Top Two Algorithms Revisited

Marc Jourdan[1]
marc.jourdan@inria.fr

Rémy Degenne[1]
remy.degenne@inria.fr

Dorian Baudry[1]
dorian.baudry@inria.fr

Rianne de Heide[2]
r.de.heide@vu.nl

Emilie Kaufmann[1]
emilie.kaufmann@univ-lille.fr

[1] Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9198-CRIStAL, F-59000 Lille, France
[2] Vrije Universiteit Amsterdam

Abstract

Top Two algorithms arose as an adaptation of Thompson sampling to best arm identification in multi-armed bandit models [38], for parametric families of arms. They select the next arm to sample from by randomizing among two candidate arms, a *leader* and a *challenger*. Despite their good empirical performance, theoretical guarantees for fixed-confidence best arm identification have only been obtained when the arms are Gaussian with known variances. In this paper, we provide a general analysis of Top Two methods, which identifies desirable properties of the leader, the challenger, and the (possibly non-parametric) distributions of the arms. As a result, we obtain theoretically supported Top Two algorithms for best arm identification with bounded distributions. Our proof method demonstrates in particular that the sampling step used to select the leader inherited from Thompson sampling can be replaced by other choices, like selecting the empirical best arm.

1 Introduction

Finding the distribution that has the largest mean by sequentially collecting samples from a pool of candidate distributions ("arms") has been extensively studied in the multi-armed bandit [6, 24] and ranking and selection [21] literature. While existing approaches often rely on parametric assumptions for the distributions, we are interested in (near) optimal and computationally efficient strategies when the distributions belong to an arbitrary class $\mathcal{F}$ of distributions.

For applications to online marketing such as A/B testing [30, 37] assuming Bernoulli or Gaussian arms is fine, but more sophisticated distributions arise in other fields such as agriculture. In Section 5 we consider a crop-management problem: a group of farmers wants to identify the best planting date for a rainfed crop. The reward (crop yield) can be modeled as a complex distribution with multiple modes, but upper bounded by a known yield potential. Therefore, sequentially identifying the best planting date calls for efficient best arm identification algorithms for the class of bounded distributions with a known range.

To tackle this problem, we build on Top Two algorithms [38, 35, 39], originally proposed for specific parametric families. We propose a generic analysis of this type of algorithms, which puts forward new possibilities for the choice of leader and challenger used by the algorithm. In particular, this work leads to the first asymptotically β-optimal strategies for bounded distributions.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

1.1 Setting and related work

A bandit problem is described by a finite number of probability distributions (K many), called arms. Let $\triangle_K$ be the K-dimensional probability simplex and $\mathcal{P}(\mathbb{R})$ the set of probability distributions over $\mathbb{R}$. Let $\mathcal{F} \subset \mathcal{P}(\mathbb{R})$ be a known family of distributions to which the arms to. We will refer to tuples of distributions in $\mathcal{F}^K$ with bold letters, e.g. $\boldsymbol{F} = (F_1, \dots, F_K) \in \mathcal{F}^K$ where F_i is the cdf of arm i. We suppose that all distributions in $\mathcal{F}$ have finite first moment and we denote the mean of $F \in \mathcal{F}$ by $m(F)$. We denote by $\mathcal{I} = \{m(F) \mid F \in \mathcal{F}\}$ the set of possible means for the arms.

The goal of a best arm identification (BAI) algorithm is to identify an arm with highest mean in the set of available arms, i.e. an arm which belongs to the set $i^\star(\boldsymbol{F}) = \arg\max_{k\in[K]} m(F_k)$. At each time $n \in \mathbb{N}$, the algorithm interacts with the environment (the set of arms) by (1) choosing an arm I_n based on previous observations, (2) observing a sample $X_{n,I_n} \sim F_{I_n}$, and (3) deciding whether to stop and return an arm $\hat{\imath}_n$ or to continue. We study the *fixed confidence* identification setting, in which we require algorithms to make mistakes with probability less than a given $\delta \in (0,1)$. To compare such algorithms we consider their *sample complexity* τ_δ, which is a stopping time counting the number of rounds before the algorithm terminates. The goal is then to minimize $\mathbb{E}[\tau_\delta]$ among the class of δ-correct algorithms.

Definition 1. *An algorithm is δ-correct*[1] *on $\mathcal{F}^K$ if $\mathbb{P}_{\boldsymbol{F}}\left(\tau_\delta < +\infty,\ \hat{\imath}_{\tau_\delta} \notin i^\star(\boldsymbol{F})\right) \le \delta$ for all $\boldsymbol{F} \in \mathcal{F}^K$.*

In order to be δ-correct on $\mathcal{F}^K$, an algorithm has to be able to distinguish problems in $\mathcal{F}^K$ with different best arms. This intuition is formalized by the lower bound provided in Lemma 1. The characteristic time defined in the lower bound depends on two functions $\mathcal{K}^+_{\inf}$ and $\mathcal{K}^-_{\inf}$, mapping $\mathcal{P}(\mathbb{R}) \times \mathbb{R}$ to $\mathbb{R}_+$, obtained by minimizing a Kullback-Leibler divergence (KL) over $\mathcal{F}$,

$$\mathcal{K}^+_{\inf}(F,u) := \inf\{\mathrm{KL}(F,G) \mid G \in \mathcal{F},\ \mathbb{E}_{X\sim G}[X] > u\}\,,$$
$$\mathcal{K}^-_{\inf}(F,u) := \inf\{\mathrm{KL}(F,G) \mid G \in \mathcal{F},\ \mathbb{E}_{X\sim G}[X] < u\}\,.$$

Lemma 1 (From [16, 3])**.** *Any algorithm which is δ-correct on $\mathcal{F}^K$ verifies, for any $\boldsymbol{F} \in \mathcal{F}^K$,*

$$\mathbb{E}_{\boldsymbol{F}}[\tau_\delta] \ge T^\star(\boldsymbol{F}) \log\left(1/(2.4\delta)\right)\,,$$

where $T^\star(\boldsymbol{F})^{-1} := \sup_{w\in\triangle_K} \min_{i\neq i^\star} \inf_{u\in\mathcal{I}} \left\{w_{i^\star}\mathcal{K}^-_{\inf}(F_{i^\star},u) + w_i\mathcal{K}^+_{\inf}(F_i,u)\right\}$.

We say that an algorithm is asymptotically optimal if its sample complexity matches that lower bound, that is if $\limsup_{\delta\to 0} \mathbb{E}_{\boldsymbol{F}}[\tau_\delta]/\log(1/\delta) \le T^\star(\boldsymbol{F})$.

A related, weaker notion of (asymptotic) optimality is (asymptotic) β-optimality [39]. An algorithm is called asymptotically β-optimal if it satisfies $\limsup_{\delta\to 0} \mathbb{E}_{\boldsymbol{F}}[\tau_\delta]/\log(1/\delta) \le T^\star_\beta(\boldsymbol{F})$, for the complexity term

$$T^\star_\beta(\boldsymbol{F})^{-1} := \sup_{w\in\triangle_K, w_{i^\star}=\beta} \min_{i\neq i^\star} \inf_{u\in\mathcal{I}} \left\{\beta\mathcal{K}^-_{\inf}(F_{i^\star},u) + w_i\mathcal{K}^+_{\inf}(F_i,u)\right\}\,.$$

An asymptotically β-optimal algorithm is asymptotically minimizing the sample complexity among algorithms which allocate a β fraction of samples to the best arm and $T^\star(\boldsymbol{F}) = \min_{\beta\in(0,1)} T^\star_\beta(\boldsymbol{F})$. As was first shown by [38] when $\mathcal{F}$ is an exponential family, an asymptotically β-optimal algorithm with $\beta = 1/2$ also has an expected sample complexity which is asymptotically optimal, up to a multiplicative factor 2. That is, $T^\star_{1/2}(\boldsymbol{F}) \le 2T^\star(\boldsymbol{F})$.

We denote by $w^\star(\boldsymbol{F})$ and $w^\star_\beta(\boldsymbol{F})$ the allocations realizing the argmax in the definition of $T^\star(\boldsymbol{F})$ and $T^\star_\beta(\boldsymbol{F})$, respectively. We will show that for common choices of $\mathcal{F}$ these allocations are unique when there is a unique best arm.

Distribution classes The characteristic time $T^\star(\boldsymbol{F})$ depends on the class of distributions $\mathcal{F}$, known to the algorithm in advance, to which $\boldsymbol{F}$ belongs to. For example, all arms could have Bernoulli distributions. We strive to provide an analysis which could easily be applied to many classes $\mathcal{F}$, but we specialize our results to two main cases:

1. distributions with bounded support, $\mathcal{F} = \{F \in \mathcal{P}(\mathbb{R}) \mid \mathrm{supp}(F) \subseteq [0,B]\}$ for $B > 0$,

[1] A stronger definition of δ-correctness has also been studied by requiring the algorithm to stop almost surely.

2. single parameter exponential families (SPEF) of sub-exponential distributions.

Given a distribution $\mathbb{P}^{(0)}$ with cumulant generating function φ, defined on an interval $\mathcal{I}_\varphi$, the SPEF defined by $\mathbb{P}^{(0)}$ is the set of distributions $\mathbb{P}^{(\lambda)}$ with density with respect to $\mathbb{P}^{(0)}$ given by $\frac{d\mathbb{P}^{(\lambda)}}{d\mathbb{P}^{(0)}}(x) = e^{\lambda x - \varphi(\lambda)}$. For example, Gaussian distributions with a known variance form a SPEF, as do Bernoulli distributions with means in $(0, 1)$. We consider SPEF of sub-exponential distributions to have a concentration property for the empirical mean estimator.

Related work The first Best Arm Identification (BAI) algorithms [14, 27, 15, 23] were proposed and analyzed for bounded rewards, but their sample complexity scales with a sum of inverse gaps between the means of arms instead of the quantity $T^\star(\boldsymbol{F})$ prescribed by the lower bound. Asymptotically optimal BAI algorithm were first designed when the arms belong to the same single-parameter exponential family. In this context, two families of asymptotically optimal algorithms have emerged. Tracking-based algorithms solve the optimization problem provided by the lower bound in every round, and track the corresponding allocation [16]. The gamification approach views the characteristic time as a min-max game between the learner and the nature, and apply a saddle-point algorithm to solve it sequentially at a lower computational cost [13].

Some Bayesian algorithms arose as another computationally appealing alternative to Track-and-Stop. Russo notably proposed the Top Two Probability Sampling (TTPS) and Top Two Thompson Sampling (TTTS) algorithms [38], that may be seen as counterparts of the popular Thompson Sampling algorithm for regret minimization [41]. Other Bayesian flavored Top Two algorithms have been proposed, Top Two Expected Improvement (TTEI, [35]) and Top Two Transportation Cost (T3C, [39]). All these algorithms sample either a *leader* with fixed probability β or a *challenger* with probability $1 - \beta$. TTTS, TTEI and T3C were proved to be asymptotically β-optimal for Gaussian bandits and perform well in practice even against asymptotically optimal algorithms [35, 39]. This motivates our investigation of Top Two algorithms to tackle bounded distributions, which led us to propose a new generic analysis of this kind of algorithms of independent interest. We prove the asymptotic β-optimality of several Top Two instances for bounded bandit models, some of which depart from their original Bayesian motivation as they don't need a sampler. An asymptotically optimal algorithm for a non-parametric class of distribution has been proposed by [3] for heavy-tailed rewards. It relies on the computationally prohibitive Track-and-Stop approach, and an adaptation to bounded distributions is mentioned, yet without an explicit calibration of the stopping rule.

1.2 Contributions

We present the first fixed-confidence analysis of Top Two algorithms for distribution classes other than Gaussian, including the non-parametric setting of bounded distributions. In Section 2, we introduce several variants of Top Two algorithms, including new ones which choose the empirical best arm as a leader instead of relying on (Thompson) sampling and/or use some penalization in the previously proposed Transportation Cost challenger.

For the class of bounded distributions, we propose in Section 3 a calibration of the stopping rule and a concrete instantiation of the Top Two algorithms, based on a Dirichlet sampler for the randomized variants. We prove in Theorem 1 that those algorithms are asymptotically β-optimal. This optimality can also be shown for deterministic instances in the case of sub-exponential single parameter exponential families (Appendix H). Our generic analysis, sketched in Section 4, provides insight on what properties the leader and challenger in a Top Two algorithm should have in order to reach asymptotic β-optimality. We show that the algorithm should ensure that all arms are explored sufficiently, and explain how to guarantee that the sampling proportions reach their optimal values once the sufficient exploration condition holds.

Finally, in Section 5 we report results from numerical experiments on a challenging non-parametric task using real-world data from a crop-management problem for various members of the Top Two family of algorithms. Most of them perform significantly better than the baselines.

2 Generic Top Two algorithms

Let $\boldsymbol{F} \in \mathcal{F}^K$ such that $|i^\star(\boldsymbol{F})| = 1$ and $\mu_i := m(F_i) \in \mathcal{I}$ for all $i \in [K]$. As for most BAI algorithms, each arm is pulled once for the initialization. At time $n + 1$, the σ-algebra $\mathcal{F}_n :=$

$\sigma(U_1, I_1, X_{1,I_1}, \cdots, I_n, X_{n,I_n}, U_{n+1})$, called history, encompasses all the information available to the agent and the internal randomization denoted by $(U_t)_{t\in[n+1]}$, which is independent of everything else. For all $\mathcal{F}_n$-measurable sets A, we denote by $\mathbb{P}_{|n}[A] := \mathbb{P}[A \mid \mathcal{F}_n]$ its probability. For an arm i, we denote its number of pulls by $N_{n,i} := \sum_{t\in[n]} \mathbb{1}(I_t = i)$, its empirical distribution by $F_{n,i} := \frac{1}{N_{n,i}} \sum_{t\in[n]} \delta_{X_{t,I_t}} \mathbb{1}(I_t = i)$ and its empirical mean by $\mu_{n,i} := m(F_{n,i})$.

Stopping and recommendation rules Our Top Two algorithms rely on the same stopping rule, which can be expressed using the (empirical) transportation cost between arm i and arm j, defined as

$$W_n(i,j) = \inf_{x\in\mathcal{I}} \left[N_{n,i} \mathcal{K}^-_{\inf}(F_{n,i}, x) + N_{n,j} \mathcal{K}^+_{\inf}(F_{n,j}, x) \right]. \tag{1}$$

In particular, using the definition of $\mathcal{K}^{\pm}_{\inf}$, it can be noted that $W_n(i,j) = 0$ if $\mu_{n,i} \leq \mu_{n,j}$. Given a threshold function $c(n,\delta)$, the stopping rule is

$$\tau_\delta = \inf\{n \in \mathbb{N} \mid \min_{j\neq \hat{\imath}_n} W_n(\hat{\imath}_n, j) > c(n,\delta)\}, \tag{2}$$

and the recommendation rule is $\hat{\imath}_n = \arg\max_i \mu_{n,i}$. Up to the choice of threshold, this stopping rule coincides with the GLR-based stopping rule proposed when $\mathcal{F}$ is an exponential family [16] and by [3] for heavy-tailed distributions with an upper bound on a non-centered moment. For a general class $\mathcal{F}$ the stopping rule can be calibrated to ensure δ-correctness under any sampling rule if the threshold is such that the following time-uniform concentration inequality holds for all $\boldsymbol{F} \in \mathcal{F}^K$:

$$\mathbb{P}_{\boldsymbol{F}} \left(\exists n,\ \exists i \neq i^\star(\boldsymbol{F}) :\ N_{n,i} \mathcal{K}^-_{\inf}(F_{n,i}, \mu_i) + N_{n,i^\star(\boldsymbol{F})} \mathcal{K}^+_{\inf}(F_{n,i^\star(\boldsymbol{F})}, \mu_{i^\star(\boldsymbol{F})}) > c(n,\delta) \right) \leq \delta. \tag{3}$$

Lemma 2 in the next section gives an explicit threshold for the class of bounded distribution. For SPEF, we can use generic stopping thresholds derived in [29].

1: **Input:** β
2: Choose a leader $B_n \in [K]$
3: $U \sim \mathcal{U}([0,1])$
4: **if** $U < \beta$ **then**
5: $\quad I_n = B_n$
6: **else**
7: $\quad$ Choose a challenger $C_n \in [K] \setminus \{B_n\}$
8: $\quad I_n = C_n$
9: **end if**
10: **Output**: next arm to sample I_n

Figure 1: Generic β-Top Two sampling rule

Choice of the leader (two propositions):

EB - $B_n^{\text{EB}} \in \arg\max_i \mu_{n-1,i}$

TS - Sample $\theta \sim \Pi_{n-1}$ then set $B_n^{\text{TS}} \in \arg\max_{i\in[K]} \theta_i$

Choice of the challenger (three propositions):

TC - $C_n^{\text{TC}} \in \arg\min_{j\neq B_n} W_{n-1}(B_n, j)$

TCI - $C_n^{\text{TCI}} \in \arg\min_{j\neq B_n} W_{n-1}(B_n, j) + \log N_{n-1,j}$

RS - repeat $\theta \sim \Pi_{n-1}$ until $C_n^{\text{RS}} \in \arg\max_{i\in[K]} \theta_i \not\ni B_n$

Figure 2: Choices of leader and challenger (uniform tie-breaking).

Sampling rule The sampling rule of a Top Two algorithm is shown in Figure 1. The method chooses a first arm B_n called leader which is then sampled with probability β. If B_n is not sampled, then a second arm C_n called challenger is chosen and sampled. Our analysis isolates properties that those two choices should fulfill in order for the Top Two algorithm to be asymptotically β-optimal.

The practical implementation of a Top Two method then requires subroutines for B_n and C_n. Two possibilities for the leader and three possibilities for the challenger are presented in Figure 2. Our analysis will apply to any combination of those and we will refer to the algorithms obtained by β-[leader]-[challenger]; for example β-EB-TCI or β-TS-TC.

We have two flavors of leaders and challengers: deterministic and randomized. The deterministic choices (EB, for Empirical Best, leader, TC and TCI challengers) rely on the empirical Transportation Costs (TC) $W_n(i,j)$ used in the stopping rule: the TC and TCI challengers are the arms which minimize the transportation cost from the leader (up to a penalization for TCI, hence TC Improved). The randomized choices (TS leader and RS challenger) rely on a *sampler*, denoted by Π_n. Π_n generates i.i.d. vectors $\theta = (\theta_1, \ldots, \theta_K) \in \mathcal{I}^K$ which are interpreted as possible means for the arms, under a distribution which depends on observations gathered in the first n rounds. The TS leader is the best arm in the sampled vector, which is inspired by Thompson Sampling. The RS (for Re-Sampling) challenger is obtained by performing repeated calls to the sampler until the best arm in the sampled vector is not B_n, then taking the best arm.

Randomized instances The samplers suggested by prior work all have a Bayesian flavor. For SPEF bandits, they use $\Pi_n = \Pi_{n,1} \times \cdots \times \Pi_{n,K}$ where $\Pi_{n,i}$ is the posterior distribution on the mean of arm i after n rounds (given some prior distribution). With this choice of sampler, β-TS-RS coincides with the TTTS algorithm [38], while β-TS-TC coincides with the T3C algorithm [39]. TTTS and T3C were only proved to be asymptotically β-optimal for Gaussian bandits with improper priors, whereas a by-product of the general analysis that we propose in this work permits to establish the necessary properties on the sampler for it to hold for more general distributions. Moreover, we extend these algorithms to bounded distributions by virtue of Dirichlet sampling and also analyze their sampler-free counterparts. As will be apparent in our analysis, the crucial property needed from the sampler in a Top Two algorithm using the RS challenger is that for all arms i, j such that $\mu_i > \mu_j$, $\mathbb{P}_{\theta\sim\Pi_n}(\theta_j > \theta_i) \simeq \exp(-W_n(i,j))$.

Deterministic instances Under the RS challenger, the probability to obtain as a challenger arm j is proportional to the probability that $\mathbb{P}_{\theta\sim\Pi_n}(\theta_j > \theta_{B_n})$. Therefore, if Π_n is a good sampler satisfying the above property, the TC challenger can be seen as replacing the randomization in the RS challenger by a computation of the mode of the distribution of C_n^{RS}. This was the motivation behind T3C [39] as Gaussian transportation costs have a simple closed form expression while re-sampling becomes more and more costly when the posterior distributions are concentrated. While our asymptotic analysis holds for deterministic algorithms, the empirical performance of fully deterministic algorithms might suffer from unlucky draws. In Section 5, we show that β-EB-TC is indeed the least robust of all our instances. To cope for this pitfall, explicit or implicit exploration mechanisms can be added. Inspired by IMED [20], the TCI challenger fosters exploration by penalizing over-sampled challengers. Randomization and forced exploration are two other examples of implicit and explicit exploration mechanisms.

3 Asymptotically β-optimal algorithms for bounded distributions

For bounded distribution, Lemma 2 provides a calibration of the stopping rule. Its proof, given in Appendix E.1, relies on a martingale construction proposed by [5].

Lemma 2. *The stopping rule (2) with threshold*

$$c(n,\delta) = \log(1/\delta) + 2\log(1 + n/2) + 2 + \log(K-1) \tag{4}$$

is δ-correct for the family of bounded distributions.

Transportation costs Both the stopping rule and the TC and TCI challengers of the sampling rule require the computation of $W_n(i,j)$ defined in (1). For single-parameter exponential families, this can be done easily since $\mathcal{K}_{\inf}^{\pm}$ are KL divergences and the transportation cost has a closed form expression [16, 38]. However, for bounded distributions, computing $\mathcal{K}_{\inf}^{\pm}$ is more challenging and we rely on the dual formulation first obtained by [18] (see Theorem 3):

$$N_{n,i}\mathcal{K}_{\inf}^{+}(F_{n,i}, x) = \sup_{\lambda\in[0,1]} \sum_{t\in[n]} \mathbb{1}(I_t = i) \log\left(1 - \lambda\frac{X_{t,i} - x}{B - x}\right) .$$

The minimization in λ can be computed using a zero-order optimization algorithm (e.g. Brent's method [10]). The same optimizer can be used to compute the minimization in $x \in [0, B]$ featured in $W_n(i,j)$. By nesting those optimizations of univariate functions on a bounded interval, the computation of $W_n(i,j)$ in the stopping rule dominates the computational cost of our Top Tow algorithms (except the RS challenger). Our experiments suggest that using (2) is twice as computationally expensive as the LUCB-based stopping rule, which is a mild price to pay for the improvement in terms of empirical stopping time. Algorithms for non-parametric distributions are bound to be computationally more expensive than their counterpart in SPEF, where a sufficient statistic can summarize $\mathcal{F}_n$.

Sampler The TS leader and RS challenger require a sampler. Our proposed sampler for bounded distributions in $[0, B]$ has a product form: $\Pi_n = \Pi_{n,1} \times \cdots \times \Pi_{n,K}$ where $\Pi_{n,i}$ leverages $\mathcal{H}_{n,i} := (X_{1,i}, \ldots, X_{N_{n,i},i})$, which is the history of samples from arm i collected in the first n rounds. Let $\tilde{F}_{n,i}$ denote the empirical cdf of $\mathcal{H}_{n,i}$ augmented by the known bounds on the support, $\{0, B\}$. For

each arm i, $\Pi_{n,i}$ outputs a random re-weighting of $\tilde{F}_{n,i}$. Concretely, letting $\boldsymbol{w} = (w_1, \dots, w_{N_{n,i}+2})$ be drawn from a Dirichlet distribution $\mathrm{Dir}(\mathbf{1}_{N_{n,i}+2})$, a call to the sampler $\Pi_{n,i}$ returns

$$\sum_{t \in [N_{n,i}]} w_t X_{t,i} + B w_{N_{n,i}+1} \,.$$

This sampler is inspired by that used in the Non Parametric Thompson Sampling (NPTS) algorithm proposed by [36] for regret minimization in bounded bandits, with the notable difference that we have to add both 0 and B in the support, while NPTS only adds the upper bound B. We will see that this is only necessary to ensure that the re-sampling procedure stops. Therefore, the TS leader could use a sampler $\tilde{\Pi}_n$ based directly on $\mathcal{H}_{n,i}$.

Theorem 1. *Combining the stopping rule (2) with threshold (4) and a Top Two algorithm with $\beta \in (0,1)$, instantiated with any pair of leader/challenger as in Figure 2, yields a δ-correct algorithm which is asymptotically β-optimal for all $\boldsymbol{F} \in \mathcal{F}^K$ with $\mu_{\boldsymbol{F}} \in (0,B)^K$ and $\Delta_{\min}(\boldsymbol{F}) := \min_{i \neq j} |\mu_{F_i} - \mu_{F_j}| > 0$.*

Theorem 1 gives the asymptotic β-optimality for six algorithms (Figure 2). Choosing our favorite Top Two instances therefore requires further empirical and computational considerations. Computing the EB leader has a constant computational cost, while the TS leader is computationally costly for large time n since it requires to sample from a Dirichlet distribution with $N_{n,i}+2$ parameters for each arm i. On the challenger side, the RS challenger is computationally very expensive for large time n as the sampler becomes concentrated around the true mean vector. On the contrary, by leveraging computations done in the stopping rule (2), the TC and TCI challengers can be computed in constant time. Based on these computational considerations, the most appealing Top Two algorithm for bounded distribution appears to be the fully deterministic β-EB-TC. But experiments performed in Section 5 reveal its lack of robustness, and for bounded distributions the best trade-off between robustness and computational complexity is β-EB-TCI. More generally, β-TS-TC can also be a good choice provided that we have access to an efficient sampler.

Distinct means Restricting to instances such that $\Delta_{\min}(\boldsymbol{F}) > 0$ (which implies $|i^\star(\boldsymbol{F})| = 1$) is an uncommon assumption in BAI. However, known Top Two algorithms [38, 35, 39] only have guarantees on those instances. Our generic analysis reveals that it is solely used to prove sufficient exploration, characterized by (7) (Appendix C.3). Experiments highlights that all our Top Two algorithms except β-EB-TC perform well on instances where $|i^\star(\boldsymbol{F})| = 1$ and $\Delta_{\min}(\boldsymbol{F}) = 0$ (Figure 4(b)). Proving theoretical guarantees in this situation is an interesting problem for future work (see Appendix D.3 for a discussion).

4 Sample complexity analysis

In this section, we sketch the proof of Theorem 1, which follows from the generic sample complexity analysis of Top Two algorithms presented in Appendix C. Our proof strategy is the same as that first introduced by [35] for the analysis of TTEI and also used by [39] for TTTS and T3C. It consists in upper bounding the expectation of the *convergence time*, defined as

$$T_\beta^\varepsilon := \inf \left\{ T \geq 1 \mid \forall n \geq T,\ \max_{i \in [K]} \left| \frac{N_{n,i}}{n} - w_i^\beta \right| \leq \varepsilon \right\} \,, \tag{5}$$

for ε small enough. Indeed, we prove in Appendix C.5 that for any sampling rule

$$\exists \varepsilon_0(\boldsymbol{F}) > 0,\ \forall \varepsilon \in (0, \varepsilon_0(\boldsymbol{F})],\ \mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] < +\infty \quad \Longrightarrow \quad \limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log(1/\delta)} \leq T_\beta^\star(\boldsymbol{F}) \,. \tag{6}$$

This implication only leverages the expression of the stopping rule and the threshold. It was previously established for Gaussian bandits by [35] and we extend this property to bounded distributions and SPEF of sub-exponential distributions. Up to technicalities ($\mathcal{K}_{\inf}$ continuity and second order terms), this implication is shown by using that if $\tau_\delta \geq n$, then

$$\log(1/\delta) \approx_{\delta \to 0} c(n,\delta) \geq \min_{j \neq \hat{\imath}_n} W_n(\hat{\imath}_n, j) \approx_{n \geq T_\beta^\varepsilon} n T_\beta^\star(\boldsymbol{F})^{-1} \,.$$

To upper bound the expected convergence time, as prior work we first establish *sufficient exploration*:

$$\exists N_1 \text{ s.t. } \mathbb{E}_{\boldsymbol{F}}[N_1] < +\infty,\ \forall n \geq N_1, \quad \min_{i \in [K]} N_{n,i} \geq \sqrt{n/K} \,. \tag{7}$$

By generalizing [39] which considered Gaussian, we identify two generic properties for the leader and the challenger under which (7) hold (Appendix C.3), provided that we assume $\Delta_{\min} > 0$.

We proceed similarly to prove convergence by identifying in Appendix C desired properties for the leader and challenger, which are satisfied by all our leaders and challengers for bounded distributions (Appendix D). We sketch these conditions below. Let $i^\star$ be the unique element of $i^\star(\boldsymbol{F})$.

The requirements on the leader and the challenger to ensure $\mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] < +\infty$ become apparent when looking at generic properties of Top Two algorithms. Under any Top Two algorithm, the probability to select arm i at round n, $\psi_{n,i} := \mathbb{P}_{|(n-1)}[I_n = i]$, can be written as

$$\psi_{n,i} = \beta \mathbb{P}_{|(n-1)}[B_n = i] + (1-\beta) \sum_{j \neq i} \mathbb{P}_{|(n-1)}[B_n = j] \mathbb{P}_{|(n-1)}[C_n = i | B_n = j] \,. \tag{8}$$

We let $\Psi_{n,i} := \sum_{t \in [n]} \psi_{t,i}$. For the leader, we can prove using (8) that

$$\forall M \in \mathbb{N}, \quad \left| \frac{\Psi_{n,i^\star}}{n} - \beta \right| \leq \frac{M-1}{n} + \frac{1}{n} \sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[B_t \neq i^\star] \,.$$

This suggests that a *good* leader should satisfy that there exists N_2 with $\mathbb{E}_{\boldsymbol{F}}[N_2] < +\infty$ s.t.

$$\forall n \geq N_2, \quad \mathbb{P}_{|n}[B_{n+1} \neq i^\star] \leq g(n) \,, \tag{9}$$

where $g(n) =_{+\infty} o(n^{-\alpha})$ for some $\alpha > 0$. For the challenger, noticing that

$$\forall M \in \mathbb{N}, \forall i \neq i^\star, \quad \frac{\Psi_{n,i}}{n} \leq \frac{M-1}{n} + \frac{1}{n} \sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[B_t \neq i^\star] + \frac{1}{n} \sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[C_t = i | B_t = i^\star] \,,$$

suggests that a *good* challenger should satisfy that there exists N_3 with $\mathbb{E}_{\boldsymbol{F}}[N_3] < +\infty$ s.t.

$$\forall n \geq N_3, \forall i \neq i^\star, \quad \frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon \;\Rightarrow\; \mathbb{P}_{|n}[C_{n+1} = i | B_{n+1} = i^\star] \leq h(n) \,, \tag{10}$$

where $h(n) =_{+\infty} o(n^{-\alpha})$ for some $\alpha > 0$. Then, Cesaro's theorem further yields

$$\exists N_4 \text{ s.t. } \mathbb{E}_{\boldsymbol{F}}[N_4] < +\infty, \forall n \geq N_4, \quad \max_{i \in [K]} \left| \frac{\Psi_{n,i}}{n} - w_i^\beta \right| \leq \varepsilon \,.$$

Using that $(N_{n,i} - \Psi_{n,i})/\sqrt{n}$ are sub-Gaussian random variables, we obtain $\mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] < +\infty$.

We now explain why (9) and (10) are satisfied for the leaders and challengers in Figure 2 when $\mathcal{F}$ is the class of bounded distributions. This follows from concentration properties. Using the fact that $\sqrt{n}\|F_{n,i} - F\|_\infty$ is sub-Gaussian, which follows for the Dvoretzky–Kiefer–Wolfowitz inequality [31], the continuity of the mean operator m on $\mathcal{F}$ and the sufficient exploration property (7), we establish that for all $\alpha > 0$, there exists a random variable N_α with finite expectation such that

$$\forall n \geq N_\alpha, \quad \max_{i \in [K]} \|F_{n,i} - F_i\|_\infty \leq \alpha \quad \text{and} \quad \max_{i \in [K]} |\mu_{n,i} - \mu_i| \leq \alpha \,. \tag{11}$$

Deterministic instances Recall that $B_{n+1}^{\mathrm{EB}} \in \arg\max_{i \in [K]} \mu_{n,i}$. Choosing α in (11) smaller than half the gap between the best and second best arm (which is possible as $|i^\star(\boldsymbol{F})| = 1$) yields that for all $n \geq N_\alpha$, $B_{n+1}^{\mathrm{EB}} = i^\star$. This proves (9) with $g(n) = 0$. Using continuity and convexity properties of $\mathcal{K}_{\inf}^{\pm}$, we then establish that there exists $\alpha > 0$ and a problem-dependent constant $C_{\boldsymbol{F}} > 0$ such that for $n \geq N_\alpha$ and for all $i \neq i^\star$,

$$\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon \quad \Longrightarrow \quad \frac{1}{n} \left(W_n(i^\star, i) - \min_{j \neq i^\star} W_n(i^\star, j) \right) \geq C_{\boldsymbol{F}} \,.$$

This implies that $i \notin \min_{j \neq i^\star} W_n(i^\star, j)$, hence $\mathbb{P}_{|n}[C_{n+1}^{\mathrm{TC}} = i \mid B_{n+1} = i^\star] = 0$ for $n \geq N_\alpha$. Therefore, (10) holds with $h(n) = 0$. A similar argument holds for C_{n+1}^{TCI}.

