2}{\big(|\mu|+\epsilon\big)^2}$.

PROOF. We first treat the case $\mu \notin (-\epsilon,\epsilon)$. In the particular case considered in this section the samples $X_1,\dots,X_t$ are i.i.d. from a Gaussian distribution. Assuming that $\mathbb{E}_\mu[\tau] < \infty$ and using Wald's lemma, the expected log-likelihood ratio takes the simple form

$$\mathbb{E}_\mu[L_\tau(\mu,\lambda)] = \mathbb{E}_\mu[\tau]\mathrm{KL}\big(\mathcal{N}(\mu,\sigma^2),\mathcal{N}(\lambda,\sigma^2)\big) = \mathbb{E}_\mu[\tau]\frac{(\mu-\lambda)^2}{2\sigma^2}\ .$$

Hence, the high-level change of measure inequality (3.4) translates into the following inequality, which obviously also holds when $\mathbb{E}_\mu[\tau] = \infty$:

$$\mathbb{E}_\mu[\tau]\frac{(\mu-\lambda)^2}{2\sigma^2} \geq \mathrm{kl}\big(\mathbb{P}_\mu(C)\ ,\mathbb{P}_\lambda(C)\big)\ . \tag{3.6}$$

For $\mu < -\epsilon$, choosing $\lambda = \epsilon$ and $C = (\hat{\imath}_{\tau_\delta} = 2)$, which is such that $\mathbb{P}_\mu(C) \le \delta$ and $\mathbb{P}_\lambda(C) \ge 1 - \delta$, directly yields

$$\mathbb{E}_\mu[\tau] \frac{\big(|\mu| + \epsilon\big)^2}{2\sigma^2} \ge \mathrm{kl}(\delta, 1-\delta) \; .$$

Similarly, for $\mu > \epsilon$, we use $\lambda = -\epsilon$ and $C = (\hat{\imath}_{\tau_\delta} = 1)$ so as to obtain the same inequality.

We now turn our attention to the (more interesting) case $\mu \in (-\epsilon, \epsilon)$. For every $\beta > 0$, we prove that

$$\lim_{\delta \to 0} \mathbb{P}_\mu \left(\tau_\delta \le \frac{2\sigma^2(1-\beta)}{(|\mu| + \epsilon)^2} \ln(1/\delta) \right) = 0 \; , \tag{3.7}$$

which entails that

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\ln(1/\delta)} \ge \frac{2\sigma^2(1-\beta)}{(|\mu| + \epsilon)^2} \; .$$

Letting β tend to zero gives the second statement of Theorem 4 for $\mu \in (-\epsilon, \epsilon)$. Note that this statement is also true for $\mu \notin (-\epsilon, \epsilon)$, as a consequence of the non-asymptotic lower bound obtained above and the fact that $\mathrm{kl}(\delta, 1-\delta) \sim \ln(1/\delta)$ when δ goes to zero.

We now prove (3.7) for a fixed $\beta > 0$. Introducing

$$n_\delta := \left\lfloor \frac{2\sigma^2(1-\beta)}{\big(|\mu| + \epsilon\big)^2} \ln \frac{1}{\delta} \right\rfloor \; ,$$

and the event $C_\delta = (\tau_\delta \le n_\delta)$, we need to prove that $\mathbb{P}_\mu(C_\delta) \to 0$ when $\delta \to 0$. To do so, we write

$$\mathbb{P}_{\boldsymbol{\mu}} \left(C_\delta \right) = \mathbb{P}_{\boldsymbol{\mu}} \left(C_\delta, \hat{\imath}_{\tau_\delta} = 1 \right) + \mathbb{P}_{\boldsymbol{\mu}} \left(C_\delta, \hat{\imath}_{\tau_\delta} = 2 \right) \; ,$$

and prove that the two terms in the RHS tend to zero when δ goes to zero, using the low-level form of Lemma 3 with two different changes of distribution.

$\boldsymbol{\lim_{\delta \to 0} \mathbb{P}_{\mu} \left(C_\delta, \hat{\imath}_{\tau_\delta} = 1 \right) = 0}$. Choosing $\lambda = \epsilon$ yields that $\mathbb{P}_\lambda(C_\delta, \hat{\imath}_{\tau_\delta} = 1) \le \mathbb{P}_\lambda(\hat{\imath}_{\tau_\delta} = 1) \le \delta$. Since τ_δ is a stopping time, the event $C_\delta \cap (\hat{\imath}_{\tau_\delta} = 1)$ belongs to $\mathcal{F}_{n_\delta}$. Hence, by Inequality (3.5), for all $x \in \mathbb{R}$,

$$\delta \ge e^{-x} \left[\mathbb{P}_\mu \left(C_\delta, \hat{\imath}_{\tau_\delta} = 1 \right) - \mathbb{P}_\mu \left(L_{n_\delta}(\mu, \epsilon) \ge x \right) \right] \; ,$$

which can be rewritten as

$$\mathbb{P}_\mu \left(C_\delta, \hat{\imath}_{\tau_\delta} = 1 \right) \le \delta e^x + \mathbb{P}_\mu \left(L_{n_\delta}(\mu, \epsilon) \ge x \right) \; .$$

The choice $x = (1-\beta/2)\ln(1/\delta)$ yields

$$\mathbb{P}_\mu\left(C_\delta, \hat{\imath}_{\tau_\delta} = 1\right) \leq \delta^{\frac{\beta}{2}} + \mathbb{P}_\mu\left(\frac{L_{n_\delta}(\mu,\epsilon)}{n_\delta} \geq \frac{1-\beta/2}{1-\beta}\frac{\left(|\mu|+\epsilon\right)^2}{2\sigma^2}\right). \tag{3.8}$$

By the law of large numbers and the fact that $n_\delta \to +\infty$ when $\delta \to 0$,

$$\frac{L_{n_\delta}(\mu,\epsilon)}{n_\delta} \xrightarrow{\delta\to 0} \mathbb{E}_\mu\left[\ln\frac{\ell(X_1;\mu)}{\ell(X_1;\epsilon)}\right] = \mathrm{KL}\left(\mathbb{P}_\mu^{X_1}, \mathbb{P}_\epsilon^{X_1}\right) = \frac{(\mu-\epsilon)^2}{2\sigma^2} < \frac{1-\beta/2}{1-\beta}\frac{\left(|\mu|+\epsilon\right)^2}{2\sigma^2},$$

and the RHS of Equation (4.5) goes to zero.

$\boldsymbol{\lim_{\delta\to 0}\mathbb{P}_\mu\left(C_\delta, \hat{\imath}_{\tau_\delta} = 2\right) = 0}$. Using the exact same reasoning as above, this time with $\lambda = -\epsilon$ which is such that $\mathbb{P}_\lambda(C_\delta, \hat{\imath}_{\tau_\delta} = 2) \leq \delta$, yields

$$\mathbb{P}_\mu\left(C_\delta, \hat{\imath}_{\tau_\delta} = 2\right) \leq \delta^{\frac{\beta}{2}} + \mathbb{P}_\mu\left(\frac{L_{n_\delta}(\mu,-\epsilon)}{n_\delta} \geq \frac{1-\beta/2}{1-\beta}\frac{\left(|\mu|+\epsilon\right)^2}{2\sigma^2}\right). \tag{3.9}$$

Writing, thanks to the law of large numbers,

$$\frac{L_{n_\delta}(\mu,-\epsilon)}{n_\delta} \xrightarrow{\delta\to 0} \frac{(\mu+\epsilon)^2}{2\sigma^2} < \frac{1-\beta/2}{1-\beta}\frac{\left(|\mu|+\epsilon\right)^2}{2\sigma^2},$$

proves that the RHS in (3.9) goes to zero.

□

4. Sample Complexity Lower Bounds for Active Testing: ϵ-Best Arm Identification. ϵ-Best-Arm Identification (BAI) was first introduced by [15, 6] in the bandit literature. Given a risk parameter $\delta \in (0,1]$ and an accuracy parameter $\epsilon > 0$, the goal is to find as quickly as possible an element of the set $\mathcal{A}_\epsilon(\boldsymbol{\mu}) := \{a : \mu_a \geq \max_i \mu_i - \epsilon\}$ of ϵ-optimal arms, with probability at least $1-\delta$, by adaptively sampling the arms. A slightly different formulation with an *indifference zone* has been studied in the ranking and selection community (see, e.g.,[12]) in which correctness guarantees are provided only when there is a unique ϵ-best arm (i.e. for $\boldsymbol{\mu}$ such that $|\mathcal{A}_\epsilon(\boldsymbol{\mu})| = 1$). In contrast, we highlight that in ϵ-BAI, in the presence of multiple ϵ-best arms (i.e., overlapping hypotheses), our goal is to identify one of those.

ϵ-best arm identification can be regarded as a particular example of *active testing* of the hypotheses

$$\mathcal{H}_1 = \left(\mu_1 > \max_i \mu_i - \epsilon\right) \quad \dots \quad \mathcal{H}_K = \left(\mu_K > \max_i \mu_i - \epsilon\right)$$

based on adaptively sampling the marginals of $f_{\boldsymbol{\mu}} = f_{\mu_1} \otimes \cdots \otimes f_{\mu_K}$.

A strategy for ϵ-BAI not only consists in a sequential test for those hypotheses, that is a pair $(\tau, \hat{\imath})$ of stopping and recommendation rule, but also on a sampling rule $(A_t)_t$ indicating which arm is selected at round t. As explained in Section 2.4, we consider the case where the arms belongs to a one-dimension canonical exponential family and their distribution is therefore parameterized by their means that belong to some interval $\mathcal{I} = (\mu^-, \mu^+)$.

DEFINITION 5. *A (ϵ, δ)-PAC strategy for ϵ-best arm identification is a triple $(A_t^\delta, \tau_\delta, \hat{\imath}_{\tau_\delta})$, where A_t^δ is $\mathcal{F}_{t-1}$-measurable for each $t \geq 1$, τ is a stopping time relative to the filtration $(\mathcal{F}_t)_t$, and $\hat{\imath}_{\tau_\delta}$ is an $\mathcal{F}_\tau$-measurable random variable such that*

$$\forall \boldsymbol{\mu} \in \mathcal{I}^K, \quad \mathbb{P}_{\boldsymbol{\mu}}\big(\tau_\delta < \infty, \hat{\imath}_{\tau_\delta} \notin \mathcal{A}_\epsilon(\boldsymbol{\mu})\big) \leq \delta \, .$$

An ϵ-PAC strategy is a family of strategies $\pi_\delta = (A_t^\delta, \tau_\delta, \hat{\imath}_{\tau_\delta})$ such that, for all δ, π_δ is (ϵ, δ)-PAC.

The goal in ϵ-BAI is to build an (ϵ, δ)-PAC strategy such that on every bandit instance $\boldsymbol{\mu}$, the expected number of samples needed before stopping $\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]$, is as close as possible to the *sample complexity*, which is the minimal number of samples needed by any (ϵ, δ)-PAC algorithm before stopping. In this section, we provide lower bounds on the sample complexity of ϵ-Best Arm Identification.

4.1. *The Characteristic Time of ϵ-Best Arm Identification.* All the lower bounds that we propose feature a quantity $T_\epsilon^*(\boldsymbol{\mu})$ which we term the characteristic time of μ.

DEFINITION 6. *The characteristic time $T_\epsilon^*(\boldsymbol{\mu})$ is defined by:*

$$T_\epsilon^*(\boldsymbol{\mu})^{-1} = \sup_{w \in \Sigma_K} \max_{a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})} \min_{b \neq a} \inf_{(\lambda_a, \lambda_b) : \lambda_a \leq \lambda_b - \epsilon} \Big[w_a d(\mu_a, \lambda_a) + w_b d(\mu_b, \lambda_b) \Big],$$

where $\Sigma_K = \{w \in [0,1]^K : \sum_i w_i = 1\}$ is the set of probability vectors.

It can be observed that the quantity $T_0^*(\boldsymbol{\mu})$ coincides with the quantity $T^*(\boldsymbol{\mu})$ introduced by [8] as the characteristic time for the best arm identification problem.

In order to illustrate the dependency of $T_\epsilon^*(\boldsymbol{\mu})$ on the means $\boldsymbol{\mu}$, we first consider the Gaussian case. Introducing the notation $\boldsymbol{\mu}^{a,\epsilon}$ for the bandit instance such that $\mu_i^{a,\epsilon} = \mu_i$ for all $i \neq a$ and $\mu_a^{a,\epsilon} = \mu_a + \epsilon$, it can be shown (see Appendix B.1) that

$$T_\epsilon^*(\boldsymbol{\mu}) = \min_{a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})} T_0\left(\boldsymbol{\mu}^{a,\epsilon}\right) = T_0\left(\boldsymbol{\mu}^{a^*,\epsilon}\right), \tag{4.1}$$

where $a^* \in \mathrm{argmax}_a \mu_a$ is an optimal arm. Hence it is sufficient to be able to compute the characteristic time for $\epsilon = 0$, for which [8] provides an efficient algorithm. Besides, [8] proposes an approximation for $T_0^*(\boldsymbol{\mu})$ from which it directly follows that

$$\sum_{a \neq a^*} \frac{2\sigma^2}{(\mu_{a^*} + \epsilon - \mu_a)^2} \leq T_\epsilon^*(\boldsymbol{\mu}) \leq 2 \times \sum_{a \neq a^*} \frac{2\sigma^2}{(\mu_{a^*} + \epsilon - \mu_a)^2} \,.$$

Finally, there is a simple expression in the two armed-case: $T_\epsilon^*(\boldsymbol{\mu}) = \frac{8\sigma^2}{(|\mu_1 - \mu_2| + \epsilon)^2}$. Explicit expressions for general (non-Gaussian) two-armed bandits can be found in Section 4.4. However, beyond 2 arms there is no closed-form expression for the characteristic time $T_\epsilon^*(\boldsymbol{\mu})$ and we refer to Section 5 for a discussion on its efficient computation.

4.2. *A First Non-Asymptotic Lower Bound.* To illustrate the difficulty of extending the lower bound technique of [8] to ϵ-BAI, we begin with a sub-optimal but simple sample complexity lower bound, obtained using the high-level change of distribution argument presented in Lemma 3.

THEOREM 7. *For any (ϵ, δ)-PAC strategy, one has*

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta] \geq T_\epsilon^*(\boldsymbol{\mu}) \left[\frac{1-\delta}{|\mathcal{A}_\epsilon(\boldsymbol{\mu})|} \ln\left(\frac{1}{\delta}\right) - \ln(2) \right] .$$

PROOF. Using the δ-PAC property and the pigeonhole principle, there exists $a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})$ such that

$$\mathbb{P}_{\boldsymbol{\mu}}(\hat{\imath}_{\tau_\delta} = a) \geq \frac{1-\delta}{|\mathcal{A}_\epsilon(\boldsymbol{\mu})|}.$$

By definition of $T_\epsilon^*(\boldsymbol{\mu})$, for the above choice of $a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})$, there exists $b \neq a$ and $\overline{\boldsymbol{\lambda}}$ such that $\overline{\lambda}_a < \overline{\lambda}_b - \epsilon$:

$$T_\epsilon^*(\boldsymbol{\mu})^{-1} \geq \frac{\mathbb{E}_{\boldsymbol{\mu}}[N_a(\tau_\delta)]}{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]} d(\mu_a, \overline{\lambda}_a) + \frac{\mathbb{E}_{\boldsymbol{\mu}}[N_b(\tau_\delta)]}{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]} d(\mu_b, \overline{\lambda}_b) \,. \tag{4.2}$$

Indeed, fixing $\overline{w}_a := \mathbb{E}_\mu[N_a(\tau_\delta)]/\mathbb{E}_\mu[\tau_\delta]$ and defining

$$F_a(w, \boldsymbol{\mu}) = \min_{b \neq a} \inf_{\lambda_a < \lambda_b - \epsilon} \left[w_a d(\mu_a, \lambda_a) + w_b d(\mu_b, \lambda_b) \right] ,$$

observe that

$$\sup_{w \in \Sigma_K} F_a(w, \boldsymbol{\mu}) \geq F_a(\overline{w}, \boldsymbol{\mu}) \,.$$

Letting $b, \overline{\lambda}_a, \overline{\lambda}_b$ be the minimizer in the definition of $F_a(\overline{w}, \boldsymbol{\mu})$ yields

$$\max_{a \in \mathcal{A}_\epsilon} \sup_{w \in \Sigma_K} F_a(w, \boldsymbol{\mu}) \geq \overline{w}_a d(\mu_a, \overline{\lambda}_a) + \overline{w}_b d(\mu_b, \overline{\lambda}_b) \,,$$

and the left hand side of this inequality is $T_\epsilon^{-1}(\boldsymbol{\mu})$.

We now assume $\mathbb{E}[\tau_\delta] < \infty$ (otherwise any lower bound on the number of samples of that algorithm obviously holds). Using Wald's Lemma, the expression of the expected log-likelihood ratio between $\boldsymbol{\mu}$ and $\overline{\boldsymbol{\lambda}}$ is

$$\mathbb{E}_{\boldsymbol{\mu}}\left[L_{\tau_\delta}(\boldsymbol{\mu},\overline{\boldsymbol{\lambda}})\right] = \mathbb{E}_{\boldsymbol{\mu}}[N_a(\tau_\delta)]d(\mu_a,\overline{\lambda}_a) + \mathbb{E}_{\boldsymbol{\mu}}[N_b(\tau_\delta)]d(\mu_b,\overline{\lambda}_b)\ .$$

Therefore, one can use Inequality (4.2) and the high-level form of Lemma 3 to write

$$\begin{aligned}\frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{T_\epsilon^*(\boldsymbol{\mu})} &\geq \mathbb{E}_{\boldsymbol{\mu}}\left[L_{\tau_\delta}(\boldsymbol{\mu},\overline{\boldsymbol{\lambda}})\right] \ \geq\ \mathrm{kl}(\mathbb{P}_{\boldsymbol{\mu}}(\hat{\imath}_{\tau_\delta}=a),\mathbb{P}_{\overline{\boldsymbol{\lambda}}}(\hat{\imath}_{\tau_\delta}=a))\\ &\geq \mathrm{kl}\left(\frac{1-\delta}{|\mathcal{A}_\epsilon(\boldsymbol{\mu})|},\delta\right) \ \geq\ \frac{1-\delta}{|\mathcal{A}_\epsilon|}\ln\left(\frac{1}{\delta}\right)-\ln(2)\ ,\end{aligned}$$

where the last inequality follows from the inequality $\mathrm{kl}(x,y) \geq x\ln(1/y) - \ln(2)$.

□

4.3. *A Tighter Lower Bound for Converging Strategies.* The lower bound obtained in Theorem 7 could incorrectly suggest that the sample complexity would be of order $\frac{T_\epsilon^*(\boldsymbol{\mu})}{|\mathcal{A}_\epsilon(\boldsymbol{\mu})|}\ln\left(\frac{1}{\delta}\right)$. As we will see, the factor $|\mathcal{A}_\epsilon(\boldsymbol{\mu})|$ in the denominator is not correct: this bound is tight only for values of $\boldsymbol{\mu}$ with a unique ϵ-best arm. We now provide a tighter lower bound, which comes with two shortcomings. First, it is asymptotic: it relies on the low-level form of change of distribution presented in Lemma 3. Second, it requires an additional assumption on the sampling rule, that needs to be *converging*.

DEFINITION 8. *A sampling rule A_t^δ is said to be* converging *if it does not depend on δ (that is, $A_t^\delta = A_t$) and for every $\boldsymbol{\mu}$, there exists $\overline{w}(\boldsymbol{\mu}) \in \Sigma_K$ such that, $\mathbb{P}_{\boldsymbol{\mu}}$- almost surely,*

$$\forall a = 1,\dots,K,\quad \frac{N_a(t)}{t} \underset{t\to\infty}{\longrightarrow} \overline{w}_a(\boldsymbol{\mu})\ .$$

THEOREM 9. *For any ϵ-PAC family of converging strategies, one has*

$$\liminf_{\delta\to 0}\frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\ln(1/\delta)} \geq T_\epsilon^*(\boldsymbol{\mu})\ .$$

PROOF. To prove Theorem 9, it is sufficient to prove that for every $\beta > 0$,

$$\lim_{\delta\to 0}\mathbb{P}\left(\tau_\delta \leq (1-\beta)T^*(\boldsymbol{\mu})\ln\left(\frac{1}{\delta}\right)\right) = 0\ . \tag{4.3}$$

Indeed, the conclusion follows by Markov inequality and letting β go to zero. We now fix $\beta > 0$ and introduce

$$n_\delta := \left\lfloor (1-\beta) T^*_\epsilon(\boldsymbol{\mu}) \ln\left(\frac{1}{\delta}\right) \right\rfloor .$$

Defining $\mathcal{C}_\delta = (\tau_\delta \leq n_\delta)$, we establish that $\lim_{\delta\to 0} \mathbb{P}_{\boldsymbol{\mu}}(C_\delta) = 0$.

First, using that the strategy is δ-correct, one can write

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\mu}}(C_\delta) &= \mathbb{P}_{\boldsymbol{\mu}}\left(C_\delta, \hat{\imath}_{\tau_\delta} \in \mathcal{A}_\epsilon(\boldsymbol{\mu})\right) + \mathbb{P}\left(C_\delta, \hat{\imath}_{\tau_\delta} \notin \mathcal{A}_\epsilon(\boldsymbol{\mu})\right) \\ &\leq \sum_{a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})} \mathbb{P}_{\boldsymbol{\mu}}(C_\delta, \hat{\imath}_{\tau_\delta} = a) + \delta \, . \end{aligned}$$

To conclude, it is sufficient to prove that $\forall a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})$, $\lim_{\delta\to 0} \mathbb{P}_{\boldsymbol{\mu}}(C_\delta, \hat{\imath}_{\tau_\delta} = a) = 0$.

Fix $a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})$. Just like in the proof of Theorem 7, we use that by definition of $T^*_\epsilon(\boldsymbol{\mu})$, there exists $b \neq a$ and an alternative model $\boldsymbol{\lambda}$ in which only arm a and b are changed and such that $\lambda_a < \lambda_b - \epsilon$ that satisfies

$$T^*_\epsilon(\boldsymbol{\mu})^{-1} \geq \overline{w}_a(\boldsymbol{\mu}) d(\mu_a, \lambda_a) + \overline{w}_b(\boldsymbol{\mu}) d(\mu_a, \lambda_b) \, , \tag{4.4}$$

where we recall that the weights vector $\overline{w}(\boldsymbol{\mu})$ contains the limit fraction of samples allocated to each arm by the converging sampling rule (A_t): for all i, it holds that $\mathbb{P}_{\boldsymbol{\mu}}\left(\lim_{t\to\infty} N_i(t)/t = \overline{w}_i(\boldsymbol{\mu})\right) = 1$.

As arm a is not ϵ-optimal in the bandit instance $\boldsymbol{\lambda}$, one has $\mathbb{P}_\lambda(C_\delta, \hat{\imath}_{\tau_\delta} = a) \leq \mathbb{P}_\lambda(\hat{\imath}_{\tau_\delta} = a) \leq \delta$. Moreover, the event $C_\delta \cap (\hat{\imath}_{\tau_\delta} = a)$ belongs to $\mathcal{F}_{n_\delta}$. Hence, using the low-level Inequality (3.5) in Lemma 3, for all $x \in \mathbb{R}$,

$$\delta \geq e^{-x} \left[\mathbb{P}_{\boldsymbol{\mu}}\left(C_\delta, \hat{\imath}_{\tau_\delta} = a\right) - \mathbb{P}_{\boldsymbol{\mu}}\left(L_{n_\delta}(\boldsymbol{\mu}, \boldsymbol{\lambda}) \geq x\right)\right]$$

which can be rewritten as

$$\mathbb{P}_{\boldsymbol{\mu}}\left(C_\delta, \hat{\imath}_{\tau_\delta} = a\right) \leq \delta e^x + \mathbb{P}_{\boldsymbol{\mu}}\left(L_{n_\delta}(\boldsymbol{\mu}, \boldsymbol{\lambda}) \geq x\right) \, .$$

The choice $x = (1-\beta/2)\ln(1/\delta)$ yields

$$\mathbb{P}_\mu\left(C_\delta, \hat{\imath}_{\tau_\delta} = a\right) \leq \delta^{\frac{\beta}{2}} + \mathbb{P}_{\boldsymbol{\mu}}\left(\frac{L_{n_\delta}(\boldsymbol{\mu}, \boldsymbol{\lambda})}{n_\delta} \geq \frac{1-\beta/2}{1-\beta} T^*_\epsilon(\boldsymbol{\mu})^{-1}\right) . \tag{4.5}$$

Because of the converging assumption, it holds $\mathbb{P}_{\boldsymbol{\mu}}$-a.s. (independently of δ) that

$$\frac{L_t(\boldsymbol{\mu}, \boldsymbol{\lambda}_a)}{t} \underset{t\to\infty}{\longrightarrow} \overline{w}_a(\boldsymbol{\mu}) d(\mu_a, \lambda_a) + \overline{w}_b(\boldsymbol{\mu}) d(\mu_b, \lambda_b) \leq T^*_\epsilon(\boldsymbol{\mu})^{-1} < \frac{1-\beta/2}{1-\beta} T^*_\epsilon(\boldsymbol{\mu})^{-1} \, ,$$

where the first inequality holds by definition of $\boldsymbol{\lambda}$, which satisfies (4.4). As $n_\delta \to \infty$ when $\delta \to 0$, it follows that

$$\lim_{\delta\to 0} \mathbb{P}_{\boldsymbol{\mu}}\left(\frac{L_{n_\delta}(\boldsymbol{\mu}, \boldsymbol{\lambda})}{n_\delta} \geq \frac{1-\beta/2}{1-\beta} T^*_\epsilon(\boldsymbol{\mu})^{-1}\right) = 0$$

and $\lim_{\delta\to 0} \mathbb{P}_\mu\left(C_\delta, \hat{\imath}_{\tau_\delta} = a\right) = 0$, which concludes the proof.

□

4.4. *An improved lower bound for two arms.* ϵ-best arm identification in a two-armed bandit model $\boldsymbol{\mu} = (\mu_1, \mu_2)$ coincides with an active testing procedure to select one of the hypotheses

$$\mathcal{H}_1 : (\mu_1 > \mu_2 - \epsilon) \quad \text{or} \quad \mathcal{H}_2 : (\mu_2 > \mu_1 - \epsilon)$$

and an (ϵ, δ)-PAC strategy should satisfy $\mathbb{P}_{\boldsymbol{\mu}}(\hat{\imath}_{\tau_\delta} = 2) \leq \delta$ for all $\boldsymbol{\mu}$ such that $\mu_1 > \mu_2 + \epsilon$ and $\mathbb{P}_{\boldsymbol{\mu}}(\hat{\imath}_{\tau_\delta} = 1) \leq \delta$ for all $\boldsymbol{\mu}$ such that $\mu_2 > \mu_1 + \epsilon$. In that case, a first observation is that the complexity term $T^*_\epsilon(\boldsymbol{\mu})$ can be made slightly more explicit. The proof of Proposition 10 can be found in Appendix B.2.