Randomized instances Let $a_{n+1,i} := \mathbb{P}_{\theta\sim\Pi_n}(i \in \arg\max_{j\in[K]} \theta_j)$ be the probability that arm i is the best arm in a sampled model at round n. Since

$$\mathbb{P}_{|n}[B_{n+1}^{\mathrm{TS}} \neq i^\star] \leq (K-1)\max_{i\neq i^\star} a_{n+1,i} \leq (K-1)\max_{i\neq i^\star} \mathbb{P}_{\theta\sim\Pi_n}(\theta_i \geq \theta_{i^\star}) \,,$$

an upper bound on $\mathbb{P}_{\theta\sim\Pi_n}[\theta_i \geq \theta_{i^\star}]$ is sufficient to prove (9). We show in Lemma 64 that this can be obtained by leveraging upper bound on the Boundary Crossing Probability (BCP) of the Dirichlet sampler, $\mathbb{P}_{\theta\sim\Pi_n}[\theta_i \geq u]$ for a fixed threshold $u \in (0, B)$. An upper bound on the BCP can be obtained using the work of [36] and is given in Theorem 5 for the sake of completeness. Putting things together yields that, for all n,

$$\mathbb{P}_{\theta\sim\Pi_n}[\theta_i \geq \theta_{i^\star}] \leq f\left(\inf_{u\in[0,B]}[(N_{n,i^\star}+2)\mathcal{K}_{\inf}^-(\tilde{F}_{n,i^\star},u) + (N_{n,i}+2)\mathcal{K}_{\inf}^+(\tilde{F}_{n,i},u)]\right) \,,$$

where $f(x) = (1+x)e^{-x}$. Using again continuity and concentration (11), we conclude that (9) holds with $g(n) = (K-1)f\left(\left(\sqrt{\frac{n}{K}}+2\right)D_{\boldsymbol{F}}\right)$, where $D_{\boldsymbol{F}} > 0$ is a problem dependent constant.

For the challenger, we first observe that

$$\mathbb{P}_{|n}[C_{n+1}^{\mathrm{RS}} = i \mid B_{n+1} = i^\star] = \frac{a_{n+1,i}}{1-a_{n+1,i^\star}} \leq \frac{\mathbb{P}_{\theta\sim\Pi_n}[\theta_i \geq \theta_{i^\star}]}{\max_{j\neq i^\star}\mathbb{P}_{\theta\sim\Pi_n}[\theta_j \geq \theta_{i^\star}]} \,.$$

Further upper bounding this quantity to prove (10) requires a lower bound on $\mathbb{P}_{\theta\sim\Pi_n}[\theta_i \geq \theta_{i^\star}]$ which can again be obtained using a lower bound on the BCP. In Appendix G.3 we provide a tight lower bound on $\mathbb{P}_{\theta\sim\Pi_n}[\theta_i \geq \theta_{i^\star}]$ featuring the $\mathcal{K}_{\inf}^{\pm}$ functions. It permits to prove that (10) holds with $-\log(h(n))/n =_{+\infty} \tilde{C}_{\boldsymbol{F}} + o(1)$ where $\tilde{C}_{\boldsymbol{F}} > 0$ is a problem dependent constant.

The above derivations all use the concentration property (11), which requires the sufficient exploration property (7). For our deterministic challengers, sufficient exploration is obtained by noticing that $W_n(i,j)$ can be upper and lower bounded by linear functions of the number of samples. Proving sufficient exploration is more challenging for a randomized challenger, and existing proofs were exploiting the symmetry of the Gaussian posterior. In our analysis we show that a coarse lower bound on the BCP is sufficient to obtain (11), and prove such lower bound for the Dirichlet sampler:

$$\mathbb{P}_{\theta\sim\Pi_n}[\theta_i \geq u] \geq (1-u/B)^{n+1} \quad \text{and} \quad \mathbb{P}_{\theta\sim\Pi_n}[\theta_i \leq u] \geq (u/B)^{n+1} \,.$$

These lower bounds ensure that any arm has some (small) probability of being the challenger thanks to re-sampling. Without adding $\{0, B\}$ to $\mathcal{H}_{n,i}$, those probabilities could be equal to zero.

Our analysis is easily amenable to tackle different families of distributions $\mathcal{F}$. This requires continuity and convexity properties for the corresponding $\mathcal{K}_{\inf}$ functions, an appropriate concentration result and further upper and lower bounds on the BCP of the sampler if one wish to analyze randomized algorithms. As an illustration, we show asymptotic β-optimality of the β-EB-TC, β-EB-TCI algorithms for SPEF with sub-exponential distributions, see Appendix H.

5 Experiments

We assess the empirical performance of our Top Two algorithms on the DSSAT real-world data and on Bernoulli instances in the moderate regime ($\delta = 0.01$). The stopping rule (2) is used with the threshold $c(n,\delta)$ defined in (4). As Top Two sampling rules, we present results for β-EB-TC, β-EB-TCI, β-TS-TC and β-TS-TCI with $\beta = 0.5$. Additional experiments are available in Appendix I.2: on the RS challenger whose computational cost prevent it to be evaluated with (4) and on larger sets of arms (up to $K = 1000$).

As benchmarks for the sampling rule, we use KL-LUCB with Bernoulli divergence [28] (whose theoretical guarantees extend to any distribution bounded in $[0,1]$), "fixed" sampling which is an oracle playing with proportions $w^\star(\boldsymbol{F})$ and uniform sampling. We also propose a heuristic adaptation of the DKM algorithm [13] (which is asymptotically optimal for SPEF) to tackle bounded distributions, which we denote by $\mathcal{K}_{\inf}$-DKM, and uses forced exploration instead of optimism. Inspired by the regret minimization algorithm $\mathcal{K}_{\inf}$-UCB [4], we propose its LUCB variant [27], named $\mathcal{K}_{\inf}$-LUCB. The upper/lower confidence indices are obtained by inverting of $\mathcal{K}_{\inf}^{\pm}$, i.e.

$$\begin{aligned} \forall i \neq \hat{\imath}_n, \quad U_{n+1,i} &= \max\left\{u \in [\mu_{n,i}, B] \mid N_{n,i}\mathcal{K}_{\inf}^+(F_{n,i},u) \leq c(n,\delta)\right\} \,, \\ L_{n+1,\hat{\imath}_n} &= \min\left\{u \in [0, \mu_{n,\hat{\imath}_n}] \mid N_{n,\hat{\imath}_n}\mathcal{K}_{\inf}^-(F_{n,\hat{\imath}_n},u) \leq c(n,\delta)\right\} \,. \end{aligned}$$

LUCB-based algorithms [27] use their own stopping rule, namely they stop when $L_{n+1,\hat{\imath}_n} \geq \max_{j \neq \hat{\imath}_n} U_{n+1,j}$. For Bernoulli distributions, $\mathcal{K}_{\inf}$-LUCB recovers KL-LUCB. While being asymptotically optimal for heavy-tailed distributions [3] with an adequate stopping threshold, the Track-and-Stop algorithm is computationally intractable for bounded distributions as it requires to compute $w^\star(\boldsymbol{F}_n)$ at each time n (or on a geometric grid). We hence omit it from our experiments.

Crop-management problem We benchmark our algorithms on the DSSAT simulator[2] [22]. Each arm corresponds to a choice of planting date and fixed soil conditions (details in Appendix I). To illustrate the problem's difficulty we represent an empirical estimate (independent of the runs of our algorithms) of the yield distributions in Figure 3(b). Since the gaps between means are small, the identification problem is hard. Moreover, $\mathcal{K}_{\inf}$ computations for non-parametric distributions are costlier than Bernoulli ones (see Appendix I.1), so we only present the results for 100 runs.

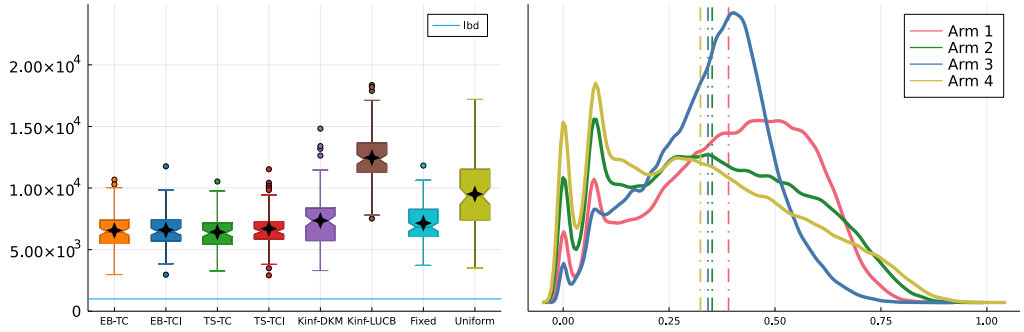

Figure 3: Empirical stopping time (a) on scaled DSSAT instances with their density and mean (b). Lower bound is $T^\star(\boldsymbol{F}) \log(1/\delta)$. "stars" equal means.

In Figure 3, β-EB-TCI, β-TS-TC and β-TS-TCI slightly outperform $\mathcal{K}_{\inf}$-DKM and the fixed (oracle) sampling rule. Moreover, $\mathcal{K}_{\inf}$-LUCB performs significantly worse than uniform sampling. Due to the small number of runs, we don't observe large outliers for β-EB-TC (see Appendix I.2). KL-LUCB performs ten times worse than $\mathcal{K}_{\inf}$-LUCB, hence we omit it from Figure 3.

Bernoulli instances Next we assess the performance on 1000 random Bernoulli instances with $K = 10$ such that $\mu_1 = 0.6$ and $\mu_i \sim \mathcal{U}([0.2, 0.5])$ for all $i \neq 1$, where we enforce that $\Delta_{\min} \geq 0.01$. We also study the instance $\mu = (0.5, 0.45, 0.45)$, in which $\Delta_{\min} = 0$, and perform 1000 runs.

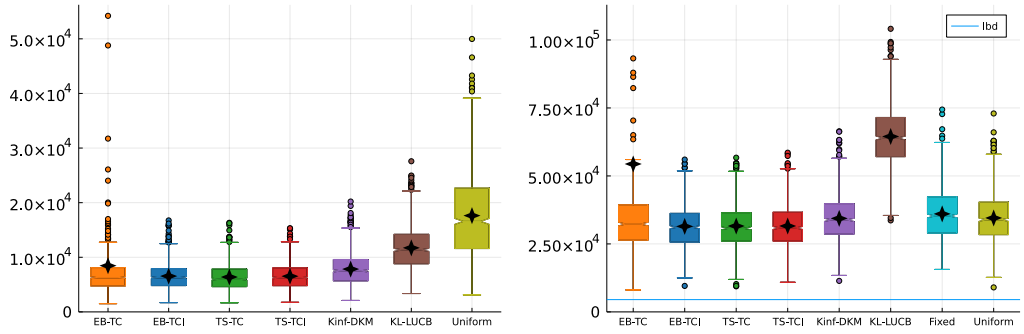

Figure 4: Empirical stopping time on Bernoulli (a) random instances with $K = 10$ and (b) instance $\mu = (0.5, 0.45, 0.45)$.

In Figure 4(a), we see that β-EB-TCI, β-TS-TC and β-TS-TCI outperform other algorithms. While this gain is slim compared to $\mathcal{K}_{\inf}$-DKM, the empirical stopping time is twice (resp. three times) as large for KL-LUCB (resp. uniform sampling). Even when $\Delta_{\min} = 0$, Figure 4(b) hints that their empirical performance might be preserved. Figure 4 confirms the lack of robustness of β-EB-TC,

[2]DSSAT is an Open-Source project maintained by the DSSAT Foundation, see https://dssat.net.

which is prone to large outliers. For the symmetric instance in Figure 4(b), uniform sampling outperforms KL-LUCB and perform on par with the "fixed" sampling.

6 Perspectives

We provided a general analysis of Top Two algorithms, including new variants using the EB leader and TCI challenger, and proved their asymptotic β-optimality on the non-parametric class of bounded distributions. On experiments on distributions coming from a real world application, several Top Two variants (in particular β-TS-TC and β-EB-TCI) proved more effective than all baselines. Furthermore, β-EB-TCI is computationally not costlier than computing the stopping rule.

As in previous work on Top Two methods our result only characterizes the asymptotic performance of the algorithms, and obtaining bounds on the sample complexity for any δ that would reflect their good empirical performance is a most pressing open question. Our work also hints at what is needed to obtain non-asymptotic guarantees: the only variant for which the empirical behavior does not reflect the asymptotic bound is β-EB-TC, which is also the most greedy variant. Algorithms using a sampler naturally explore, and the penalized version β-EB-TCI successfully corrects the shortcomings of β-EB-TC by penalizing over-sampling. Quantifying the amount of exploration required by Top Two algorithms should also allow the removal of the hypothesis $\Delta_{\min} > 0$ from Theorem 1.

Finally, Top Two algorithms are promising algorithms to tackle the setting of fixed budget identification, in which the algorithms have to stop at a given time and should then make as few mistakes as possible. As their sampling rule is anytime (i.e. independent of δ), Top Two algorithms might also have theoretical guarantees for BAI in the fixed-budget setting or even the anytime one, in which guarantees on the error probability should be given at all time.

Acknowledgments and Disclosure of Funding

Experiments presented in this paper were carried out using the Grid'5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see https://www.grid5000.fr). This work has been partially supported by the THIA ANR program "AI_PhD@Lille". The authors acknowledge the funding of the French National Research Agency under the project BOLD (ANR-19-CE23-0026-04), and the Dutch Research Council (NWO) Rubicon grant number 019.202EN.004.

References

[1] S. Agrawal and N. Goyal. Analysis of Thompson Sampling for the multi-armed bandit problem. In *Proceedings of the 25th Conference On Learning Theory*, 2012.

[2] S. Agrawal and N. Goyal. Further Optimal Regret Bounds for Thompson Sampling. In *Proceedings of the 16th Conference on Artificial Intelligence and Statistics*, 2013.

[3] S. Agrawal, S. Juneja, and P. W. Glynn. Optimal δ-correct best-arm selection for heavy-tailed distributions. In *Algorithmic Learning Theory (ALT)*, 2020.

[4] S. Agrawal, S. K. Juneja, and W. M. Koolen. Regret minimization in heavy-tailed bandits. In *Conference on Learning Theory*, pages 26–62. PMLR, 2021.

[5] S. Agrawal, W. M. Koolen, and S. Juneja. Optimal best-arm identification methods for tail-risk measures. *Advances in Neural Information Processing Systems*, 34, 2021.

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

[7] D. Baudry, R. Gautron, E. Kaufmann, and O. Maillard. Optimal thompson sampling strategies for support-aware cvar bandits. In *Proceedings of the 38th International Conference on Machine Learning*, 2021.

[8] D. Baudry, P. Saux, and O. Maillard. From optimality to robustness: Dirichlet sampling strategies in stochastic bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[9] C. Berge. *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*. Courier Corporation, 1997.

[10] R. P. Brent. *Algorithms for minimization without derivatives*. Courier Corporation, 2013.

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

[12] R. Degenne and W. M. Koolen. Pure exploration with multiple correct answers. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[13] R. Degenne, W. M. Koolen, and P. Ménard. Non-asymptotic pure exploration by solving games. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

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

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

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

[17] A. Garivier, H. Hadiji, P. Menard, and G. Stoltz. Kl-ucb-switch: optimal regret bounds for stochastic bandits from both a distribution-dependent and a distribution-free viewpoints. *arXiv preprint arXiv:1805.05071*, 2018.

[18] J. Honda and A. Takemura. An Asymptotically Optimal Bandit Algorithm for Bounded Support Models. In *Proceedings of the 23rd Conference on Learning Theory*, 2010.

[19] J. Honda and A. Takemura. An asymptotically optimal policy for finite support models in the multiarmed bandit problem. *Machine Learning*, 85(3):361–391, 2011.

[20] J. Honda and A. Takemura. Non-asymptotic analysis of a new bandit algorithm for semi-bounded rewards. *Journal of Machine Learning Research*, 16:3721–3756, 2015.

[21] L. Hong, W. Fan, and J. Luo. Review on ranking and selection: A new perspective. *Frontiers of Engineering Management*, 8:321–343, 2021.

[22] G. Hoogenboom, C. Porter, K. Boote, V. Shelia, P. Wilkens, U. Singh, J. White, S. Asseng, J. Lizaso, L. Moreno, et al. The dssat crop modeling ecosystem. *Advances in crop modelling for a sustainable agriculture*, pages 173–216, 2019.

[23] K. Jamieson, M. Malloy, R. Nowak, and S. Bubeck. lil'UCB: an Optimal Exploration Algorithm for Multi-Armed Bandits. In *Proceedings of the 27th Conference on Learning Theory*, 2014.

[24] K. G. Jamieson and R. D. Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *CISS*, pages 1–6. IEEE, 2014.

[25] M. Jourdan and R. Degenne. Choosing answers in ε-best-answer identification for linear bandits. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, 2022.

[26] M. Jourdan, M. Mutnỳ, J. Kirschner, and A. Krause. Efficient pure exploration for combinatorial bandits with semi-bandit feedback. In *Algorithmic Learning Theory (ALT)*, 2021.

[27] S. Kalyanakrishnan, A. Tewari, P. Auer, and P. Stone. PAC subset selection in stochastic multi-armed bandits. In *International Conference on Machine Learning (ICML)*, 2012.

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

[29] E. Kaufmann and W. M. Koolen. Mixture martingales revisited with applications to sequential tests and confidence intervals. *Journal of Machine Learning Research*, 22(246):1–44, 2021.

[30] E. Kaufmann, O. Cappé, and A. Garivier. On the Complexity of A/B Testing. In *Proceedings of the 27th Conference On Learning Theory*, 2014.

[31] P. Massart. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *Annals of Probability*, 18, 1990.

[32] P. Ménard and A. Garivier. A minimax and asymptotically optimal algorithm for stochastic bandits. In *International Conference on Algorithmic Learning Theory*, pages 223–237. PMLR, 2017.

[33] A. Mukherjee and A. Tajer. Sprt-based efficient best arm identification in stochastic bandits. 2022.

[34] E. Posner. Random coding strategies for minimum entropy. *IEEE Transactions on Information Theory*, 21(4):388–391, 1975.

[35] C. Qin, D. Klabjan, and D. Russo. Improving the expected improvement algorithm. In *Advances in Neural Information Processing Systems 30 (NIPS)*, 2017.

[36] C. Riou and J. Honda. Bandit algorithms based on thompson sampling for bounded reward distributions. In *Algorithmic Learning Theory (ALT)*, 2020.

[37] Y. Russac, C. Katsimerou, D. Bohle, O. Cappé, A. Garivier, and W. M. Koolen. A/b/n testing with control in the presence of subpopulations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[38] D. Russo. Simple Bayesian algorithms for best arm identification. In *Proceedings of the 29th Conference on Learning Theory (COLT)*, 2016.

[39] X. Shang, R. de Heide, E. Kaufmann, P. Ménard, and M. Valko. Fixed-confidence guarantees for bayesian best-arm identification. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

[40] R. K. Sundaram et al. *A first course in optimization theory*. Cambridge university press, 1996.

[41] W. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25:285–294, 1933.

[42] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.

[43] Z. Wang, S. Yang, and W. You. Optimality conditions and algorithms for top-k arm identification. *arXiv preprint arXiv:2205.12086*, 2022.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] In the supplementary material.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] See Appendix I.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Appendix I.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No] We are using an internal cluster. As giving more details would break anonymity, we will include them in the camera-ready version.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] See Appendix I.
   (b) Did you mention the license of the assets? [Yes] See Appendix I.
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A] No new assets.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] Open source data on agricultural production.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] Idem.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Outline

The appendices are organized as follows:

- Notation are summarized in Appendix B.
- We present a unified analysis of Top Two algorithms in Appendix C, which highlights key properties on the leader and challenger mechanisms.
- In Appendix D, we analyze several instances for the leader and the challenger mechanisms.
- In Appendix E, we show Lemma 2 and derive results from concentration on sub-Gaussian random variables.
- Appendix F gathers key properties on $\mathcal{K}_{\inf}^{\pm}$, including new ones which are required for BAI.
- In Appendix G, we show lower and upper bounds on Boundary Crossing Probability (BCP) and on $\mathbb{P}_n[\theta_i \geq \theta_j]$ for the Dirichlet sampler.
- The generalization to single-parameter exponential families is done in Appendix H.
- Implementation details and additional experiments are presented in Appendix I.

Table 1: Notation for the setting.

Notation	Type	Description
K	$\mathbb{N}$	Number of arms
B	$\mathbb{R}_+^\star$	Upper bound for bounded distributions
$\mathcal{P}(\mathbb{R})$		Probability distributions over $\mathbb{R}$
$\mathcal{F}$		Set of distributions, e.g. bounded distributions on $[0, B]$
F_i	$\mathcal{F}$	CDF of the distribution of arm $i \in [K]$
$\boldsymbol{F}$	$\mathcal{F}^K$	$\boldsymbol{F} := (F_i)_{i\in[K]}$
m	$\mathcal{F} \to \mathbb{R}$	Mean operator, $m(F) := \mathbb{E}_{X\sim F}[X]$
$\mathcal{I} \subseteq \mathbb{R}$		Interval of means $\mathcal{I} := \{m(F) \mid F \in \mathcal{F}\}$, e.g. $[0, B]$ for bounded
μ_i	$\mathring{\mathcal{I}}$	Mean of arm $i \in [K]$, i.e. $\mu_i := m(F_i)$
μ	$(\mathring{\mathcal{I}})^K$	Vector of means, $\mu := (\mu_i)_{i\in[K]}$
$i^\star$	$\mathcal{F}^K \to [K]$	Best arm operator, $i^\star(\boldsymbol{F}) \in \arg\max_{i\in[K]} \mu_i$
$T^\star(\boldsymbol{F})$	$\mathbb{R}_+^\star$	Asymptotic characteristic time
$T_\beta^\star(\boldsymbol{F})$	$\mathbb{R}_+^\star$	Asymptotic β-characteristic time
$w^\star(\boldsymbol{F})$	$\triangle_K$	Asymptotic optimal allocation
$w_\beta^\star(\boldsymbol{F})$	$\triangle_K$	Asymptotic β-optimal allocation

B Notation

We recall some commonly used notation: the set of integers $[n] := \{1, \cdots, n\}$, the complement $\overline{X}$ and interior $\mathring{X}$ of a set X, the Kullback-Leibler (KL) divergence $\mathrm{KL}(F, G)$ between two distributions F and G, the KL for Bernoulli distributions kl, the Kinf $\mathcal{K}_{\inf}^{\pm}(F, u)$ between a distribution F and a scalar u, Landau's notation o and $\mathcal{O}$ and the K-dimensional probability simplex $\triangle_K := \left\{w \in \mathbb{R}_+^K \mid w \geq 0,\ \sum_{i\in[K]} w_i = 1\right\}$, the infinity norm $\|\cdot\|_\infty$, i.e. $\|f\|_\infty = \sup_{x\in\mathbb{R}} f(x)$. For all $\mathcal{F}_n$-measurable set A, we denote by $\mathbb{P}_{|n}[A] := \mathbb{P}[A \mid \mathcal{F}_n]$ its probability. For all $\mathcal{F}_n$-measurable set A_θ depending on $\theta \sim \Pi_n$, we denote by $\mathbb{P}_n[A_\theta] := \mathbb{P}_{\theta\sim\Pi_n}[A_\theta \mid \mathcal{F}_n]$. In Table 1, we summarize problem-specific notation. Table 2 gathers notation for the algorithms.

Table 2: Notation for algorithms.

Notation	Type	Description
B_n	$[K]$	Leader at time n
C_n	$[K]$	Challenger at time n
I_n	$[K]$	Arm sampled at time n
β	$(0,1)$	Probability of sampling the leader instead of the challenger
X_{n,I_n}	$\mathcal{I}$	Sample observed at the end of time n, i.e. $X_{n,I_n} \sim F_{I_n}$
U_n		Internal randomization at time n
$\mathcal{F}_n$		History at time n, i.e. $\mathcal{F}_n := \sigma(U_1, I_1, X_{1,I_1}, \cdots, I_n, X_{n,I_n}, U_{n+1})$
$\hat{\imath}_n$	$[K]$	Arm recommended after time n, i.e. $\hat{\imath}_n \in \arg\max_{i\in[K]} \mu_{n,i}$
τ_δ	$\mathbb{N}$	Sample complexity (stopping time of the algorithm)
$\hat{\imath}$	$[K]$	Arm recommended by the algorithm
$c(n,\delta)$	$\mathbb{N}\times(0,1)\to\mathbb{R}_+^\star$	Stopping threshold function
$N_{n,i}$	$\mathbb{N}$	Number of pulls of arm i at time n, i.e. $N_{n,i} := \sum_{t\in[n]} \mathbb{1}(I_t = i)$
$F_{n,i}$	$\mathcal{F}$	Empirical distribution, i.e. $F_{n,i} := \frac{1}{N_{n,i}} \sum_{t\in[n]} \delta_{X_{t,I_t}} \mathbb{1}(I_t = i)$
$\mu_{n,i}$	$\mathcal{I}$	Empirical mean, i.e. $\mu_{n,i} := m(F_{n,i})$
$W_n(i,j)$	$\mathbb{R}_+$	Empirical transportation between arms i and j, defined in (1)
Π_n		Sampler at time n, e.g. Dirichlet sampler for bounded
θ	$\mathcal{I}^K$	Observation from the sampler, i.e. $\theta \sim \Pi_n$
$a_{n,i}$	$[0,1]$	$a_{n,i} := \mathbb{P}_{n-1}[i \in \arg\max_{j\in[K]} \theta_j]$
$\psi_{n,i}$	$[0,1]$	Probability of sampling arm i at time n: $\psi_{n,i} := \mathbb{P}_{\vert(n-1)}[I_n = i]$
$\Psi_{n,i}$	$\mathbb{R}_+^\star$	Cumulative sampling probability: $\Psi_{n,i} := \sum_{t\in[n]} \psi_{t,i}$
$\widehat{B}_n$	$[K]$	Effective leader at time n
$\widehat{C}_n$	$[K]$	Effective challenger at time n

C Unified analysis of Top Two algorithms

In this section, we present a unified analysis of Top Two algorithms (Appendix C.2). The analysis is split into three parts that will highlight how to explore (Appendix C.3), how to converge towards the β-optimal allocation (Appendix C.4) and finally proving asymptotic optimality (Appendix C.5).

In this section, we identify the required properties that the leader and the challenger should satisfy. In Appendix D, we prove that those properties are verified by the EB and TS leader and the TC, TCI and RS challenger for bounded distributions. In Appendix H, we discuss the proofs of those properties for single-parameter exponential families.

The general proof strategy follows that first proposed by [35] for the TTEI algorithm and later also used by [39] for TTTS and T3C. However we contribute with a new, modular proof structure which furthermore get rids of several Gaussian-specific arguments.

Striving to tackle simultaneously bounded distributions (Appendix F) and single-parameter exponential families (Appendix H), we need to introduce some notation to unify both formulation (Appendix C.1).

C.1 Generic asymptotic β-optimality

In the case of single-parameter exponential families, a distribution $F \in \mathcal{F}$ is characterized by its mean parameter $m(F) \in \mathbb{R}$. Therefore, convergence/continuity results can be formulated directly with the $|\cdot|$ norm.

For general bounded distributions, it is not possible to characterize them by using a scalar (or vector) parameter. Therefore, we need to consider the space of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with the topology of weak convergence of measures, denoted by $\mathcal{P}(\mathbb{R})$. Recall that weak convergence is equivalent to the convergence of the respective cdfs for the infinity norm $\|\cdot\|_\infty$, defined as $\|f\|_\infty = \sup_{x\in\mathbb{R}} f(x)$.

To unify both approach, we introduce an operator $\mathcal{T}$ from $\mathcal{F}$ to a topological space, which associates the distribution $F \in \mathcal{F}$ with a transformation $\mathcal{T}(F)$ that characterizes it. When considering single-parameter exponential families, $\mathcal{T}$ coincides with the mean operator $m : F \mapsto \mathbb{E}_F[X]$. For bounded distributions, $\mathcal{T}$ will be the identity. Moreover, we define $\mathcal{T}(\boldsymbol{F}) := (\mathcal{T}(F_i))_{i\in[K]}$ for all $\boldsymbol{F} \in \mathcal{F}^K$ and $\mathcal{T}(\mathcal{F}^K) := \left\{\mathcal{T}(\boldsymbol{F}) \mid \boldsymbol{F} \in \mathcal{F}^K\right\}$.

Let $\mathcal{I} \subseteq \mathbb{R}$ be the interval of means $\mathcal{I} := \{m(F) \mid F \in \mathcal{F}\}$. The functions $(F, u) \mapsto \mathcal{K}_{\inf}^{\pm}(F, u)$ are defined on $\mathcal{F} \times \mathcal{I}$. Two archetypal examples for $\mathcal{I}$ are $\mathcal{I} = [0, B]$ (bounded, Bernoulli, Beta, etc) and $\mathcal{I} = \mathbb{R}$ (Gaussian, etc).

Condition on the means We detail the assumptions on the means of $\boldsymbol{F} \in \mathcal{F}^K$ under which Top Two algorithms can be studied.

Assumption 1. *There is a unique best arm denoted by $i^\star(\boldsymbol{F})$ and the means are away from the boundary, i.e. $\mu_i := m(F_i) \in \mathring{\mathcal{I}}$ for all $i \in [K]$.*

The first part of Assumption 1 is a standard assumption in BAI problem, where a unique best arm has to be identified. Indeed, in the presence of two best arms existing algorithms would only stop with a small probability as they would try to statistically distinguish two identical distributions. In order to circumvent this hurdle, one can relax the BAI problem in which the goal is to find one arm which is ε-close to the best arm, for some parameter $\varepsilon > 0$. When it comes to asymptotic optimality, this setting is known to be much more complex than standard BAI [12, 25].

The second part of Assumption 1 is also standard, as we require the distribution to have mass away from the boundary. When $\mathcal{I} = \mathbb{R}$, the condition $\mu_i \in \mathring{\mathcal{I}} = \mathbb{R}$ is always satisfied since we consider finite means. When $\mathcal{I} = [0, B]$, the assumption $\mu_i \in (0, B)$ is often made when studying Bernoulli or bounded distributions. For the bounded setting, this excludes δ_B and δ_0, where δ_x denotes the Dirac distributions in x. Since those requirements are mild, we consider that Assumption 1 holds in the following, without mentioning it further.

Assumption 2. *All the arms have distinct means, i.e.* $\Delta_{\min}(\boldsymbol{F}) := \min_{i\neq j} |\mu_{F_i} - \mu_{F_j}| > 0$.

Requiring $\Delta_{\min} > 0$, is stronger than the unique best arm condition from Assumption 1. While this is a unusual requirement to study BAI problem, previous works on Top Two algorithms [38, 35, 39] also supposed that $\Delta_{\min} > 0$. Our unified analysis of Top Two algorithms highlights the role of this condition in the analysis. It is solely used to prove sufficient exploration (Appendix C.3). Provided enough exploration, the convergence towards the β-optimal allocation (Appendix C.4) only relies on Assumption 1.

As we aim to shed light on the role of Assumption 2 in the analysis, we will explicitly highlight where it is used in the proof of sufficient exploration. The empirical performance of Top Two algorithms on instances where $\Delta_{\min} = 0$ is assessed in Appendix I.2.2. In Appendix D.3, we discuss possible relaxations of this Assumption for some leaders and challengers.