PROPOSITION 10. *If $\mu_a \geq \mu_b - \epsilon$, define $\mu^*_\epsilon(\mu_a, \mu_b)$ as the unique solution in $\lambda \in (\mu^-, \mu^+ - \epsilon)$ to*

$$d(\mu_a, \lambda) = d(\mu_b, \lambda + \epsilon) \, .$$

Then

$$T^*_\epsilon(\boldsymbol{\mu})^{-1} := \begin{cases} d(\mu_1, \mu^*_\epsilon(\mu_1, \mu_2)) & \text{if} \quad \mu_1 > \mu_2 + \epsilon \, , \\ \max[d(\mu_1, \mu^*_\epsilon(\mu_1, \mu_2)), d(\mu_2, \mu^*_\epsilon(\mu_2, \mu_1))] & \text{if} \quad \mu_1 - \mu_2 \in (-\epsilon, \epsilon) \, , \\ d(\mu_2, \mu^*_\epsilon(\mu_2, \mu_1)) & \text{if} \quad \mu_2 > \mu_1 + \epsilon \, . \end{cases}$$

Note that the quantity $\mu^*_\epsilon(\mu_a, \mu_b)$ defined in Proposition 10 is indeed well-defined when $\mu_a \geq \mu_b - \epsilon$, as the the mapping $g : \mu \mapsto d(\mu_a, \mu) - d(\mu_b, \mu + \epsilon)$ is decreasing on $[\mu_b - \epsilon, \mu_a]$ and satisfies $g(\mu_b - \epsilon) = d(\mu_a, \mu_b - \epsilon) > 0$ and $g(\mu_a) = -d(\mu_b, \mu_a + \epsilon) < 0$.

For two armed bandits, we provide below an asymptotic sample complexity lower bound that relax the converging assumption required for Theorem 9. Theorem 11 given below still requires the sampling rule A^δ_t to be independent from δ (we write A_t instead of A^δ_t), but doesn't need the fraction of selection of each arm to converge to a certain proportion under that sampling rule.

THEOREM 11. *Every ϵ-PAC family of* anytime *strategies (i.e., for which the sampling rule (A_t) is independent from δ) satisfies, for all $\boldsymbol{\mu}$,*

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\ln(1/\delta)} \geq T^*_\epsilon(\boldsymbol{\mu}) \, .$$

PROOF. Assume to fix the ideas that $\mu_1 \geq \mu_2$.

Case 1: $\mu_1 \geq \mu_2 + \epsilon$. To simplify the notation we use the notation μ_ϵ^* to denote the quantity $\mu_\epsilon^*(\mu_1, \mu_2)$ defined in Proposition 10. Considering the alternative bandit model $\boldsymbol{\lambda}$ such that $\lambda_1 = \mu_\epsilon^*$ and $\lambda_2 = \mu_\epsilon^* + \epsilon$, one has $\lambda_1 - \lambda_2 = -\epsilon$, hence $\mathbb{P}_{\boldsymbol{\lambda}}(\hat{\imath}_{\tau_\delta} = 1) \leq \delta$, while $\mathbb{P}_{\boldsymbol{\mu}}(\hat{\imath}_{\tau_\delta} = 1) \geq 1 - \delta$. Inequality (3.5) in Lemma 3 yields

$$\begin{aligned} \mathbb{E}[N_1(\tau_\delta)]d(\mu_1, \lambda_1) + \mathbb{E}[N_2(\tau_\delta)]d(\mu_2, \lambda_2) &\geq \mathrm{kl}(\delta, 1-\delta) \\ (\mathbb{E}[N_1(\tau_\delta)] + \mathbb{E}[N_2(\tau_\delta)])d(\mu_1, \mu_\epsilon^*) &\geq \mathrm{kl}(\delta, 1-\delta) \\ \mathbb{E}[\tau_\delta] &\geq \frac{1}{d(\mu_1, \mu_\epsilon^*)} \mathrm{kl}(\delta, 1-\delta) \ , \end{aligned}$$

and the conclusion follows by letting δ go to zero and using $\mathrm{kl}(\delta, 1-\delta) \sim \ln(1/\delta)$.

Case 2: $0 < \mu_1 - \mu_2 < \epsilon$. As in the proof of Theorem 11, we fix $\beta \in (0,1)$ and prove that

$$\lim_{\delta \to 0} \mathbb{P}_{\boldsymbol{\mu}}\left(\tau_\delta \leq \frac{1-\beta}{d_\epsilon^*(\mu_1, \mu_2)} \ln\left(\frac{1}{\delta}\right)\right) = 0 \ . \tag{4.6}$$

Introducing the notation $n_\delta = \left\lfloor \frac{1-\beta}{d_\epsilon^*(\mu_1,\mu_2)} \ln\left(\frac{1}{\delta}\right) \right\rfloor$ and $C_\delta = (\tau_\delta \leq n_\delta)$, we prove that $\lim_{\delta \to 0} \mathbb{P}_{\boldsymbol{\mu}}(C_\delta) = 0$. To prove this, we show that $\mathbb{P}_{\boldsymbol{\mu}}(\mathcal{C}_\delta, \hat{\imath}_{\tau_\delta} = 1)$ and $\mathbb{P}_{\boldsymbol{\mu}}(\mathcal{C}_\delta, \hat{\imath}_{\tau_\delta} = 2)$ tend to zero when δ goes to zero, by invoking Inequality 3.4 in Lemma 3 with a different alternative model $\boldsymbol{\lambda}$ in each case.

To control $\mathbb{P}_{\boldsymbol{\mu}}(\mathcal{C}_\delta, \hat{\imath}_{\tau_\delta} = 1)$ we consider the alternative model $\boldsymbol{\lambda}$ in which $\lambda_1 = \mu_\epsilon^*(\mu_1, \mu_2)$ and $\lambda_2 = \mu_\epsilon^*(\mu_1, \mu_2) + \epsilon$, under which $\mathbb{P}_{\boldsymbol{\lambda}}(\hat{\imath}_{\tau_\delta} = 1) \leq \delta$. We let $L_t(\boldsymbol{\mu}, \boldsymbol{\lambda})$ be the log-likelihood ratio of the observations under the two models:

$$L_t(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \sum_{s=1}^{N_1(t)} \ln \frac{f_{\mu_1}(Y_{1,s})}{f_{\lambda_1}(Y_{1,s})} + \sum_{s=1}^{N_2(t)} \ln \frac{f_{\mu_2}(Y_{2,s})}{f_{\lambda_2}(Y_{2,s})}.$$

The choice of $\boldsymbol{\lambda}$ permits to write

$$\mathbb{E}_{\boldsymbol{\mu}}\left[\ln \frac{f_{\mu_1}(Y_{1,s})}{f_{\lambda_1}(Y_{1,s}))}\right] = \mathbb{E}_{\boldsymbol{\mu}}\left[\ln \frac{f_{\mu_2}(Y_{2,s})}{f_{\lambda_2}(Y_{2,s}))}\right] = d(\mu_1, \mu_\epsilon^*(\mu_1, \mu_2)) \ ,$$

thus $M_t = L_t(\boldsymbol{\mu}, \boldsymbol{\lambda}) - t d(\mu_1, \mu_\epsilon^*(\mu_1, \mu_2))$ is a martingale whose increments have bounded variance, and that does not depend on δ (as the sampling rule A_t is independent from δ). By the law of large number for martingales, it holds that

$$\frac{L_t(\boldsymbol{\mu}, \boldsymbol{\lambda})}{t} \underset{t \to \infty}{\overset{\mathbb{P}_{\boldsymbol{\mu}}-a.s.}{\longrightarrow}} d(\mu_1, \mu_\epsilon^*(\mu_1, \mu_2)). \tag{4.7}$$

Proceeding as in the previous proofs, it follows from Inequality (3.4) with $x = (1 - \beta/2)\ln(1/\delta)$ that

$$\mathbb{P}_\mu(C_\delta, \hat{\imath}_{\tau_\delta} = 1) \leq \delta^{\frac{\beta}{2}} + \mathbb{P}_\mu\left(\frac{L_{n_\delta}(\mu, \epsilon)}{n_\delta} \geq \frac{1 - \beta/2}{1-\beta} d_\epsilon^*(\mu_1, \mu_2)\right).$$

The second term in the RHS tends to zero when δ goes to zero using (4.7) and the fact that $n_\delta \to \infty$ and $\frac{1-\beta/2}{1-\beta} d^*_\epsilon(\mu_1,\mu_2) > d(\mu_1,\mu^*_\epsilon(\mu_1,\mu_2))$. Hence, one obtains $\lim_{\delta\to 0} \mathbb{P}_\mu\left(C_\delta, \hat{\imath}_{\tau_\delta} = 1\right) = 0$.

To control $\mathbb{P}_{\boldsymbol{\mu}}\left(\mathcal{C}_\delta, \hat{\imath}_{\tau_\delta} = 2\right)$, we proceed similarly and consider the alternative model $\boldsymbol{\lambda}'$ in which $\lambda'_1 = \mu^*_\epsilon(\mu_2,\mu_1)$ and $\lambda'_2 = \mu^*_\epsilon(\mu_2,\mu_1) - \epsilon$, under which $\mathbb{P}_{\boldsymbol{\lambda}'}(\hat{\imath}_{\tau_\delta} = 2) \le \delta$. For this particular choice, the log-likelihood ratio satisfies

$$\lim_{t\to\infty} \frac{L_t(\boldsymbol{\mu},\boldsymbol{\lambda}')}{t} \underset{t\to\infty}{\overset{\mathbb{P}_{\boldsymbol{\mu}}-a.s.}{\longrightarrow}} d(\mu_2,\mu^*_\epsilon(\mu_2,\mu_1)) \,. \tag{4.8}$$

Applying Inequality 3.4 as above yields

$$\mathbb{P}_\mu\left(C_\delta, \hat{\imath}_{\tau_\delta} = 2\right) \le \delta^{\frac{\beta}{2}} + \mathbb{P}_\mu\left(\frac{L_{n_\delta}(\mu,\epsilon)}{n_\delta} \ge \frac{1-\beta/2}{1-\beta} d^*_\epsilon(\mu_1,\mu_2)\right)$$

and noting that $\frac{1-\beta/2}{1-\beta} d^*_\epsilon(\mu_1,\mu_2) > d(\mu_2,\mu^*_\epsilon(\mu_2,\mu_1))$ one can prove using (4.8) that the second term in the RHS goes to zero. Hence $\lim_{\delta\to 0} \mathbb{P}_\mu\left(C_\delta, \hat{\imath}_{\tau_\delta} = 2\right) = 0$.

□

5. Asymptotically Optimal Strategies for ϵ-Best Arm Identification. The strategy described in this section hinges on the same ideas than the Track-and-Stop strategy proposed by [8] for the case $\epsilon = 0$. The sampling rule builds on the knowledge of "optimal weights" under which the arms should be sampled, whereas the stopping is a parallel GLRT test with an appropriate choice of threshold. We prove that the resulting strategy, called ϵ-Track-and-Stop is asymptotically optimal for some instances of ϵ-best arm identification.

5.1. *Optimal Weights and their Computation.* From the proof of our lower bounds, it appears that a converging strategy matching the lower bound should have its empirical proportions of draws converge to a vector that belongs to the set

$$\mathcal{W}^*_\epsilon(\boldsymbol{\mu}) = \operatorname*{argmax}_{w\in\Sigma_K} \max_{a\in\mathcal{A}_\epsilon} \min_{b\neq a} \inf_{(\lambda_a,\lambda_b):\lambda_a\le\lambda_b-\epsilon} \left[w_a d(\mu_a,\lambda_a) + w_b d(\mu_b,\lambda_b)\right] \,.$$

To ease the notation, we assume that μ_1 is an optimal arm, i.e. $\mu_1 \ge \mu_b$ for all $b \in \{1,\dots,K\}$. We define

$$T^{*,a}_\epsilon(\boldsymbol{\mu})^{-1} := \sup_{w\in\Sigma_K} \min_{b\neq a} \inf_{(\lambda_a,\lambda_b):\lambda_a\le\lambda_b-\epsilon} \left[w_a d(\mu_a,\lambda_a) + w_b d(\mu_b,\lambda_b)\right] \,.$$

Observe that for a such that there exists $b \neq a$ with $\mu_a = \mu_b - \epsilon$, then $T^{*,a}_\epsilon(\boldsymbol{\mu}) = +\infty$. For a such that $\mu_a > \mu_b - \epsilon$ for all $b \neq a$, we argue below that the argmax in w is unique and define

$$w^{*,a}_\epsilon(\boldsymbol{\mu}) := \operatorname*{argmax}_{w\in\Sigma_K} \min_{b\neq a} \inf_{(\lambda_a,\lambda_b):\lambda_a\le\lambda_b-\epsilon} \left[w_a d(\mu_a,\lambda_a) + w_b d(\mu_b,\lambda_b)\right] \,. \tag{5.1}$$

With this notation, for $\epsilon > 0$, one has

$$\mathcal{W}_\epsilon^\star(\boldsymbol{\mu}) = \left\{ w_\epsilon^{\star,a}(\boldsymbol{\mu}) \text{ for } a \in \operatorname*{argmin}_{i \in \mathcal{A}_\epsilon(\boldsymbol{\mu}) : \mu_i > \max_{j \neq i} \mu_j - \epsilon} T_\epsilon^{\star,i}(\boldsymbol{\mu}) \right\} . \tag{5.2}$$

For $\epsilon > 0$, this set if always non-empty as arm 1 satisfy $\mu_1 > \min_{j \neq 1} \mu_i - \epsilon$. For $\epsilon = 0$, in the presence of a unique optimal arm, $\mathcal{W}_\epsilon(\boldsymbol{\mu}) = \{w_0^{\star,1}(\boldsymbol{\mu})\}$, but in the presence of multiple arms with mean μ_1, the optimal weights are not well defined. In this degenerate case, we define $\mathcal{W}_\epsilon(\boldsymbol{\mu}) = \{|\mathcal{A}_0(\boldsymbol{\mu})|^{-1} \mathbb{1}_{\mathcal{A}_0(\boldsymbol{\mu})}\}$.

We now fix $a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})$ such that $\mu_a > \max_{b \neq a} \mu_b - \epsilon$ and explain why $w_\epsilon^{\star,a}(\boldsymbol{\mu})$ is well-defined and how to compute this probability vector. Introducing the interval

$$\mathcal{I}_{a,b}^\epsilon = \left[\mu^- \vee (\mu_b - \epsilon), \mu_a \wedge (\mu^+ - \epsilon) \right] ,$$

it follows from Proposition 18 (stated in Appendix B) that

$$\begin{aligned} w_\epsilon^{\star,a}(\boldsymbol{\mu}) &= \operatorname*{argmax}_{w \in \Sigma_K} \min_{b \neq a} \inf_{\lambda \in \mathcal{I}_{a,b}^\epsilon} \left[w_a d(\mu_a, \lambda) + w_b d(\mu_b, \lambda + \epsilon) \right] \\ &= \operatorname*{argmax}_{w \in \Sigma_K} \min_{b \neq a} w_a g_b\left(\frac{w_b}{w_a} \right) \end{aligned}$$

where for $b \neq a$ the application g_b (also depending on a, $\boldsymbol{\mu}$ and ϵ) is defined by

$$g_b(x) = \inf_{\lambda \in \mathcal{I}_{a,b}^\epsilon} \left[d(\mu_a, \lambda) + x d(\mu_b, \lambda + \epsilon) \right] . \tag{5.3}$$

In the particular case in which $\mu_a > \mu^+ - \epsilon$ and $\mu_b = \mu^+$, the interval $\mathcal{I}_{a,b}^\epsilon$ is reduced to the point $\mu^+ - \epsilon$ and the mapping $g_b(x) = d(\mu_a, \mu^+ - \epsilon)$ is constant. Besides this degenerate case, $\mathcal{I}_{a,b}^\epsilon$ is a non-degenerate compact interval, and one can leverage this property to show that the mapping $x \mapsto g_b(x)$ is increasing. Moreover, it satisfies $g_b(0) = d(\mu_a, \mu_a \wedge (\mu^+ - \epsilon))$ and $\lim_{x \to \infty} g_b(x) = d(\mu_a, (\mu_b - \epsilon) \vee \mu^-)$. Therefore, one can define the inverse mapping

$$x_b(y) = g_b^{-1}(y) \text{ for } y \in \Big[d(\mu_a, \mu_a \wedge (\mu^+ - \epsilon)), d(\mu_a, (\mu_b - \epsilon) \vee \mu^-) \Big).$$

We furthermore define the constant mapping $x_a(y) = 1$ for all $y \in \Big[d(\mu_a, \mu_a \wedge (\mu^+ - \epsilon)), d(\mu_a, (\max_{b \neq a} \mu_b - \epsilon) \vee \mu^-) \Big)$.

To state Proposition 12 below, which provides a more explicit expression of the weights $w_\epsilon^{\star,a}(\boldsymbol{\mu})$, we also introduce the notation $\lambda_b(x)$ for the minimizer in (5.3):

$$g_b(x) = d\big(\mu_a, \lambda_b(x)\big) + x d\big(\mu_b, \lambda_b(x) + \epsilon\big)$$

The proof of Proposition 12 is given in Appendix C. It hinges on the fact that at the optimum all the values of $g_b(w_b/w_a)$ are identical, which allows to re-parameterize the optimization problem by their common value, y.

PROPOSITION 12. *Fix a bandit instance $\boldsymbol{\mu}$ and $a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})$ and let*

$$\mathcal{I}_a = \Big[d(\mu_a, \mu_a \wedge (\mu^+ - \epsilon)), d(\mu_a, (\max_{b \neq a} \mu_b - \epsilon)) \vee \mu^- \Big) .$$

1. *For degenerate instances $(\boldsymbol{\mu}, \epsilon)$ such that $\mu_a > \mu^+ - \epsilon$ and there exists $b \neq a$ such that $\mu_b = \mu^+$,*
$$(w_\epsilon^{*,a}(\boldsymbol{\mu}))_b = \mathbb{1}_{(b=a)} .$$
2. *For other instances $(\boldsymbol{\mu}, \epsilon)$, let y^* be the unique solution on $\mathcal{I}_a$ to the equation*

$$\sum_{b \neq a} \frac{d(\mu_a, \lambda_b(x_b(y)))}{d(\mu_b, \lambda_b(x_b(y)) + \epsilon)} = 1 . \tag{5.4}$$

Then the argmax in (5.1) is unique and is defined by $\big(w_\epsilon^{*,a}(\boldsymbol{\mu})\big)_b = \frac{x_b(y^*)}{\sum_{b=1}^k x_b(y^*)}$.

In practice, the optimal weights can be computed by solving (5.4) using binary search. Each computation of the left-hand side requires to compute $\lambda_b(x_b(y))$. For each x, $\lambda_b(x)$ is the minimizer of a twice differentiable convex function and can be approximated numerically using for example Newton's method. In the Gaussian case, we may also use the closed form $\lambda_b(x) = \frac{\mu_a + x(\mu_b - \epsilon)}{1+x}$. Then the value of the inverse function $x_b(y)$ can be obtained by solving $g_b(x) = y$ using binary search.

5.2. *The ϵ-Tracking rule.* The Tracking sampling rule was originally proposed by [8] for best arm identification ($\epsilon = 0$), for which the set of optimal weights $\mathcal{W}_0(\boldsymbol{\mu})$ is reduced to a single vector, $w^*(\boldsymbol{\mu})$. Building on our ability to compute those optimal weights, the Tracking rule is a mechanism that forces the empirical proportions of arm selections to converge to this oracle allocation, by relying on plug-in estimates $w^*(\hat{\boldsymbol{\mu}}(t))$ and ensuring enough exploration.

We present below an extension of this rule, called ϵ-Tracking, that can be used in any ϵ-best arm identification problem. The difference with Tracking is that ties need to be broken when $\mathcal{W}_*(\hat{\boldsymbol{\mu}}(t))$ contains multiple vectors. We do by relying on a fixed ordering of the arms, introducing the notation

$$w_\epsilon^*(\boldsymbol{\mu}) = w_\epsilon^{*,a^*}(\boldsymbol{\mu}) \quad \text{where} \quad a^* = \min \Big\{ a \in \{1, \dots, K\} : w_\epsilon^{*,a}(\boldsymbol{\mu}) \in \mathcal{W}_\epsilon(\boldsymbol{\mu}) \Big\}.$$

At round $t+1$, ϵ-Tracking first computes the set $U_t = \{a : N_a(t) < \sqrt{t} - K/2\}$ of arms that are currently under-sampled. Then it selects

$$A_{t+1} \in \begin{cases} \operatorname{argmin}_{a \in U_t} N_a(t) & \text{if } U_t \neq \varnothing \quad (\textit{forced exploration}), \text{or otherwise} \\ \operatorname{argmax}_{1 \le a \le K} \big[t \times (w_\epsilon^*)_a(\hat{\boldsymbol{\mu}}(t)) - N_a(t) \big] & (\textit{tracking the plug-in estimate}). \end{cases}$$

Observe that this sampling rule does not depend on δ. Other Tracking rules exist, for example tracking the cumulative proportions (C-Tracking), which forces $N_a(t)$ to be close to $\sum_{s=1}^{t} w_a^*(\hat{\boldsymbol{\mu}}_s(t))$. Their relative merits is discussed in [5].

DEFINITION 13. *An instance $(\boldsymbol{\mu}, \epsilon)$ is regular if the set $\mathcal{W}_\epsilon^*(\boldsymbol{\mu})$ is of cardinality one. For a regular instance, $\mathcal{W}_\epsilon^*(\boldsymbol{\mu}) = \{w_\epsilon^*(\boldsymbol{\mu})\}$.*

Using the same arguments as those used by [8] for the Tracking rule, one can show that for regular instances ϵ-Tracking is a converging strategy (in the sense of Definition 8), which converges to the optimal proportions.

LEMMA 14. *Let $(\boldsymbol{\mu}, \epsilon)$ be a regular instance of almost optimal best arm identification. Then, under the ϵ-Tracking sampling rule,*

$$\mathbb{P}_{\boldsymbol{\mu}}\left(\forall a = 1, \dots, K, \ \lim_{t\to\infty} \frac{N_a(t)}{t} = \left(w_\epsilon^*(\boldsymbol{\mu})\right)_a\right) = 1 \ .$$

5.3. *The Parallel GLRT Stopping Rule for ϵ-Best Arm Identification.* Introducing for all a, b the statistic

$$\hat{Z}_{a,b}^{\epsilon}(t) = \inf_{\boldsymbol{\lambda}:\lambda_a<\lambda_b-\epsilon} \left[N_a(t)d(\hat{\mu}_a(t), \lambda_a) + N_b(t)d(\hat{\mu}_b(t), \lambda_b)\right]$$

and $\hat{\mathcal{A}}_\epsilon(t) = \{a : \hat{\mu}_a(t) \geq \max_j \hat{\mu}_j(t) - \epsilon\}$, the parallel GLRT stopping rule can be written

$$\tau_\delta = \inf\left\{t \in \mathbb{N} : \max_{a\in\hat{\mathcal{A}}_\epsilon(t)} \min_{b\neq a} \hat{Z}_{a,b}^{\epsilon}(t) > \beta(t,\delta)\right\} \tag{5.5}$$

and the associated recommendation rule is

$$\hat{\imath}_{\tau_\delta} \in \operatorname*{argmax}_{a\in\hat{\mathcal{A}}_\epsilon(\tau_\delta)} \min_{b\neq a} \hat{Z}_{a,b}^{\epsilon}(\tau_\delta). \tag{5.6}$$

In the Gaussian case, a more explicit expression of $\hat{Z}_{a,b}^{\epsilon}(t)$ can be given for a and b such that $\hat{\mu}_a(t) \geq \hat{\mu}_b(t) - \epsilon$ (which holds when $a \in \hat{\mathcal{A}}_\epsilon(t)$):

$$\hat{Z}_{a,b}^{\epsilon}(t) = \frac{1}{2\sigma^2}\frac{N_a(t)N_b(t)}{N_a(t)+N_b(t)}\left(|\hat{\mu}_a(t) - \hat{\mu}_b(t)| + \epsilon\right)^2 \ ,$$

Beyond the Gaussian case, when $\hat{\mu}_a(t) \geq \hat{\mu}_b(t) - \epsilon$ one can write

$$\hat{Z}_{a,b}^{\epsilon}(t) \quad = \quad \inf_{\lambda\in]\mu^-,\mu^+-\epsilon[} \left[N_a(t)d(\hat{\mu}_a(t), \lambda) + N_b(t)d(\hat{\mu}_b(t), \lambda+\epsilon)\right] ,$$

and the solution of this minimization problem has to be numerically approximated.

Lemma 2 ensures that the parallel GLRT stopping rule defined in (5.5) based on the threshold

$$\beta(t,\delta) = 3K\ln(1+\ln t) + K\mathcal{T}\left(\frac{\ln(1/\delta)}{K}\right)$$

yields an (ϵ,δ)-PAC strategy when combined with any sampling rule. Using a more refined analysis tailored to particular structure of the parallel GLRT test for ϵ-best arm identification permit to guarantee the (ϵ,δ)-PAC property for a smaller threshold. Note that the threshold $\beta(t,\delta) = \log\left(\frac{2(K-1)t}{\delta}\right)$ proposed by [8] for $\epsilon = 0$ (and that rely on different deviation inequalities) can also be used here, but the one given in Lemma 15 can be smaller for large values of t.

LEMMA 15. *For any sampling rule that parallel GLRT stopping rule with threshold*

$$\beta(t,\delta) = 6\ln(1+\ln t) + 2\mathcal{T}\left(\frac{\ln((K-1)/\delta)}{2}\right) \tag{5.7}$$

satisfies $\mathbb{P}\left(\tau_\delta < \infty, \hat{\imath}_{\tau_\delta} \notin \mathcal{A}_\epsilon\right) \le \delta$.

PROOF.

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\mu}}\left(\mu_{\hat{a}_\tau} < \mu_1 - \epsilon\right) &\le \sum_{b\notin\mathcal{A}_\epsilon} \mathbb{P}_{\boldsymbol{\mu}}(\hat{a}_\tau = b, \tau < \infty)\\
&\le \sum_{b\notin\mathcal{A}_\epsilon} \mathbb{P}_{\boldsymbol{\mu}}\left(\exists t\in\mathbb{N} : \inf_{\mu'_b-\mu'_1<-\epsilon}\left[N_1(t)d(\hat{\mu}_1(t),\mu'_1) + N_b(t)d(\hat{\mu}_b(t),\mu'_b)\right] > \beta(t,\delta)\right)\\
&\le \sum_{b\notin\mathcal{A}_\epsilon} \mathbb{P}_{\boldsymbol{\mu}}\left(\exists t\in\mathbb{N} : N_1(t)d(\hat{\mu}_1(t),\mu_1) + N_b(t)d(\hat{\mu}_b(t),\mu_b) > \beta(t,\delta)\right)\\
&\le \sum_{b\notin\mathcal{A}_\epsilon} \frac{\delta}{K-1} = \delta\ ,
\end{aligned}$$

where the last inequality relies on Theorem 14 in [11].

□

5.4. *The ϵ-Track-and-Stop Strategy.* We define the ϵ-Track-and-Stop strategy with parameter δ as the following strategy for ϵ-Best Arm Identification:

- its sampling rule is the ϵ-Tracking rule
- its stopping rule is the parallel GLRT with threshold $\beta(t,\delta)$ given in (5.7)
- its recommendation rule is the one associated to the parallel GLRT, that is $\hat{\imath}_{\tau_\delta}$ defined in (5.6).

This strategy, that we denote by ϵ-TaS(δ) has the following properties.

THEOREM 16. *Fix $\delta \in (0,1]$. Then ϵ-TaS(δ) is (ϵ,δ)-PAC. Moreover, for every* regular *instance $(\boldsymbol{\mu},\epsilon)$, if τ_δ denotes the stopping rule of ϵ-TaS(δ),*

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\ln(1/\delta)} \le T^*_\epsilon(\boldsymbol{\mu}) \ .$$

This result shows that the family of ϵ-TaS(δ) strategies for δ in $(0,1)$, which is a family of converging strategies by Lemma 14, is ϵ-PAC and is matching the lower bound given in Theorem 9, for regular instances.