Transportation costs and β-optimal allocation With the notation introduced above, the transportation cost between arms $(i, j) \in [K]^2$ for an allocation $w \in \triangle_K$ rewrites as

$$C_{i,j}(\mathcal{T}(\boldsymbol{F}), w) := \inf_{u\in\mathcal{I}} \left\{w_i \mathcal{K}_{\inf}^{-}(\mathcal{T}(F_i), u) + w_j \mathcal{K}_{\inf}^{+}(\mathcal{T}(F_j), u)\right\} , \tag{12}$$

and the empirical transportation cost rewrites as

$$\frac{1}{n} W_n(i, j) = C_{i,j}\left(\mathcal{T}(\boldsymbol{F}_n), \frac{N_n}{n}\right) . \tag{13}$$

Similarly, the β-characteristic time and β-optimal allocation

$$T_\beta^\star(\boldsymbol{F})^{-1} := \max_{w\in\triangle_K : w_{i^\star(\boldsymbol{F})} = \beta} \min_{j\neq i^\star(\boldsymbol{F})} C_{i^\star(\boldsymbol{F}),j}(\mathcal{T}(\boldsymbol{F}), w) ,$$
$$w_\beta^\star(\boldsymbol{F}) := \operatorname*{arg\,max}_{w\in\triangle_K : w_{i^\star(\boldsymbol{F})} = \beta} \min_{j\neq i^\star(\boldsymbol{F})} C_{i^\star(\boldsymbol{F}),j}(\mathcal{T}(\boldsymbol{F}), w) .$$

Property 1 requires $w_\beta^\star(\boldsymbol{F})$ to be a singleton. For single-parameter exponential families, it is well known that Property 1 holds [38]. For bounded distribution, we showed it in Lemma 61. As Property 1 holds for the distributions of interest, we won't mention it further.

Property 1. *For all $\boldsymbol{F} \in \mathcal{F}^K$ satisfying Assumption 1, $w^\star_\beta(\boldsymbol{F})$ is a singleton.*

To ease the notation, in the sequel we denote the unique β-optimal allocation by $w^\beta = (w^\beta_i)_{i\in[K]}$. For an algorithm to be asymptotically β-optimal, its empirical allocation $(N_{n,i}/n)_{i\in[K]}$ should converge towards w^β.

C.2 Generic Top Two algorithms

The σ-algebra $\mathcal{F}_n := \sigma(U_1, I_1, X_{1,I_1}, \cdots, I_n, X_{n,I_n}, U_{n+1})$, called history, encompasses all the information available to the agent at time n and the internal randomization denoted by $(U_t)_{t\in[n+1]}$, which is independent of everything else. For all $\mathcal{F}_n$-measurable set A, we denote by $\mathbb{P}_{|n}[A] := \mathbb{P}[A \mid \mathcal{F}_n]$ its probability. As most BAI algorithms, out methods pull each arm once for the initialization. At time $n+1$, a Top Two sampling rule outputs an arm I_{n+1} which is $\mathcal{F}_n$-measurable. The choice I_{n+1} is defined by two mechanisms: the choice of a leader $B_{n+1} \in [K]$ which is $\mathcal{F}_n$-measurable and the choice of the challenger $C_{n+1} \in [K] \setminus \{B_{n+1}\}$ is $\mathcal{F}_n$-measurable.

Following the proof strategy first introduced by [35], our goal is to upper bound the expectation of the *convergence time*. For $\varepsilon > 0$, the random variable T^ε_β (already defined in (5)) quantifies the number of samples required for the empirical allocations $\frac{N_n}{n}$ to be ε-close to w^β:

$$T^\varepsilon_\beta := \inf\left\{ T \geq 1 \mid \forall n \geq T,\ \left\| \frac{N_n}{n} - w^\beta \right\|_\infty \leq \varepsilon \right\} .$$

To this end, we first leverage generic properties of Top Two algorithms to understand how the average probability to select an arm can converge to w^β. We denote by $\psi_{n,i} := \mathbb{P}_{|(n-1)}[I_n = i]$, the probability that an arm is sampled at round n, and by $\Psi_{n,i} := \sum_{t\in[n]} \psi_{t,i}$ its cumulative version. For a Top Two sampling rule, we have

$$\psi_{n,i} = \beta \mathbb{P}_{|(n-1)}[B_n = i] + (1-\beta) \sum_{j\neq i} \mathbb{P}_{|(n-1)}[B_n = j]\mathbb{P}_{|(n-1)}[C_n = i | B_n = j] . \tag{14}$$

Mean probability of being sampled Even before specifying the leader and the challenger mechanisms, we can study the general properties of Top Two algorithms, given in Lemmas 3 and 4. While being obtained by simple algebra, they highlight quite naturally the respective roles of the leader and the challenger mechanisms in order to achieve asymptotic β-optimality.

Lemma 3 upper bounds the deviation between the fixed allocation β and the mean probability of sampling the optimal arm $\frac{\Psi_{n,i^\star(\boldsymbol{F})}}{n}$.

Lemma 3. *For all $M \in \mathbb{N}^*$,*

$$\left| \frac{\Psi_{n,i^\star(\boldsymbol{F})}}{n} - \beta \right| \leq \frac{M-1}{n} + \frac{1}{n} \sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[B_t \neq i^\star(\boldsymbol{F})] . \tag{15}$$

Proof. Let $i^\star = i^\star(\boldsymbol{F})$ and $M \in \mathbb{N}^*$. Summing (14) for $i^\star$ and using $\mathbb{P}_{|(t-1)}[B_t = i^\star] = 1 - \mathbb{P}_{|(t-1)}[B_t \neq i^\star]$ yields

$$\frac{\Psi_{n,i^\star}}{n} - \beta = \frac{1-\beta}{n} \sum_{t\in[n]} \sum_{j\neq i^\star} \mathbb{P}_{|(t-1)}[B_t = j]\mathbb{P}_{|(t-1)}[C_t = i^\star | B_t = j] - \frac{\beta}{n} \sum_{t\in[n]} \mathbb{P}_{|(t-1)}[B_t \neq i^\star] .$$

Dropping the second negative term, splitting the sum into two and using that $\mathbb{P}_{|(t-1)}[C_t = i^\star | B_t = j] \leq 1$ and $\sum_{j\neq i^\star} \mathbb{P}_{|(t-1)}[B_t = j] = \mathbb{P}_{|(t-1)}[B_t \neq i^\star]$, we obtain the following upper bound

$$\frac{\Psi_{n,i^\star}}{n} - \beta \leq (1-\beta) \left(\frac{M-1}{n} + \frac{1}{n} \sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[B_t \neq i^\star] \right) .$$

Dropping the first positive term and splitting the sum into two, we obtain the following lower bound

$$\frac{\Psi_{n,i^\star}}{n} - \beta \geq -\beta \left(\frac{M-1}{n} + \frac{1}{n} \sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[B_t \neq i^\star] \right) .$$

Combining the upper and the lower bound and using that $\max\{\beta, 1-\beta\} \leq 1$ yields the result. □

Since an asymptotically β-optimal algorithm should allocate a proportion β of its samples to the best arm, the right-hand side of (15) should vanish. Cesaro's theorem yields the result when

$$\lim_{t\to+\infty} \mathbb{P}_{|(t-1)}[B_t \neq i^\star(\boldsymbol{F})] = 0\,.$$

This means that a *good* leader should asymptotically identify $i^\star(\boldsymbol{F})$. As we will see, the convergence almost surely won't be enough to obtain an upper bound on $\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]$. To that end, we will need to specify the rate of convergence.

Lemma 4 upper bounds the probability of sampling an arm different from the optimal one.

Lemma 4. *For all $M \in \mathbb{N}^*$ and $i \neq i^\star(\boldsymbol{F})$,*

$$\frac{\Psi_{n,i}}{n} \leq \frac{M-1}{n} + \frac{1}{n}\sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[B_t \neq i^\star(\boldsymbol{F})] + \frac{1}{n}\sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[C_t = i | B_t = i^\star(\boldsymbol{F})]\,. \tag{16}$$

Proof. Let $i^\star = i^\star(\boldsymbol{F})$. Using (14) for $i \neq i^\star$ and $\mathbb{P}[C_n = i | B_n = j] \leq 1$, we have

$$\begin{aligned}
\psi_{n,i} &\leq \beta \mathbb{P}_{|(n-1)}[B_n = i] + (1-\beta) \sum_{j \notin \{i, i^\star\}} \mathbb{P}_{|(n-1)}[B_n = j] \\
&\quad + (1-\beta)\mathbb{P}_{|(n-1)}[B_n = i^\star]\mathbb{P}_{|(n-1)}[C_n = i | B_t = i^\star] \\
&\leq \max\{\beta, 1-\beta\}\mathbb{P}_{|(n-1)}[B_n \neq i^\star] + (1-\beta)\mathbb{P}_{|(n-1)}[B_n = i^\star]\mathbb{P}_{|(n-1)}[C_n = i | B_n = i^\star] \\
&\leq \mathbb{P}_{|(n-1)}[B_n \neq i^\star] + \mathbb{P}_{|(n-1)}[C_n = i | B_n = i^\star]
\end{aligned}$$

where we used that $\sum_{j \neq i^\star} \mathbb{P}_{|(n-1)}[B_n = j] = \mathbb{P}_{|(n-1)}[B_n \neq i^\star]$, $\max\{\beta, 1-\beta\} \leq 1$ and $\mathbb{P}_{|(n-1)}[B_n = i^\star] \leq 1$. Summing over $t \in [n]$ after splitting the sum into two, we obtain

$$\frac{\Psi_{n,i}}{n} \leq \frac{M-1}{n} + \frac{1}{n}\sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[B_t \neq i^\star] + \frac{1}{n}\sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[C_t = i | B_t = i^\star]$$

□

Given a good leader, the first two terms on the right-hand side of (16) vanish. An asymptotically β-optimal algorithm should allocate a proportion w_i^β of its samples to the sub-optimal arms. As $\sum_{i\in[K]} \frac{\Psi_{n,i}}{n} = 1$, Cesaro's theorem yields the result when

$$\lim_{t\to+\infty} \mathbb{P}_{|(t-1)}[C_t = i | B_t = i^\star(\boldsymbol{F})] = w_i^\beta\,.$$

Given a good leader, a *good* challenger should asymptotically have a probability w_i^β of pulling a sub-optimal arm i. Likewise, a rate of convergence will be necessary to upper bound $\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]$.

From $\Psi_{n,i}$ to $N_{n,i}$ The above results feature $\frac{\Psi_{n,i}}{n}$, which is the mean probability of an arm to be sampled. Thanks to Lemma 5, it can be linked to the empirical allocation $\frac{N_{n,i}}{n}$. Its proof, deferred to Appendix E.2, is a direct consequence of concentration inequalities for sub-Gaussian random variables. A similar result was already derived in the work of [35], who first introduced this style of W-based concentration results, which we will use also in Appendix D.

Lemma 5. *There exists a sub-Gaussian random variable W_1 such that for all $(n, i) \in \mathbb{N} \times [K]$*

$$|N_{n,i} - \Psi_{n,i}| \leq W_1\sqrt{(n+1)\log(e+n)} \quad a.s.\,. \tag{17}$$

In particular, $\mathbb{E}\left[e^{\lambda W_1}\right] < +\infty$ for all $\lambda > 0$.

In the following, we take W_1 as in Lemma 5. Since $\mathbb{E}\left[e^{\lambda W_1}\right] < +\infty$ for all $\lambda > 0$, we have in particular that for all $N = Poly(W_1)$, $\mathbb{E}[N] < +\infty$.

Convergence towards β-optimal allocation In Appendix C.5, we show that if $\mathbb{E}_{\boldsymbol{F}}\left[T_\beta^\varepsilon\right] < +\infty$ for all ε small enough, then

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log\left(\frac{1}{\delta}\right)} \leq T_\beta^\star(\boldsymbol{F}) \,.$$

The proof of $\mathbb{E}_{\boldsymbol{F}}\left[T_\beta^\varepsilon\right] < +\infty$ can be naturally split into two distinct parts. In Appendix C.3, under Assumption 2, we show that any Top Two algorithm ensures sufficient exploration provided its leader and challenger each satisfy one property. In Appendix C.4, given sufficient exploration, the convergence of the empirical allocation towards the β-optimal one is proven for any Top Two algorithm provided its leader and challenger pairs each satisfy one property.

C.3 How to explore

In this section, we identify one property for the leader (Property 2) and one property for the challenger (Property 3) under which we prove that the corresponding Top Two algorithm ensures sufficient exploration, when Assumption 2 holds. More precisely, we prove in Lemma 7 that

$$\exists N_1 \text{s.t. } \mathbb{E}_{\boldsymbol{F}}[N_1] < +\infty : \quad \forall n \geq N_1, \min_{i \in [K]} N_{n,i} \geq \sqrt{n/K}$$

We discuss other algorithmic choices that could ensure sufficient exploration without Assumption 2 in Appendix D.3. This section borrows several elements from existing proofs of sufficient exploration for Top Two algorithms in Gaussian bandits [35, 39] but we managed to simplify the argument in order to put forward the key properties needed from a leader and a challenger. First, our generic analysis needs to define an appropriate notion of effective leader and challenger.

Effective leader and challenger For an algorithm to alleviate under-sampling some arms, it should have a strictly positive probability of sampling them. In Top Two algorithms, the choice of the arm to pull I_n is defined by the leader B_n and the challenger C_n. Due to possible randomization, it is not trivial to manipulate B_n and C_n. Therefore, we define the *effective* leader $\widehat{B}_n$ and the *effective* challenger $\widehat{C}_n$ as the arms maximizing the respective probability of being sampled:

$$\widehat{B}_n \in \operatorname*{arg\,max}_{i \in [K]} \mathbb{P}_{|(n-1)}[B_n = i] \quad \text{and} \quad \widehat{C}_n \in \operatorname*{arg\,max}_{i \neq \widehat{B}_n} \mathbb{P}_{|(n-1)}[C_n = i | B_n = \widehat{B}_n] \,, \tag{18}$$

where $\widehat{C}_n$ is defined conditioned on the effective leader $\widehat{B}_n$. We assume that ties are broken uniformly at random. Note that they are fully determined by the leader and challenger mechanisms.

Lemma 6 gives a strictly positive lower bound on the probability of sampling $\widehat{B}_n$ and $\widehat{C}_n$.

Lemma 6. *Let* $\psi_{\min} := \frac{1}{K} \min\{\beta, \frac{1-\beta}{K-1}\}$. *Then,* $\psi_{n,i} \geq \psi_{\min}$ *for all* $i \in \{\widehat{B}_n, \widehat{C}_n\}$.

Proof. Since $\sum_{i \in [K]} \mathbb{P}_{|(n-1)}[B_n = i] = 1$ and $\widehat{B}_n \in \arg\max_{i \in [K]} \mathbb{P}_{|(n-1)}[B_n = i]$, we have

$$\psi_{n,\widehat{B}_n} \geq \beta \mathbb{P}_{|(n-1)}[B_n = \widehat{B}_n] = \frac{\beta}{K} \geq \psi_{\min} \,.$$

Similarly, $\sum_{i \in [K]} \mathbb{P}_{|(n-1)}[C_n = i | B_n = \widehat{B}_n] = 1$ and $\widehat{C}_n \in \arg\max_{i \neq \widehat{B}_n} \mathbb{P}_{|(n-1)}[C_n = i | B_n = \widehat{B}_n]$ yields that $\mathbb{P}_{|(n-1)}[C_n = \widehat{C}_n | B_n = \widehat{B}_n] \geq \frac{1}{K-1}$. Therefore,

$$\psi_{n,\widehat{C}_n} \geq (1-\beta) \mathbb{P}_{|(n-1)}[B_n = \widehat{B}_n] \mathbb{P}_{|(n-1)}[C_n = \widehat{C}_n | B_n = \widehat{B}_n] \geq \frac{1-\beta}{K(K-1)} \geq \psi_{\min} \,.$$

□

In light of Lemma 6, the sufficient exploration can be proven if we show that either $\widehat{B}_n$ or $\widehat{C}_n$ is among the under-sampled arms if some still exists. Before formalizing the properties required by the leader and challenger pair to ensure sufficient exploration, we introduce the relevant notation.

Given an arbitrary threshold $L \in \mathbb{R}_+^*$, we define the sampled enough set and its arms with highest mean (when not empty) as

$$S_n^L := \{i \in [K] \mid N_{n,i} \geq L\} \quad \text{and} \quad \mathcal{I}_n^\star := \arg\max_{i \in S_n^L} \mu_i \,. \tag{19}$$

Assumption 2 ensures that $\mathcal{I}_n^\star$ is unique. To highlight why it is necessary, we view $\mathcal{I}_n^\star$ as a set with potentially multiple values, and derive properties without assuming that $\mathcal{I}_n^\star$ is a singleton. At time n, S_n^L can only be non-empty for $L \leq n$, hence it depends explicitly on n.

To prove sufficient exploration, we aim at finding a threshold $L(n)$ such that $S_n^{L(n)} = [K]$ for $n \geq \tilde{N}_0$ where $\mathbb{E}_{\boldsymbol{F}}[\tilde{N}_0] < +\infty$. We proceed by contradiction. The idea is to show that if some arms are still highly under-sampled, then either $\widehat{B}_n$ or $\widehat{C}_n$ will be mildly under-sampled. Since they have a strictly positive probability of being sampled (Lemma 6), this would yield a contradiction by the pigeonhole principle. We define the highly and the mildly under-sampled sets

$$U_n^L := \{i \in [K] \mid N_{n,i} < \sqrt{L}\} \quad \text{and} \quad V_n^L := \{i \in [K] \mid N_{n,i} < L^{3/4}\} \,. \tag{20}$$

The choice of $\sqrt{L}$ and $L^{3/4}$ is arbitrary and we could consider instead L^{α_1} and L^{α_2} with $0 < \alpha_1 < \alpha_2 < 1$. Note that $U_n^L = \overline{S_n^{\sqrt{L}}}$ and $V_n^L = \overline{S_n^{L^{3/4}}}$ where S_n^L is the set defined in (19). We are now ready to state the properties that the leader and the challenger should satisfy in order to show sufficient exploration under Assumption 2.

Exploring with leader and challenger pair We describe the properties that a good leader/challenger should have to ensure sufficient exploration. In order for the challenger to explore, a *good* leader should first identify the best arm among the arms that are sampled enough (Property 2). Then, given a good leader, a *good* challenger should enforce exploration on the arms that are not sampled enough yet when the leader doesn't do it already (Property 3).

Property 2 states that if $\widehat{B}_{n+1}$ is sampled enough, then $\widehat{B}_{n+1}$ is an arm with highest mean among the sampled enough arms.

Property 2. *There exists L_0 with $\mathbb{E}_{\boldsymbol{F}}[(L_0)^\alpha] < +\infty$ for all $\alpha > 0$ such that if $L \geq L_0$, for all n such that $S_n^L \neq \emptyset$, $\widehat{B}_{n+1} \in S_n^L$ implies $\widehat{B}_{n+1} \in \mathcal{I}_n^\star$.*

Property 2 holds for the EB leader (Lemma 17) and the TS leader (Lemma 26).

Property 3 states that if some arms are still highly under-sampled, i.e. $U_n^L \neq \emptyset$, then having sampled $\widehat{B}_{n+1}$ enough implies that $\widehat{C}_{n+1}$ is mildly under-sampled or has highest true mean among the sampled enough arms.

Property 3. *Let B_{n+1} be a leader satisfying Property 2 and C_n the associated challenger. Let $\mathcal{J}_n^\star = \arg\max_{i \in \overline{V_n^L}} \mu_i$. There exists L_1 with $\mathbb{E}_{\boldsymbol{F}}[L_1] < +\infty$ such that if $L \geq L_1$, for all n such that $U_n^L \neq \emptyset$, $\widehat{B}_{n+1} \notin V_n^L$ implies $\widehat{C}_{n+1} \in V_n^L \cup \left(\mathcal{J}_n^\star \setminus \left\{\widehat{B}_{n+1}\right\}\right)$.*

Property 3 holds for the TC challenger (Lemma 19), the TCI challenger (Lemma 21) and the RS challenger (Lemma 28).

Provided Assumption 2 holds, Lemma 7 shows that sufficient exploration is achieved for any Top Two algorithm satisfying Properties 2 and 3.

Lemma 7. *Assume Assumption 2 holds. Under a Top Two algorithm whose leader B_{n+1} and challenger C_{n+1} satisfy Properties 2 and 3, there exist N_0 with $\mathbb{E}_{\boldsymbol{F}}[N_0] < +\infty$ such that for all $n \geq N_0$ and all $i \in [K]$, $N_{n,i} \geq \sqrt{\frac{n}{K}}$.*

Proof of Lemma 7 When Assumption 2 holds, combining Properties 2 and 3 and Lemma 6 yields Lemma 8.

Lemma 8. *Assume Assumption 2 holds. Under a Top Two algorithm whose leader B_{n+1} and challenger C_{n+1} satisfy Properties 2 and 3, there exists L_2 with $\mathbb{E}_{\boldsymbol{F}}[L_2] < +\infty$ such that if $L \geq L_2$, for all n, $U_n^L \neq \emptyset$ implies that there exists $J_{n+1} \in V_n^L$ such that $\psi_{n+1,J_{n+1}} \geq \psi_{\min}$.*

Proof. Let $\mathcal{J}_n^\star = \arg\max_{i \in \overline{V_n^L}} \mu_i$. Under Assumption 2, we know that $|\mathcal{J}_n^\star| = 1$. Let L_0 as in Property 2. If $L \geq L_0^{4/3}$, for all n, $\widehat{B}_{n+1} \in \overline{V_n^L}$ implies $\mathcal{J}_n^\star = \{\widehat{B}_{n+1}\}$. Let L_1 as in Property 3.

Therefore, we have if $L \geq L_2 := \max\{L_1, L_0^{4/3}\}$, for all n such that $U_n^L \neq \emptyset$, $\widehat{B}_{n+1} \notin V_n^L$ implies $\widehat{C}_{n+1} \in V_n^L$. By Lemma 6, we know that $\psi_{n+1,i} \geq \psi_{\min}$ for all $i \in \{\widehat{B}_{n+1}, \widehat{C}_{n+1}\}$. Since $\mathbb{E}_{\boldsymbol{F}}[L_2] \leq \mathbb{E}_{\boldsymbol{F}}[L_1] + \mathbb{E}_{\boldsymbol{F}}[L_0^{4/3}] < +\infty$, this concludes the proof. □

Using concentration on $\|T_n - \Psi_n\|_\infty$ (Lemma 5) and the pigeonhole principle yield a contradiction for a large enough L. Therefore, the set of highly under-sampled arms is empty (Lemma 9). This technical result was proven in [39] for TTTS (Lemma 11) and T3C (Lemma 18). For the sake of completeness, we include the proof.

Lemma 9. *Assume Assumption 2 holds. Under a Top Two algorithm whose leader B_{n+1} and challenger C_{n+1} satisfy Properties 2 and 3, there exists L_3 with $\mathbb{E}_{\boldsymbol{F}}[L_3] < +\infty$ such that for all $L \geq L_3, U^L_{\lfloor KL \rfloor}$ is empty.*

Proof. Assume Assumption 2 holds and we are given a Top Two algorithm whose leader B_{n+1} and challenger C_{n+1} satisfy Properties 2 and 3. Let L_2 as in Lemma 8, with $\mathbb{E}_{\boldsymbol{F}}[L_2] < +\infty$. We proceed by contradiction, and we assume that $U^L_{\lfloor KL \rfloor}$ is not empty. Then for any $1 \leq \ell \leq \lfloor KL \rfloor$, U_ℓ^L and V_ℓ^L are non empty as well. There exists a deterministic L_4 such that for all $L \geq L_4$, $\lfloor L \rfloor \geq KL^{3/4}$. In particular, $\mathbb{E}_{\boldsymbol{F}}[L_4] = L_4 < +\infty$. In the following, we consider $L \geq \max\{L_2, L_4\}$.

Using the pigeonhole principle, there exists some $i \in [K]$ such that $N_{\lfloor L \rfloor, i} \geq L^{3/4}$. Thus, we have $\left|V^L_{\lfloor L \rfloor}\right| \leq K - 1$. Next, we prove $\left|V^L_{\lfloor 2L \rfloor}\right| \leq K - 2$. Otherwise, since U_ℓ^L is non-empty for any $\lfloor L \rfloor + 1 \leq \ell \leq \lfloor 2L \rfloor$, thus by Lemma 8, there exists $J_{\ell+1} \in V_\ell^L$ such that $\psi_{\ell+1, J_{\ell+1}} \geq \psi_{\min}$. Since $V_\ell^L \subset V^L_{\lfloor L \rfloor}$, we have

$$\sum_{i \in V_\ell^L} \psi_{\ell+1,i} \geq \psi_{\min} \quad \text{and} \quad \sum_{i \in V^L_{\lfloor L \rfloor}} \psi_{\ell+1,i} \geq \psi_{\min} \,.$$

Therefore,

$$\sum_{i \in V^L_{\lfloor L \rfloor}} \left(\Psi_{\lfloor 2L \rfloor + 1, i} - \Psi_{\lfloor L \rfloor + 1, i}\right) = \sum_{\ell = \lfloor L \rfloor + 1}^{\lfloor 2L \rfloor} \sum_{i \in V^L_{\lfloor L \rfloor}} \psi_{\ell+1,i} \geq \psi_{\min} \lfloor L \rfloor$$

Then, using Lemma 5, there exists $L_5 = Poly\,(W_1)$ such that for all $L \geq \max\{L_2, L_4, L_5\}$, we have

$$\begin{aligned}
&\sum_{i \in V^L_{\lfloor L \rfloor}} \left(N_{\lfloor 2L \rfloor + 1, i} - N_{\lfloor L \rfloor + 1, i}\right) \\
&\geq \sum_{i \in V^L_{\lfloor L \rfloor}} \left(\Psi_{\lfloor 2L \rfloor + 1, i} - \Psi_{\lfloor L \rfloor + 1, i} - 2W_1 \sqrt{(\lfloor 2L \rfloor + 1) \log\left(e + \lfloor 2L \rfloor + 1\right)}\right) \\
&\geq \psi_{\min} \lfloor L \rfloor - 2KW_1 \sqrt{(\lfloor 2L \rfloor + 1) \log\left(e + \lfloor 2L \rfloor + 1\right)} \,.
\end{aligned}$$

Then, there exists $L_3 = Poly\,(W_1)$ such that for all $L \geq L_3 := \max\{L_2, L_4, L_5, L_6\}$,

$$\sum_{i \in V^L_{\lfloor L \rfloor}} \left(N_{\lfloor 2L \rfloor + 1, i} - N_{\lfloor L \rfloor + 1, i}\right) \geq KL^{3/4} \,,$$

which implies that we have one arm in $V^L_{\lfloor L \rfloor}$ that is pulled at least $L^{3/4}$ times between $\lfloor L \rfloor + 1$ and $\lfloor 2L \rfloor$, thus $\left|V^L_{\lfloor 2L \rfloor}\right| \leq K - 2$.

By induction, for any $1 \leq k \leq K$, we have $\left|V^L_{\lfloor kL \rfloor}\right| \leq K - k$, and finally $U^L_{\lfloor KL \rfloor} = \emptyset$ for all $L \geq L_3$. Since $\mathbb{E}[e^{\lambda W_1}] < +\infty$ for all $\lambda > 0$, we have in particular that $\mathbb{E}_{\boldsymbol{F}}[Poly(W_1)] < +\infty$. Since

$$\mathbb{E}_{\boldsymbol{F}}[L_3] \leq \sum_{i \in \{2,4,5,6\}} \mathbb{E}_{\boldsymbol{F}}[L_i] < +\infty \,,$$

this concludes the proof. □

Let L_3 as Lemma 9. Defining $N_0 = KL_3$, we have $\mathbb{E}_{\boldsymbol{F}}[N_0] = K\mathbb{E}_{\boldsymbol{F}}[L_3] < +\infty$. For all $n \geq N_0$, we let $L = \frac{n}{K}$, then by Lemma 9, we have $U^L_{\lfloor KL \rfloor} = U_n^{n/K}$ is empty, which concludes the proof of Lemma 7.

□

C.4 How to converge

In this section, we identify one property for the leader (Property 5) and one property for the challenger (Property 6) under which we prove that the convergence time of the corresponding Top Two algorithm has finite expectation expectation (Lemma 10), provided sufficient exploration occurs. Sufficient exploration is formalized by Property 4.

Property 4. *There exist N_1 with $\mathbb{E}_{\boldsymbol{F}}[N_1] < +\infty$ such that for all $n \geq N_1$ and all $i \in [K]$, $N_{n,i} \geq \sqrt{\frac{n}{K}}$.*

For a Top Two algorithm whose leader B_n and challenger C_n satisfy Properties 2 and 3, Lemma 7 shows that Property 4 holds provided that Assumption 2 holds. While Assumption 2 allows to ensure sufficient exploration, other algorithmic choices could ensure this property without having the constraint that all means are distinct. We discuss algorithmic fixes in Appendix D.3. As mentioned above the dependency $\sqrt{n}$ is arbitrary and we could consider n^α with $\alpha \in (0, 1)$. A similar Lemma 7 could be obtained for this choice.

Let $S_n^L := \{i \in [K] \mid N_{n,i} \geq L\}$ as in (19) and N_1 as in Property 4. Using Assumption 1, we have that $\arg\max_{i \in S_n^{\sqrt{n/K}}} \mu_i = i^\star(\boldsymbol{F})$ for all $n \geq N_1$.

Converging with leader and challenger pair We describe the properties that a good leader/challenger should have to ensure convergence towards the β-optimal allocation, assuming Property 4 holds. In order for the challenger to converge, a *good* leader should first identify the best arm $i^\star(\boldsymbol{F})$ (Property 5). Then, given a good leader, a *good* challenger should sample each sub-optimal arm with probability w_i^β: in particular, when the mean probability of sampling a sub-optimal arm i exceeds w_i^β, this arm should have a small probability of being sampled again (Property 6).

Property 5. *Assume Property 4 holds. There exists N_2 with $\mathbb{E}_{\boldsymbol{F}}[N_2] < +\infty$ such that for all $n \geq N_2$,*

$$\mathbb{P}_{|n}[B_{n+1} \neq i^\star(\boldsymbol{F})] \leq g(n) \,,$$

where $g : \mathbb{N}^\star \to (0, +\infty)$ such that $g(n) =_{+\infty} o(n^{-\alpha})$ with $\alpha > 0$.

Property 5 holds for the EB leader (Lemma 18) and the TS leader (Lemma 27).

Property 6. *Assume Property 4 holds. Let B_{n+1} be a leader satisfying Property 5 and C_{n+1} the associated challenger. Let $\varepsilon \in (0, \varepsilon_0(\boldsymbol{F})]$ where $\varepsilon_0(\boldsymbol{F}) > 0$ is a problem dependent constant. There exists N_3 with $\mathbb{E}_{\boldsymbol{F}}[N_3] < +\infty$ such that for all $n \geq N_3$ and all $i \neq i^\star(\boldsymbol{F})$,*

$$\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon \quad \Longrightarrow \quad \mathbb{P}_{|n}[C_{n+1} = i \mid B_{n+1} = i^\star(\boldsymbol{F})] \leq h(n) \,, \tag{21}$$

where $h : \mathbb{N}^\star \to (0, +\infty)$ such that $h(n) =_{+\infty} o(n^{-\alpha})$ with $\alpha > 0$.

Property 6 holds for the TC challenger (Lemma 20), the TCI challenger (Lemma 22) and the RS challenger (Lemma 29).

Provided that Property 4 holds, Lemma 10 shows that $\mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] < +\infty$ for any Top Two algorithm satisfying Properties 5 and 6.

Lemma 10. *Assume Property 4 holds. Let $\varepsilon \in (0, \varepsilon_1(\boldsymbol{F})]$ where $\varepsilon_1(\boldsymbol{F}) > 0$ is a problem dependent constant. Let T_β^ε as in (5). Under a Top Two algorithm whose leader B_{n+1} and challenger C_{n+1} satisfy Properties 5 and 6, we have $\mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] < +\infty$.*

Proof of Lemma 10 We first establish in Lemma 11 the convergence towards the optimal allocation for the best arm, $w^\beta_{i^\star(\boldsymbol{F})} = \beta$.

Lemma 11. *Let $\varepsilon > 0$. Assume Property 4 holds. Under a Top Two algorithm whose leader B_n satisfies Property 5, there exists N_4 with $\mathbb{E}_{\boldsymbol{F}}[N_4] < +\infty$ such that for all $n \geq N_4$,*

$$\left| \frac{N_{n,i^\star(\boldsymbol{F})}}{n} - \beta \right| \leq \varepsilon \,.$$

Proof. Let $i^\star = i^\star(\boldsymbol{F})$ and $\varepsilon > 0$. Let N_1 as in Property 4 and N_2 as in Property 5.