Beyond regular instances, the picture is intriguingly more complex, and was recently studied in details by [5], who consider a general active testing framework with overlapping hypotheses. They show that ϵ-Tracking can fail to converge to an element in $\mathcal{W}^*_\epsilon(\boldsymbol{\mu})$ and propose a "sticky" variant fixing this problem. This alternative sampling rule turns out not to be a converging strategy anymore, as under two different runs it may converge to a different element in $\mathcal{W}^*_\epsilon(\boldsymbol{\mu})$. However, the authors manage to extend the lower bound of Theorem 9 to *any* strategy (using a nice game-theoretic argument). Hence the quantity $T^*_\epsilon(\boldsymbol{\mu})$ is indeed the right characteristic number of samples for ϵ-best arm identification.

Sketch of proof. To understand why the parallel GLRT attains the lower bound of Theorem 9 if coupled with a good sampling rule, we first rewrite its stopping rule

$$\tau_\delta = \inf\left\{t \in \mathbb{N}^* : \hat{Z}(t) \ge \beta(t,\delta)\right\} \ ,$$

where we introduce

$$\begin{aligned}\hat{Z}(t) &= \max_{a\in\hat{\mathcal{A}}_\epsilon(t)} \min_{b\neq a} \inf_{\lambda\in(\mu^-,\mu^+-\epsilon)} \left[N_a(t)d(\hat{\mu}_a(t),\lambda) + N_b(t)d(\hat{\mu}_b(t),\lambda+\epsilon)\right] \\ &= t\times\left(\max_{a\in\hat{\mathcal{A}}_\epsilon(t)} \min_{b\neq a} \inf_{\lambda\in(\mu^-,\mu^+-\epsilon)} \left[\frac{N_a(t)}{t}d(\hat{\mu}_a(t),\lambda) + \frac{N_b(t)}{t}d(\hat{\mu}_b(t),\lambda+\epsilon)\right]\right) \ .\end{aligned}$$

By Lemma 14, under a regular instance for which $\mathcal{W}^*_\epsilon(\boldsymbol{\mu}) = \{w^*_\epsilon(\boldsymbol{\mu})\}$ the ϵ-Tracking sampling rule satisfies that, for every arm a, $\hat{\mu}_a(t)$ converges almost surely to μ_a and the sampling proportion $N_a(t)/t$ converges almost surely to the target $(w^*_\epsilon(\boldsymbol{\mu}))_a$. Using the fact that $\boldsymbol{\mu} \mapsto w^*_\epsilon(\boldsymbol{\mu})$ is continuous in any regular instance $\boldsymbol{\mu}$, for large values of t it holds that

$$\begin{aligned}\hat{Z}(t) &\simeq t\times\left(\max_{a\in\mathcal{A}_\epsilon} \min_{b\neq a} \inf_{\lambda\in(\mu^-,\mu^+-\epsilon)} \left[(w^*_\epsilon(\boldsymbol{\mu}))_a d(\mu_a,\lambda) + (w^*_\epsilon(\boldsymbol{\mu}))_b d(\mu_b,\lambda+\epsilon)\right]\right) \\ &= t\times(T^*_\epsilon(\boldsymbol{\mu}))^{-1} \ .\end{aligned}$$

Therefore one obtains

$$\tau_\delta \simeq \inf\left\{t \in \mathbb{N}^* : t \geq T^*_\epsilon(\boldsymbol{\mu})\beta(t,\delta)\right\} \ . \tag{5.8}$$

If the threshold function were chosen to be $\beta(t,\delta) = \ln(1/\delta)$ one would immediately get the upper bound $T^*_\epsilon(\boldsymbol{\mu})\ln(1/\delta)$. However, a slightly larger threshold is needed to ensure the (ϵ,δ)-PAC property of ϵ-TaS(δ). Still, the threshold defined in (5.7) satisfies $\beta(t,\delta) \leq \ln(1/\delta) + \ln(t)$ for large values of t hence Lemma 17 can be applied to show that the right hand side of (5.8) can be upper bounded by

$$T^*_\epsilon(\boldsymbol{\mu})\ln\left(\frac{1}{\delta}\right) + 2T^*_\epsilon(\boldsymbol{\mu})\ln\left(T^*_\epsilon(\boldsymbol{\mu})\ln\left(\frac{1}{\delta}\right)\right) \ .$$

This argument can be made rigorous to prove that on every regular instance $\boldsymbol{\mu}$,

$$\mathbb{P}_{\boldsymbol{\mu}}\left(\lim_{\delta\to 0}\frac{\tau_\delta}{\ln(1/\delta)} \leq T^*_\epsilon(\boldsymbol{\mu})\right) = 1 \ ,$$

in the spirit of the proof of Proposition 13 in [8]. The proof of Theorem 16 is a bit more complex as it requires to control the expectation $\mathbb{E}[\tau_\delta]$. It can be obtained by following similar arguments as the proof of Theorem 14 in [8].

6. Numerical Experiments. We report results of some experiments comparing ϵ-Track-and-Stop (TaS) to some state-of-the-art algorithms for ϵ-best arm identification for Bernoulli distributed arms. These competitors are KL-LUCB and KL-Racing [10] and UGapE [7]. For instances having a unique optimal arm, we also include 0-TaS in our study (designed for finding the unique best arm) in order to assess the gain in sample complexity achieved when studying the ϵ relaxation.

We experiments with two regular instances,

$$\begin{array}{rclcrcl} \boldsymbol{\mu}_1 & = & [0.2 \ \ 0.4 \ \ 0.5 \ \ 0.55 \ \ 0.7] & & \epsilon_1 & = & 0.1 \\ \boldsymbol{\mu}_2 & = & [0.4 \ \ 0.5 \ \ 0.6 \ \ 0.7 \ \ 0.75 \ \ 0.8] & & \epsilon_2 & = & 0.15 \ , \end{array}$$

and one non-regular instance

$$\boldsymbol{\mu}_3 = [0.2 \ \ 0.3 \ \ 0.45 \ \ 0.55 \ \ 0.6 \ \ 0.6] \quad \epsilon_3 = 0.1 \ .$$

In the experiments we set $\delta = 0.1$ and perform $N = 1000$ independent replications in order to estimate the probability of error $\mathbb{P}_{\boldsymbol{\mu}_i}(\boldsymbol{\mu}_{\hat{\imath}_{\tau_\delta}} \notin \mathcal{A}_{\epsilon_i}(\boldsymbol{\mu}_i))$, and the expected number of samples $\mathbb{E}_{\boldsymbol{\mu}_i}[\tau_\delta]$ for each algorithm and problem instance.

Table 1 report the estimated values of $\mathbb{P}_{\boldsymbol{\mu}_i}(\hat{\imath}_{\tau_\delta} = k)$ for each arm k in for different algorithms and the three problem instances. Errors (non-zero values of $\mathbb{P}_{\boldsymbol{\mu}_i}(\hat{\imath}_{\tau_\delta} = k)$ for k that is not ϵ-optimal) are indicated in bold in this table. We see that the empirical error of all algorithms is (much) smaller than the prescribed

$\boldsymbol{\mu}_1$	0.2	0.4	0.5	0.55	0.7
ϵ-TaS	0	0	**0.002**	**0.013**	0.985
KL-LUCB	0	0	0	**0.003**	0.997
UGap-E	0	0	0	**0.002**	0.998
KL-Racing	0	0	**0.001**	**0.002**	0.997
0-TaS	0	**0.001**	**0.001**	**0.001**	0.997

$\boldsymbol{\mu}_2$	0.4	0.5	0.6	0.7	0.75	0.8
ϵ-TaS	0	0	0	0.044	0.217	0.739
KL-LUCB	0	0	0	0.009	0.104	0.887
UGap-E	0	0	0	0.012	0.13	0.858
KL-Racing	0	0	0	0.009	0.138	0.853
0-TaS	0	0	0	0	0.004	0.996

$\boldsymbol{\mu}_3$	0.2	0.3	0.45	0.55	0.6	0.6
ϵ-TaS	0	0	**0.001**	0.05	0.493	0.456
KL-LUCB	0	0	0	0.011	0.49	0.499
UGap-E	0	0	0	0.01	0.49	0.5
KL-Racing	0	0	0	0.006	0.479	0.515

TABLE 1

Empirical distribution of the recommendation $\hat{a}_{\tau_\delta}$ based on $N = 1000$ repetitions.

	$T^*(\boldsymbol{\mu}) \ln(1/\delta)$	ϵ-TaS	KL-LUCB	UGapE	KL-Racing	0-TaS
$\boldsymbol{\mu}_1$, $\epsilon_1 = 0.1$	97	171 (104)	322 (137)	324 (143)	372 (159)	433 (277)
$\boldsymbol{\mu}_2$, $\epsilon_2 = 0.15$	108	162 (83)	345 (135)	344 (141)	402 (146)	2659 (1863)
$\boldsymbol{\mu}_3$, $\epsilon_3 = 0.1$	531	501 (261)	1236 (403)	1199 (414)	1348 (436)	- -

TABLE 2

Estimated values of $\mathbb{E}_{\boldsymbol{\mu}_i}[\tau_\delta]$ based on $N = 1000$ repetitions for different instances and algorithms (standard deviation indicated in parenthesis).

$(\boldsymbol{\mu}_2)_a$	0.4	0.5	0.6	**0.7**	**0.75**	**0.8**
$\left(w^\star_{\epsilon_2}(\boldsymbol{\mu}_2)\right)_a$	0.024	0.036	0.060	0.136	0.275	0.469
$\mathbb{E}_{\boldsymbol{\mu}_2}[N_a(\tau_\delta)/\tau_\delta]$	0.079	0.077	0.099	0.156	0.235	0.353

$(\boldsymbol{\mu}_3)_a$	0.2	0.3	0.45	**0.55**	**0.6**	**0.6**
$\left(w^{\star,5}_{\epsilon_3}(\boldsymbol{\mu}_3)\right)_a$	0.008	0.0133	0.035	0.114	0.436	0.393
$\left(w^{\star,4}_{\epsilon_3}(\boldsymbol{\mu}_3)\right)_a$	0.008	0.0133	0.035	0.114	0.393	0.436
$\mathbb{E}_{\boldsymbol{\mu}_3}[N_a(\tau_\delta)/\tau_\delta]$	0.049	0.049	0.081	0.194	0.317	0.309

TABLE 3

Optimal proportions and empirical proportions achieved by ϵ-TaS (ϵ-optimal arms in bold)

δ for all algorithms. In those experiments, the stopping rule of Track-and-Stop is the parallel GLRT with threshold $\beta(t,\delta) = \ln((1+\ln(t))/\delta)$. This threshold is an approximation of the theoretical threshold given in Lemma 15 that we recommend to use in practice.

All the other algorithms rely on confidence bounds, and we implement them with upper and lower confidence bounds based on the Kullback-Leibler divergence

$$\begin{aligned} u_a(t) &= \max\{q \in [0,1] : N_a(t) d(\hat{\mu}_a(t), q) \leq \beta(t,\delta)\} \quad \text{and} \\ \text{and } \ell_a(t) &= \min\ \{q \in [0,1] : N_a(t) d(\hat{\mu}_a(t), q) \leq \beta(t,\delta)\}\ , \end{aligned}$$

for the same choice of $\beta(t,\delta) = \ln((1+\ln(t))/\delta)$. The stopping rule of KL-LUCB or UGapE consists in waiting for the lower confidence bound of the candidate best arm to overlap by at most ϵ with the upper-confidence bound of all other arms. This confidence-based stopping rule seem to be far to conservative compared to the parallel GLRT stopping rule, as the sample complexity results now reveal.

Sample complexity results are reported in Table 2. It can be seen that ϵ-TaS requires about half the number of samples needed by other methods before stopping, and is sometimes getting very close to the ideal sample complexity $T^*(\boldsymbol{\mu})\ln(1/\delta)$. In the first instance, in which there is a single ϵ-best arm, one can see that ϵ-TaS slightly improves over 0-TaS due to the ϵ-relaxation. This improvement is much more flagrant in the second instance, which contains 3 ϵ-best arms.

For the two instances with 3 ϵ-best arms $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$, we now report in Table 3 the optimal weights in $\mathcal{W}_\epsilon(\boldsymbol{\mu})$. There is one such weight for $\boldsymbol{\mu}_2$ and two for $\boldsymbol{\mu}_3$ in which the fraction allocation to the two best arms are exchanged. We also report in this table the empirical fraction of selection of each arm by ϵ-TaS. For the regular instance $\boldsymbol{\mu}_2$, according to Lemma 14 the empirical proportions converge to the true proportions. However, as can be seen from Table 3 the convergence is slow and did not yet occur at the stopping time τ_δ (all sub-optimal arms are slightly over sampled which may be a consequence of the forced exploration scheme). For the non-regular instance $\boldsymbol{\mu}_3$ in which there are two candidate optimal weights, ϵ-TaS has no convergence guarantee towards one of these weights (and [5] propose a fix). Just like for $\boldsymbol{\mu}_2$, we see that all sub-optimal arms are slightly over-sampled under $\boldsymbol{\mu}_3$. On average over many simulations, the strategy spends about the same time on the two optimal arms, which could indicate that "convergence" occurs towards $w^{*,4}_{\epsilon_3}(\boldsymbol{\mu}_3)$ or $w^{*,5}_{\epsilon_3}(\boldsymbol{\mu}_3)$ half of the time. Even without proper convergence guarantees, ϵ-TaS performs very well on $\boldsymbol{\mu}_3$ as can be noted from Table 2.

7. Conclusion. We proposed a non-asymptotic way to analyze parallel Generalized Likelihood Ratio Tests, which can be used is various adaptive testing settings, from sequential to active testing, even in the presence of overlapping hypotheses. We also presented a way to derive lower bounds providing the exact

asymptotic sample complexity of some testing problems. Remarkably, for problem instances that belong to multiple hypotheses, non-asymptotic lower bounds are still out of reach, maybe because of the use of the low-level change of low formula. An interesting open question is thus to know whether lower bounds can always be derived using only the high-level form, in a way that provides non-asymptotic results. So far, we were not able to achieve this goal, nor to show that it is impossible.

Natural extensions of our approach can be obtained for other sequential and active testing problems. For example, in bandit models, our approach is likely to apply also for thresholding bandits, or for best-arm identification with a structure on the set of arms. Moreover, if a regularity constraint like $\mu_{i+1} - \mu_i \leq L$ could be handled, this would open the way to the study of the sample complexity of optimizing a Lipschitz continuous function. In that context, one can hope to exhibit a optimal density of sampling (corresponding to the optimal weights in the finite case), depending on the function only, which an optimal optimization procedure would need to track. All these developments are left for future investigations.

Acknowledgment. Emilie Kaufmann acknowledges the support of the Agence Nationale de la Recherche (ANR) under grant ANR-16-CE40-0002 (BADASS project) and the European CHIST-ERA project DELTA. Aurélien Garivier acknowledges the support of the Project IDEXLYON of the University of Lyon, in the framework of the Programme Investissements d'Avenir (ANR-16-IDEX-0005).

References.

[1] Audibert, J.-Y., Bubeck, S., and Munos, R. (2010). Best Arm Identification in Multi-armed Bandits. In *Proceedings of the 23rd Conference on Learning Theory*.

[2] Bubeck, S. and Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Fondations and Trends in Machine Learning*, 5(1):1–122.

[3] Cappé, O., Garivier, A., Maillard, O.-A., Munos, R., and Stoltz, G. (2013). Kullback-Leibler upper confidence bounds for optimal sequential allocation. *Annals of Statistics*, 41(3):1516–1541.

[4] Chernoff, H. (1959). Sequential design of Experiments. *The Annals of Mathematical Statistics*, 30(3):755–770.

[5] Degenne, R. and Koolen, W. M. (2019). Pure Exploration with Multiple Correct Answers. *arXiv:1902.03475*.

[6] Even-Dar, E., Mannor, S., and Mansour, Y. (2006). Action Elimination and Stopping Conditions for the Multi-Armed Bandit and Reinforcement Learning Problems. *Journal of Machine Learning Research*, 7:1079–1105.

[7] Gabillon, V., Ghavamzadeh, M., and Lazaric, A. (2012). Best Arm Identification: A Unified Approach to Fixed Budget and Fixed Confidence. In *Advances in Neural Information Processing Systems*.

[8] Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference On Learning Theory*.

[9] Garivier, A., Ménard, P., and Stoltz, G. (2016). Explore first, exploit next: The true shape of regret in bandit problems. *arXiv preprint arXiv:1602.07182*.

[10] Kaufmann, E. and Kalyanakrishnan, S. (2013). Information complexity in bandit subset selection. In *Proceeding of the 26th Conference On Learning Theory.*

[11] Kaufmann, E. and Koolen, W. (2018). Mixture Martingales Revisited with Applications to Sequential Tests and Confidence Intervals. *arXiv:1811.11419*.
[12] Kim, S. and Nelson, B. L. (2001). A fully sequential procedure for indifference-zone selection in simulation. *ACM Trans. Model. Comput. Simul.*, 11(3):251–273.
[13] Lai, T. (1988). Boundary Crossing problems for samples means. *Annals of Probability*, 16(1):375–396.
[14] Lai, T. and Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22.
[15] Mannor, S. and Tsitsiklis, J. (2004). The Sample Complexity of Exploration in the Multi-Armed Bandit Problem. *Journal of Machine Learning Research*, pages 623–648.
[16] Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535.
[17] Robbins, H. and Siegmund, D. (1974). The expected sample size of some tests of power one. *The Annals of Statistics*, 2(3):415–436.
[18] Russo, D. (2016). Simple bayesian algorithms for best arm identification. In *Proceedings of the 29th Conference On Learning Theory*.
[19] Wald, A. (1945). Sequential Tests of Statistical Hypotheses. *Annals of Mathematical Statistics*, 16(2):117–186.
[20] Wilks, S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62.

APPENDIX A: NON-ASYMPTOTIC SAMPLE COMPLEXITY ANALYSIS FOR A GAUSSIAN TEST

We provide a general sample complexity analysis of the parallel GLRT test using a generic threshold function $\beta(t,\delta)$. Introducing the function

$$f(t,\delta) := \frac{|\mu|+\epsilon}{\sigma}\sqrt{t} - \sqrt{2\beta(t,\delta)}\,,$$

we make the following assumptions:

1. $\beta(t,\delta) \le \ell(\delta) + c\ln(t)$ where $\ell(\delta) \ge \ln(1/\delta)$. This permits to define $T_0(\delta)$ to be any constant such that

$$\forall t \ge T_0(\delta),\quad f(t,\delta) \ge 0\,.$$

2. For all $\delta \in (0,1]$, $\forall t \ge T_0(\delta)$,

$$\frac{\partial}{\partial t}\left(\sqrt{2\beta(t,\delta)}\right) \le \frac{1}{2}\frac{\partial}{\partial t}\sqrt{t\frac{(|\mu|+\epsilon)^2}{\sigma^2}} = \frac{|\mu|+\epsilon}{4\sigma^2}\frac{1}{\sqrt{t}}\,. \tag{A.1}$$

A consequence of Assumption 2 is that $t \mapsto f(t,\delta)$ is non-decreasing, which permits to upper bound the expectation of T_δ as follows, introducing (Z_t) a sequence

of standard normal random variables.

$$\begin{aligned}
\mathbb{E}[T_\delta] &\leq \sum_{t=1}^{\infty} \mathbb{P}\left(t\frac{(|\hat{\mu}_t| + \epsilon)^2}{2\sigma^2} \leq \beta(t,\delta) \right) \\
&\leq \sum_{t=1}^{\infty} \mathbb{P}\left(\sqrt{t}\left(\left| \mu + \frac{\sigma Z_t}{\sqrt{t}} \right| + \epsilon \right) \leq \sqrt{2\sigma^2\beta(t,\delta)} \right) \\
&\leq \sum_{t=1}^{\infty} \mathbb{P}\left(\sqrt{t}\left(|\mu| - |\sigma Z_t/\sqrt{t}| + \epsilon \right) \leq \sqrt{2\sigma^2\beta(t,\delta)} \right) \\
&\leq \sum_{t=1}^{\infty} \mathbb{P}\left(|Z_t| \geq \frac{|\mu| + \epsilon}{\sigma}\sqrt{t} - \sqrt{2\beta(t,\delta)} \right) \\
&\leq \lceil T_0(\delta) \rceil + \sum_{t=\lceil T_0(\delta) \rceil + 1}^{\infty} \mathbb{P}\left(|Z_t| \geq f(t,\delta) \right) dt \\
&\leq \lceil T_0(\delta) \rceil + \int_{T_0(\delta)}^{\infty} \mathbb{P}\left(|Z_t| \geq f(t,\delta) \right) dt \, .
\end{aligned}$$

Using assumption (A.1) one can write

$$\begin{aligned}
f(t,\delta) &= f(T_0(\delta),\delta) + \int_{T_0(\delta)}^{t} f'(s,\delta) ds \\
&\geq 0 + \frac{1}{2}\int_{T_0(\delta)}^{t} \left(\sqrt{t\frac{(|\mu| + \epsilon)^2}{\sigma^2}} \right)' ds \\
&= \int_{T_0(\delta)}^{t} \frac{|\mu| + \epsilon}{4\sigma} \frac{1}{\sqrt{s}} ds \, .
\end{aligned}$$

For $t \in [T_0(\delta), 2T_0(\delta)]$, observe that

$$f(t,\delta) \geq (t - T_0(\delta)) \frac{|\mu| + \epsilon}{4\sqrt{2}\sigma} \frac{1}{\sqrt{T_0(\delta)}} \, ,$$

while for $t > 2T_0(\delta)$,

$$f(t,\delta) \geq (t - T_0(\delta)) \frac{|\mu| + \epsilon}{4\sigma} \frac{1}{\sqrt{t}} \geq \sqrt{t - T_0(\delta)} \frac{|\mu| + \epsilon}{4\sqrt{2}\sigma} \, .$$

Hence

$$\begin{aligned}
\mathbb{E}[T_\delta] &\le \lceil T_0(\delta)\rceil + \int_{T_0(\delta)}^{2T_0(\delta)} 2\overline{\Phi}(f(t,\delta))dt + \int_{2T_0(\delta)}^{\infty} 2\overline{\Phi}(f(t,\delta))dt \\
&\le \lceil T_0(\delta)\rceil + 2\int_{T_0(\delta)}^{\infty} \overline{\Phi}\left((t-T_0(\delta))\frac{|\mu|+\epsilon}{4\sqrt{2}\sigma}\frac{1}{\sqrt{T_0(\delta)}}\right)dt + 2\int_{2T_0(\delta)}^{\infty} \overline{\Phi}\left(\sqrt{t-T_0(\delta)}\frac{|\mu|+\epsilon}{4\sqrt{2}\sigma}\right)dt \\
&\le \lceil T_0(\delta)\rceil + \frac{8\sqrt{2}\sigma}{|\mu|+\epsilon}\sqrt{T_0(\delta)}\int_0^{\infty}\overline{\Phi}(u)\,du + \frac{64\sigma^2}{(|\mu|+\epsilon)^2}\int_{T_0(\delta)\frac{(|\mu|+\epsilon)^2}{8\sigma^2}}^{\infty}\overline{\Phi}\left(\sqrt{u}\right)du \\
&\le \lceil T_0(\delta)\rceil + \frac{8\sqrt{2}\sigma}{|\mu|+\epsilon}\sqrt{T_0(\delta)}\frac{1}{\sqrt{2\pi}} + \frac{64\sigma^2}{(|\mu|+\epsilon)^2}\int_{T_0(\delta)\frac{(|\mu|+\epsilon)^2}{8\sigma^2}}^{\infty} e^{-\frac{u}{2}}du \\
&\le 1 + T_0(\delta) + \frac{8\sqrt{2}\sigma}{|\mu|+\epsilon}\sqrt{T_0(\delta)}\frac{1}{\sqrt{2\pi}} + \frac{64\sigma^2}{(|\mu|+\epsilon)^2}e^{-T_0(\delta)\frac{(|\mu|+\epsilon)^2}{8\sqrt{2}\sigma^2}} .
\end{aligned}$$

If the exploration function β satisfies

$$\beta(t,\delta) \le \ell(\delta) + c\ln(t) ,$$

Lemma 17 below permits to prove that one can pick

$$T_0(\delta) = \frac{2\sigma^2}{(|\mu|+\epsilon)^2}\ell(\delta) + \frac{4c\sigma^2}{(|\mu|+\epsilon)^2}\ln\left(\frac{2\sigma^2}{(|\mu|+\epsilon)^2}\ell(\delta)\right) .$$

In that case,

$$\begin{aligned}
\mathbb{E}[T_\delta] \le \frac{2\sigma^2}{(|\mu|+\epsilon)^2}\Bigg[&\ell(\delta) + 2c\ln\left(\frac{2\sigma^2}{(|\mu|+\epsilon)^2}\ell(\delta)\right) \\
&+ 8\sqrt{\ell(\delta) + 2c\ln\left(\frac{2\sigma^2}{(|\mu|+\epsilon)^2}\ell(\delta)\right) + 32\delta^{1/8}}\Bigg] + 1 .
\end{aligned}$$

Theorem 3.1 follows from observing that the threshold (3.3) for which the parallel GLRT is δ-correct satisfies

$$\beta(t,\delta) \le \mathcal{T}\left(\ln(1/\delta)\right) + 3\frac{1+\ln(t)}{e} ,$$

and thus Assumption 1 is satisfied with $c = 3/e$ and $\ell(\delta) = \mathcal{T}\left(\ln(1/\delta)\right) + 3/e$. It can be checked by calculus that Assumption 2 is also satisfied.

LEMMA 17. *Fix $\alpha \ge 0$ and $\gamma \ge 1+\alpha$. Then for all $t > 0$,*

$$t \ge \gamma + 2\alpha\ln(\gamma) \quad \Rightarrow \quad t \ge \gamma + \alpha\ln(t) .$$

PROOF. Let $f(t) = t - \gamma - \alpha \ln(t)$. Then $f'(t) = 1 - \alpha/t \geq 0$ for $t \geq \alpha$. Hence, for all $t \geq t_0 := \gamma + 2\alpha \ln(\gamma) \geq \alpha$,

$$\begin{aligned} f(t) \geq f(t_0) &= \gamma + 2\alpha \ln(\gamma) - \gamma - \alpha \ln\big(\gamma + 2\alpha \ln(\gamma)\big) \\ &= \alpha \ln(\gamma) - \alpha \ln\left(1 + \frac{2\alpha}{\gamma} \ln(\gamma)\right) \end{aligned}$$

which has the same sign as

$$\gamma - 1 - \frac{2\alpha}{\gamma} \ln(\gamma) \geq \gamma - 1 - \frac{2\alpha\gamma}{e\,\gamma} \geq \gamma - 1 - \alpha \geq 0 \,.$$

□

APPENDIX B: COMPUTING THE CHARACTERISTIC TIME

The following result is useful at several places to provide more explicit expressions of the characteristic time $T_\epsilon^*(\boldsymbol{\mu})$. It follows easily from monotonicity properties of $\lambda \mapsto d(\mu_a, \lambda)$ and $\lambda \mapsto d(\mu_b, \lambda)$.