For all $n \geq \max\{N_1, N_2\}$, we have $\mathbb{P}_{|(n-1)}[B_n \neq i^\star] \leq g(n-1)$. Let $M \geq \max\{N_1, N_2\}$. Using Lemma 3, we have

$$\left|\frac{\Psi_{n,i^\star}}{n} - \beta\right| \leq \frac{M-1}{n} + \frac{1}{n}\sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[B_t \neq i^\star] \leq \frac{M-1}{n} + \frac{1}{n}\sum_{t=M}^{n} g(t-1)$$

By Property 5, $g(n) = o(n^{-\alpha})$ with $\alpha > 0$. Using Cesaro's theorem, there exists $N_5 = Poly(W_1)$ such that for all $n \geq \max\{N_1, N_2, N_5\}$,

$$\frac{M-1}{n} \leq \frac{\varepsilon}{4} \quad \text{and} \quad \frac{1}{n}\sum_{t=M}^{n} g(t-1) \leq \frac{\varepsilon}{4}\,.$$

Therefore, we have shown that $\left|\frac{\Psi_{n,i^\star}}{n} - \beta\right| \leq \frac{\varepsilon}{2}$ for all $n \geq \max\{N_1, N_2, N_5\}$. Using Lemma 5, we have

$$\forall n \in \mathbb{N}, \quad \left|\frac{N_{n,i^\star}}{n} - \frac{\Psi_{n,i^\star}}{n}\right| \leq W_1\sqrt{\frac{n+1}{n^2}\log(e+n)}\,.$$

Therefore, there exists $N_6 = Poly\,(W_1)$ such that for all $n \geq N_4 := \max\{N_1, N_2, N_5, N_6\}$,

$$\left|\frac{N_{n,i^\star}}{n} - \frac{\Psi_{n,i^\star}}{n}\right| \leq \frac{\varepsilon}{2}\,.$$

Using the triangular inequality, we obtain

$$\left|\frac{N_{n,i^\star}}{n} - \beta\right| \leq \varepsilon\,.$$

Finally N_4 verifies the condition $\mathbb{E}_{\boldsymbol{F}}[N_4] < \infty$ since

$$\mathbb{E}_{\boldsymbol{F}}[N_4] \leq \sum_{i \in \{1,2,5,6\}} \mathbb{E}_{\boldsymbol{F}}[N_i] < +\infty\,.$$

□

We then prove in Lemma 12 the convergence towards the optimal allocation for all arms. We notably use the previous convergence result established for the optimal arm.

Lemma 12. *Assume Property 4 holds. Let $\varepsilon \in (0, \varepsilon_1(\boldsymbol{F})]$ where $\varepsilon_1(\boldsymbol{F}) > 0$ is a problem dependent constant. Under a Top Two algorithm whose leader B_n and challenger C_n satisfy Properties 5 and 6, there exists N_5 with $\mathbb{E}_{\boldsymbol{F}}[N_5] < +\infty$ such that for all $n \geq N_5$,*

$$\forall i \in [K], \quad \left|\frac{N_{n,i}}{n} - w_i^\beta\right| \leq \varepsilon\,.$$

Proof. Let $i^\star = i^\star(\boldsymbol{F})$. Let $\varepsilon_0 = \varepsilon_0(\boldsymbol{F})$ and N_3 as in Property 6 and $\varepsilon \in (0, \varepsilon_0]$. Let N_1 and N_2 as in Properties 4 and 5. Let N_4 as in Lemma 11. For all $n \geq \max_{i\in[4]} N_i$, we have $\left|\frac{N_{n,i^\star}}{n} - \beta\right| \leq \varepsilon$ and for all $i \neq i^\star$,

$$\frac{\Psi_{n-1,i}}{n-1} \geq w_i^\beta + \varepsilon \implies \mathbb{P}_{|(n-1)}[C_n = i \mid B_n = i^\star] \leq h(n-1)\,.$$

Let $M \geq \max_{i\in[4]} N_i$. By Properties 5 and 6, $g(n) = o(n^{-\alpha})$ with $\alpha > 0$ and $h(n) = o(n^{-\alpha})$ with $\alpha > 0$. Using Cesaro's theorem, there exists a deterministic N_6 such that for all $n \geq N_6$,

$$\frac{M-1}{n} \leq \varepsilon \quad , \quad \frac{1}{n}\sum_{t=M}^{n} g(t-1) \leq \varepsilon \quad \text{and} \quad \frac{1}{n}\sum_{t=M}^{n} h(t-1) \leq \varepsilon\,.$$

In particular, $\mathbb{E}_{\boldsymbol{F}}[N_6] = N_6 < +\infty$. Let $t_{n-1,i}(\varepsilon) = \max\left\{t \leq n \mid \frac{\Psi_{t-1,i}}{n-1} \leq w_i^\beta + \varepsilon\right\}$. Using Lemma 4 and $\frac{\Psi_{t-1,i}}{n-1} \leq \frac{\Psi_{t-1,i}}{t-1}$ for $t \leq n$, we obtain for all $n \geq \max_{i\in[4]\cup\{6\}} N_i$

$$\begin{aligned}
\frac{\Psi_{n,i}}{n} &\leq \frac{M-1}{n} + \frac{1}{n}\sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[B_t \neq i^\star] + \frac{1}{n}\sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[C_t = i | B_t = i^\star] \\
&\leq \varepsilon + \frac{1}{n}\sum_{t=M}^{n} g(t-1) + \frac{1}{n}\sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[C_t = i | B_t = i^\star] \mathbb{1}\left(\frac{\Psi_{t-1,i}}{n-1} \geq w_i^\beta + \varepsilon\right) \\
&\quad + \frac{1}{n}\sum_{t=M}^{n} \mathbb{P}_{|(t-1)}[C_t = i | B_t = i^\star] \mathbb{1}\left(\frac{\Psi_{t,i}}{n-1} \leq w_i^\beta + \varepsilon\right) \\
&\leq 2\varepsilon + \frac{1}{n}\sum_{t=M}^{n} h(t-1) + \frac{1}{n}\sum_{t=M}^{t_{n,i}(\varepsilon)} \mathbb{P}_{|(t-1)}[C_t = i | B_t = i^\star] \mathbb{1}\left(\frac{\Psi_{t-1,i}}{n-1} \leq w_i^\beta + \varepsilon\right) \\
&\leq 3\varepsilon + \frac{\Psi_{t_{n-1,i}(\varepsilon),i}}{n-1} \\
&\leq w_i^\beta + 4\varepsilon
\end{aligned}$$

As a similar upper bound was already shown in the proof of Lemma 11, we obtain $\frac{\Psi_{n,i}}{n} \leq w_i^\beta + 4\varepsilon$ for all $i \in [K]$ and all $n \geq \max_{i\in[4]\cup\{6\}} N_i$.

Since $\frac{\Psi_{n,i}}{n}$ and w_i^β sum to 1, we obtain for all $n \geq \max_{i\in[4]\cup\{6\}} N_i$ and all $i \in [K]$,

$$\frac{\Psi_{n,i}}{n} = 1 - \sum_{j\neq i} \frac{\Psi_{n,j}}{n} \geq 1 - \sum_{j\neq i}\left(w_j^\beta + 4\varepsilon\right) = w_i^\beta - 4(K-1)\varepsilon\,.$$

Therefore, for all $n \geq \max_{i\in[4]\cup\{6\}} N_i$ and all $i \in [K]$, $\left|\frac{\Psi_{n,i}}{n} - w_i^\beta\right| \leq 4(K-1)\varepsilon$.

Using Lemma 5, we have for all $n \in \mathbb{N}$ and all $i \in [K]$, $\left|\frac{N_{n,i}}{n} - \frac{\Psi_{n,i}}{n}\right| \leq W_1\sqrt{\frac{n+1}{n^2}\log(e+n)}$, hence there exist $N_7 = Poly\,(W_1)$ such that for all $n \geq N_5 := \max_{i\in[4]\cup\{6,7\}} N_i$ and all $i \in [K]$,

$$\left|\frac{N_{n,i}}{n} - \frac{\Psi_{n,i}}{n}\right| \leq \varepsilon\,.$$

Using the triangular inequality, we obtain that for all $n \geq N_5$ and all $i \in [K]$,

$$\left|\frac{N_{n,i}}{n} - w_i^\beta\right| \leq (4K-3)\varepsilon\,.$$

Since

$$\mathbb{E}_{\boldsymbol{F}}[N_5] \leq \sum_{i\in[4]\cup\{6,7\}} \mathbb{E}_{\boldsymbol{F}}[N_i] < +\infty\,,$$

taking $\varepsilon_1 = \frac{\varepsilon_0}{4K-3}$ yields the result for all $\varepsilon \in (0, \varepsilon_1]$. □

Let N_5 as in Lemma 12. By definition of T_β^ε in (5), we have $\mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] \leq \mathbb{E}_{\boldsymbol{F}}[N_5] < +\infty$. This concludes the proof of Lemma 10.

C.5 Asymptotic β-optimality

Provided some regularity assumption on the class of distribution $\mathcal{F}$, we show that asymptotic β-optimality is a direct consequence of $\mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] < +\infty$. More precisely, we show (6):

$$\exists \varepsilon_1(\boldsymbol{F}) > 0,\ \forall \varepsilon \in (0, \varepsilon_1(\boldsymbol{F})],\ \mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] < +\infty \quad \Longrightarrow \quad \limsup_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log(1/\delta)} \leq T_\beta^\star(\boldsymbol{F})\,.$$

In [35], (6) was proven for Gaussian. We generalize the proof from [35] to hold provided we have joint continuity of the minimal transportation cost (Property 7) and rate of convergence for

$\mathcal{T}(F_{n,i})$ (Property 8). Those properties hold for bounded distributions and SPEF with sub-exponential distributions, and potentially many other distributions.

Before stating the adequate properties, we recall the notation introduced in Appendix C.1. Let $C_{i,j}(\mathcal{T}(\boldsymbol{F}), w)$ be transportation costs defined in (12) as

$$C_{i,j}(\mathcal{T}(\boldsymbol{F}), w) := \inf_{u \in \mathcal{I}} \left\{ w_i \mathcal{K}_{\inf}^{-}(\mathcal{T}(F_i), u) + w_j \mathcal{K}_{\inf}^{+}(\mathcal{T}(F_j), u) \right\} ,$$

where $\mathcal{I} \subseteq \mathbb{R}$, and their empirical version in (13) as

$$\frac{1}{n} W_n(i,j) = C_{i,j}\left(\mathcal{T}(\boldsymbol{F}_n), \frac{N_n}{n} \right) .$$

Similarly, the β-characteristic time and β-optimal allocation

$$T_\beta^\star(\boldsymbol{F})^{-1} := \max_{w \in \triangle_K : w_{i^\star(\boldsymbol{F})} = \beta} \min_{j \neq i^\star(\boldsymbol{F})} C_{i^\star(\boldsymbol{F}),j}(\mathcal{T}(\boldsymbol{F}), w) ,$$

$$w_\beta^\star(\boldsymbol{F}) := \operatorname*{arg\,max}_{w \in \triangle_K : w_{i^\star(\boldsymbol{F})} = \beta} \min_{j \neq i^\star(\boldsymbol{F})} C_{i^\star(\boldsymbol{F}),j}(\mathcal{T}(\boldsymbol{F}), w) .$$

Property 7. *$\mathcal{T}(F) \mapsto m(F)$ is continuous on $\mathcal{T}(\mathcal{F})$ and $(\mathcal{T}(\boldsymbol{F}), w) \mapsto \min_{j \neq i^\star(\boldsymbol{F})} C_{i^\star(\boldsymbol{F}),j}(\mathcal{T}(\boldsymbol{F}), w)$ is continuous on $\mathcal{T}(\mathcal{F}^K) \times \triangle_K$. If $|i^\star(\boldsymbol{F})| = 1$, then $w_\beta^\star(\boldsymbol{F}) = \{w^\beta\}$ is a singleton such that $\min_{i \in [K]} w_i^\beta > 0$.*

For single-parameter exponential families, Property 7 is a known result from the literature [38] as $\mathcal{T}(F) = m(F)$. Property 7 holds for bounded distributions: $F \mapsto m(F)$ continuous (bounded), using proof of Lemma 58 (consequence of Lemma 54) and by Lemmas 61 and 60.

Property 8. *For all $\varepsilon > 0$, there exists N_ε with $\mathbb{E}_{\boldsymbol{F}}[N_\varepsilon] < +\infty$ such that*

$$\forall i \in [K],\ \forall N_{n,i} \geq N_\varepsilon, \quad \|\mathcal{T}(F_{n,i}) - \mathcal{T}(F_i)\| \leq \varepsilon ,$$

where $\|\cdot\|$ is the norm on $\mathcal{T}(\mathcal{F})$.

For SPEF, we have $\mathcal{T}(F_{n,i}) = \mu_{n,i}$, hence Property 8 holds for any SPEF with sub-exponential distributions (see Lemma 73). For bounded distributions, Property 8 is a direct corollary of Lemma 14. Since $N_\varepsilon = Poly(\frac{1}{\varepsilon}, W_2)$ and $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_2}] < +\infty$ for all $\lambda > 0$, we have directly that $\mathbb{E}_{\boldsymbol{F}}[N_\varepsilon] < +\infty$.

Using the empirical transportation defined in (13), generalizing the stopping time in (2) yields

$$\tau_\delta = \inf \left\{ n \in \mathbb{N} \mid \min_{j \neq \hat{\imath}_n} W_n(\hat{\imath}_n, j) > c(n, \delta) \right\} . \tag{22}$$

Calibrating the stopping threshold to obtain δ-correctness of the stopping rule (22) highly depends on the considered $\mathcal{F}$. Definition 2 introduces *asymptotically tight* thresholds, whose (n, δ) dependencies ensure asymptotic (β-)optimality.

Definition 2 (Asymptotically tight threshold)**.** *A threshold $c : \mathbb{N} \times (0,1] \to \mathbb{R}_+$ is said to be asymptotically tight if there exists $\alpha \in [0,1)$, $\delta_0 \in (0,1]$, functions $f, \bar{T} : (0,1] \to \mathbb{R}_+$ and C independent of δ satisfying: (1) for all $\delta \in (0, \delta_0]$ and $n \geq \bar{T}(\delta)$, then $c(n,\delta) \leq f(\delta) + Cn^\alpha$, (2) $\limsup_{\delta \to 0} f(\delta)/\log(1/\delta) \leq 1$ and (3) $\limsup_{\delta \to 0} \bar{T}(\delta)/\log(1/\delta) = 0$.*

For bounded distributions, the stopping threshold defined in (4) is asymptotically tight, e.g. take $(\alpha, \delta_0, C) = (1/2, 1, 1)$, $f(\delta) = \log\left(\frac{K-1}{\delta}\right) + 2$ and $\bar{T}(\delta) = 1$. Lemma 2 shows that it is also δ-correct for bounded distributions.

For single-parameter exponential families, the thresholds for which δ-correctness has been proved are also asymptotically tight, e.g. the ones derived in [29]. Those thresholds are all upper bounded by some threshold of the form $c(n,\delta) = \log\left(\frac{Dn^\kappa}{\delta}\right)$. This stylized threshold is asymptotically tight, e.g. by taking $(\alpha, \delta_0, C) = (1/2, 1, \kappa)$, $f(\delta) = \log\left(\frac{D}{\delta}\right)$ and $\bar{T}(\delta) = 1$.

Theorem 2 shows (6) when using the stopping rule (22) with an asymptotically tight threshold.

Theorem 2. *Assume that Properties 7 and 8 hold on $\mathcal{F}^K$. Let $(\delta, \beta) \in (0,1)^2$. Assume that there exists $\varepsilon_1(\boldsymbol{F}) > 0$ such that for all $\varepsilon \in (0, \varepsilon_1(\boldsymbol{F})]$, $\mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] < +\infty$. Combining the stopping rule (22) with an asymptotically tight threshold yields an algorithm such that for all $\boldsymbol{F} \in \mathcal{F}^K$, with $|i^\star(\boldsymbol{F})| = 1$ and $\mu_{\boldsymbol{F}} \in \left(\mathring{\mathcal{I}}\right)^K$,*

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log\left(\frac{1}{\delta}\right)} \leq T_\beta^\star(\boldsymbol{F}) .$$

Proof. Let $i^\star = i^\star(\boldsymbol{F})$ and $\varepsilon_1 = \varepsilon_1(\boldsymbol{F})$. Let $c_\beta = \frac{1}{2}\min_{i\in[K]} w_i^\beta > 0$ and $\Delta = \min_{j\neq i^\star} |\mu_{i^\star} - \mu_i| > 0$. Let $\zeta > 0$. Using Property 7, the continuity of

$$(\mathcal{T}(\boldsymbol{F}), w) \mapsto \min_{j\neq i^\star(\boldsymbol{F})} C_{i^\star(\boldsymbol{F}),j}(\mathcal{T}(\boldsymbol{F}), w) \quad \text{and} \quad \mathcal{T}(F) \mapsto m(F)$$

yields that there exists $\varepsilon_2 > 0$ such that

$$\begin{aligned} &\max_{i\in[K]} \left| \frac{N_{n,i}}{n} - w_i^\beta \right| \leq \varepsilon_2 \quad \text{and} \quad \max_{i\in[K]} \|\mathcal{T}(F_{n,i}) - \mathcal{T}(F_i)\| \leq \varepsilon_2 \\ \implies \quad &\max_{i\in[K]} |\mu_{n,i} - \mu_i| \leq \frac{\Delta}{4} \quad \text{and} \quad \frac{1}{n}\min_{j\neq i^\star} W_n(i^\star, j) \geq \frac{1-\zeta}{T_\beta^\star(\boldsymbol{F})} \,. \end{aligned}$$

Choosing such a ε_2, we take $\varepsilon \in (0, \min\{\varepsilon_1, \varepsilon_2, c_\beta\})$. By assumption, we have $\mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] < +\infty$, hence $\frac{N_{n,i}}{n} \geq w_i^\beta - \varepsilon \geq c_\beta$ for all $i \in [K]$.

Let N_ε as in Property 8. Using Property 8, for all $n \geq c_\beta^{-1} N_\varepsilon$, we have $\max_{i\in[K]} \|\mathcal{T}(F_{n,i}) - \mathcal{T}(F_i)\| \leq \varepsilon \leq \varepsilon_2$ as $\min_{i\in[K]} N_{n,i} \geq N_\varepsilon$. Therefore, we have $\hat{\imath}_n \in \arg\max_{i\in[K]} \mu_{n,i} = \arg\max_{i\in[K]} \mu_i$ as $\max_{i\in[K]} |\mu_{n,i} - \mu_i| \leq \frac{\Delta}{4}$.

Let $\alpha \in [0,1)$, $\delta_0 \in (0,1]$, functions $f, \bar{T} : (0,1] \to \mathbb{R}_+$ and C as in the definition of an asymptotically tight family of thresholds. In the following, we consider $\delta \leq \delta_0$. Let $\kappa > 0$. Let $T \geq \frac{1}{\kappa}\max\{T_\beta^\varepsilon, c_\beta^{-1}N_\varepsilon, \bar{T}(\delta)\}$. Using the definition of the stopping rule (2) with a family of asymptotically tight threshold, we have

$$\begin{aligned} \min\{\tau_\delta, T\} &\leq \kappa T + \sum_{n=\kappa T}^{T} \mathbb{1}(\tau_\delta > n) \leq \kappa T + \sum_{n=\kappa T}^{T} \mathbb{1}\left(\min_{j\neq i^\star} W_n(i^\star, j) \leq c(n,\delta)\right) \\ &\leq \kappa T + \sum_{n=\kappa T}^{T} \mathbb{1}\left(n\frac{1-\zeta}{T_\beta^\star(\boldsymbol{F})} \leq f(\delta) + CT^\alpha\right) \\ &\leq \kappa T + \frac{T_\beta^\star(\boldsymbol{F})}{1-\zeta}\left(f(\delta) + CT^\alpha\right) \,. \end{aligned}$$

Let $T_\zeta(\delta)$ defined as

$$T_\zeta(\delta) := \inf\left\{ T \geq 1 \mid \frac{T_\beta^\star(\boldsymbol{F})}{(1-\zeta)(1-\kappa)}\left(f(\delta) + CT^\alpha\right) \leq T \right\} \,.$$

For every $T \geq \max\{T_\zeta(\delta), \frac{1}{\kappa}\max\{T_\beta^\varepsilon, c_\beta^{-1}N_\varepsilon, \bar{T}(\delta)\}\}$, we have $\tau_\delta \leq T$, hence

$$\mathbb{E}_{\boldsymbol{F}}[\tau_\delta] \leq \frac{1}{\kappa}\mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] + \frac{1}{\kappa c_\beta}\mathbb{E}_{\boldsymbol{F}}[N_\varepsilon] + \frac{1}{\kappa}\bar{T}(\delta) + T_\zeta(\delta) \,.$$

As $\mathbb{E}_{\boldsymbol{F}}[T_\beta^\varepsilon] + c_\beta^{-1}\mathbb{E}_{\boldsymbol{F}}[N_\varepsilon] < +\infty$ and $\lim_{\delta\to 0} \frac{\bar{T}(\delta)}{\log(1/\delta)}$, we obtain for all $\zeta, \kappa > 0$

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log(1/\delta)} \leq \limsup_{\delta\to 0} \frac{T_\zeta(\delta)}{\log(1/\delta)} \leq \frac{T_\beta^\star(\boldsymbol{F})}{(1-\zeta)(1-\kappa)} \,,$$

where the last inequality uses Lemma 13, which is an inversion result. Letting ζ and κ go to zero yields that

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log(1/\delta)} \leq T_\beta^\star(\boldsymbol{F}) \,.$$

□

Corollary 1 is a direct consequence of Lemma 7, Lemma 10 and Theorem 2.

Corollary 1. *Assume that Properties 7 and 8 hold on $\mathcal{F}^K$. Let $(\delta, \beta) \in (0,1)^2$. Combining the stopping rule (22) with an asymptotically tight threshold and a Top Two algorithm, whose leader B_n and challenger C_n satisfy Properties (2, 5) and (3, 6), yields an algorithm such that for all $\boldsymbol{F} \in \mathcal{F}^K$, with $\Delta_{\min}(\boldsymbol{F}) := \min_{j \neq i} |\mu_i - \mu_j| > 0$ and $\mu_{\boldsymbol{F}} \in \left(\mathring{\mathcal{I}}\right)^K$,*

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{F}}[\tau_\delta]}{\log\left(\frac{1}{\delta}\right)} \leq T^\star_\beta(\boldsymbol{F}) \,.$$

Proof. Having $\Delta_{\min}(\boldsymbol{F}) > 0$ yields that $i^\star(\boldsymbol{F})$ is a singleton. Since $\Delta_{\min}(\boldsymbol{F}) > 0$ and leader B_n and challenger C_n satisfies Properties 2 and 3, Lemma 7 shows that Property 4 holds. As leader B_n and challenger C_n satisfies Properties 5 and 6, we can use Lemma 10. Directly applying Theorem 2 yields the result. □

Lemma 13. *Let $C, D \in \mathbb{R}^\star_+$, $\alpha \in [0,1)$, $f : (0,1] \to \mathbb{R}_+$ such that $\lim_{\delta \to 0} \frac{f(\delta)}{\log(1/\delta)} \leq 1$ and*

$$T_D(\delta) := \inf\left\{T \geq 1 \mid D\left(f(\delta) + CT^\alpha\right) \leq T\right\} \,. \tag{23}$$

Then,

$$\limsup_{\delta \to 0} \frac{T_D(\delta)}{\log(1/\delta)} \leq D \,.$$

Proof. Let $\gamma > 0$. Since $\alpha \in [0,1)$, there exists T_γ (depending on D) such that for all $T \geq T_\gamma$,

$$T\frac{1}{D} - CT^\alpha \geq T\frac{1}{D(1+\gamma)} \,.$$

Then,

$$T_D(\delta) \leq T_\gamma + \inf\left\{T \geq 1 \mid f(\delta) \leq T\frac{1}{D(1+\gamma)}\right\} \leq T_\gamma + D(1+\gamma)f(\delta) + 1 \,.$$

Since $\limsup_{\delta \to 0} \frac{f(\delta)}{\log(1/\delta)} \leq 1$, we obtain for all $\gamma > 0$

$$\limsup_{\delta \to 0} \frac{T_D(\delta)}{\log(1/\delta)} \leq D(1+\gamma) \,.$$

Letting γ go to zero yields the result. □

D Top Two instances for bounded distributions

While we provided a unified analysis of Top Two algorithms in Appendix C, we are interested in specific instances. We distinguish between the deterministic mechanisms in Appendix D.1 and the randomized mechanisms in Appendix D.2. After introducing them, we will show they each satisfies the properties required on the leader and the challenger to ensure sufficient exploration (Appendix C.3) and to converge towards the β-optimal allocation (Appendix C.4).

As deterministic mechanisms, we study the EB leader (Appendix D.1.1), the TC challenger (Appendix D.1.2) and the TCI challenger (Appendix D.1.3). For the randomized mechanisms which are based on a sampler Π_n (Appendix D.2.1), we consider the TS leader (Appendix D.2.2) and the RS challenger (Appendix D.2.3). While those leaders and challengers are defined and analyzed for bounded distributions, we will also discuss why the analysis still hold for single-parameter exponential families (Appendix H). This is especially simple for deterministic mechanisms. For randomized mechanisms, a natural sampler is the posterior distribution. However, proving the properties on Π_n (Appendix D.2.1) in all generality is more cumbersome.

By the end of Appendix D, we will have shown that Properties 2 and 5 hold for the EB and TS leaders, and that Properties 3 and 6 hold for the TC, TCI and RS challengers, which leads to Theorem 1.

Proof of Theorem 1 The threshold (4) is asymptotically tight (Definition 2), e.g. $(\alpha, \delta_0, C) = (1/2, 1, 1)$, $f(\delta) = \log\left(\frac{K-1}{\delta}\right) + 2$ and $\bar{T}(\delta) = 1$. Lemma 2 shows that it is also δ-correct for bounded distributions. Therefore, Theorem 1 is obtained by applying Corollary 1.

Proof of Property 8 Lemma 14 gives the convergence rate of empirical cdfs $(F_{n,i})_{i\in[K]}$ towards the true cdfs $(F_i)_{i\in[K]}$. Deferred to Appendix E.2, its proof is a direct consequence of concentration inequalities for sub-Gaussian random variables.

Lemma 14. *There exists a sub-Gaussian random variable W_2 such that for all $(n,i)\in\mathbb{N}\times[K]$*

$$\|F_{n,i}-F_i\|_\infty \le W_2\sqrt{\frac{\log(e+N_{n,i})}{1+N_{n,i}}} \quad a.s.\ . \tag{24}$$

In particular, $\mathbb{E}\left[e^{\lambda W_2}\right]<+\infty$ for all $\lambda>0$.

In the following, we take W_2 as in Lemma 14. Let $\varepsilon>0$. Using Lemma 14, there exists $N_\varepsilon = Poly(\frac{1}{\varepsilon}, W_2)$ such that for all $i\in[K]$ and all $N_{n,i}\ge N_\varepsilon$,

$$\max_{i\in[K]}\|F_{n,i}-F_i\|_\infty \le \varepsilon\ .$$

As $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_2}]<+\infty$ for all $\lambda>0$, we have $\mathbb{E}_{\boldsymbol{F}}[N_\varepsilon]<+\infty$. Therefore, Property 8 holds for bounded distributions.

Property 7 holds for bounded distributions: $F\mapsto m(F)$ continuous (bounded), using proof of Lemma 58 (consequence of Lemma 54) and by Lemmas 61 and 60.

D.1 Deterministic mechanisms

Conditioned on the history $\mathcal{F}_n$, deterministic mechanisms for the leader and the challenger don't depend on a sampler Π_n. The sole randomness in those mechanisms occurs in case of ties, which are broken uniformly at random. In Appendix D.1.1, we define the EB leader and shows that it satisfies Properties 2 and 5. In Appendix D.1.2, we define the TC challenger and proves that Properties 3 and 6 hold. In Appendix D.1.3, we define the TCI challenger and proves that Properties 3 and 6 hold.

Rates for empirical transportation costs Analyzing deterministic mechanisms heavily relies on properties of the empirical transportation costs. Given two arms having distinct mean, Lemma 15 shows that the transportation cost is strictly positive and increases linearly.

Lemma 15. *Let S_n^L and $\mathcal{I}_n^\star$ as in (19). There exists L_4 with $\mathbb{E}_{\boldsymbol{F}}[(L_4)^\alpha]<+\infty$ for all $\alpha>0$ such that if $L\ge L_4$, for all n such that $S_n^L\neq\emptyset$,*

$$\forall (i,j)\in\mathcal{I}_n^\star\times\left(S_n^L\setminus\mathcal{I}_n^\star\right),\quad W_n(i,j)\ge LD_{\boldsymbol{F}}\ ,$$

where $D_{\boldsymbol{F}}>0$ is a problem dependent constant.

Proof. Let S_n^L and $\mathcal{I}_n^\star$ as in (19). Assume that $S_n^L\neq\emptyset$. If $S_n^L\setminus\mathcal{I}_n^\star$ is empty, then the statement is not informative. Assume $S_n^L\setminus\mathcal{I}_n^\star$ is not empty. Let $(i,j)\in\mathcal{I}_n^\star\times\left(S_n^L\setminus\mathcal{I}_n^\star\right)$.

By definition of W_n in (1) and using $\{i,j\}\subseteq S_n^L$, we obtain

$$\begin{aligned}W_n(i,j) &= \inf_{u\in[0,B]}\left\{N_{n,i}\mathcal{K}_{\inf}^-(F_{n,i},u)+N_{n,j}\mathcal{K}_{\inf}^+(F_{n,j},u)\right\}\\ &\ge L\inf_{u\in[0,B]}\left\{\mathcal{K}_{\inf}^-(F_{n,i},u)+\mathcal{K}_{\inf}^+(F_{n,j},u)\right\}\ .\end{aligned}$$

Using Lemma 30, there exists $\alpha>0$ such that

$$D_{\boldsymbol{F}} = \min_{(i,j):m(F_i)>m(F_j)}\ \inf_{\substack{G_i,G_j:\\ \forall k\in\{i,j\},\|G_k-F_k\|_\infty\le\alpha}}\ \inf_{u\in[0,B]}\left\{\mathcal{K}_{\inf}^-(G_i,u)+\mathcal{K}_{\inf}^+(G_j,u)\right\}>0\ .$$

Using Lemma 14, there exists $L_4=Poly(W_2)$ such that for all $L\ge L_4$ and all $i\in S_n^L$,

$$\|F_{n,i}-F_i\|_\infty\le\alpha\ .$$

Further lower bounding by using that $\mu_i>\mu_j$, we obtain

$$W_n(i,j)\ge L\ \inf_{\substack{G_i,G_j:\\ \forall k\in\{i,j\},\|G_k-F_k\|_\infty\le\alpha}}\ \inf_{u\in[0,B]}\left\{\mathcal{K}_{\inf}^-(G_i,u)+\mathcal{K}_{\inf}^+(G_j,u)\right\}\ge LD_{\boldsymbol{F}}\ .$$

As $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_2}]<+\infty$ for all $\lambda>0$, we have $\mathbb{E}_{\boldsymbol{F}}[(L_4)^\alpha]<+\infty$ for all $\alpha>0$ since $(L_4)^\alpha = Poly(W_2)$. This concludes the proof. □

Lemma 16 gives an upper bound on the transportation costs between a sampled enough arm and an under-sampled one.

Lemma 16. *Let S_n^L as in (19). There exists L_5 with $\mathbb{E}_{\boldsymbol{F}}[(L_5)^\alpha] < +\infty$ for all $\alpha > 0$ such that for all $L \geq L_5$ and all $n \in \mathbb{N}$,*

$$\forall (i,j) \in S_n^L \times \overline{S_n^L}, \quad W_n(i,j) \leq LD_1 \,,$$

where $D_1 > 0$ is a problem dependent constant.

Proof. For bounded distributions, $F \mapsto m(F)$ is continuous on $\mathcal{F}$ for the weak convergence. Since $\mu_i \in (0,B)$ for all $i \in [K]$ (Assumption 1), Lemma 14 yields that there exists $L_6 = Poly(W_1)$ such that for all $L \geq L_6$ and all $i \in S_n^L$, we have $\mu_{n,i} \in (0,B)$. In the following, we consider $L \geq L_6$.