PROPOSITION 18. *Fix $w_a, w_b \in \mathbb{R}$ and μ_a, μ_b such that $\mu_a \geq \mu_b - \epsilon$.*

$$\begin{aligned} &\inf_{\lambda_a < \lambda_b - \epsilon} \big[w_a d(\mu_a, \lambda_a) + w_b d(\mu_b, \lambda_b) \big] \\ &\quad = \inf_{\lambda \in \big[\mu^- \vee (\mu_b - \epsilon), \mu_a \wedge (\mu^+ - \epsilon)\big]} \big[w_a d(\mu_a, \lambda) + w_b d(\mu_b, \lambda + \epsilon) \big] \\ &\quad = \inf_{\lambda \in (\mu^-, \mu^+ - \epsilon)} \big[w_a d(\mu_a, \lambda) + w_b d(\mu_b, \lambda + \epsilon) \big] \,. \end{aligned}$$

This result follows from the monotonicity properties of the mapping $\lambda \mapsto d(\mu_a, \lambda)$ and $\lambda \mapsto d(\mu_b, \lambda)$ that are illustrated in Figure 1.

First, we show that the minimizer in (λ_a, λ_b) of

$$g(\lambda_a, \lambda_b) = w_a d(\mu_a, \lambda_a) + w_b d(\mu_b, \lambda_b)$$

under the constraints $\lambda_a \leq \lambda_b - \epsilon$ is such that $\lambda_a \in \big[\mu^- \vee (\mu_b - \epsilon), \mu_a \wedge (\mu^+ - \epsilon)\big]$. Indeed, if $\mu^- < \lambda_a < \mu_b - \epsilon$,

$$\inf_{\lambda : \lambda_b > \lambda_a + \epsilon} g(\lambda_a, \lambda) = g(\lambda_a, \mu_b) \geq g(\mu_b - \epsilon, \mu_b) \,.$$

Similarly, if $\lambda_a > \mu_a$ (and $\lambda_a < \mu^+ - \epsilon$),

$$\inf_{\lambda : \lambda > \lambda_a + \epsilon} g(\lambda_a, \lambda_b) = g(\lambda_a, \lambda_a + \epsilon) \geq g(\mu_a, \mu_a + \epsilon) \,.$$

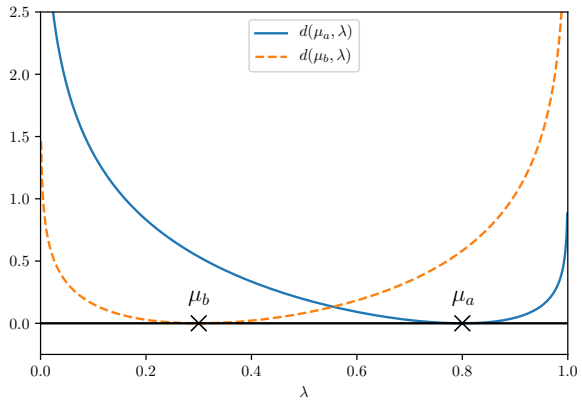

FIGURE 1. *Monotonicity properties of the Kullback-Leibler divergence in the Bernoulli case*

Now for $\lambda_a \in \left[\mu^- \vee (\mu_b - \epsilon), \mu_a \wedge (\mu^+ - \epsilon)\right]$, λ_b satisfying the constraint $\lambda_b \geq \lambda_a + \epsilon$ is larger than μ_b. As $d(\mu_b, \lambda)$ is decreasing for $\lambda \geq \mu_b$, the value λ_b that minimizes $w_b d(\mu_b, \lambda_b)$ is the smallest admissible value, that is $\lambda_b = \lambda_a + \epsilon$. This justifies the first equality in Proposition 18.

The second equality follows from the same monotonicity properties: letting $g(\lambda) = w_a d(\mu_a, \lambda) + w_b d(\mu_b, \lambda + \epsilon)$, one can easily show that the minimum of this function on $(\mu^-, \mu^+ - \epsilon)$ is attained on $\left[\mu^- \vee (\mu_b - \epsilon), \mu_a \wedge (\mu^+ - \epsilon)\right]$. Indeed, for all $\lambda < \mu_b - \epsilon$, $g(\lambda) > g(\mu_b - \epsilon)$ (as the two mappings $\lambda \mapsto d(\mu_a, \lambda)$ and $\lambda \mapsto d(\mu_b, \lambda + \epsilon)$ are decreasing for $\lambda < \mu_b - \epsilon$) and for all $\lambda > \mu_a$, $g(\lambda) > g(\mu_a)$ (the two mappings are increasing for $\lambda > \mu_a$).

B.1. Computing the Characteristic Time for Gaussian Distributions. In the Gaussian case, there is no edge effect as $\mu^- = -\infty$ and $\mu^+ = +\infty$. Therefore one can write

$$
\begin{aligned}
T_\epsilon^*(\boldsymbol{\mu})^{-1} &= \sup_{w \in \Sigma_K} \max_{a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})} \min_{b \neq a} \inf_{\lambda \in \mathbb{R}} \left[w_a d(\mu_a, \lambda) + w_b d(\mu_b, \lambda + \epsilon) \right] \\
&= \sup_{w \in \Sigma_K} \max_{a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})} \min_{b \neq a} \inf_{\lambda \in \mathbb{R}} \left[w_a d(\mu_a, \lambda - \epsilon) + w_b d(\mu_b, \lambda) \right] \\
&= \max_{a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})} \sup_{w \in \Sigma_K} \min_{b \neq a} \inf_{\lambda \in \mathbb{R}} \left[w_a d(\mu_a + \epsilon, \lambda) + w_b d(\mu_b, \lambda) \right] \\
&= \max_{a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})} T_0^*\left(\boldsymbol{\mu}^{a,\epsilon}\right)^{-1} ,
\end{aligned}
$$

where the last but one equality uses that $d(\mu_a, \lambda - \epsilon) = d(\mu_a + \epsilon, \lambda)$, which is a special feature of the divergence function $d(x, y) = (x - y)^2/(2\sigma^2)$. Recall that $\boldsymbol{\mu}^{a,\epsilon}$ is a bandit instance identical to $\boldsymbol{\mu}$ expect for arm a that is set to $\mu_a + \epsilon$.

Now, exploiting further the explicit form of the Gaussian divergence, one can write for any $\boldsymbol{\lambda}$ with a unique optimal arm $a^* = a^*(\boldsymbol{\lambda})$

$$T_0^*(\boldsymbol{\lambda})^{-1} = \sup_{w\in\Sigma_K} \inf_{b\neq a^*} \frac{w_{a^*}w_b}{w_{a^*}+w_b}\frac{\Delta_b(\boldsymbol{\lambda})}{2\sigma^2} \quad \text{with} \quad \Delta_b(\boldsymbol{\lambda}) = \lambda^{a^*(\boldsymbol{\lambda})} - \lambda_b \,.$$

Moreover, if λ does not have a unique optimal arm, $T_\epsilon(\boldsymbol{\lambda})^{-1} = 0$. Letting 1 be an optimal arm in $\boldsymbol{\mu}$, it can be observed that for all a such that $\mu_a > \mu_1 - \epsilon$, the instance $\boldsymbol{\mu}^{a,\epsilon}$ as a unique optimal arm and the vector of gaps $(\Delta_b(\boldsymbol{\mu}^{a,\epsilon}))_{b\neq a}$ is dominated by the vector of gaps $(\Delta_b(\boldsymbol{\mu}^{1,\epsilon}))_{b\neq 1}$, in the sense that there exists a permutation σ such that $\Delta_b(\boldsymbol{\mu}^{a,\epsilon}) \leq \Delta_{\sigma(b)}(\boldsymbol{\mu}^{1,\epsilon})$ (intuitively, adding ϵ to μ_1 creates the overall largest gaps in the bandit instance compared to adding it elsewhere). This yields $T_\epsilon(\boldsymbol{\mu}^{a,\epsilon})^{-1} \leq T_\epsilon(\boldsymbol{\mu}^{1,\epsilon})^{-1}$. For a such that $\mu_a = \mu_1 - \epsilon$, $T_\epsilon(\boldsymbol{\mu}^{a,\epsilon})^{-1} = 0 \leq T_\epsilon(\boldsymbol{\mu}^{1,\epsilon})^{-1}$.

Putting things together, we proved that

$$T_\epsilon^*(\boldsymbol{\mu})^{-1} = \max_{a\in\mathcal{A}_\epsilon(\boldsymbol{\mu})} T_0^*(\boldsymbol{\mu}^{a,\epsilon})^{-1} = T_0^*(\boldsymbol{\mu}^{1,\epsilon})^{-1} \,,$$

which easily yields the equality (4.1) stated in Section 4.2.

B.2. Computing the Characteristic Time for Two-Armed bandits: proof of Proposition 10. The optimization problem defining $T_\epsilon^*(\boldsymbol{\mu})^{-1}$ can rewritten

$$\begin{aligned} T_\epsilon^*(\boldsymbol{\mu})^{-1} &= \sup_{w\in\Sigma_K} \max_{a\in\mathcal{A}_\epsilon(\boldsymbol{\mu})} \min_{b\neq a} \inf_{(\lambda_a,\lambda_b):\lambda_a\leq\lambda_b-\epsilon} \Big[w_a d(\mu_a,\lambda_a) + w_b d(\mu_b,\lambda_b)\Big] \\ &= \sup_{w\in\Sigma_K} \max_{a\in\mathcal{A}_\epsilon(\boldsymbol{\mu})} \min_{b\neq a} \inf_{\lambda\in(\mu^-,\mu^+-\epsilon)} \Big[w_a d(\mu_a,\lambda) + w_b d(\mu_b,\lambda+\epsilon)\Big] \,, \end{aligned} \tag{B.1}$$

where the second equality follows from Proposition 18.

The proof relies on the following key observation: if $\mu_a \geq \mu_b - \epsilon$,

$$\sup_{\alpha\in[0,1]} \inf_{\lambda\in(\mu^-,\mu^+-\epsilon)} \Big[\alpha d(\mu_a,\lambda) + (1-\alpha) d(\mu_b,\lambda+\epsilon)\Big] = d\Big(\mu_a, \mu_\epsilon^*(\mu_a,\mu_b)\Big) \,. \tag{B.2}$$

Using the definition of $T_\epsilon^*(\boldsymbol{\mu})$, Equality (B.1) and the notation $b = 3 - a$ it holds that

$$\begin{aligned} T_\epsilon^*(\boldsymbol{\mu})^{-1} &= \sup_{w\in\Sigma_2} \max_{a\in\mathcal{A}_\epsilon} \inf_{\lambda\in(\mu^-,\mu^+-\epsilon)} \Big[w_a d(\mu_a,\lambda) + w_b d(\mu_b,\lambda+\epsilon)\Big] \\ &= \max_{a\in\mathcal{A}_\epsilon} \sup_{\alpha\in[0,1]} \inf_{\lambda\in(\mu^-\mu^+-\epsilon)} \Big[\alpha d(\mu_a,\lambda) + (1-\alpha) d(\mu_b,\lambda+\epsilon)\Big] \,. \end{aligned}$$

The result follows by using Equation (B.2) for all $a \in \mathcal{A}_\epsilon$ (that is either equal to $\{1\}, \{2\}$ or $\{1,2\}$).

We now prove (B.2) as a double inequality. First,

$$\sup_{\alpha\in[0,1]} \inf_{\lambda\in(\mu^-,\mu^+-\epsilon)} \Big[\alpha d(\mu_a,\lambda) + (1-\alpha) d(\mu_b,\lambda+\epsilon)\Big]$$
$$\leq \inf_{\lambda\in(\mu^-,\mu^+-\epsilon)} \max\Big[d(\mu_a,\lambda), d(\mu_b,\lambda+\epsilon)\Big] = d(\mu_a,\mu_\epsilon^*(\mu_a,\mu_b)) \ .$$

To prove the second inequality, we denote by

$$\lambda_\alpha^*(\mu_a,\mu_b) = \operatorname*{argmin}_{\lambda\in(\mu^-,\mu^+-\epsilon)} \Big[\alpha d(\mu_a,\lambda) + (1-\alpha) d(\mu_b,\lambda+\epsilon)\Big]$$

and prove that there exists $\alpha \in [0,1]$ such that $\lambda_\alpha^*(\mu_a,\mu_b) = \mu_\epsilon^*(\mu_a,\mu_b)$. We first prove that the mapping $\alpha \mapsto \lambda_\alpha^*(\mu_a,\mu_b)$ is continuous. This can be established using that the minimizer of the convex function $\lambda \mapsto \alpha d(\mu_a,\lambda) + (1-\alpha) d(\mu_b,\lambda+\epsilon)$ which is attained on the compact set $[\mu_b-\epsilon,\mu_a]$, is unique together with the continuity of $(\alpha,\lambda) \mapsto \alpha d(\mu_a,\lambda) + (1-\alpha) d(\mu_b,\lambda+\epsilon)$. As $\lambda_0^*(\mu_a,\mu_b) = \mu_b - \epsilon$, $\lambda_1^*(\mu_a,\mu_b) = \mu_a$ and $\mu_\epsilon^*(\mu_a,\mu_b) \in (\mu_b-\epsilon,\mu_a)$, from the intermediate values theorem there exists $\alpha^* \in [0,1]$ such that $\lambda_{\alpha^*}^*(\mu_a,\mu_b) = \mu_\epsilon^*(\mu_a,\mu_b) \in (\mu_b-\epsilon,\mu_a)$. With this value α^* one can write

$$\begin{aligned}
&\sup_{\alpha\in[0,1]} \inf_{\lambda\in(\mu^-,\mu^+-\epsilon)} \Big[\alpha d(\mu_a,\lambda) + (1-\alpha) d(\mu_b,\lambda+\epsilon)\Big] \\
&\qquad \geq \alpha^* d(\mu_a,\lambda_{\alpha^*}(\mu_a,\mu_b)) + (1-\alpha^*) d(\mu_b,\lambda_{\alpha^*}(\mu_a,\mu_b)+\epsilon) \\
&\qquad = \alpha^* d(\mu_a,\mu_\epsilon^*(\mu_a,\mu_b)) + (1-\alpha^*) d(\mu_b,\mu_\epsilon^*(\mu_a,\mu_b)+\epsilon) \\
&\qquad = d(\mu_a,\mu_\epsilon^*(\mu_a,\mu_b)) \ ,
\end{aligned}$$

which together with the first inequality proves (B.2).