Let $(i,j) \in S_n^L \times \overline{S_n^L}$. By definition and taking $u = \mu_{n,i} \in (0,B)$ yields

$$\begin{aligned} W_n(i,j) &= \inf_{u \in [0,B]} \left\{ N_{n,i} \mathcal{K}_{\inf}^-(F_{n,i}, u) + N_{n,j} \mathcal{K}_{\inf}^+(F_{n,j}, u) \right\} \\ &\leq N_{n,j} \mathcal{K}_{\inf}^+(F_{n,j}, \mu_{n,i}) \leq L \mathcal{K}_{\inf}^+(F_{n,j}, \mu_{n,i}) \leq -L \log\left(1 - \frac{\mu_{n,i}}{B}\right) \,, \end{aligned}$$

where we used that $j \in \overline{S_n^L}$ and Lemma 42. By continuity of $F \mapsto m(F)$, Lemma 14 yields that there exists $L_7 = Poly(W_1)$ such that for all $L \geq L_5 := \max\{L_6, L_7\}$ and all $i \in S_n^L$

$$-\log\left(1 - \frac{\mu_{n,i}}{B}\right) \leq -2\log\left(1 - \frac{\mu_i}{B}\right) \leq D_1 \,,$$

where $D_1 = -2\log\left(1 - \frac{\max_{k \in [K]} \mu_k}{B}\right)$. As $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_2}] < +\infty$ for all $\lambda > 0$, we have $\mathbb{E}_{\boldsymbol{F}}[(L_5)^\alpha] \leq \mathbb{E}_{\boldsymbol{F}}[(L_6)^\alpha] + \mathbb{E}_{\boldsymbol{F}}[(L_7)^\alpha] < +\infty$ since $(L_6)^\alpha = Poly(W_2)$ and $(L_7)^\alpha = Poly(W_2)$. This concludes the proof. □

D.1.1 EB leader

Conditioned on $\mathcal{F}_n$, the Empirical Best (EB) leader is defined as an arm with highest empirical mean

$$B_{n+1}^{\mathrm{EB}} \in \operatorname*{arg\,max}_{i \in [K]} \mu_{n,i} \quad , \quad \mathbb{P}_{|n}[B_{n+1}^{\mathrm{EB}} = i] = \frac{\mathbb{1}\left(i \in \arg\max_{i \in [K]} \mu_{n,i}\right)}{|\arg\max_{i \in [K]} \mu_{n,i}|} \,. \tag{25}$$

and $\widehat{B}_{n+1}^{\mathrm{EB}} = B_{n+1}^{\mathrm{EB}}$.

Property 2 Lemma 17 shows that Property 2 is satisfied by B_{n+1}^{EB}.

Lemma 17. *Let S_n^L and $\mathcal{I}_n^\star$ as in (19). Let L_4 in Lemma 15. Then, for all $L \geq L_4$, for all n such that $S_n^L \neq \emptyset$, $\widehat{B}_{n+1}^{EB} \in S_n^L$ implies $\widehat{B}_{n+1}^{EB} \in \mathcal{I}_n^\star$.*

Proof. Let S_n^L and $\mathcal{I}_n^\star$ as in (19). Assume that $S_n^L \neq \emptyset$. If $S_n^L \setminus \mathcal{I}_n^\star$ is empty, then the result is true. Assume $S_n^L \setminus \mathcal{I}_n^\star$ is not empty. Let L_4 in Lemma 15. Then,

$$\forall (i,j) \in \mathcal{I}_n^\star \times \left(S_n^L \setminus \mathcal{I}_n^\star\right), \quad W_n(i,j) \geq LD_{\boldsymbol{F}} \,,$$

Assume that $\widehat{B}_{n+1}^{\mathrm{EB}} \in S_n^L$. Suppose towards contradiction that $\widehat{B}_{n+1}^{\mathrm{EB}} \notin \mathcal{I}_n^\star$. Therefore, $W_n(i, \widehat{B}_{n+1}^{\mathrm{EB}}) \geq LD_{\boldsymbol{F}} > 0$ for all $i \in \mathcal{I}_n^\star$. Since the choice of the leader is deterministic, we have $B_{n+1}^{\mathrm{EB}} = \widehat{B}_{n+1}^{\mathrm{EB}}$. Since $B_{n+1}^{\mathrm{EB}} \in \arg\max_{i \in [K]} \mu_{n,i}$, we have $W_n(i, \widehat{B}_{n+1}^{\mathrm{EB}}) = 0$. This is a contradiction, hence $\widehat{B}_{n+1}^{\mathrm{EB}} \in \mathcal{I}_n^\star$. □

Property 5 Lemma 18 shows that Property 5 is satisfied by B_{n+1}^{EB}. More precisely, we show that after enough time, the leader is the best arm almost surely.

Lemma 18. *Assume Property 4 holds. There exists N_6 with $\mathbb{E}_{\boldsymbol{F}}[N_6] < +\infty$ such that for all $n \geq N_6$,*

$$\mathbb{P}_{|n}[B_{n+1}^{EB} \neq i^\star(\boldsymbol{F})] = 0 \,.$$

Proof. Let $i^\star = i^\star(\boldsymbol{F})$. Let N_1 as in Property 4, then $N_{n,i} \geq \sqrt{\frac{n}{K}}$ for all $n \geq N_1$. Since $i^\star$ is unique, we have $\Delta := \min_{j \neq i^\star} |\mu_{i^\star} - \mu_j| > 0$. For bounded distributions, $F \mapsto m(F)$ is continuous on $\mathcal{F}$ for the weak convergence. Lemma 14 yields that there exists $N_7 = Poly(W_2)$ such that for all $n \geq N_6 := \max\{N_1, N_7\}$ and all $i \in [K]$, we have $|\mu_{n,i} - \mu_i| \leq \frac{\Delta}{4}$. Therefore, for all $n \geq N_6$, $\arg\max_{i\in[K]} \mu_{n,i} = \arg\max_{i\in[K]} \mu_i = i^\star$ and

$$\mathbb{P}_{|n}[B_{n+1}^{\mathrm{EB}} \neq i^\star] = 1 - \mathbb{P}_{|n}[B_{n+1}^{\mathrm{EB}} = i^\star] = 1 - \frac{\mathbb{1}\left(i^\star \in \arg\max_{i\in[K]} \mu_{n,i}\right)}{|\arg\max_{i\in[K]} \mu_{n,i}|} = 0\,.$$

As $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_2}] < +\infty$ for all $\lambda > 0$, we have $\mathbb{E}_{\boldsymbol{F}}[N_7] < +\infty$. Therefore, $\mathbb{E}_{\boldsymbol{F}}[N_6] \leq \mathbb{E}_{\boldsymbol{F}}[N_1] + \mathbb{E}_{\boldsymbol{F}}[N_7] < +\infty$ yields the result. □

D.1.2 TC challenger

Conditioned on $\mathcal{F}_n$ and given a leader B_{n+1}, the Transportation Cost (TC) challenger is defined as the arm with smallest transportation cost compared to the leader

$$C_{n+1}^{\mathrm{TC}} \in \operatorname*{arg\,min}_{j \neq B_{n+1}} W_n(B_{n+1}, j) \quad , \quad \mathbb{P}_{|n}[C_{n+1}^{\mathrm{TC}} = j | B_{n+1} = i] = \frac{\mathbb{1}\left(j \in \arg\min_{k\neq i} W_n(i,k)\right)}{|\arg\min_{k\neq i} W_n(i,k)|}\,, \tag{26}$$

and $\widehat{C}_{n+1}^{\mathrm{TC}} \in \arg\min_{j \neq \widehat{B}_{n+1}} W_n(\widehat{B}_{n+1}, j)$.

Property 3 We prove Property 3 for C_{n+1}^{TC} in Lemma 19 by comparing the rates at which W_n increases (Lemmas 15 and 16). The effective challenger $\widehat{C}_{n+1}^{\mathrm{TC}}$ is taken as an arm minimizing the transportation cost compared to the leader $\widehat{B}_{n+1}$. Therefore, it is sufficient to show that the sampled enough arms have higher transportation costs than the mildly under-sampled ones. This implies that $\widehat{C}_{n+1}^{\mathrm{TC}}$ has to be mildly under-sampled or be an arm with highest mean among the sampled enough arms.

Lemma 19. *Let B_{n+1} be a leader satisfying Property 2. Given $(B_{n+1}, \widehat{B}_{n+1})$, let $(C_{n+1}^{TC}, \widehat{C}_{n+1}^{TC})$ as in (26). Let U_n^L and V_n^L as in (20) and $\mathcal{J}_n^\star = \arg\max_{i\in\overline{V_n^L}} \mu_i$. There exists L_6 with $\mathbb{E}_{\boldsymbol{F}}[L_6] < +\infty$ such that if $L \geq L_6$, for all n such that $U_n^L \neq \emptyset$, $\widehat{B}_{n+1} \notin V_n^L$ implies $\widehat{C}_{n+1}^{TC} \in V_n^L \cup \left(\mathcal{J}_n^\star \setminus \left\{\widehat{B}_{n+1}\right\}\right)$.*

Proof. Let $\mathcal{J}_n^\star = \arg\max_{i\in\overline{V_n^L}} \mu_i$. In the following, we consider $U_n^L \neq \emptyset$ (hence $V_n^L \neq \emptyset$) and $\widehat{B}_{n+1} \in V_n^L$. Let B_{n+1} be a leader satisfying Property 2, and L_0 defined therein. Then, for $L \geq L_0^{4/3}$, we have $\widehat{B}_{n+1} \in \mathcal{J}_n^\star$. If $\widehat{C}_{n+1}^{\mathrm{TC}} \in \mathcal{J}_n^\star \setminus \left\{\widehat{B}_{n+1}\right\}$, we are done. Assume that $\widehat{C}_{n+1}^{\mathrm{TC}} \notin \mathcal{J}_n^\star \setminus \left\{\widehat{B}_{n+1}\right\}$.

Let $(L_4, D_{\boldsymbol{F}})$ and (L_5, D_1) as in Lemmas 15 and 16. Then, for all $L \geq \max\{L_0^{4/3}, L_4^{4/3}, L_5^2\}$,

$$\begin{aligned} \widehat{B}_{n+1} &\in \mathcal{J}_n^\star\,, \\ \forall (i,j) \in \mathcal{J}_n^\star \times \left(\overline{V_n^L} \setminus \mathcal{J}_n^\star\right), \quad & W_n(i,j) \geq L^{3/4} D_{\boldsymbol{F}}\,, \\ \forall (i,j) \in \overline{U_n^L} \times U_n^L, \quad & W_n(i,j) \leq \sqrt{L} D_1\,. \end{aligned}$$

There exists a deterministic L_7 such that for all $L \geq L_7$,

$$L^{3/4} D_{\boldsymbol{F}} > \sqrt{L} D_1\,.$$

Since $\mathcal{J}_n^\star \subseteq \overline{V_n^L} \subseteq \overline{U_n^L}$, for all $L \geq L_6 := \max\{L_0^{4/3}, L_4^{4/3}, L_5^2, L_7\}$ we have

$$\forall (i,k,j) \in \mathcal{J}_n^\star \times U_n^L \times \left(\overline{V_n^L} \setminus \mathcal{J}_n^\star\right), \quad W_n(i,j) > W_n(i,k)\,.$$

As $\widehat{B}_{n+1} \in \mathcal{J}_n^\star$ and $\widehat{C}_{n+1}^{\mathrm{TC}} \notin \mathcal{J}_n^\star \setminus \left\{\widehat{B}_{n+1}\right\}$, the definition $\widehat{C}_{n+1}^{\mathrm{TC}} \in \arg\max_{j\neq\widehat{B}_{n+1}} W_n(\widehat{B}_{n+1}, j)$ yields that $\widehat{C}_{n+1}^{\mathrm{TC}} \in V_n^L$. Otherwise the above strict inequality would wield a contradiction. Since

$$\mathbb{E}_{\boldsymbol{F}}[L_6] \leq L_7 + \mathbb{E}_{\boldsymbol{F}}[(L_0)^{4/3}] + \mathbb{E}_{\boldsymbol{F}}[(L_4)^{4/3}] + \mathbb{E}_{\boldsymbol{F}}[(L_5)^2] < +\infty\,,$$

this concludes the proof. □

Property 6 Lemma 20 shows that the Property 6 is satisfied by C^{TC}_{n+1}. More precisely, it shows that if the mean probability of sampling a sub-optimal arm overshoots its β-optimal allocation, then it won't be sampled almost surely if the leader is the best arm.

Lemma 20. *Assume Property 4 holds. Let $\varepsilon > 0$. Let B_{n+1} be a leader satisfying Property 5 and C^{TC}_{n+1} as in (26). There exists N_7 with $\mathbb{E}_{\boldsymbol{F}}[N_7] < +\infty$ such that for all $n \geq N_7$ and all $i \neq i^\star(\boldsymbol{F})$,*

$$\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon \quad \Longrightarrow \quad \mathbb{P}_{|n}[C^{TC}_{n+1} = i \mid B_{n+1} = i^\star(\boldsymbol{F})] = 0\,. \tag{27}$$

Proof. Let $\varepsilon > 0$ and $i^\star = i^\star(\boldsymbol{F})$. Let N_1 as in Property 4, then $N_{n,i} \geq \sqrt{\frac{n}{K}}$ for all $n \geq N_1$. Since $i^\star$ is unique, we have $\Delta := \min_{j \neq i^\star} |\mu_{i^\star} - \mu_j| > 0$. For bounded distributions, $F \mapsto m(F)$ is continuous on $\mathcal{F}$ for the weak convergence. Lemma 14 yields that there exists $N_8 = Poly(W_2)$ such that for all $n \geq \max\{N_1, N_8\}$ and all $i \in [K]$, we have $|\mu_{n,i} - \mu_i| \leq \frac{\Delta}{4}$. Therefore, for all $n \geq \max\{N_1, N_8\}$, $\arg\max_{i\in[K]} \mu_{n,i} = \arg\max_{i\in[K]} \mu_i = i^\star$.

Let $\xi > 0$. Since Property 4 holds and B_{n+1} satisfies Property 5, we can use the results from Lemma 11. Let N_4 defined in Lemma 11, we have $\left|\frac{N_{n,i^\star}}{n} - \beta\right| \leq \xi$ for all $n \geq \max\{N_1, N_4\}$.

Using the definition of C^{TC}_{n+1} in (26), we have

$$\begin{aligned}
\mathbb{P}_{|n}[C^{\mathrm{TC}}_{n+1} = i \mid B_{n+1} = i^\star] = 0 \quad &\Longleftrightarrow \quad i \notin \arg\min_{k \neq i^\star} W_n(i^\star, k) \\
&\Longleftrightarrow \quad \frac{1}{n}\left(W_n(i^\star, i) - \min_{j \neq i^\star} W_n(i^\star, j)\right) > 0\,.
\end{aligned}$$

Let $i \neq i^\star$ such that $\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon$. Using Lemma 5, there exists $N_9 = Poly(W_1)$, such that for all $n \geq \max\{N_1, N_9\}$, we have $\frac{N_{n,i}}{n} \geq w_i^\beta + \frac{\varepsilon}{2}$. Therefore, for all $n \geq \max\{N_1, N_4, N_8, N_9\}$,

$$\begin{aligned}
&\frac{1}{n}\left(W_n(i^\star, i) - \min_{j \neq i^\star} W_n(i^\star, j)\right) \\
&\geq \inf_{u\in[0,B]} \left\{\frac{N_{n,i^\star}}{n}\mathcal{K}^-_{\inf}(F_{n,i^\star}, u) + \left(w_i^\beta + \frac{\varepsilon}{2}\right)\mathcal{K}^+_{\inf}(F_{n,i}, u)\right\} \\
&\quad - \min_{j\neq i^\star} \inf_{u\in[0,B]} \left\{\frac{N_{n,i^\star}}{n}\mathcal{K}^-_{\inf}(F_{n,i^\star}, u) + \frac{N_{n,j}}{n}\mathcal{K}^+_{\inf}(F_{n,j}, u)\right\} \\
&\geq \inf_{u\in[0,B]} \left\{\frac{N_{n,i^\star}}{n}\mathcal{K}^-_{\inf}(F_{n,i^\star}, u) + \left(w_i^\beta + \frac{\varepsilon}{2}\right)\mathcal{K}^+_{\inf}(F_{n,i}, u)\right\} \\
&\quad - \sup_{w\in\triangle_K : w_{i^\star} = \frac{N_{n,i^\star}}{n}} \min_{j\neq i^\star} \inf_{u\in[0,B]} \left\{w_{i^\star}\mathcal{K}^-_{\inf}(F_{n,i^\star}, u) + w_j\mathcal{K}^+_{\inf}(F_{n,j}, u)\right\} \\
&\geq \inf_{\tilde\beta : |\tilde\beta - \beta| \leq \xi} G_i(\boldsymbol{F}_n, \tilde\beta)
\end{aligned}$$

where

$$\begin{aligned}
G_i(\boldsymbol{F}, \tilde\beta) = &\inf_{u\in[0,B]} \left\{\tilde\beta\mathcal{K}^-_{\inf}(F_{i^\star}, u) + \left(w_i^\beta + \frac{\varepsilon}{2}\right)\mathcal{K}^+_{\inf}(F_i, u)\right\} \\
&- \sup_{w\in\triangle_K : w_{i^\star} = \tilde\beta} \min_{j\neq i^\star} \inf_{u\in\mathcal{I}} \left\{w_{i^\star}\mathcal{K}^-_{\inf}(F_{i^\star}, u) + w_j\mathcal{K}^+_{\inf}(F_j, u)\right\}\,,
\end{aligned}$$

where we lower bounded by considering the best possible allocation such that $w_{i^\star} = \frac{N_{n,i^\star}}{n}$.

Using Lemma 31, the functions $(\boldsymbol{F}, \tilde\beta) \mapsto G_i(\boldsymbol{F}, \tilde\beta)$ and $\boldsymbol{F} \mapsto \inf_{\tilde\beta : |\tilde\beta - \beta| \leq \xi} G_i(\boldsymbol{F}, \tilde\beta)$ are continuous. Therefore, there exists $N_{10} = Poly(W_2)$ and ξ_0 such that for $n \geq N_7 := \{N_1, N_4, N_8, N_9, N_{10}\}$ and all $\xi \leq \xi_0$,

$$\inf_{\tilde\beta : |\tilde\beta - \beta| \leq \xi} G_i(\boldsymbol{F}_n, \tilde\beta) \geq \frac{1}{2} \inf_{\tilde\beta : |\tilde\beta - \beta| \leq \xi} G_i(\boldsymbol{F}, \tilde\beta) \geq \frac{1}{4} G_i(\boldsymbol{F}, \beta)\,.$$

At the β-equilibrium all transportation costs are equal (Lemma 61). Therefore, by definition of w^β,

$$\begin{aligned}
&\sup_{w\in\triangle_K : w_{i^\star}=\beta} \min_{j\neq i^\star} \inf_{u\in[0,B]} \left\{ w_{i^\star}\mathcal{K}_{\inf}^-(F_{i^\star},u) + w_j \mathcal{K}_{\inf}^+(F_j,u)\right\} \\
&\quad = \min_{j\neq i^\star} \inf_{u\in[0,B]} \left\{ \beta\mathcal{K}_{\inf}^-(F_{i^\star},u) + w_j^\beta \mathcal{K}_{\inf}^+(F_j,u)\right\} \\
&\quad = \inf_{u\in[0,B]} \left\{ \beta\mathcal{K}_{\inf}^-(F_{i^\star},u) + w_i^\beta \mathcal{K}_{\inf}^+(F_i,u)\right\} \\
&\quad < \inf_{u\in[0,B]} \left\{ \beta\mathcal{K}_{\inf}^-(F_{i^\star},u) + \left(w_i^\beta + \frac{\varepsilon}{2}\right) \mathcal{K}_{\inf}^+(F_i,u)\right\}
\end{aligned}$$

where the strict inequality is obtained because the transportation costs are increasing in their allocation arguments (Lemma 56). Therefore, we have $G_i(\boldsymbol{F},\beta) > 0$. This yields that $W_n(i^\star,i) > \min_{j\neq i^\star} W_n(i^\star,j)$. As $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_1}] < +\infty$ and $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_2}] < +\infty$ for all $\lambda > 0$, we have $\mathbb{E}_{\boldsymbol{F}}[N_i] < +\infty$ for $i \in \{8,9,10\}$. Since

$$\mathbb{E}_{\boldsymbol{F}}[N_7] \leq \sum_{i\in\{1,4,8,9,10\}} \mathbb{E}_{\boldsymbol{F}}[N_i] < +\infty\,,$$

this concludes the proof. □

D.1.3 TCI challenger

Conditioned on $\mathcal{F}_n$ and given a leader B_{n+1}, the Transportation Cost Improved (TCI) challenger is defined as the arm with smallest penalized transportation cost compared to the leader

$$C_{n+1}^{\mathrm{TCI}} \in \operatorname*{arg\,min}_{j\neq B_{n+1}} \left\{W_n(B_{n+1},j) + \log(N_{n,j})\right\} \quad , \quad \widehat{C}_{n+1}^{\mathrm{TCI}} \in \operatorname*{arg\,min}_{j\neq \widehat{B}_{n+1}} \left\{W_n(\widehat{B}_{n+1},j) + \log(N_{n,j})\right\}, \tag{28}$$

and

$$\mathbb{P}_{|n}[C_{n+1}^{\mathrm{TCI}} = j | B_{n+1} = i] = \frac{\mathbb{1}\left(j \in \arg\min_{k\neq i}\left\{W_n(i,k) + \log(N_{n,k})\right\}\right)}{\left|\arg\min_{k\neq i}\left\{W_n(i,k) + \log(N_{n,k})\right\}\right|}$$

The TCI challenger is inspired by IMED [20]. As we will see in Appendix I.2, it is more stable than the TC challenger. In Appendix D.3, we explain intuitively why. The analysis of the TCI challenger is very close to the one of the TC challenger.

Property 3 With similar arguments as in Lemma 19, Lemma 21 shows that the Property 3 is satisfied by C_{n+1}^{TCI}.

Lemma 21. *Let B_{n+1} be a leader satisfying Property 2. Let $(C_{n+1}^{TCI}, \widehat{C}_{n+1}^{TCI})$ as in (28). Let U_n^L and V_n^L as in (20) and $\mathcal{J}_n^\star = \arg\max_{i\in\overline{V_n^L}} \mu_i$. There exists $\tilde{L}_6$ with $\mathbb{E}_{\boldsymbol{F}}[\tilde{L}_6] < +\infty$ such that if $L \geq \tilde{L}_6$, for all n such that $U_n^L \neq \emptyset$, $\widehat{B}_{n+1} \notin V_n^L$ implies $\widehat{C}_{n+1}^{TCI} \in V_n^L \cup \left(\mathcal{J}_n^\star \setminus \left\{\widehat{B}_{n+1}\right\}\right)$.*

Proof. In the following, we consider $U_n^L \neq \emptyset$ (hence $V_n^L \neq \emptyset$) and $\widehat{B}_{n+1} \in V_n^L$. Let L_0 be defined as in Property 2. Then, for $L \geq L_0^{4/3}$, we have $\widehat{B}_{n+1} \in \mathcal{J}_n^\star$. If $\widehat{C}_{n+1}^{\mathrm{TCI}} \in \mathcal{J}_n^\star \setminus \left\{\widehat{B}_{n+1}\right\}$, we are done. Assume that $\widehat{C}_{n+1}^{\mathrm{TCI}} \notin \mathcal{J}_n^\star \setminus \left\{\widehat{B}_{n+1}\right\}$.

Let $(L_4, D_{\boldsymbol{F}})$ and (L_5, D_1) as in Lemmas 15 and 16. Then, for all $L \geq \max\{L_0^{4/3}, L_4^{4/3}, L_5^2\}$,

$$\begin{aligned}
&\widehat{B}_{n+1} \in \mathcal{J}_n^\star\,, \\
\forall (i,j) \in \mathcal{J}_n^\star \times \left(\overline{V_n^L} \setminus \mathcal{J}_n^\star\right), \quad & W_n(i,j) + \log(N_{n,j}) \geq L^{3/4} D_{\boldsymbol{F}} + \frac{3}{4}\log L\,, \\
\forall (i,j) \in \overline{U_n^L} \times U_n^L, \quad & W_n(i,j) + \log(N_{n,j}) \leq \sqrt{L} D_1 + \frac{1}{2}\log L\,.
\end{aligned}$$

There exists a deterministic $\tilde{L}_7$ such that for all $L \geq \tilde{L}_7$,

$$L^{3/4} D_{\boldsymbol{F}} + \frac{3}{4} > \sqrt{L} D_1 + \frac{1}{2}\log L\,.$$

Since $\mathcal{J}_n^\star \subseteq \overline{V_n^L} \subseteq \overline{U_n^L}$, for all $L \geq \tilde{L}_6 := \max\{L_0^{4/3}, L_4^{4/3}, L_5^2, \tilde{L}_7\}$ we have

$$\forall (i,k,j) \in \mathcal{J}_n^\star \times U_n^L \times \left(\overline{V_n^L} \setminus \mathcal{J}_n^\star\right), \quad W_n(i,j) + \log(N_{n,j}) > W_n(i,k) + \log(N_{n,k}) \,.$$

As $\widehat{B}_{n+1} \in \mathcal{J}_n^\star$ and $\widehat{C}_{n+1}^{\mathrm{TCI}} \notin \mathcal{J}_n^\star \setminus \left\{\widehat{B}_{n+1}\right\}$, the definition $\widehat{C}_{n+1}^{\mathrm{TCI}} \in \arg\max_{j \neq \widehat{B}_{n+1}} \left\{W_n(\widehat{B}_{n+1}, j) + \log(N_{n,j})\right\}$ yields that $\widehat{C}_{n+1}^{\mathrm{TCI}} \in V_n^L$. Otherwise the above strict inequality would wield a contradiction. Since

$$\mathbb{E}_{\boldsymbol{F}}[\tilde{L}_6] \leq \tilde{L}_7 + \mathbb{E}_{\boldsymbol{F}}[L_0^{4/3}] + \mathbb{E}_{\boldsymbol{F}}[L_4^{4/3}] + \mathbb{E}_{\boldsymbol{F}}[L_5^2] < +\infty \,,$$

this concludes the proof. □

Property 6 With similar arguments as in Lemma 20, Lemma 22 shows that the Property 6 is satisfied by C_{n+1}^{TCI}.

Lemma 22. *Assume Property 4 holds. Let $\varepsilon > 0$. Let B_{n+1} be a leader satisfying Property 5 and C_{n+1}^{TCI} as in (28). There exists $\tilde{N}_7$ with $\mathbb{E}_{\boldsymbol{F}}[\tilde{N}_7] < +\infty$ such that for all $n \geq \tilde{N}_7$ and all $i \neq i^\star(\boldsymbol{F})$,*

$$\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon \quad \Longrightarrow \quad \mathbb{P}_{|n}[C_{n+1}^{TCI} = i \mid B_{n+1} = i^\star(\boldsymbol{F})] = 0 \,. \tag{29}$$

Proof. Let $\varepsilon > 0$ and $i^\star = i^\star(\boldsymbol{F})$. Using the definition of C_{n+1}^{TCI} in (28), we have

$$\begin{aligned}
&\mathbb{P}_{|n}[C_{n+1}^{\mathrm{TCI}} = i \mid B_{n+1} = i^\star] = 0 \\
\iff\ & i \notin \arg\min_{k \neq i^\star} \{W_n(i^\star, k) + \log(N_{n,k})\} \\
\iff\ & \frac{1}{n}\left(W_n(i^\star, i) + \log(N_{n,i}) - \min_{j \neq i^\star}\{W_n(i^\star, j) + \log(N_{n,j})\}\right) > 0 \\
\impliedby\ & \frac{1}{n}\left(W_n(i^\star, i) - \min_{j \neq i^\star} W_n(i^\star, j)\right) > \frac{\log(nK)}{2n} \,,
\end{aligned}$$

where we used that $N_{n,i} \geq \sqrt{\frac{n}{K}}$ and $N_{n,j} \leq n$.

Let N_7 as in Lemma 22. In the proof of Lemma 22, we showed that there exists $C_{\boldsymbol{F}} > 0$ such that for all $n \geq N_7$ and all $i \neq i^\star$,

$$\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon \quad \Longrightarrow \quad \frac{1}{n}\left(W_n(i^\star, i) - \min_{j \neq i^\star} W_n(i^\star, j)\right) \geq C_{\boldsymbol{F}} \,.$$

Since $\frac{\log(nK)}{2n} \to_\infty 0$, there exists a deterministic N_8 such that for all $n \geq \tilde{N}_8$,

$$\frac{\log(nK)}{2n} < C_{\boldsymbol{F}} \,.$$

Therefore, for all $n \geq \tilde{N}_7 := \max\{N_8, N_7\}$ and all $i \neq i^\star$,

$$\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon \quad \Longrightarrow \quad \mathbb{P}_{|n}[C_{n+1}^{\mathrm{TCI}} = i \mid B_{n+1} = i^\star] = 0 \,.$$

Since $\mathbb{E}_{\boldsymbol{F}}[\tilde{N}_7] = N_8 + \mathbb{E}_{\boldsymbol{F}}[N_7] < +\infty$, this concludes the proof. □

D.2 Randomized mechanisms

Conditioned on the history $\mathcal{F}_n$, randomized mechanisms for the leader and the challenger depend on a sampler Π_n. In addition, ties will be broken uniformly at random. In Appendix D.2.1, we introduce the general properties that a *good* sampler should verify. Depending on whether we are using the sampler for the leader or for the challenger different properties are necessary. In Appendix D.2.2, we define the TS leader and shows that it satisfies Properties 2 and 5. In Appendix D.2.3, we define the RS challenger and proves that Properties 3 and 6 hold.

D.2.1 How to sample

As both TS leader and RS challenger rely on a sampler Π_n, it is crucial for this sampler to be tailored to the considered set of distributions $\mathcal{F}^K$. For bounded distributions, the sampler Π_n under scrutiny is the Dirichlet sampler introduced in Section 3, which produces for each arm a random re-weighting of the current history of rewards, augmented with $\{0, B\}$. Yet, aiming for a unified analysis of Top Two algorithms relying on a sampler, we put forward some general properties that the sampler should satisfy. Those properties are only expressed in terms of the Boundary Crossing Probability (BCP) associated to each arm, i.e. the probabilities $\mathbb{P}_n[\theta_i \geq u]$ and $\mathbb{P}_n[\theta_i \leq u]$ where u is a fixed threshold. For all measurable sets A_θ, we denote by $\mathbb{P}_n[A_\theta] := \mathbb{P}_{\theta\sim\Pi_n}[A_\theta \mid \mathcal{F}_n]$.

From $a_{n+1,i}$ to BCP Let $a_{n,i} := \mathbb{P}_{n-1}[i \in \arg\max_{j\in[K]} \theta_j]$. From the definition of the TS leader in (32) and the RS challenger in (33), it becomes apparent that we should control the quantity $a_{n,i}$. We will need both upper and lower bounds on $a_{n+1,i}$. To derive those, we can write

$$\begin{aligned} \forall j \neq i, \quad & a_{n+1,i} \leq \mathbb{P}_n[\theta_i \geq \theta_j] \,, \\ & a_{n+1,i} \leq 1 - \max_{j\neq i} \mathbb{P}_n[\theta_j \geq \theta_i] \,, \\ \forall u \in (0,B), \quad & a_{n+1,i} \geq \mathbb{P}_n\left[\theta_i \geq u\right] \prod_{j\neq i} \mathbb{P}_n\left[\theta_j \leq u\right] \,. \end{aligned}$$

The lower bound is already expressed using BCPs, however the upper bound requires to control $\mathbb{P}_n[\theta_i \geq \theta_j]$. In Lemma 64 stated in Appendix G, we provide upper bounds on this probability featuring only BCPs (for some well-chosen threshold).

As we will see, a sampler Π_n is tailored to the considered set of distributions when the upper and lower bounds on the BCP involve $\mathcal{K}_{\inf}^{\pm}$. Those bounds will be referred to as *tight*, the ones without $\mathcal{K}_{\inf}^{\pm}$ will be referred as *coarse*. To show that the TS leader satisfies Properties 2 and 5, we need a tight upper bound on the BCP. Proving Property 3 for the RS challenger requires a tight upper bound and a coarse lower bound on the BCP. However, the proof of Property 6 for the RS challenger relies on a tight upper *and* lower bound on the BCP.

In Appendix G, we prove the corresponding bounds on the BCP for the Dirichlet sampler. These bounds use some ingredients from BPCs bounds obtained for variants of Non-Parametric Thompson Sampling [36, 7] in the regret minimization literature. Our new Lemma 64 is instrumental to bring those to the best arm identification literature.

Coarse lower bound on $a_{n+1,i}$ Recall that $F_{n,i}$ is the empirical cdf of arms i. To ensure that the sampling stops, the sampler Π_n should rely on modified cdfs instead of simply using $\boldsymbol{F}_n$. Those probability measures are denoted by $\tilde{F}_{n,i}$ and their means by $\tilde{\mu}_{n,i}$. For single-parameter exponential family, this modification corresponds to the posterior update based on the prior. For bounded distributions (and Bernoulli), this modification amounts to adding $\{0, B\}$ to the support. This ensures that $\tilde{\mu}_{n,i} \in (0, B)$, hence it is not a Dirac in 0 or B. Alternatively, we can view this step as mixing $F_{n,i}$ with the distribution $G = \frac{1}{2}(\delta_0 + \delta_B)$ which is in the interior of the domain, i.e.

$$\tilde{F}_{n,i} = \left(1 - \frac{2}{n+2}\right) F_{n,i} + \frac{1}{n+2}(\delta_0 + \delta_B) \,. \tag{30}$$

The necessity of adding B to the support was already known [8] to obtain a lower bound on $\mathbb{P}_n[\theta_i \geq u]$. Since we also need to control $\mathbb{P}_n[\theta_i \leq u]$, we should add 0 in the support (by symmetry).

In all generality, the considered modified cdfs should be close to the empirical cdf, i.e.

$$\forall n \in \mathbb{N}, \quad \max_{i\in[K]} \left\| \tilde{F}_{n,i} - F_{n,i} \right\|_\infty \leq d_0(n) \,, \tag{31}$$

where the function $d_0 : \mathbb{N}^* \to \mathbb{N}^*$ satisfies $d_0(n) =_{+\infty} o(n^{-\alpha})$ with $\alpha > 0$. For bounded distributions (and Bernoulli), $\tilde{\boldsymbol{F}}_n$ defined in (30) verifies Property (31) for $d_0(n) = \frac{3}{n+2}$.

Property 9 states that the sampler Π_n based on $\tilde{\boldsymbol{F}}_n$ yields an exponential lower bound on the BCP. It is coarse as it doesn't involve $\mathcal{K}_{\inf}^{\pm}$.

Property 9. *There exists functions $\kappa^+, \kappa^- : (0,B) \to \mathbb{R}_+^*$ such that for all $u \in (0,B)$, all $n \geq 1$ and all $i \in [K]$*

$$\mathbb{P}_n[\theta_i \geq u] \geq e^{-c_0(N_{n,i})\kappa^+(u)} \qquad \text{and} \qquad \mathbb{P}_n[\theta_i \leq u] \geq e^{-c_0(N_{n,i})\kappa^-(u)} .$$

The function $c_0 : \mathbb{N}^* \to \mathbb{N}^*$ is increasing with $c_0(x) \sim_{+\infty} x$. Moreover, $\kappa^-(B) = \kappa^+(0) = 0$ and $\lim_{u\to B} \kappa^+(u) = \lim_{u\to 0} \kappa^-(u) = +\infty$.

Property 9 plays a role in the proof of sufficient exploration for the RS challenger. For bounded distributions (and Bernoulli), Lemma 65 in Appendix G shows that Property 9 holds with

$$\kappa^-(u) = -\log\left(\frac{u}{B}\right) \quad , \quad \kappa^+(u) = -\log\left(1 - \frac{u}{B}\right) \quad , \quad c_0(n) = n+1 .$$

Tight upper bound on $a_{n+1,i}$ Property 10 states that the sampler Π_n based on $\tilde{\boldsymbol{F}}_n$ yields an exponential upper bound on the BCP. Importantly, this upper bound is tailored to the considered family of distributions $\mathcal{F}$ as it involves the $\mathcal{K}_{\inf}^{\pm}$ for the set of distributions $\mathcal{F}$.

Property 10. *For all $u \in (0,B)$, all $n \geq 1$ and all $i \in [K]$,*

$$\mathbb{P}_n[\theta_i \geq u] \leq e^{-c_1(N_{n,i})\mathcal{K}_{\inf}^+(\tilde{F}_{n,i},u)} \qquad \text{and} \qquad \mathbb{P}_n[\theta_i \leq u] \leq e^{-c_1(N_{n,i})\mathcal{K}_{\inf}^-(\tilde{F}_{n,i},u)} ,$$

where $c_1 : \mathbb{N}^* \to \mathbb{N}^*$ is an increasing function such that $c_1(x) \sim_{+\infty} x$.

Property 10 plays an important role for the TS leader and the RS challenger, both in the proof of sufficient exploration and convergence towards the β-optimal allocation. For bounded distributions (and Bernoulli), Theorem 5 in Appendix G shows that Property 10 holds with $c_1(n) = n+2$.

Lemma 23 is a direct corollary of Property 10 by using Lemma 64. For bounded distributions, it is exactly Corollary 2.

Lemma 23. *Let Π_n satisfying Property 10. Then, for all $n \geq 1$ and all $(i,j) \in [K]^2$*

$$\mathbb{P}_n[\theta_j \geq \theta_i] \leq f\left(\inf_{u\in[0,B]} \left\{ c_1(N_{n,i})\mathcal{K}_{\inf}^-(\tilde{F}_{n,i},u) + c_1(N_{n,j})\mathcal{K}_{\inf}^+(\tilde{F}_{n,j},u) \right\}\right) ,$$

where $f : x \mapsto (1+x)e^{-x}$ is decreasing on $\mathbb{R}_+$ with values in $(0,1]$.

Proof. Using Property 10 and Lemma 64, we obtain for all $n \geq 1$ and all $(i,j) \in [K]^2$

$$\begin{aligned}\mathbb{P}_n[\theta_j \geq \theta_i] &\leq f\left(c_1(N_{n,i})\mathcal{K}_{\inf}^-(\tilde{F}_{n,i},u_{i,j}) + c_1(N_{n,j})\mathcal{K}_{\inf}^+(\tilde{F}_{n,j},u_{i,j})\right) \\ &\leq f\left(\inf_{u\in[0,B]} \left\{ c_1(N_{n,i})\mathcal{K}_{\inf}^-(\tilde{F}_{n,i},u) + c_1(N_{n,j})\mathcal{K}_{\inf}^+(\tilde{F}_{n,j},u) \right\}\right) ,\end{aligned}$$

where $u_{i,j} = \arg\max_{c\in[0,B]} \mathbb{P}_n[\theta_j \geq c]\mathbb{P}_n[\theta_i \leq c]$. When $\tilde{\mu}_{n,i} \leq \tilde{\mu}_{n,j}$, this result is non informative as $f(0) = 1$. □

While the proof of Property 9 heavily relies on the transformed cdfs, Property 10 also holds for the empirical cdfs $\boldsymbol{F}_n$. There is no need to add $\{0,B\}$ in the support for this property. We aim at presenting a unified sampler Π_n, which could be used both for the TS leader and the RS challenger. Therefore, we present all the results with the modified cdfs $\tilde{\boldsymbol{F}}_n$ instead of differentiating between a sampler Π_n for the TS leader and a sampler $\tilde{\Pi}_n$ for the RS challenger.

Tight lower bound on $a_{n+1,i}$ Property 11 states that the sampler Π_n based on $\tilde{\boldsymbol{F}}_n$ yields an exponential lower bound on $\mathbb{P}_n[\theta_i \geq \theta_{i^\star}]$. Importantly, this lower bound is tailored to the considered family of distributions $\mathcal{F}$ as it involves the $\mathcal{K}_{\inf}^{\pm}$ for the set of distributions $\mathcal{F}$.

Property 11. *Let $\varepsilon > 0$ and $i^\star = i^\star(\boldsymbol{F})$. There exists N_8, with $\mathbb{E}_{\boldsymbol{F}}[(N_8)^\alpha] < +\infty$ for all $\alpha > 0$, such that for all n with $\min_{i\in[K]} N_{n,i} \geq N_8$ and all $i \neq i^\star$,*

$$\mathbb{P}_n\left[\theta_i \geq \theta_{i^\star}\right] \geq \frac{e^{-\varepsilon(N_{n,i^\star}+N_{n,i})}}{h_\varepsilon(N_{n,i^\star},N_{n,i})} \exp\left(-\inf_{x\in[0,B]} \left\{ N_{n,i^\star}\mathcal{K}_{\inf}^-(F_{i^\star},x) + N_{n,i}\mathcal{K}_{\inf}^+(F_i,x) \right\}\right) ,$$

where $h_\varepsilon : (\mathbb{N}^\star)^2 \to (0,+\infty)$ is an increasing function of both arguments, such that $h_\varepsilon(n,m) =_{+\infty} o\left(e^{(n+m)^\alpha}\right)$ where $\alpha < 1$.

For bounded distributions (and Bernoulli), Property 11 is a direct corollary of Theorem 8 given in Appendix G. Let $\varepsilon > 0$ and $\eta > 0$ as in Theorem 8. Using Lemma 14, there exists $N_8 = Poly(W_2)$ such that for all n such that $\min_{i\in[K]} N_{n,i} \geq N_8$, we have $\max_{i\in[K]} \|F_{n,i} - F_i\|_\infty \leq \eta$. As $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_2}] < +\infty$ for all $\lambda > 0$, we have $\mathbb{E}_{\boldsymbol{F}}[(N_8)^\alpha] < +\infty$ for all $\alpha > 0$ since $(N_8)^\alpha = Poly(W_2)$. Therefore, Property 11 holds for bounded distributions with

$$h_\varepsilon(n,m) = (nm)^{\frac{M_\varepsilon+1}{2}} C_\varepsilon \quad \text{and} \quad C_\varepsilon = \frac{4(8\pi)^{M_\varepsilon-1}}{M_\varepsilon^{M_\varepsilon}} ,$$

where h_ε satisfies the conditions from Property 11.

Using Lemma 64, Theorem 8 was shown thanks to tight lower bound on the BCP for bounded distributions (Lemma 67). Given a family of distribution for which a tight lower bound on the BCP exists, similar manipulations would yield a tight lower bound on $\mathbb{P}_n[\theta_i \geq \theta_j]$, hence showing Property 11. The proof of Theorem 8 heavily relies on the modified empirical cdf featured in the tight BCP lower bound. Indeed, for bounded distributions Lemma 67 follows from a discretization of the empirical cdf, which allows to use results on multinomial distributions. Since the technicalities depend on the considered distribution, we don't provide a general proof of Property 11 given a tight lower bound on the BCP.

Rates for $a_{n+1,i}$ Analyzing randomized mechanisms heavily relies on properties of $a_{n+1,i}$. Lemma 24 shows that $a_{n+1,i}$ decreases exponentially with a linear rate for the arms not having highest means.

Lemma 24. *Let Π_n satisfying Property 10, and c_1 therein. Let S_n^L and $\mathcal{I}_n^\star$ as in (19). There exists L_7 with $\mathbb{E}_{\boldsymbol{F}}[(L_7)^\alpha] < +\infty$ for all $\alpha > 0$ such that if $L \geq L_7$, for all n such that $S_n^L \neq \emptyset$,*

$$\forall i \in S_n^L \setminus \mathcal{I}_n^\star, \quad a_{n+1,i} \leq f(c_1(L) D_{\boldsymbol{F}}) ,$$

where $f(x) = (1+x)e^{-x}$ and $D_{\boldsymbol{F}} > 0$ is the problem dependent constant from Lemma 15.

Proof. Let S_n^L and $\mathcal{I}_n^\star$ as in (19). Assume that $S_n^L \neq \emptyset$. If $S_n^L \setminus \mathcal{I}_n^\star$ is empty, then the statement is not informative. Assume $S_n^L \setminus \mathcal{I}_n^\star$ is not empty. Let $(i,j) \in \mathcal{I}_n^\star \times \left(S_n^L \setminus \mathcal{I}_n^\star\right)$.

Since Π_n satisfies Property 10, using Lemma 23 yields

$$\begin{aligned} a_{n+1,j} \leq \mathbb{P}_n[\theta_j \geq \theta_i] &\leq f\left(\inf_{u\in[0,B]} \left\{ c_1(N_{n,i}) \mathcal{K}_{\inf}^-(\tilde{F}_{n,i}, u) + c_1(N_{n,j}) \mathcal{K}_{\inf}^+(\tilde{F}_{n,j}, u) \right\} \right) \\ &\leq f\left(c_1(L) \inf_{u\in[0,B]} \left\{ \mathcal{K}_{\inf}^-(\tilde{F}_{n,i}, u) + \mathcal{K}_{\inf}^+(\tilde{F}_{n,j}, u) \right\} \right) , \end{aligned}$$

where we used that $\{i,j\} \subset S_n^L$, c_1 increasing and f decreasing.

Using Lemma 30, there exists $\alpha > 0$ such that

$$D_{\boldsymbol{F}} = \min_{(i,j):m(F_i)>m(F_j)} \inf_{\substack{G_i,G_j: \\ \forall k\in\{i,j\}, \|G_k - F_k\|_\infty \leq \alpha}} \inf_{u\in[0,B]} \left\{ \mathcal{K}_{\inf}^-(G_i,u) + \mathcal{K}_{\inf}^+(G_j,u) \right\} > 0 .$$

Using Lemma 14 and (31), i.e. $\max_{i\in[K]} \left\| \tilde{F}_{n,i} - F_{n,i} \right\|_\infty \leq d_0(n)$ where $d_0(n) =_{+\infty} o(n^{-\alpha})$, there exists $L_7 = Poly(W_2)$ such that for all $L \geq L_7$ and all $i \in S_n^L$, $\left\| \tilde{F}_{n,i} - F_i \right\|_\infty \leq \alpha$.

Since f is decreasing, further upper bounding yields directly that for all $L \geq L_7$ and all $j \in S_n^L \setminus \mathcal{I}_n^\star$, we have

$$a_{n+1,j} \leq f\left(c_1(L) D_{\boldsymbol{F}}\right) .$$

As $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_2}] < +\infty$ for all $\lambda > 0$, we have $\mathbb{E}_{\boldsymbol{F}}[(L_7)^\alpha] < +\infty$ for all $\alpha > 0$ since $(L_7)^\alpha = Poly(W_2)$. This concludes the proof. □

Lemma 25 gives a lower bound on $a_{n,i}$ for under-sampled arms.

Lemma 25. *Let Π_n satisfying Properties 9 and 10. Let S_n^L as in (19). There exists L_8 with $\mathbb{E}_{\boldsymbol{F}}[(L_8)^\alpha] < +\infty$ for all $\alpha > 0$ such that for all $L \geq L_8$ and all $n \in \mathbb{N}$,*

$$\forall i \in \overline{S_n^L}, \quad a_{n+1,i} \geq \frac{e^{-D_0 c_0(L)}}{2^{K-1}} ,$$

where c_0 is defined in Property 9 and $D_0 > 0$ is a problem dependent constant.

Proof. Let $i \in \overline{S_n^L}$ and $u \in (0, B)$. As explained above, we have

$$a_{n+1,i} \geq \mathbb{P}_n\left[\theta_i \geq u\right] \prod_{j \in S_n^L} \mathbb{P}_n\left[\theta_j \leq u\right] \prod_{j \in \overline{S_n^L} \setminus \{i\}} \mathbb{P}_n\left[\theta_j \leq u\right] .$$

For all $j \in S_n^L$, Property 10 yields

$$\mathbb{P}_n\left[\theta_j \leq u\right] = 1 - \mathbb{P}_n\left[\theta_j \geq u\right] \geq 1 - e^{-c_1(N_{n,j}) \mathcal{K}_{\inf}^+(\tilde{F}_{n,j}, u)} \geq 1 - e^{-c_1(L) \mathcal{K}_{\inf}^+(\tilde{F}_{n,j}, u)} ,$$

where we used that $N_{n,j} \geq L$ for all $j \in S_n^L$ and c_1 increasing.

By Theorem 4, the function $F \mapsto \mathcal{K}_{\inf}^+(F, u)$ is continuous on $\mathcal{F}$. Using Lemma 14 and (31), i.e. $\max_{i \in [K]} \left\| \tilde{F}_{n,i} - F_{n,i} \right\|_\infty \leq d_0(n)$ where $d_0(n) =_{+\infty} o(n^{-\alpha})$, there exists $L_9 = Poly(W_2)$ such that for all $L \geq L_9$ and all $j \in S_n^L$,

$$\mathcal{K}_{\inf}^+(\tilde{F}_{n,j}, u) \geq \frac{1}{2} \mathcal{K}_{\inf}^+(F_j, u) \geq \frac{1}{2} \min_{j \in [K]} \mathcal{K}_{\inf}^+(F_j, u) .$$

Since $\mu_j \in (0, B)$ for all $j \in [K]$, there exists $u \in (0, B)$ such that $\min_{j \in [K]} \mathcal{K}_{\inf}^+(F_j, u) > 0$. Choosing such a u, there exists a deterministic L_{10} such that for all $L \geq L_8 := \max\{L_9, L_{10}\}$

$$\forall j \in S_n^L, \quad \mathbb{P}_n\left[\theta_j \leq u\right] \geq \frac{1}{2} .$$

For the under-sampled arms $j \in \overline{S_n^L}$, Property 9 yields that

$$\begin{aligned} \forall j \in \overline{S_n^L} \setminus \{i\}, \qquad & \mathbb{P}_n\left[\theta_j \leq u\right] \geq e^{-c_0(N_{n,j}) \kappa^-(u)} \geq e^{-c_0(L) \kappa^-(u)} , \\ & \mathbb{P}_n\left[\theta_i \geq u\right] \geq e^{-c_0(N_{n,i}) \kappa^+(u)} \geq e^{-c_0(L) \kappa^+(u)} , \end{aligned}$$

where we used that $N_{n,i} < L$ for $j \in \overline{S_n^L}$, c_0 increasing and κ^-, κ^+ strictly positive.

Combining the above and further lower bounding, we have shown that for $L \geq L_8$,

$$\forall i \in \overline{S_n^L}, \quad a_{n+1,i} \geq \frac{e^{-D_0 c_0(L)}}{2^{K-1}} ,$$

where $D_0 = \kappa^+(u) + (K-1)\kappa^-(u)$. As $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_2}] < +\infty$ for all $\lambda > 0$, we have $\mathbb{E}_{\boldsymbol{F}}[(L_9)^\alpha] < +\infty$ for all $\alpha > 0$. Since $\mathbb{E}_{\boldsymbol{F}}[(L_8)^\alpha] \leq (L_{10})^\alpha + \mathbb{E}_{\boldsymbol{F}}[(L_9)^\alpha] < +\infty$, this concludes the proof. □

D.2.2 TS leader

Conditioned on $\mathcal{F}_n$, the internal randomness of the Thompson Sampling (TS) leader is parameterized by a sampler Π_n, where $a_{n+1,i} = \mathbb{P}_n[i \in \arg\max_{j \in [K]} \theta_j]$. Given an observation $\theta \sim \Pi_n$, the TS leader is defined as an arm with highest mean for θ,

$$B_{n+1}^{\text{TS}} \in \arg\max_{i \in [K]} \theta_i \quad , \quad \mathbb{P}_{|n}[B_{n+1}^{\text{TS}} = i] = a_{n+1,i} \quad \text{and} \quad \widehat{B}_{n+1}^{\text{TS}} \in \arg\max_{i \in [K]} a_{n+1,i} , \tag{32}$$

where $\widehat{B}_{n+1}^{\text{TS}}$ is defined as an arm with highest $a_{n,i}$.

Property 2 Lemma 26 shows that Property 2 is satisfied by B^{TS}_{n+1}.

Lemma 26. *Let Π_n satisfying Property 10. Let S^L_n and $\mathcal{I}^\star_n$ as in (19). There exists $\tilde{L}_7$ with $\mathbb{E}_{\boldsymbol{F}}[(\tilde{L}_7)^\alpha] < +\infty$ for all $\alpha > 0$ such that if $L \geq \tilde{L}_7$, for all n such that $S^L_n \neq \emptyset$, $\widehat{B}^{TS}_{n+1} \in S^L_n$ implies $\widehat{B}^{TS}_{n+1} \in \mathcal{I}^\star_n$.*

Proof. Let S^L_n and $\mathcal{I}^\star_n$ as in (19). Assume that $S^L_n \neq \emptyset$. If $S^L_n \setminus \mathcal{I}^\star_n$ is empty, then the result is true. Assume $S^L_n \setminus \mathcal{I}^\star_n$ is not empty. Let $j \in S^L_n \setminus \mathcal{I}^\star_n$ and L_7 as in Lemma 24, hence

$$a_{n+1,j} \leq f\left(c_1(L) D_{\boldsymbol{F}}\right) .$$

As $c_1(x) \sim_{+\infty} x$ and $\lim_{+\infty} f(x) = 0$, there exists a deterministic L_8 such that for all $L \geq L_8$,

$$f\left(c_1(L) D_{\boldsymbol{F}}\right) < \frac{1}{K} .$$

Therefore, for all $L \geq \tilde{L}_7 := \max\{L_7, L_8\}$ and all $j \in S^L_n \setminus \mathcal{I}^\star_n$, we have $a_{n+1,j} < \frac{1}{K}$.

Assume that $\widehat{B}^{\mathrm{TS}}_{n+1} \in S^L_n$. Suppose towards contradiction that $\widehat{B}^{\mathrm{TS}}_{n+1} \notin \mathcal{I}^\star_n$. Then, the above shows that $a_{n+1,\widehat{B}^{\mathrm{TS}}_{n+1}} < \frac{1}{K}$. This is a contradiction with $\widehat{B}^{\mathrm{TS}}_{n+1} \in \arg\max_{i\in[K]} a_{n+1,i}$, hence $\widehat{B}^{\mathrm{TS}}_{n+1} \in \mathcal{I}^\star_n$. Since $\mathbb{E}_{\boldsymbol{F}}[(\tilde{L}_7)^\alpha] \leq (L_8)^\alpha + \mathbb{E}_{\boldsymbol{F}}[(L_7)^\alpha]$, this concludes the proof. □

Property 5 Lemma 27 shows that Property 5 is satisfied by B^{TS}_{n+1}. More precisely, we show that after enough time, the probability for the leader to not be the best arm is decreasing exponentially fast.

Lemma 27. *Assume Property 4 holds. Let Π_n satisfying Property 10, and c_1 therein. There exists N_9 with $\mathbb{E}_{\boldsymbol{F}}[N_9] < +\infty$ such that for all $n \geq N_9$,*

$$\mathbb{P}_{|n}[B^{TS}_{n+1} \neq i^\star(\boldsymbol{F})] \leq (K-1) f\left(c_1\left(\sqrt{\frac{n}{K}}\right) D_{\boldsymbol{F}}\right) ,$$

where $f(x) = (1+x)e^{-x}$ and $D_{\boldsymbol{F}} > 0$ is the problem dependent constant from Lemma 15.

Proof. Let $i^\star = i^\star(\boldsymbol{F})$. Let N_1 as in Property 4, then $N_{n,i} \geq \sqrt{\frac{n}{K}}$ for all $n \geq N_1$. Let L_7 as in Lemma 24. For all $n \geq N_9 = \max\{N_1, K(L_7)^2\}$, Lemma 26 and Property 4 yields that

$$\forall i \neq i^\star, \quad a_{n+1,i} \leq f\left(c_1\left(\sqrt{\frac{n}{K}}\right) D_{\boldsymbol{F}}\right) .$$

Using the definition of B^{TS}_{n+1} in (32), we obtain

$$\mathbb{P}_{|n}[B^{\mathrm{TS}}_{n+1} \neq i^\star] = \sum_{i\neq i^\star} \mathbb{P}_{|n}[B^{\mathrm{TS}}_{n+1} = i] \leq (K-1) \max_{i\neq i^\star} a_{n+1,i} \leq (K-1) f\left(c_1\left(\sqrt{\frac{n}{K}}\right) D_{\boldsymbol{F}}\right) .$$

Since $\mathbb{E}_{\boldsymbol{F}}[N_9] \leq \mathbb{E}_{\boldsymbol{F}}[N_1] + K\mathbb{E}_{\boldsymbol{F}}[(L_7)^2] < +\infty$, this concludes the result. □

D.2.3 RS challenger

Conditioned on $\mathcal{F}_n$, the internal randomness of the Re-Sampling (RS) challenger is parameterized by a sampler Π_n, where $a_{n+1,i} := \mathbb{P}_n[i \in \arg\max_{j\in[K]} \theta_j]$. Given a leader B_{n+1}, the RS challenger is defined by repeatedly sampling $\tilde{\theta} \sim \Pi_n$ until $B_{n+1} \notin \arg\max_{i\in[K]} \tilde{\theta}_i$ and by taking an arm with highest mean for this $\tilde{\theta}$

$$C^{\mathrm{RS}}_{n+1} \in \arg\max_{i\in[K]} \tilde{\theta}_i \not\ni B_{n+1} \quad \text{and} \quad \widehat{C}^{\mathrm{RS}}_{n+1} \in \arg\max_{j\neq \widehat{B}_{n+1}} a_{n+1,j} , \tag{33}$$

where

$$\mathbb{P}_{|n}[C^{\mathrm{RS}}_{n+1} = j | B_{n+1} = i] = \sum_{k=0}^{+\infty} a^k_{n+1,i} a_{n+1,j} = \frac{a_{n+1,j}}{1 - a_{n+1,i}} .$$

Property 3 We prove Property 3 for C^{RS}_{n+1} in Lemma 28 by comparing the rates with which $a_{n+1,i}$ decreases. The effective challenger $\widehat{C}^{\text{RS}}_{n+1}$ is taken as an arm different from $\widehat{B}_{n+1}$ which maximizes $a_{n+1,i}$. Therefore, it is sufficient to show that the sampled enough arms have lower $a_{n+1,i}$ than the mildly under-sampled ones. This will imply that $\widehat{C}^{\text{RS}}_{n+1}$ has to be mildly under-sampled or be an arm with highest true mean among the sampled enough arms.

Lemma 28. *Let Π_n satisfying Properties 9 and 10. Let B_{n+1} be a leader satisfying Property 2. Given $(B_{n+1}, \widehat{B}_{n+1})$, let $(C^{RS}_{n+1}, \widehat{C}^{RS}_{n+1})$ as in (33). Let U^L_n and V^L_n as in (20) and $\mathcal{J}^\star_n = \arg\max_{i\in\overline{V^L_n}} \mu_i$. There exists L_9 with $\mathbb{E}_{\boldsymbol{F}}[L_9] < +\infty$ such that if $L \geq L_9$, for all n such that $U^L_n \neq \emptyset$, $\widehat{B}_{n+1} \notin V^L_n$ implies $\widehat{C}^{RS}_{n+1} \in V^L_n \cup \left(\mathcal{J}^\star_n \setminus \left\{\widehat{B}_{n+1}\right\}\right)$.*

Proof. Let $\mathcal{J}^\star_n = \arg\max_{i\in\overline{V^L_n}} \mu_i$. In the following, we consider $U^L_n \neq \emptyset$ (hence $V^L_n \neq \emptyset$) and $\widehat{B}_{n+1} \in V^L_n$. Let B_{n+1} be a leader satisfying Property 2, and L_0 defined therein. Then, for $L \geq L_0^{4/3}$, we have $\widehat{B}_{n+1} \in \mathcal{J}^\star_n$. If $\widehat{C}^{\text{RS}}_{n+1} \in \mathcal{J}^\star_n \setminus \left\{\widehat{B}_{n+1}\right\}$, we are done. Assume that $\widehat{C}^{\text{RS}}_{n+1} \notin \mathcal{J}^\star_n \setminus \left\{\widehat{B}_{n+1}\right\}$.

Since Π_n satisfies Properties 9 and 10, let L_7 and L_8 as in Lemmas 24 and 25. Then, for all $L \geq \max\{L_0^{4/3}, L_7^{4/3}, L_8^2\}$,

$$\begin{aligned} &\widehat{B}_{n+1} \in \mathcal{J}^\star_n \,, \\ \forall i \in \overline{V^L_n} \setminus \mathcal{J}^\star_n, \quad &a_{n+1,i} \leq f(c_1(L^{3/4})D_{\boldsymbol{F}}) \,, \\ \forall j \in U^L_n, \quad &a_{n+1,j} \geq \frac{e^{-D_0 c_0(\sqrt{L})}}{2^{K-1}} \,. \end{aligned}$$

Since $f(x) = (1+x)e^{-x}$, $c_0(x) \sim_{+\infty} x$ and $c_1(x) \sim_{+\infty} x$, there exists a deterministic L_{10} such that for all $L \geq L_{10}$,

$$f(c_1(L^{3/4})D_{\boldsymbol{F}}) < \frac{e^{-D_0 c_0(\sqrt{L})}}{2^{K-1}} \,.$$

Therefore, for all $L \geq L_9 := \max\{L_0^{4/3}, L_7^{4/3}, L_8^2, L_{10}\}$,

$$\forall (j,i) \in U^L_n \times \left(\overline{V^L_n} \setminus \mathcal{J}^\star_n\right), \quad a_{n+1,j} > a_{n+1,i} \,.$$

As $\widehat{B}_{n+1} \in \mathcal{J}^\star_n$ and $\widehat{C}^{\text{RS}}_{n+1} \notin \mathcal{J}^\star_n \setminus \left\{\widehat{B}_{n+1}\right\}$, the definition $\widehat{C}^{\text{RS}}_{n+1} \in \arg\max_{j\neq\widehat{B}_{n+1}} a_{n+1,j}$ yields that $\widehat{C}^{\text{RS}}_{n+1} \in V^L_n$. Otherwise the above strict inequality would wield a contradiction. Since

$$\mathbb{E}_{\boldsymbol{F}}[L_9] \leq L_{10} + \mathbb{E}_{\boldsymbol{F}}[(L_0)^{4/3}] + \mathbb{E}_{\boldsymbol{F}}[(L_7)^{4/3}] + \mathbb{E}_{\boldsymbol{F}}[(L_8)^2] < +\infty \,,$$

this concludes the proof. □

Property 6 Lemma 29 shows that Property 6 is satisfied by C^{RS}_{n+1}.

Lemma 29. *Assume Property 4 holds. Let Π_n satisfying Properties 10 and 11. Let B_{n+1} be a leader satisfying Property 5. Let $\varepsilon \in (0, \varepsilon_0]$ where ε_0 is a problem dependent constant. Given B_{n+1}, let C^{RS}_{n+1} as in (33). There exists N_{10} with $\mathbb{E}_{\boldsymbol{F}}[N_{10}] < +\infty$ such that for all $n \geq N_{10}$ and all $i \neq i^\star(\boldsymbol{F})$,*

$$\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon \quad \Longrightarrow \quad \mathbb{P}_{|n}[C^{RS}_{n+1} = i \mid B_{n+1} = i^\star(\boldsymbol{F})] \leq h(n) \,, \tag{34}$$

where $h : \mathbb{N}^\star \to (0, +\infty)$ such that $h(n) =_{+\infty} o(n^{-\alpha})$ with $\alpha > 0$.

Proof. Let $\varepsilon > 0$ and $i^\star = i^\star(\boldsymbol{F})$. Let N_1 as in Property 4, then $N_{n,i} \geq \sqrt{\frac{n}{K}}$ for all $n \geq N_1$. Since $i^\star$ is unique, we have $\Delta := \min_{j\neq i^\star} |\mu_{i^\star} - \mu_j| > 0$. For bounded distributions, $F \mapsto m(F)$ is continuous on $\mathcal{F}$ for the weak convergence. Lemma 14 yields that there exists $N_{11} = Poly(W_2)$ such that for all $n \geq \max\{N_1, N_{11}\}$ and all $i \in [K]$, we have $|\mu_{n,i} - \mu_i| \leq \frac{\Delta}{4}$. Therefore, for all $n \geq \max\{N_1, N_8\}$, $\arg\max_{i\in[K]} \mu_{n,i} = \arg\max_{i\in[K]} \mu_i = i^\star$.

Let $\xi > 0$. Since Property 4 holds and B_{n+1} satisfies Property 5, we can use the results from Lemma 11. Let N_4 defined in Lemma 11, we have $\left|\frac{N_{n,i^\star}}{n} - \beta\right| \leq \xi$ for all $n \geq \max\{N_1, N_4\}$.

Using the definition of C_{n+1}^{RS} in (33), we have

$$\mathbb{P}_{|n}[C_{n+1}^{\mathrm{RS}} = i \mid B_{n+1} = i^\star] = \frac{a_{n+1,i}}{1 - a_{n+1,i^\star}} \leq \frac{\mathbb{P}_n[\theta_i \geq \theta_{i^\star}]}{\max_{j \neq i^\star} \mathbb{P}_n[\theta_j \geq \theta_{i^\star}]} ,$$

where we used that $\{\theta_j > \theta_{i^\star}\} \subseteq \bigcup_{j \neq i^\star} \{\theta_j > \theta_{i^\star}\} = \{i^\star \notin \arg\max_{j \in [K]} \theta_j\}$.