APPENDIX C: COMPUTATION OF THE OPTIMAL WEIGHTS

In this section, we prove Proposition 12. Let $a \in \mathcal{A}_\epsilon(\boldsymbol{\mu})$ be an ϵ-optimal such that $\mu_a > \mu_b - \epsilon$ and recall the definition of the interval

$$\mathcal{I}_{a,b}^\epsilon = \Big[\mu^- \vee (\mu_b-\epsilon), \mu_a \wedge (\mu^+-\epsilon)\Big]$$

and of the optimal weights

$$\begin{aligned}
w_\epsilon^{*,a}(\boldsymbol{\mu}) &= \operatorname*{argmax}_{w\in\Sigma_K} \min_{b\neq a} \inf_{\lambda\in\mathcal{I}_{a,b}^\epsilon} \Big[w_a d(\mu_a,\lambda) + w_b d(\mu_b,\lambda+\epsilon)\Big]. \\
&= \operatorname*{argmax}_{w\in\Sigma_K} \min_{b\neq a} w_a g_b\left(\frac{w_b}{w_a}\right) \ , \qquad \text{(C.1)}
\end{aligned}$$

where

$$\begin{aligned}
g_b(x) &= d\big(\mu_a,\lambda_b(x)\big) + x d(\mu_b,\lambda_b(x)+\epsilon\big) \ \text{ and} \\
\lambda_b(x) &= \operatorname*{argmin}_{\lambda\in\mathcal{I}_{a,b}^\epsilon} \Big[d(\mu_a,\lambda) + x d(\mu_b,\lambda+\epsilon)\Big] \ .
\end{aligned}$$

A particular case. We first study the particular case in which $\mu_a \geq \mu^+ - \epsilon$ and there exists $\tilde{b} \neq a$ such that $\mu_{\tilde{b}} = \mu^+$. Among common exponential families, this may happen for Bernoulli distribution with $\mu_a = \mu_{\tilde{b}} = 1$. In that case $\mathcal{I}^\epsilon_{a,\tilde{b}}$ is reduced to the singleton $\{\mu^+ - \epsilon\}$ and

$$\lambda_{\tilde{b}}(x) = d(\mu_a, \mu^+ - \epsilon) + xd(\mu^+, \mu^+) = d(\mu_a, \mu^+ - \epsilon)$$

is a constant mapping.

Observe that, for all $b \neq a$, as $\lambda \mapsto d(\mu_a, \lambda)$ is decreasing on $(\mu^-, \mu^+ - \epsilon)$,

$$\begin{aligned} w_a d(\mu_a, \mu^+ - \epsilon) &\leq \inf_{\lambda \in (\mu^-, \mu^+ - \epsilon)} \left[w_a d(\mu_a, \lambda) + w_b d(\mu_b, \lambda + \epsilon) \right] \\ &= \inf_{\lambda \in \mathcal{I}^\epsilon_{a,b}} \left[w_a d(\mu_a, \lambda) + w_b d(\mu_b, \lambda + \epsilon) \right]. \end{aligned}$$

Therefore, $\min_{b \neq a} w_a g_b(w_a/w_b) = w_a g_{\tilde{b}}(w_a/w_{\tilde{b}}) = w_a d(\mu_a, 1 - \epsilon)$ and

$$w^{*,a}_\epsilon(\boldsymbol{\mu}) = \operatorname*{argmax}_{w \in \Sigma_K} w_a d(\mu_a, \mu^+ - \epsilon) = (\mathbb{1}_{(b=a)})_{b=1,\dots,K}.$$

This proves the first statement of Proposition 12.

General case. In the general case, $\mathcal{I}^\epsilon_{a,b}$ is a non-degenerated compact interval, and one can define the following inverse mapping

$$\begin{aligned} x_b : [d(\mu_a, \mu_a \wedge (\mu^+ - \epsilon)), d(\mu_a, (\mu_b - \epsilon) \vee \mu^-)) &\longrightarrow \mathbb{R}^+ \\ y &\mapsto g_b^{-1}(y), \end{aligned}$$

which is increasing. Let w^* be an optimum in (C.1), and define $x^*_b = w^*_b/w^*_a$ if $b \neq a$ and $x^*_a = 1$. Then

$$(x^*_b)_{b \neq a} \in \operatorname*{argmax}_{\boldsymbol{x} \in (\mathbb{R}^+)^{K-1}} \frac{\min_{b \neq a} g_b(x_b)}{1 + \sum_{b \neq a} x_b}$$

Since all the g_b are increasing, such an optimum equalizes the values of $g_b(x^*_b)$: there exists y such that $\forall b \neq a, \quad g_b(x^*_b) = y$, and y necessarily belongs to $\mathcal{I}_a = [d(\mu_a, \mu_a \wedge (\mu^+ - \epsilon)), d(\mu_a, (\max_{b \neq a} \mu_b - \epsilon) \vee \mu^-))$, the intersection of the image of all functions g_b.

By re-parameterizing the optimization problem by the common value y of $g_b(x_b)$ (which yields $x_b = x_b(y)$ by definition of the inverse function x_b), one obtains $x^*_b = x_b(y^*)$ where

$$y^* \in \operatorname*{argmax}_{y \in \mathcal{I}_a} \frac{y}{1 + \sum_{b \neq a} x_b(y)}.$$

We now justify that this maximum is unique, and provide a more explicit expression for it. The derivative of $G : \mathcal{I}_a \to \mathbb{R}$ defined by $G(y) = \frac{y}{1+\sum_{b\neq a} x_b(y)}$ is

$$G'(y) = \frac{1 + \sum_{b\neq a} x_b(y) - y \sum_{b\neq a} x_b'(y)}{\big(1 + \sum_{b\neq a} x_b(y)\big)^2} \,. \tag{C.2}$$

In order to compute $x_b'(y)$ and simplify the expression of this derivative, we rely on the following properties of the minimizer $\lambda_b(x)$.

PROPOSITION 19. *The mapping $x \mapsto \lambda_b(x)$ is differentiable on $\mathbb{R}^+$ and such that for all $x > 0$, $\lambda_b'(x) < 0$. Moreover,*

$$\lim_{x\to 0} \lambda_b(x) = \mu_a \wedge (\mu^+ - \epsilon) \quad \textit{and} \quad \lim_{x\to\infty} \lambda_b(x) = (\mu_b - \epsilon) \vee \mu^- \,.$$

PROOF. For every x, the function $h_x(\lambda) := d(\mu_a, \lambda) + x d(\mu_b, \lambda + \epsilon)$ is strictly convex: its derivative $h_x'(\lambda)$ is increasing and crosses 0 at $\lambda_b(x)$. The differentiability of λ_b follows from that of d by the implicit function theorem. Since for every $\lambda > \mu_b - \epsilon$, $d_2'(\mu_b, \lambda + \epsilon) > 0$, for every $y > x$

$$\begin{aligned} h_y'(\lambda_b(x)) &= d_2'\big(\mu_a, \lambda_b(x)\big) + y d_2'\big(\mu_b, \lambda_b(x) + \epsilon\big) \\ &= d_2'\big(\mu_a, \lambda_b(x)\big) + x d_2'\big(\mu_b, \lambda_b(x) + \epsilon\big) + (y - x) d_2'\big(\mu_b, \lambda_b(x) + \epsilon\big) \\ &= 0 + (y - x) d_2'\big(\mu_b, \lambda_b(x) + \epsilon\big) > 0 \,. \end{aligned}$$

Thus, for every $\alpha > 0$ small enough,

$$h_y'\big(\lambda_b(x) - \alpha\big) \geq (y - x) d_2'\big(\mu_b, \lambda_b(x) + \epsilon\big) - \alpha \sup_{[\lambda_b(x)-\alpha, \lambda_b(x)]} h_y'' > 0 \,.$$

Since $\lambda_b(y)$ is the zero of the increasing function h_y', this entails that

$$\lambda_b(y) \leq \lambda_b(x) - \frac{(y - x) d_2'(\mu_b, \lambda_b(x) + \epsilon)}{\sup_{[\lambda_b(y), \lambda_b(x)]} h_y''} \,,$$

which shows that $\lambda_b'(x) < 0$. The limit when x goes to 0 follows from the continuity of λ_b, and when $x \to \infty$ it appears that

$$\begin{aligned} \operatorname*{argmin}_{\lambda \in \mathcal{I}_{a,b}^\epsilon} \big[d(\mu_a, \lambda) + x d(\mu_b, \lambda + \epsilon) \big] &= \operatorname*{argmin}_{\lambda \in \mathcal{I}_{a,b}^\epsilon} \big[d(\mu_a, \lambda)/x + d(\mu_b, \lambda + \epsilon) \big] \\ &\to \operatorname*{argmin}_{\lambda \in \mathcal{I}_{a,b}^\epsilon} \big[d(\mu_b, \lambda + \epsilon) \big] = (\mu_b - \epsilon) \vee \mu^- \,. \end{aligned}$$

□

As the minimum of the convex function $\lambda \mapsto d(\mu_a, \lambda) + d(\mu_b, \lambda + \epsilon)$ on $(\mu^-, \mu^+ - \epsilon)$, $\lambda_b(x)$ cancels its derivative, hence

$$d_2'(\mu_a, \lambda_b(x)) + x d_2'(\mu_b, \lambda_b(x) + \epsilon) = 0 \ ,$$

where $d_2'(x, y) := \partial d(x, y)/\partial y$ denotes the partial derivative of $d(x, y)$ with respect to the second variable. Using this equality, the derivative of g_b simplifies to

$$\begin{aligned} g_b'(x) &= d(\mu_b, \lambda_b(x) + \epsilon) + \lambda_b'(x)\big[d_2'(\mu_a, \lambda_b(x)) + x d_2'(\mu_b, \lambda_b(x) + \epsilon)\big] \\ &= d(\mu_b, \lambda_b(x) + \epsilon) \ . \end{aligned}$$

Hence for every $y \in \big[d(\mu_a, \mu_a \wedge (\mu^+ - \epsilon)), d(\mu_a, (\mu_b - \epsilon) \vee \mu^-)\big)$,

$$x_b'(y) = \frac{1}{d\big(\mu_b, \lambda_b(x_b(y)) + \epsilon\big)} \ .$$

Using this expression in the derivative (C.2) yields

$$\begin{aligned} G'(y) \geq 0 &\Leftrightarrow 1 + \sum_{b \neq a} x_b(y) \geq y \sum_{b \neq a} \frac{1}{d\big(\mu_b, \lambda_b(x_b(y)) + \epsilon\big)} \\ &\Leftrightarrow \sum_{b \neq a} \frac{d\big(\mu_a, \lambda_b(x_b(y))\big)}{d\big(\mu_b, \lambda_b(x_b(y)) + \epsilon\big)} \leq 1 \ , \end{aligned}$$

where we use that by definition, for all b,

$$y = g_b(x_b(y)) = d\big(\mu_a, \lambda_b(x_b(y))\big) + x_b(y) d\big(\mu_b, \lambda_b(x_b(y)) + \epsilon\big) \ .$$

The mapping

$$F(y) := \sum_{b \neq a} \frac{d\big(\mu_a, \lambda_b(x_b(y))\big)}{d\big(\mu_b, \lambda_b(x_b(y)) + \epsilon\big)}$$

is increasing as a composition of the two decreasing mappings $y \mapsto \lambda_b\big(x_b(y)\big)$ (decreasing by Proposition 19) and $\lambda \mapsto \sum_{b \neq a} \frac{d(\mu_a, \lambda)}{d(\mu_b, \lambda + \epsilon)}$. With $b^* \in \operatorname*{argmax}_{b \neq a} \mu_a$, Proposition 19 permits to show that

$$\begin{aligned} &\lim_{y \to d(\mu_a, \mu_a \wedge (\mu^+ - \epsilon)))} \lambda_b(x_b(y)) = \mu_a \wedge (\mu^+ - \epsilon) \quad \text{and} \\ &\lim_{y \to d(\mu_a, (\mu_{b^*} - \epsilon) \vee \mu^-)} \lambda_b(x_b(y)) = (\mu_{b^*} - \epsilon) \vee \mu^- \ . \end{aligned}$$

Thus F tends to 0 in $d\big(\mu_a, \mu_a \wedge (\mu^+ - \epsilon)\big)$ and to $+\infty$ in $d\big(\mu_a, (\mu_{b^*} - \epsilon) \vee \mu^-\big)$. Hence, there is a unique y^* that satisfies $F(y^*) = 1$ and G' is positive for $y < y^*$, negative

for $y > y^*$ and satisfies $G'(y^*) = 0$. This proves that G as a unique maximum on $\mathcal{I}_a$, given by the unique solution y^* to the equation

$$\sum_{b \neq a} \frac{d\big(\mu_a, \lambda_b(x_b(y))\big)}{d\big(\mu_b, \lambda_b(x_b(y)) + \epsilon\big)} = 1 \ ,$$

which justifies the second statement of Proposition 12.

AURÉLIEN GARIVIER
UNITÉ DE MATHÉMATIQUES PURES ET APPLIQUÉES ET LABORATOIRE DE L'INFORMATIQUE DU PARALLÉLISME
ÉCOLE NORMALE SUPÉRIEURE DE LYON, UNIVERSITÉ DE LYON
LYON, FRANCE
E-MAIL: aurelien.garivier@ens-lyon.fr
EMILIE KAUFMANN
INRIA LILLE, EQUIPE SEQUEL
40 AVENUE HALLEY
VILLENEUVE D'ASCQ, FRANCE
E-MAIL: emilie.kaufmann@univ-lille.fr