Let $i \neq i^\star$ such that $\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon$. Using Lemma 5, there exists $N_{12} = Poly(W_1)$, such that for all $n \geq \max\{N_1, N_{12}\}$, we have $\frac{N_{n,i}}{n} \geq w_i^\beta + \frac{\varepsilon}{2}$. Therefore, for all $n \geq \max_{i \in \{1,4,11,12\}} N_i$,

Let $f(x) = (1+x)e^{-x}$. Since Π_n satisfies Property 10, Lemma 23 yields

$$\mathbb{P}_n[\theta_i \geq \theta_{i^\star}] \leq f\left(n \inf_{u \in [0,B]} \left\{ \frac{c_1(N_{n,i^\star})}{n} \mathcal{K}_{\inf}^-(\tilde{F}_{n,i^\star}, u) + \frac{c_1(N_{n,i})}{n} \mathcal{K}_{\inf}^+(\tilde{F}_{n,i}, u) \right\} \right) .$$

Let $\tilde{\varepsilon} > 0$. Since $f(x) =_{+\infty} \mathcal{O}\left(e^{-(1-\tilde{\varepsilon})x}\right)$ and $c_1(x) \sim_{+\infty} x$, there exists deterministic $C_{\tilde{\varepsilon}}$ and N_{13} such that for all $n \geq \max_{i \in \{1,4,11,12,13\}} N_i$,

$$\begin{aligned}
\mathbb{P}_n[\theta_i \geq \theta_{i^\star}] &\leq C_{\tilde{\varepsilon}} \exp\left(-n(1-\tilde{\varepsilon}) \inf_{u \in [0,B]} \left\{ \frac{N_{n,i^\star}}{n} \mathcal{K}_{\inf}^-(\tilde{F}_{n,i^\star}, u) + \frac{N_{n,i}}{n} \mathcal{K}_{\inf}^+(\tilde{F}_{n,i}, u) \right\} \right) \\
&\leq C_{\tilde{\varepsilon}} \exp\left(-n(1-\tilde{\varepsilon}) \inf_{u \in [0,B]} \left\{ \frac{N_{n,i^\star}}{n} \mathcal{K}_{\inf}^-(\tilde{F}_{n,i^\star}, u) + \left(w_i^\beta + \frac{\varepsilon}{2}\right) \mathcal{K}_{\inf}^+(\tilde{F}_{n,i}, u) \right\} \right) .
\end{aligned}$$

Let (h_ε, N_8) as in Property 11. Since h_ε is increasing in both its arguments, we have $h_\varepsilon(N_{n,i^\star}, N_{n,i}) \leq h_\varepsilon(n,n)$ and $N_{n,i^\star} + N_{n,i} \leq n$. Therefore, for all $n \geq \max\{KN_8^2, \max_{i \in \{1,4,11,12\}} N_i\}$,

$$\begin{aligned}
&\max_{j \neq i^\star} \mathbb{P}_n[\theta_j \geq \theta_{i^\star}] \\
&\geq \frac{e^{-\varepsilon n}}{h_\varepsilon(n,n)} \exp\left(-n \min_{j \neq i^\star} \inf_{x \in [0,B]} \left\{ \frac{N_{n,i^\star}}{n} \mathcal{K}_{\inf}^-(F_{i^\star}, x) + \frac{N_{n,j}}{n} \mathcal{K}_{\inf}^+(F_j, x) \right\} \right) \\
&\geq \frac{e^{-\varepsilon n}}{h_\varepsilon(n,n)} \exp\left(-n \sup_{w \in \triangle_K : w_{i^\star} = \frac{N_{n,i^\star}}{n}} \min_{j \neq i^\star} \inf_{x \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, x) + w_j \mathcal{K}_{\inf}^+(F_j, x) \right\} \right) ,
\end{aligned}$$

where we lower bounded by considering the best possible allocation such that $w_{i^\star} = \frac{N_{n,i^\star}}{n}$. For $(\boldsymbol{G}, \tilde{\beta}) \in \mathcal{F}^2 \times [0,1]$, let

$$\begin{aligned}
H_{\tilde{\varepsilon}}(\boldsymbol{G}, \tilde{\beta}) = (1-\tilde{\varepsilon}) \inf_{u \in [0,B]} &\left\{ \tilde{\beta} \mathcal{K}_{\inf}^-(G_1, u) + \left(w_i^\beta + \frac{\varepsilon}{2}\right) \mathcal{K}_{\inf}^+(G_2, u) \right\} \\
&- \sup_{w \in \triangle_K : w_{i^\star} = \tilde{\beta}} \min_{j \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_j \mathcal{K}_{\inf}^+(F_j, u) \right\} .
\end{aligned}$$

Let $\tilde{\boldsymbol{G}}_{n,i^\star,i} = (\tilde{F}_{n,i^\star}, \tilde{F}_{n,i})$. Combining the upper and the lower bound, we obtain for all $n \geq \max\{KN_8^2, \max_{i \in \{1,4,11,12,13\}} N_i\}$,

$$\begin{aligned}
\mathbb{P}_{|n}[C_{n+1}^{\mathrm{RS}} = i \mid B_{n+1} = i^\star] &\leq C_{\tilde{\varepsilon}} h_\varepsilon(n,n) e^{\varepsilon n} \exp\left(-n H_{\tilde{\varepsilon}}\left(\tilde{\boldsymbol{G}}_{n,i^\star,i}, \frac{N_{n,i^\star}}{n} \right) \right) \\
&\leq C_{\tilde{\varepsilon}} h_\varepsilon(n,n) e^{\varepsilon n} \exp\left(-n \inf_{\tilde{\beta} : |\beta - \tilde{\beta}| \leq \xi} H_{\tilde{\varepsilon}}\left(\tilde{\boldsymbol{G}}_{n,i^\star,i}, \tilde{\beta} \right) \right) .
\end{aligned}$$

Using Lemma 31, the functions $(\boldsymbol{G}, \tilde{\beta}) \mapsto H_{\tilde{\varepsilon}}(\boldsymbol{G}, \tilde{\beta})$ and $\boldsymbol{G} \mapsto \inf_{\tilde{\beta} : |\beta - \tilde{\beta}| \leq \xi} H_{\tilde{\varepsilon}}(\boldsymbol{G}, \tilde{\beta})$ are continuous. Let $\boldsymbol{G}_{i^\star,i} = (F_{i^\star}, F_i)$. Therefore, there exists $N_{14} = Poly(W_1)$, $\xi_0 > 0$ and $\tilde{\varepsilon}_0 > 0$ such that for all $n \geq N_{10} := \max\{KN_8^2, \max_{i \in \{1,4,11,12,13,14\}} N_i\}$, all $\xi \in (0, \xi_0]$ and $\tilde{\varepsilon} \in (0, \tilde{\varepsilon}_0]$, we have

$$\inf_{\tilde{\beta} : |\beta - \tilde{\beta}| \leq \xi} H_{\tilde{\varepsilon}}\left(\tilde{\boldsymbol{G}}_{n,i^\star,i}, \tilde{\beta} \right) \geq \frac{1}{2} \inf_{\tilde{\beta} : |\beta - \tilde{\beta}| \leq \xi} H_{\tilde{\varepsilon}}(\boldsymbol{G}_{i^\star,i}, \tilde{\beta}) \geq \frac{1}{4} H_{\tilde{\varepsilon}}(\boldsymbol{G}_{i^\star,i}, \beta) \geq \frac{1}{8} H_0(\boldsymbol{G}_{i^\star,i}, \beta) .$$

In the following, we take such $\xi_0 > 0$ and $\tilde{\varepsilon}_0 > 0$ and $\varepsilon \in (0, \varepsilon_0]$ where $\varepsilon_0 = \frac{1}{16} H_0(\boldsymbol{G}_{i^\star,i}, \beta)$ is a problem dependent constant.

At the β-equilibrium all transportation costs are equal (Lemma 61). Therefore, by definition of w^β,

$$\begin{aligned}
&\sup_{w \in \triangle_K : w_{i^\star} = \beta} \min_{j \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}^-_{\inf}(F_{i^\star}, u) + w_j \mathcal{K}^+_{\inf}(F_j, u) \right\} \\
&\quad = \min_{j \neq i^\star} \inf_{u \in [0,B]} \left\{ \beta \mathcal{K}^-_{\inf}(F_{i^\star}, u) + w_j^\beta \mathcal{K}^+_{\inf}(F_j, u) \right\} \\
&\quad = \inf_{u \in [0,B]} \left\{ \beta \mathcal{K}^-_{\inf}(F_{i^\star}, u) + w_i^\beta \mathcal{K}^+_{\inf}(F_i, u) \right\} \\
&\quad < \inf_{u \in [0,B]} \left\{ \beta \mathcal{K}^-_{\inf}(F_{i^\star}, u) + \left(w_i^\beta + \frac{\varepsilon}{2} \right) \mathcal{K}^+_{\inf}(F_i, u) \right\}
\end{aligned}$$

where the strict inequality is obtained because the transportation costs are strictly increasing in their allocation arguments (Lemma 56). Therefore, we have $H_0(\boldsymbol{G}_{i^\star,i}, \beta) > 0$.

As $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_1}] < +\infty$ and $\mathbb{E}_{\boldsymbol{F}}[e^{\lambda W_2}] < +\infty$ for all $\lambda > 0$, we have $\mathbb{E}_{\boldsymbol{F}}[N_i] < +\infty$ for $i \in \{11, 12, 14\}$ and

$$\mathbb{E}_{\boldsymbol{F}}[N_{10}] \leq N_{13} + K \mathbb{E}_{\boldsymbol{F}}[(N_8)^2] + \sum_{i \in \{1,4,11,12,14\}} \mathbb{E}_{\boldsymbol{F}}[N_i] < +\infty \,.$$

Summarizing, we have shown that for all $\varepsilon \in (0, \varepsilon_0]$, there exists N_{10} with $\mathbb{E}_{\boldsymbol{F}}[N_{10}] < +\infty$ such that for all $n \geq N_{10}$,

$$\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon \quad \Longrightarrow \quad \mathbb{P}_{|n}[C^{\mathrm{RS}}_{n+1} = i \mid B_{n+1} = i^\star(\boldsymbol{F})] \leq h(n) \,,$$

where

$$h(n) := C_{\tilde{\varepsilon}_0} h_\varepsilon(n,n) \exp\left(-\frac{n}{16} H_0(\boldsymbol{G}_{i^\star,i}, \beta) \right) \,.$$

Since $H_0(\boldsymbol{G}_{i^\star,i}, \beta) > 0$, $n \mapsto h_\varepsilon(n,n)$ is decreasing and $h_\varepsilon(n,n) =_{+\infty} o\left(e^{(2n)^\alpha}\right)$ where $\alpha < 1$, we obtain that $h(n) =_{+\infty} o(n^{-\alpha})$ with $\alpha > 0$. It is obvious by definition that $h(n) \in (0, +\infty)$ for all $n \in \mathbb{N}^\star$. □

D.3 Relaxing the distinct means assumption

In Appendix C, we highlighted that Assumption 2 ($\Delta_{\min}(\boldsymbol{F}) > 0$) was only used to show sufficient exploration (see Appendix C.3). We also remarked that the proofs in Appendices C.3 and C.4 work similarly when the amount of exploration $\sqrt{\frac{n}{K}}$ in Lemma 7 and Property 4 is replaced by $\left(\frac{n}{K}\right)^\alpha$ for some arbitrary $\alpha \in (0,1)$. We conjecture that, besides β-EB-TC, all the Top Two algorithms studied in this paper are also asymptotically β-optimal when $\Delta_{\min}(\boldsymbol{F}) = 0$, as detailed below. Let $\Delta_{\min} := \Delta_{\min}(\boldsymbol{F})$.

Lack of robustness of β-EB-TC for $\Delta_{\min} = 0$ For the EB-TC sampling rule, a simple explanation hints that it can dramatically fail empirically, which is confirmed experimentally in Appendix I.2. Let $\boldsymbol{F}$ be a bandit instance in which there are two arms with equal mean that are closest to $\mu_{i^\star}$. At small time, it can happen that the best arm is under-estimated (e.g. when under-sampled) and the two second-best arms have higher empirical mean. In that case, is is very hard for β-EB-TC to recover as it will mostly sample the two second-best arms instead of the best arm. The EB leader will alternate between one of the two second-best arms, depending on the collected samples. Then, given the EB leader, the TC challenger will output the arm with smallest transportation cost. When both second-best arms have higher empirical mean and the best arm is under-estimated, the transportation cost will be smaller between the two second-best arms. Therefore, the TC challenger will propose the second of the two second-best arms. As neither the leader nor the challenger propose to sample the true best arm, it is very hard for β-EB-TC to recover from unlucky first draws.

The condition $\Delta_{\min} > 0$ asymptotically prevents the above situation. When $\mu_i > \mu_j$, the transportation cost between (i,j) grows linearly with $N_{n,i} + N_{n,j}$. Therefore, the transportation cost between

the over-sampled arms will become larger than between the current leader and the best arm, even if it is under-estimated. This ensures that the challenger will propose to sample the best arm, hence allowing the algorithm to eventually recover from unlucky first draws. Based on our analysis, the number of samples required by β-EB-TC to recover from unlucky first draws is a function of $(D_{\boldsymbol{F}})^{-1}$, where $D_{\boldsymbol{F}}$ is a problem dependent constant defined in (35). Extrapolating from results on Gaussian, it is intuitive to expect that small $\Delta_{\min}$ yields small $D_{\boldsymbol{F}}$. Therefore, for small $\Delta_{\min}$, β-EB-TC can need a large number of samples before recovering from unlucky first draws. This undesirable behavior in moderate confidence regime is hidden in the asymptotic analysis. Therefore, we expect β-EB-TC to also suffer from large outliers in the moderate regime, even when $\Delta_{\min} > 0$.

On asymptotic β-optimality for $\Delta_{\min} = 0$ Experiments reported in Appendix I.2.2 reveal that on some instance with $\Delta_{\min} = 0$, the other Top Two instances still have a good performance. We conjecture that using either regularization in the TCI challenger or randomization in the TS leader or RS challenger is adding the right amount of exploration to avoid the undesirable behavior of β-EB-TC described above, and ensure asymptotic β-optimality. More precisely, we conjecture that this amount of exploration is actually logarithmic, and that logarithmic exploration is sufficient to prove β-optimality (which is currently not supported by our analysis).

In particular, for the TS leader it is known from the literature on regret minimization that Thompson Sampling is selecting sub-optimal arms a logarithmic amount of time (at least in expectation) [1]. As for the TCI challenger, we observe that it is designed to avoid the situation described above in which β-EB-TC fails when there are two equal second best arms. When choosing the challenger, we penalize the highly over-sampled arms by adding $\log(N_{n,j})$. While the transportation cost can be very small for two highly sampled arms having similar means, the penalization makes sure that the under-sampled best arm will be selected as the challenger. We conjecture that the TCI challenger ensures an implicit logarithmic exploration.

On forced exploration Another natural idea to prove asymptotic β-optimality when $\Delta_{\min} = 0$ is to add some small amount of forced exploration to the algorithm. A round n, if there exists an arm i such that $N_{n,i} < n^{\alpha}$ (for some small value of α), we draw this arm. This will make Property 4 hold for an exploration level $(n/K)^{\alpha}$. However, forced exploration can be wasteful as it is agnostic to $\mathcal{F}_n$ and all under-sampled arms should not be drawn equally. Our experiments confirm that it is actually not needed for most Top Two algorithms.

Concurrently to our work, [33] introduces and studies the TT-SPRT algorithm for general SPEF. In our terminology, it corresponds to the β-EB-TC algorithm with an added forced exploration in $\sqrt{n/K}$. As expected, adding forced exploration allows to obtain asymptotic β-optimality even for instances where $\Delta_{\min} = 0$. By adding forced exploration, their result also holds for SPEF which are not sub-exponential distributions. In our work, the sub-exponential assumption is made to control the concentration towards the mean parameter. Controlling the concentration rate is of the upmost importance to prove sufficient exploration. Therefore, while this fact is not a direct consequence of our unified analysis, it is not surprising.

D.4 Technicalities

We present some technical results used in the above proofs. Those technicalities are direct corollaries of properties on $\mathcal{K}_{\inf}^{\pm}$ obtained in the Appendix F.

Lemma 30. *There exists $\alpha > 0$ such that*

$$D_{\boldsymbol{F}} = \min_{(i,j):m(F_i)>m(F_j)} \inf_{\substack{G_i,G_j:\\ \forall k\in\{i,j\},\|G_k-F_k\|_\infty\le\alpha}} \inf_{u\in[0,B]} \left\{\mathcal{K}_{\inf}^{-}(G_i,u) + \mathcal{K}_{\inf}^{+}(G_j,u)\right\} > 0\,. \quad (35)$$

Proof. Using Lemma 54 for $w_1 = w_2 = 1$, we have that

$$\boldsymbol{F} \mapsto \inf_{u\in[0,B]} \left\{\mathcal{K}_{\inf}^{-}(F_i,u) + \mathcal{K}_{\inf}^{+}(F_j,u)\right\}$$

is continuous on $\mathcal{F}^K$. Since it has strictly positive values when $m(F_i) > m(F_j)$ (Lemma 55), there exists α such that

$$\inf_{\substack{G_i,G_j:\\ \forall k\in\{i,j\},\|G_k-F_k\|_\infty\le\alpha}} \inf_{u\in[0,B]} \left\{\mathcal{K}_{\inf}^{-}(G_i,u) + \mathcal{K}_{\inf}^{+}(G_j,u)\right\} > 0\,.$$

Further lower bounding by a finite number of strictly positive constants yields the result. □

For all $i \in [K]$, we define the distributions for which i is among the best arm

$$\mathcal{F}_i^K := \left\{ \boldsymbol{F} \in \mathcal{F}^K \mid i \in i^\star(\boldsymbol{F}) \right\} .$$

Lemma 31. *Let $i^\star \in [K]$, $\boldsymbol{F} \in \mathcal{F}_{i^\star}^K$, $i \neq i^\star$ and $\varphi \in [0,1]$. Define for $\beta \in [0,1]$,*

$$G_i(\boldsymbol{F}, \beta) = \inf_{u \in [0,B]} \left\{ \beta \mathcal{K}_{\inf}^-(F_{i^\star}, u) + \varphi \mathcal{K}_{\inf}^+(F_i, u) \right\} - \sup_{w \in \triangle_K : w_{i^\star} = \beta} \min_{j \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_j \mathcal{K}_{\inf}^+(F_j, u) \right\} .$$

Then, $(\boldsymbol{F}, \beta) \mapsto G_i(\boldsymbol{F}, \beta)$ is continuous on $\mathcal{F}^K \times [0,1]$. Moreover, the function $\boldsymbol{F} \mapsto \inf_{\tilde{\beta} : |\beta - \tilde{\beta}| \leq \xi} G_i(\boldsymbol{F}, \tilde{\beta})$ is continuous on $\mathcal{F}^K$.

Let $\boldsymbol{\nu} \in \mathcal{F}_1^K$, $\boldsymbol{F} \in \mathcal{F}^2$ such that $m(F_1) > m(F_2)$, $\alpha > 0$ and $\varphi \in [0,1]$. Define for $\beta \in [0,1]$,

$$H(\boldsymbol{F}, \beta) = \alpha \inf_{u \in [0,B]} \left\{ \beta \mathcal{K}_{\inf}^-(F_1, u) + \varphi \mathcal{K}_{\inf}^+(F_2, u) \right\} - \sup_{w \in \triangle_K : w_1 = \beta} \min_{i \neq 1} \inf_{u \in [0,B]} \left[w_1 \mathcal{K}_{\inf}^-(\nu_1, u) + w_i \mathcal{K}_{\inf}^+(\nu_i, u) \right] .$$

Then, $(\boldsymbol{F}, \beta) \mapsto H(\boldsymbol{F}, \beta)$ is continuous on $\mathcal{F}^2 \times [0,1]$. Moreover, the function $\boldsymbol{F} \mapsto \inf_{\tilde{\beta} : |\beta - \tilde{\beta}| \leq \xi} H(\boldsymbol{F}, \tilde{\beta})$ is continuous on $\mathcal{F}^2$.

Proof. Since $\bigcup_{i \in [K]} \mathcal{F}_i^K = \mathcal{F}^K$, it is enough to show the property for all $i \in [K]$. Let $i^\star \in [K]$ and $i \neq i^\star$. In the proof of Lemma 58, we have obtained that

$$(\boldsymbol{F}, w) \mapsto \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_i \mathcal{K}_{\inf}^+(F_i, u) \right\}$$

$$\text{and} \quad (\boldsymbol{F}, w) \mapsto \min_{i \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_i \mathcal{K}_{\inf}^+(F_i, u) \right\}$$

are continuous on $\mathcal{F}_{i^\star}^K \times \triangle_K$.

Let $\Phi_{i^\star} : (\boldsymbol{F}, \beta) \mapsto \{ w \in \triangle_K \mid w_{i^\star} = \beta \}$, it is compact valued and non-empty for all $\beta \in [0,1]$. It is also continuous (both lower and upper hemicontinuous). Using the continuity proven above, Berge's theorem yields that

$$(\boldsymbol{F}, \beta) \mapsto \sup_{w \in \triangle_K : w_1 = \beta} \min_{i \neq 1} \inf_{u \in [0,B]} \left[w_1 \mathcal{K}_{\inf}^-(\nu_1, u) + w_i \mathcal{K}_{\inf}^+(\nu_i, u) \right] ,$$

is continuous on $\mathcal{F}_{i^\star}^K \times [0,1]$. Combining with the above continuity results, we obtain that $(\boldsymbol{F}, \beta) \mapsto G_i(\boldsymbol{F}, \beta)$ is continuous on $\mathcal{F}_{i^\star}^K \times [0,1]$ for all $i^\star \in [K]$, hence on $\mathcal{F}^K \times [0,1]$.

Let $\Phi : \boldsymbol{F} \mapsto \left\{ \tilde{\beta} : |\beta - \tilde{\beta}| \leq \xi \right\}$, it is a continuous (constant), compact valued and non-empty correspondence. Using the continuity proven above, Berge's theorem yields that $\boldsymbol{F} \mapsto \inf_{\tilde{\beta} : |\beta - \tilde{\beta}| \leq \xi} G_i(\boldsymbol{F}, \tilde{\beta})$ is continuous on $\mathcal{F}^K$.

Using exactly the same arguments, we obtain that $(\boldsymbol{F}, \beta) \mapsto H(\boldsymbol{F}, \beta)$ is continuous on $\mathcal{F}^2 \times [0,1]$ and $\boldsymbol{F} \mapsto \inf_{\tilde{\beta} : |\beta - \tilde{\beta}| \leq \xi} H(\boldsymbol{F}, \tilde{\beta})$ is continuous on $\mathcal{F}^2$. □

E Concentration

In Appendix E.1, we leverage results on martingales [5] to prove δ-correctness of the threshold (4) from Lemma 2. In Appendix E.2, we derive technical results needed in our analysis of Top Two algorithms based on concentration for sub-Gaussian random variables.

E.1 Calibration for bounded distributions

After proving Lemma 32, we give a threshold for bounded distributions (Lemma 2). The concentration is obtained as a direct corollary of recent work on martingales [5]. We apply their technical result to the case of bounded distributions.

Calibration by concentration Lemma 32 states that δ-correct thresholds can be obtained by concentration results.

Lemma 32. *If with probability $1-\delta$, for all $n \in \mathbb{N}$ and all $i \neq i^\star(\boldsymbol{F})$,*

$$N_{n,i}\mathcal{K}_{\inf}^{-}(F_{n,i},\mu_i) + N_{n,i^\star(\boldsymbol{F})}\mathcal{K}_{\inf}^{+}(F_{n,i^\star(\boldsymbol{F})},\mu_{i^\star(\boldsymbol{F})})) \leq c(n,\delta)\,, \tag{36}$$

then the stopping rule (2) using $c(n,\delta)$ is δ-correct on $\mathcal{F}^K$.

Proof. Let $i^\star = i^\star(\boldsymbol{F})$ and $\hat{\imath}_n = i^\star(\boldsymbol{F}_n)$. The empirical transportation costs in (1) can be rewritten as

$$W_n(\hat{\imath}_n, j) = \inf_{x \leq y} \left[N_{n,\hat{\imath}_n}\mathcal{K}_{\inf}^{-}(F_{n,\hat{\imath}_n},x) + N_{n,j}\mathcal{K}_{\inf}^{+}(F_{n,j},y) \right]\,.$$

Using $j = i^\star$, $x = \mu_i$ and $y = \mu_{i^\star}$, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\tau_\delta < +\infty, \hat{\imath}_{\tau_\delta} \neq i^\star\right) \\
&\leq \mathbb{P}\left(\exists n \in \mathbb{N},\ \exists i \neq i^\star,\ i = \hat{\imath}_n,\ \min_{j \neq i} W_n(i,j) > c(n,\delta)\right) \\
&\leq \mathbb{P}\left(\exists n \in \mathbb{N},\ \exists i \neq i^\star,\ N_{n,i}\mathcal{K}_{\inf}^{-}(F_{n,i},\mu_i) + N_{n,i^\star}\mathcal{K}_{\inf}^{+}(F_{n,i^\star},\mu_{i^\star}) > c(n,\delta)\right)\,.
\end{aligned}$$

□

Concentration of $\mathcal{K}_{\inf}$ The key technical result, which was extracted from [5], is reproduced in Lemma 33.

Lemma 33 (Lemma E.1 in [5])**.** *Let a compact and convex set $\Lambda \subseteq \mathbb{R}^d$, and q be the uniform distribution on Λ. Let $g_t : \Lambda \mapsto \mathbb{R}$ be any series of exp-concave functions. Then,*

$$\max_{\boldsymbol{\lambda} \in \Lambda} \sum_{k=1}^{n} g_k(\boldsymbol{\lambda}) \leq \log \mathbb{E}_{\boldsymbol{\lambda} \sim q}\left[e^{\sum_{k=1}^{n} g_t(\boldsymbol{\lambda})}\right] + d\log(n+1) + 1$$

We are now ready to prove Lemma 2.

Proof. For all $(n,i) \in \mathbb{N} \times [K]$, we denote by $(X_{k,i})_{k \in [N_{n,i}]}$ the samples collected on arm i. Let $i^\star = i^\star(\boldsymbol{F})$ and $i \in [K] \setminus \{i^\star\}$. Using Theorem 3, we obtain

$$\begin{aligned}
N_{n,i^\star}\mathcal{K}_{\inf}^{+}(F_{n,i^\star},\mu_{i^\star})) &= \max_{\lambda \in \left[0, \frac{1}{B-\mu_{i^\star}}\right]} \sum_{k \in [N_{n,i^\star}]} \log(1 - \lambda(X_{k,i^\star} - \mu_{i^\star}))\,, \\
N_{n,i}\mathcal{K}_{\inf}^{-}(F_{n,i},\mu_i)) &= \max_{\lambda \in \left[0, \frac{1}{\mu_i}\right]} \sum_{k \in [N_{n,i}]} \log(1 + \lambda(X_{k,i} - \mu_i))\,.
\end{aligned}$$

Let q_i^+ and q_i^- be the uniform distributions over $\left[0, \frac{1}{B-\mu_i}\right]$ and $\left[0, \frac{1}{\mu_i}\right]$, which are compact and convex sets of $\mathbb{R}$. Define

$$\begin{aligned}
L_{n,i} &= \mathbb{E}_{\lambda \sim q_i^-}\left[\prod_{k \in [N_{n,i}]} (1 + \lambda(X_{k,i} - \mu_i)) \mid X_{1,i}, \cdots, X_{N_{n,i},i}\right]\,, \\
U_{n,i} &= \mathbb{E}_{\lambda \sim q_i^+}\left[\prod_{k \in [N_{n,i}]} (1 - \lambda(X_{i,k} - \mu_i)) \mid X_{1,i}, \cdots, X_{N_{n,i},i}\right]\,, \\
Y_{n,i}^- &= N_{n,i}\mathcal{K}_{\inf}^{-}(F_{n,i},\mu_i) - \log(N_{n,i}+1) - 1\,, \\
Y_{n,i}^+ &= N_{n,i}\mathcal{K}_{\inf}^{+}(F_{n,i},\mu_i) - \log(N_{n,i}+1) - 1\,.
\end{aligned}$$

With $d = 1$, using Lemma 33 with the exp-concave functions $g_{k,i}^+(\lambda) = \log(1 - \lambda(X_{k,i} - \mu_i))$ for $k \in [N_{n,i}]$, and $g_{k,i}^-(\lambda) = \log(1 + \lambda(X_{k,i} - \mu_i))$ for $k \in [N_{n,i}]$, yields

$$e^{Y_{n,i}^-} \leq L_{n,i} \qquad \text{and} \qquad e^{Y_{n,i}^+} \leq U_{n,i} \qquad \text{a.s.}$$

Furthermore, it is easy to verify that for each arm $i \in [K]$, $L_{n,i}$ and $U_{n,i}$ are non-negative martingales with unit initial value $L_{0,i} = 1$ and $U_{0,i} = 1$ almost surely. The martingale property is shown directly

by the tower rule (conditioned on the arm sampled at time n) and $\mathbb{E}\left[1 \pm \lambda(X_{N_{n,i},i} - \mu_i)\right] = 1$. Furthermore, they satisfy $\mathbb{E}[U_{n,i}] \leq 1$ and $\mathbb{E}[L_{n,i}] \leq 1$. Thus, $U_{n,i^\star} L_{n,i}$ is a non-negative martingale with unit initial value.

By concavity of log and using $\sum_{j \in \{i,i^\star\}} N_{n,j} \leq n$, we have

$$c(n,\delta) \geq \log\left(\frac{K-1}{\delta}\right) + 2 + \sum_{j \in \{i,i^\star\}} \log\left(N_{n,j}+1\right) .$$

Taking a union bound over $i \neq i^\star$ and using Ville's inequality, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\exists t \in \mathbb{N},\ \exists i \neq i^\star,\ N_{n,i}\mathcal{K}_{\inf}^-(F_{n,i},\mu_i) + N_{n,i^\star}\mathcal{K}_{\inf}^+(F_{n,i^\star},\mu_{i^\star}) > c(n,\delta)\right) \\
&\leq \sum_{i \neq i^\star} \mathbb{P}\left(\exists t \in \mathbb{N},\ Y_{n,i}^- + Y_{n,i^\star}^+ > \log\left(\frac{K-1}{\delta}\right)\right) \\
&\leq \sum_{i \neq i^\star} \mathbb{P}\left(\exists t \in \mathbb{N},\ U_{n,i^\star} L_{n,i} > \frac{K-1}{\delta}\right) \leq \delta\,.
\end{aligned}$$

Combining the above concentration with Lemma 32 yields the result. □

E.2 Sub-Gaussian random variables

We want to exhibit a sub-Gaussian random variables which controls the deviation of various random variables to their means. More precisely, we will prove the existence of W_1 in Lemma 5 and W_2 in Lemma 14.

Definition 3. *A random variable X is said to be sub-Gaussian with constant c if for all $x \geq 0$, $\mathbb{P}(X \geq x) \leq e^{-cx^2/2}$ and for all $x \leq 0$, $\mathbb{P}(X \leq x) \leq e^{-cx^2/2}$.*

We are interested in sub-Gaussian random variable mainly due to the following property.

Lemma 34. *If X is sub-Gaussian, then for all $\lambda \in \mathbb{R}$, $\mathbb{E}[e^{\lambda X}] < \infty$.*

The proof can be found in any textbook dealing with sub-Gaussian random variables, e.g. [42]. We will furthermore use the following classical properties:

- If X and Y are sub-Gaussian and $\alpha \in \mathbb{R}$ then $X + Y$ is sub-Gaussian and αX is sub-Gaussian.
- Bounded random variables are sub-Gaussian.
- If X verifies that for all $x \geq x_1 \geq 0$, $\mathbb{P}(X \geq x) \leq a_1 e^{-c_1 x^2/2}$ and for all $x \leq x_2 \leq 0$, $\mathbb{P}(X \leq x) \leq a_2 e^{-c_2 x^2/2}$, then X is sub-Gaussian.
- The maximum (or minimum) of a finite number of sub-Gaussian random variables is sub-Gaussian.

Lemma 35. *If $(X_n)_{n\in\mathbb{N},n\geq 1}$ are sub-Gaussian random variables with constants (c_n) such that $\inf_n c_n > 0$, then $\sup_n \frac{X_n}{\sqrt{n\log(e+n)/(1+n)}}$ is sub-Gaussian.*

Proof. For $x \geq \sqrt{\frac{8}{\inf_n c_n}}$,

$$\begin{aligned}
\mathbb{P}(\xi \geq x) &\leq \sum_{n=1}^{\infty} \mathbb{P}\left(X_n \geq x\sqrt{\frac{n\log(e+n)}{n+1}}\right) \\
&\leq \sum_{n=1}^{\infty} \exp\left(-(\inf_n c_n)x^2\frac{n\log(e+n)}{n+1}\right) \\
&\leq \sum_{n=1}^{\infty} \exp\left(-\left[2\log(e+n) + \frac{\inf_n c_n}{2}\frac{x^2 n}{n+1}\right]\right) \\
&\leq \sum_{n=1}^{\infty} \exp\left(-\left[2\log(e+n) + \frac{\inf_n c_n}{4}x^2\right]\right) \\
&= \left[\sum_{n=1}^{\infty}\frac{1}{(e+n)^2}\right] e^{-\frac{\inf_n c_n}{4}x^2},
\end{aligned}$$

where we have used that $\frac{n}{n+1} \geq \frac{1}{2}$ and that $\alpha\beta \geq \alpha + \beta$ for $\alpha, \beta \geq 2$. Now for the lower tail, for $x \leq 0$,

$$\mathbb{P}(\xi \leq x) \leq \mathbb{P}\left(X_1 \leq x\sqrt{\frac{\log(1+e)}{2}}\right) \leq \exp\left(-\frac{\log(1+e)}{2}c_1 x^2/2\right).$$

□

Application to our work We will use the following two examples of sub-Gaussian variables. Lemma 36 is a consequence of the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality [31] while Lemma 37 follows from Azuma's inequality.

Lemma 36. *Let $(X_n)_{n\geq 1}$ be i.i.d. random variables with cdf F and let F_n be the empirical distribution function of $(X_i)_{i\in[n]}$. Then for all n, $\sqrt{n}\|F_n - F\|_\infty$ is sub-Gaussian with a constant which does not depend on n.*

Lemma 37. *Let $(X_n)_{n\geq 1}$ be a martingale with mean 0 and c-sub-Gaussian increments and $\mu_n = X_n/n$. Then for all n, $\sqrt{n}\mu_n$ is sub-Gaussian with constant c.*

Theses results permit to establish the concentration results that are used in Appendix C and D.

Lemma (Lemma 5 and 14)**.** *There exists sub-Gaussian random variables W_1 and W_2 such that for all $(n, i) \in \mathbb{N} \times [K]$ with $n \geq K + 1$ (such that all arms are pulled at least once)*

$$|N_{n,i} - \Psi_{n,i}| \leq W_1\sqrt{(n+1)\log(e+n)} \quad a.s.,$$

$$\|F_{n,i} - F_i\|_\infty \leq W_2\sqrt{\frac{\log(e+N_{n,i})}{1+N_{n,i}}} \quad a.s..$$

In particular, $\mathbb{E}\left[e^{\lambda W_i}\right] < +\infty$ for all $\lambda > 0$ and $i \in \{1, 2\}$.

Proof. For the first inequality, we use that $N_{n,i} - \Psi_{n,i}$ is a martingale and combine Lemma 37 with Lemma 35 to get that for all $i \in [K]$

$$W_{1,i} := \sup_{n\geq 1}\frac{|N_{n,i} - \Psi_{n,i}|/\sqrt{n}}{\sqrt{n\log(e+n)/(1+n)}}$$

is sub-Gaussian. Therefore $W_1 = \max_{i\in[K]} W_{1,i}$ is sub-Gaussian and we have, for all (n, i),

$$|N_{n,i} - \Psi_{n,i}| \leq W_1\sqrt{\frac{n^2}{n+1}\log(e+n)} \leq W_1\sqrt{(n+1)\log(e+n)}.$$

For the second inequality, we define $W_{2,i} := \sup_{n\geq K+1}\|F_{n,i} - F_i\|_\infty\sqrt{\frac{1+N_{n,i}}{\log(e+N_{n,i})}}$. Letting $\hat{F}_{n,i}$ be the empirical distribution fo the first n samples from arm i (while $F_{n,i}$ is the empirical distribution of the samples collected up to time n), one can rewrite

$$W_{2,i} = \sup_{n\geq 1}\left\|\hat{F}_{n,i} - F_i\right\|_\infty\sqrt{\frac{1+n}{\log(e+n)}} = \sup_{n\geq 1}\frac{\sqrt{n}\left\|\hat{F}_{n,i} - F_i\right\|_\infty}{\sqrt{n\log(e+n)/(1+n)}}.$$

Combining Lemma 36 with Lemma 35 yields that $W_{2,i}$ is sub-Gaussian for all $i \in [K]$. Then $W_2 = \max_{i\in[K]} W_{2,i}$ is sub-Gaussian and we have, for all (n, i),

$$\|F_{n,i} - F_i\|_\infty \leq W_1 \sqrt{\frac{1 + N_{n,i}}{\log(e + N_{n,i})}}.$$

□

F Kinf for bounded distributions

Here, $\mathcal{F}$ is the set of probability distributions with support in the interval $[0, B]$. The goal of this section is to study the properties of $\mathcal{K}_{\inf}^+$ and $\mathcal{K}_{\inf}^-$, which are functions $\mathcal{F} \times [0, B] \to \mathbb{R}_+$ defined by

$$\mathcal{K}_{\inf}^+(F, \mu) = \inf\{\mathrm{KL}(F, G) \mid G \in \mathcal{F}, \mathbb{E}_G[X] > \mu\} ,$$
$$\mathcal{K}_{\inf}^-(F, \mu) = \inf\{\mathrm{KL}(F, G) \mid G \in \mathcal{F}, \mathbb{E}_G[X] < \mu\} .$$

As a first step, we remark that for $\mu \in (0, B)$ we can rewrite the $\mathcal{K}_{\inf}$ functions using non-strict inequalities, which will be more convenient [18, 17]. We do so and will work in this section on

$$\mathcal{K}_{\inf}^+(F, \mu) = \inf\{\mathrm{KL}(F, G) \mid G \in \mathcal{F}, \mathbb{E}_G[X] \geq \mu\} ,$$
$$\mathcal{K}_{\inf}^-(F, \mu) = \inf\{\mathrm{KL}(F, G) \mid G \in \mathcal{F}, \mathbb{E}_G[X] \leq \mu\} .$$

There is a strong link between these two definitions, which we will use to transport results from one function to the other.

Lemma 38. *Let $F \in \mathcal{F}$, $\mu \in [0, B]$ and let $f : [0, B] \to [0, B]$ be defined by $f(x) = B - x$. Let F^{B-X} be the pushforward measure of F through f. Then*

$$\mathcal{K}_{\inf}^+(F^{B-X}, B - \mu) = \mathcal{K}_{\inf}^-(F, \mu) \quad \text{and} \quad \mathcal{K}_{\inf}^-(F^{B-X}, B - \mu) = \mathcal{K}_{\inf}^+(F, \mu) .$$

Proof. The function f is measurable, bijective and involutive. We have $\mathrm{KL}(F, G) = \mathrm{KL}(F^{B-X}, G^{B-X})$ for all $F, G \in \mathcal{F}$.

$$\begin{aligned}
\mathcal{K}_{\inf}^+(F^{B-X}, B - \mu) &= \inf\{\mathrm{KL}(F^{B-X}, G) \mid G \in \mathcal{F}, \mathbb{E}_G[X] \geq B - \mu\} \\
&= \inf\{\mathrm{KL}(F^{B-X}, G) \mid G \in \mathcal{F}, \mathbb{E}_{G^{B-X}}[X] \leq \mu\} \\
&= \inf\{\mathrm{KL}(F^{B-X}, (G^{B-X})^{B-X}) \mid G \in \mathcal{F}, \mathbb{E}_{G^{B-X}}[X] \leq \mu\} \\
&= \inf\{\mathrm{KL}(F, G^{B-X}) \mid G \in \mathcal{F}, \mathbb{E}_{G^{B-X}}[X] \leq \mu\} \\
&= \inf\{\mathrm{KL}(F, G) \mid G \in \mathcal{F}, \mathbb{E}_G[X] \leq \mu\} \\
&= \mathcal{K}_{\inf}^-(F, \mu) .
\end{aligned}$$

□

Let $\mathcal{F}^+(\mu) = \{G \in \mathcal{F} \mid \mathbb{E}_G[X] \geq \mu\}$ and define $\mathcal{F}^-(\mu)$ similarly.

Lemma 39. *For all $\mu \in [0, B]$, $\mathcal{F}^+(\mu)$ is a nonempty compact convex set (for the weak convergence of measures).*

Proof. It is nonempty since the Dirac distribution at B belongs to the set.

The set $\mathcal{F}$ is compact, hence a sequence of distributions in $\mathcal{F}^+(\mu)$ admits a convergent subsequence. Suppose then that we have a convergent sequence $(F_n)_{n\in\mathbb{N}}$, converging to F, and let's show that $F \in \mathcal{F}^+(\mu)$. We can rewrite $\mathcal{F}^+(\mu) = \{G \in \mathcal{F} \mid \mathbb{E}_G[\max\{X, B\}] \geq \mu\}$. This is useful since the function $x \mapsto \max\{x, B\}$ is bounded from above and continuous. We can thus apply the Portmanteau theorem to write

$$\mathbb{E}_F[\max\{X, B\}] \geq \limsup_n \mathbb{E}_{F_n}[\max\{X, B\}] \geq \mu .$$

We conclude that $F \in \mathcal{F}^+(\mu)$, which is then compact.

To prove convexity, let $F, G \in \mathcal{F}^+(\mu)$ and let $\alpha \in [0, 1]$: $\alpha F + (1 - \alpha)G \in \mathcal{F}$ and

$$\mathbb{E}_{\alpha F+(1-\alpha)G}[X] = \alpha\mathbb{E}_F[X] + (1 - \alpha)\mathbb{E}_G[X] \geq \alpha\mu + (1 - \alpha)\mu = \mu .$$

□

Lemma 40. *$\mu \mapsto \mathcal{F}^+(\mu)$ is an upper hemicontinuous correspondence.*

Proof. Since $\mathcal{F}^+$ is a compact-valued correspondence (Lemma 39), it suffices to show that for all sequences (μ_n) and (Q_n) with $Q_n \in \mathcal{F}^+(\mu_n)$, if $\mu_n \to \mu$ then there exists a convergent subsequence of (Q_n), which converges to $Q \in \mathcal{F}^+(\mu)$ [40, Proposition 9.8, p. 231]. The existence of a convergent subsequence comes from the compactness of $\mathcal{F}$. The limit Q then belongs to $\mathcal{F}$ since $\mathcal{F}$ is closed. We need to show that Q belongs to $\{G \mid \mathbb{E}_G[X] \geq \mu\}$.

We can rewrite $\mathcal{F}^+(\mu) = \{G \in \mathcal{F} \mid \mathbb{E}_G[\max\{X, B\}] \geq \mu\}$ and prove that Q belongs to $\{G \mid \mathbb{E}_G[\max\{X, B\}] \geq \mu\}$. This is useful since the function $x \mapsto \max\{x, B\}$ is bounded from above and continuous. We can thus apply the Portmanteau theorem to write

$$\mathbb{E}_Q[\max\{X, B\}] \geq \limsup_n \mathbb{E}_{Q_n}[\max\{X, B\}] \geq \limsup_n \mu_n = \mu \,.$$

We conclude that $Q \in \mathcal{F}^+(\mu)$. We have proved upper hemicontinuity. □

Lemma 41. *The infimum in the definition of $\mathcal{K}^+_{\inf}$ is attained at a distribution in $\mathcal{F}^+(\mu)$.*

Proof. $G \mapsto \mathrm{KL}(F, G)$ is lower semicontinuous wrt the topology of weak convergence of measures and the set $\mathcal{F}^+(\mu)$ over which the minimization is performed is compact, hence the functions attains its infimum at a point in $\mathcal{F}^+(\mu)$. □

F.1 Duality

For $(\lambda, F, u) \in \mathbb{R}_+ \times \mathcal{F} \times [0, B]$, let $H^+(\lambda, F, u) = \mathbb{E}_F[\log(1 - \lambda(X - u))]$, where we define $\log(x) = -\infty$ for $x \leq 0$. Let $H^-(\lambda, F, u) = \mathbb{E}_F[\log(1 + \lambda(X - u))]$.

Theorem 3. *For all $F \in \mathcal{F}$ and $u \in [0, B]$,*

$$\mathcal{K}^+_{\inf}(F, u) = \sup_{\lambda \in [0, (B-u)^{-1}]} H^+(\lambda, F, u) \,,$$
$$\mathcal{K}^-_{\inf}(F, u) = \sup_{\lambda \in [0, u^{-1}]} H^-(\lambda, F, u) \,.$$

Proof. A proof of this statement for $\mathcal{K}^+_{\inf}$ can be found in any one of [18, 17]. The result for $\mathcal{K}^-_{\inf}$ then follows from Lemma 38. □

Lemma 42 ([18], Lemma 14)**.** *For all $(F, u) \in \mathcal{F} \times [0, B)$, $\mathcal{K}^+_{\inf}(F, u) \leq -\log\left(1 - \frac{u}{B}\right)$.*

Proof. Let $(F, u) \in \mathcal{F} \times [0, B)$. The proof relies on Theorem 3 and $X \geq 0$ for all $X \in \mathrm{supp}(F)$. Using that log is increasing on $(0, +\infty)$ and $\mathbb{E}_F[1] = 1$,

$$\mathcal{K}^+_{\inf}(F, u) = \sup_{\lambda \in [0, (B-u)^{-1}]} H^+(\lambda, F, u) \leq \sup_{\lambda \in [0, (B-u)^{-1}]} \log(1 + \lambda u) = -\log\left(1 - \frac{u}{B}\right) \,.$$

□

F.2 Continuity and differentiability

Lemma 43. *The function $\lambda, F, u \mapsto H^+(\lambda, F, u)$ is upper semicontinuous (jointly in all arguments) on $\mathbb{R}_+ \times \mathcal{F} \times [0, B]$.*

Proof. Let $(\lambda_n, F_n, u_n) \in \mathbb{R}_+ \times \mathcal{F} \times [0, B]$ be a sequence converging to (λ, F, u). We want to prove that

$$\limsup \mathbb{E}_{F_n}[\log(1 - \lambda_n(X - u_n))] \leq \mathbb{E}_F[\log(1 - \lambda(X - u))] \,.$$

By Skorokhod's representation theorem, there exists real random variables $(X_n)_{n \in \mathbb{N}}, X$ defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that the law of X_n is F_n for all $n \in \mathbb{N}$, the law of X is F and (X_n) converges to X almost surely (hence also in probability).

The family (X_n) has compact support, hence it is uniformly integrable. By Vitali's theorem, since X_n is uniformly integrable and converges in probability, it also converges in L^1.

We get that $\lambda_n(X_n - u_n) \xrightarrow{a.s.} \lambda(X-u)$ and $\mathbb{E}[\lambda_n(X_n-u_n)] \to \mathbb{E}[\lambda(X-u)]$. Now we want to translate this into a statement about the log.

Since $-\lambda_n(X_n-u_n) - \log(1-\lambda_n(X_n-u_n)) \geq 0$, by Fatou's lemma,

$$\begin{aligned}
&\mathbb{E}\left[-\lambda(X-u)\right] - \mathbb{E}\left[\limsup \log(1-\lambda_n(X_n-u_n))\right] \\
&= \mathbb{E}\left[\liminf\left(-\lambda_n(X_n-u_n) - \log(1-\lambda_n(X_n-u_n))\right)\right] \\
&\leq \liminf\left(\mathbb{E}\left[-\lambda_n(X_n-u_n) - \log(1-\lambda_n(X_n-u_n))\right]\right) \\
&= \mathbb{E}\left[-\lambda(X-u)\right] - \limsup \mathbb{E}\left[\log(1-\lambda_n(X_n-u_n))\right] .
\end{aligned}$$

Canceling the first term, we get the inequality we were after. □

Theorem 4. *The function $\mathcal{K}^+_{\inf}$ (resp. $\mathcal{K}^-_{\inf}$) is continuous on $\mathcal{F} \times [0, B)$ (resp. $\mathcal{F} \times (0, B]$).*

Proof. We follow the proof method of [5] (which applied to a slightly different setting).

We first prove lower semicontinuity. We want to apply Berge's Maximum Theorem [9, Theorem 2, p. 116] to the correspondence $C(F,u) = \mathcal{F}^+(u)$ and the function $f((F,u),G) = -\mathrm{KL}(F,G)$. We will obtain the upper semicontinuity of $f^*(F,u) := \inf\{f((F,u),G) \mid G \in C(F,u)\} = -\inf\{\mathrm{KL}(F,G) \mid G \in \mathcal{F}^+(u)\}$, which gives us the lower semicontinuity we are after. We need to show that

- $F, u \mapsto C(F,u) = \mathcal{F}^+(u)$ is upper hemicontinuous: this is proved in Lemma 40,
- $F, u, G \mapsto f(F,u,G) = -\mathrm{KL}(F,G)$ is upper semicontinuous (jointly in all arguments): this is true since KL is jointly lower semicontinuous [34].

We have lower semicontinuity on $\mathcal{F} \times [0, B]$.

To prove upper semicontinuity, we first use duality (Theorem 3) to write $\mathcal{K}^+_{\inf}(F,u) = \sup_{\lambda \in [0,(B-u)^{-1}]} H^+(\lambda, F, u)$. Since we want to prove semicontinuity on $\mathcal{F} \times [0, B)$, we can take any $\varepsilon > 0$ and prove it on $\mathcal{F} \times [0, B-\varepsilon]$.

We want to apply Berge's Maximum Theorem [9, Theorem 2, p. 116] to the correspondence $C(F,u) = [0,(B-u)^{-1}]$ and the function $f((F,u),\lambda) = \mathbb{E}_F[\log(1-\lambda(X-u))]$. We will obtain the upper semicontinuity of $f^*(F,u) = \sup\{f((F,u),G) \mid G \in C(F,u)\}$, which is exactly what we are after. We need to show that

- $F, u \mapsto C(F,u) = [0,(B-u)^{-1}]$ is upper hemicontinuous, nonempty and compact,
- $F, u, \lambda \mapsto f(F,u,\lambda) = \mathbb{E}_F[\log(1-\lambda(X-u))]$ is upper semicontinuous (jointly in all arguments): this is true by Lemma 43.

For the first point, nonempty compact values are obvious for $u \leq B - \varepsilon$. The upper hemicontinuity comes from the continuity of the upper bound of the interval. □

Lemma 44. *$\lambda \mapsto H^+(\lambda, F, u)$ is strictly concave on $[0, (B-u)^{-1})$.*

Proof. For $\alpha \in (0,1)$, by strict concavity of the logarithm,

$$\begin{aligned}
H^+(\alpha\lambda + (1-\alpha)\eta, F, u) &= \mathbb{E}_F[\log(\alpha(1-\lambda(X-u)) + (1-\alpha)(1-\eta(X-u)))] \\
&> \mathbb{E}_F[\alpha\log(1-\lambda(X-u)) + (1-\alpha)\log(1-\eta(X-u))] \\
&= \alpha H^+(\lambda, F, u) + (1-\alpha) H^+(\eta, F, u) .
\end{aligned}$$

The restriction to the interval $[0, (B-u)^{-1})$ guarantees that all quantities appearing in logarithms above are finite. □

Lemma 45. *$(\lambda, u) \mapsto H^+(\lambda, F, u)$ is continuous on $\{(\lambda, u) \in \mathbb{R}_+ \times [0, B] \mid \lambda < (B-u)^{-1}\}$.*

Proof. We already have upper semicontinuity by Lemma 43. We only need lower semicontinuity. That is, we need that for $\lambda_n \to \lambda$ and $u_n \to u$ in that set, we have

$$\liminf_n \mathbb{E}_F[\log(1-\lambda_n(X-u_n))] \geq \mathbb{E}_F[\log(1-\lambda(X-u))] .$$

There exists $\varepsilon > 0$ such that $\lambda \leq (B-u)^{-1}(1-\varepsilon)$. Then for n big enough, $\lambda_n \leq (B-u_n)^{-1}(1-\varepsilon/2)$. For all n large enough, we get $\log(1-\lambda_n(X-u_n)) - \log(\varepsilon/2) \geq 0$. By Fatou's lemma,

$$\liminf_n \mathbb{E}_F[\log(1-\lambda_n(X-u_n)) - \log(\varepsilon/2)] \geq \mathbb{E}_F[\log(1-\lambda(X-u)) - \log(\varepsilon/2)] \,.$$

We cancel the $\log(\varepsilon/2)$ term and get the lower semicontinuity. □

Let $\lambda_\star^+(F,u) = \arg\max_{\lambda\in[0,(B-u)^{-1}]} \mathbb{E}_F[\log(1-\lambda(X-u))] = \arg\max_{\lambda\in[0,(B-u)^{-1}]} H^+(\lambda,F,u)$ and $\lambda_\star^-(F,u) = \arg\max_{\lambda\in[0,u^{-1}]} H^-(\lambda,F,u)$.

Lemma 46. *$u \mapsto \lambda_\star^+(F,u)$ is continuous over the set $\{u \in [0,B) \mid \lambda_\star^+(F,u) < (B-u)^{-1}\}$.*

Proof. We first show that for any $\varepsilon \in (0, B^{-1}]$, the function $u \mapsto \arg\max_{\lambda\in[0,(B-u)^{-1}-\varepsilon]} H^+(\lambda,F,u)$ is continuous on $[0,B)$ and the argmax is unique. This is not exactly continuity of $u \mapsto \lambda_\star^+(F,u)$ because of the $[0,(B-u)^{-1}-\varepsilon]$ interval instead of $[0,(B-u)^{-1}]$.

We will apply Berge's Maximum theorem [9, page 116]. For $\varepsilon \in (0, B^{-1}]$, let

$$\begin{aligned}
\varphi(\lambda,u) &= H(\lambda,F,u) \,, \\
\Gamma(u) &= [0,(B-u)^{-1}-\varepsilon] \,, \\
M(u) &= \max\{H(\lambda,u) \mid \lambda \in \Gamma(u)\} \,, \\
\Phi(u) &= \arg\max\{\varphi(\lambda,u) \mid \lambda \in \Gamma(u)\} \,.
\end{aligned}$$

We verify the hypotheses of the theorem for any $\varepsilon' > 0$ and $u < B - \varepsilon'$:

- H is continuous on $\{(\lambda,u) \in \mathbb{R}_+ \times [0,B] \mid \lambda < (B-u)^{-1}\}$, by Lemma 45, which is a domain containing $\Gamma(u) \times \{u\}$ for all u.
- Γ is nonempty (since $(B-u)^{-1} - \varepsilon \geq 0$), compact-valued (since $u \leq B - \varepsilon'$) and continuous.

We obtain that M is continuous on $[0,B)$ and that Φ is upper hemicontinuous.

Now since φ is a strictly concave function of λ (by Lemma 44) and Γ is convex, we can argue as in [40, Theorem 9.17] to prove that Φ is a single-valued upper hemicontinuous correspondence, hence a continuous function.

Now that we have proved the continuity of the argmax restricted to the interval $[0,(B-u)^{-1}-\varepsilon]$, let's prove the continuity of $u \mapsto \lambda_\star^+(F,u)$ over the set $\{u \in [0,B) \mid \lambda_\star^+(F,u) < (B-u)^{-1}\}$.

let $u \in [0,B)$ such that $\lambda_\star^+(F,u) < (B-u)^{-1}$. Then there exists $\varepsilon > 0$ such that

$$\begin{aligned}
\lambda_\star^+(F,u) &= \arg\max_{\lambda\in[0,(B-u)^{-1}-\varepsilon]} H^+(\lambda,F,u) \,, \\
\text{and } \lambda_\star^+(F,u) &\leq (B-u)^{-1} - 3\varepsilon \,.
\end{aligned}$$

Remark that by concavity of H in λ, for all u, if $\arg\max_{\lambda\in[0,(B-u)^{-1}-\varepsilon]} H^+(\lambda,F,u) \neq (B-u)^{-1} - \varepsilon$ then $\lambda_\star^+(F,u) = \arg\max_{\lambda\in[0,(B-u)^{-1}-\varepsilon]} H^+(\lambda,F,u)$.

For v in some neighborhood of u, we have both $(B-v)^{-1} - \varepsilon > (B-u)^{-1} - 2\varepsilon$ and $\arg\max_{\lambda\in[0,(B-v)^{-1}-\varepsilon]} H^+(\lambda,F,v) < (B-u)^{-1} - 2\varepsilon$. This means that for all v in that neighborhood, $\lambda_\star^+(F,v) = \arg\max_{\lambda\in[0,(B-v)^{-1}-\varepsilon]} H^+(\lambda,F,v)$. The continuity of the ε version then gives continuity of $\lambda_\star^+$ at u. □

Lemma 47 (Theorem 6 in [18])**.** *For all $F \in \mathcal{F}$ and $u \in (m(F),B]$, $u \mapsto \mathcal{K}_{\inf}^+(F,u)$ is differentiable and*

$$\frac{\partial \mathcal{K}_{\inf}^+(F,u)}{\partial u} = \lambda_\star^+(F,u) \,. \tag{37}$$

For all $F \in \mathcal{F}$ and $u \in [0,m(F))$, $u \mapsto \mathcal{K}_{\inf}^-(F,u)$ is differentiable and

$$\frac{\partial \mathcal{K}_{\inf}^-(F,u)}{\partial u} = -\lambda_\star^-(F,u) \,. \tag{38}$$

F.3 Convexity

Lemma 48. *The functions $\mathcal{K}_{\inf}^{+}$ and $\mathcal{K}_{\inf}^{-}$ are jointly convex on $\mathcal{F} \times [0, B]$.*

Proof. We prove the result for $\mathcal{K}_{\inf}^{+}$, but the proof for $\mathcal{K}_{\inf}^{-}$ is identical. Let $F_1, F_2 \in \mathcal{F}$, $u_1, u_2 \in [0, B]$ and let $G_1, G_2 \in \mathcal{F}$ be distributions at which the infimum is attained in $\mathcal{K}_{\inf}^{+}(F_1, u_1)$ and $\mathcal{K}_{\inf}^{+}(F_2, u_2)$ respectively (which exist by Lemma 41). For all $\alpha \in [0, 1]$, $\alpha G_1 + (1-\alpha)G_2$ has expectation $\alpha u_1 + (1-\alpha)u_2$. Hence for all $\alpha \in [0, 1]$,

$$\begin{aligned}\mathcal{K}_{\inf}^{+}(\alpha F_1 + (1-\alpha)F_2, \alpha u_1 + (1-\alpha)u_2) &\leq \mathrm{KL}(\alpha F_1 + (1-\alpha)F_2, \alpha G_1 + (1-\alpha)G_2) \\ &\leq \alpha \mathrm{KL}(F_1, G_1) + (1-\alpha)\mathrm{KL}(F_2, G_2) \\ &= \alpha \mathcal{K}_{\inf}^{+}(F_1, u_1) + (1-\alpha)\mathcal{K}_{\inf}^{+}(F_2, u_2)\end{aligned}$$

The first inequality follows from the definition of $\mathcal{K}_{\inf}^{+}$ as an infimum and the second inequality comes from the joint convexity of the Kullback-Leibler divergence. We have proved joint convexity. □

Lemma 49 (Theorem 5 in [18])**.** *Let $F \in \mathcal{F}$ and $u^+(F) = B - \frac{1}{\mathbb{E}_F\left[\frac{1}{B-X}\right]} \geq m(F)$. We have*

$$\lambda_\star^+(F, u) = 0 \iff u \leq m(F)\,, \tag{39}$$

$$u \in (m(F), u^+(F)] \implies \mathbb{E}_F\left[\frac{1}{1 - \lambda_\star^+(F, u)(X-u)}\right] = 1\,, \tag{40}$$

and

$$\lambda_\star^+(F, u) = \frac{1}{B-u} \iff u \geq u^+(F)\,. \tag{41}$$

Proof. First, we have

$$\lambda_\star^+(F, u) = 0 \iff u \leq m(F)\,.$$

If $\lambda_\star^+(F, u) = 0$, then $\mathcal{K}_{\inf}^{+}(F, u) = H^+(0, F, u) = 0$, hence $u \leq m(F)$. The other direction is obtained in Theorem 5 in [18].

Let $u^+(F) = B - \frac{1}{\mathbb{E}_F\left[\frac{1}{B-X}\right]}$ and $u^-(F) = \frac{1}{\mathbb{E}_F\left[\frac{1}{X}\right]}$. Then,

$$u \in (m(F), u^+(F)] \implies \mathbb{E}_F\left[\frac{1}{1 - \lambda_\star^+(F, u)(X-u)}\right] = 1\,,$$

Moreover,

$$\lambda_\star^+(F, u) = \frac{1}{B-u} \iff u \geq u^+(F)\,.$$

If $u \geq u^+(F)$, then $\mathbb{E}_F\left[\frac{B-u}{B-X}\right] \leq 1$, hence $\lambda_\star^+(F, u) = \frac{1}{B-u}$ by Theorem 5 in [18]. Assume that there exists $u < u^+(F)$, such that $\lambda_\star^+(F, u) = \frac{1}{B-u}$. Then, by equation (40), we obtain that $1 = \mathbb{E}_F\left[\frac{1}{1 - \frac{X-u}{B-u}}\right] = \mathbb{E}_F\left[\frac{B-u}{B-X}\right]$. This condition can be rewritten as $u = u(F)$, hence contradicting $u < u^+(F)$. Therefore, we have shown $\lambda_\star^+(F, u) = \frac{1}{B-u}$ implies that $u \geq u^+(F)$. □

Lemma 50. *The function $u \mapsto \mathcal{K}_{\inf}^{+}(F, u)$ is strictly convex on $(m(F), B]$. The function $u \mapsto \mathcal{K}_{\inf}^{-}(F, u)$ is strictly convex on $[0, m(F))$.*

Proof. We prove the result for $\mathcal{K}_{\inf}^{+}$. The proof for $\mathcal{K}_{\inf}^{-}$ is similar.

Using Lemma 47, $u \mapsto \mathcal{K}_{\inf}^{+}(F, u)$ is strictly convex for $u > m(F)$ if and only if $\lambda_\star^+(F, u)$ is increasing for $u > m(F)$.

Using (39), $\lambda_\star^+(F, u)$ is null for $u \leq m(F)$. Therefore, $u \mapsto \mathcal{K}_{\inf}^{\pm}(F, u)$ is not strictly convex on those intervals.

Using (41), we obtain directly that $\lambda_\star^+(F, u)$ is increasing for $u \geq u^+(F)$.

Suppose towards contradiction that $u \mapsto \lambda_\star^+(F,u)$ is not increasing for $(m(F), u^+(F))$. Therefore, there exists an open $\mathcal{O} \subseteq (m(F), u^+(F))$, such that $u \mapsto \lambda_\star^+(F,u)$ is constant on $\mathcal{O}$, i.e. there exists $c_\mathcal{O} \in \left[0, \frac{1}{B - \inf_{u\in\mathcal{O}} u}\right]$ such that $\lambda_\star^+(F,u) = c_\mathcal{O}$. Using (39-41), we know that $c_\mathcal{O} \in \left(0, \frac{1}{B - \inf_{u\in\mathcal{O}} u}\right)$. On $\mathcal{O}$, $u \mapsto \lambda_\star^+(F,u)$ is constant, hence it is continuously differentiable with null derivative. Since $\mathcal{O} \subseteq (m(F), u^+(F))$, (40) defines implicitly $\lambda_\star^+(F,u)$ as satisfying

$$\mathbb{E}_F\left[\frac{1}{1-\lambda_\star^+(F,u)(X-u)}\right] = 1\,.$$

Since $\mathcal{O} \subseteq (m(F), u^+(F))$, we have $\lambda_\star^+(F,u) \in \left(0, \frac{1}{B - \inf_{u\in\mathcal{O}} u}\right)$. Therefore, the function $(u,x) \mapsto \frac{1}{1-\lambda_\star^+(F,u)(x-u)}$ is bounded on $[0,B]\times\mathcal{O}$, hence integrable, and the function $u \mapsto \frac{1}{1-\lambda_\star^+(F,u)(x-u)}$ is continuously differentiable. Moreover, the function $x \mapsto \frac{1}{\left(1-\lambda_\star^+(F,u)(x-u)\right)^2}$ is strictly positive and bounded on $[0,B]$, hence integrable with strictly positive integrable. Having checked all the conditions to interchange the derivative with the expectation, differentiating the above yields

$$0 = \mathbb{E}_F\left[-\frac{\lambda_\star^+(F,u) + (u-X)\frac{\partial\lambda_\star^+(F,u)}{\partial u}}{\left(1-(X-u)\lambda_\star^+(F,u)\right)^2}\right] = -c_\mathcal{O}\mathbb{E}_F\left[\frac{1}{\left(1-(X-u)c_\mathcal{O}\right)^2}\right] < 0\,,$$

where the strict inequality is obtained since we show that $c_\mathcal{O} > 0$ and $\mathbb{E}_F\left[\frac{1}{(1-(X-u)c_\mathcal{O})^2}\right] > 0$. This is a contradiction, hence such $\mathcal{O} \subset (m(F), u^+(F))$ doesn't exist. Therefore, $u \mapsto \lambda_\star^+(F,u)$ is increasing on $(m(F), u^+(F))$.

Since the convexity already gave that $u \mapsto \lambda_\star^+(F,u)$ is increasing on $(m(F), B]$. The fact that $u \mapsto \lambda_\star^+(F,u)$ is increasing on $(m(F), u^+(F))$ and on $[u^+(F), B]$, yields that $u \mapsto \lambda_\star^+(F,u)$ is increasing on $(m(F), B]$. □

Lemma 51. *$u \mapsto \mathcal{K}_{\inf}^+(F,u)$ is equal to zero on $[0, m(F)]$ and increasing on $(m(F), B]$.*

Proof. We already proved that $\mathcal{K}_{\inf}^+(F,u)$ is equal to zero on $[0,m(F)]$. Since $\mathcal{K}_{\inf}^+(F,m(F)) = 0$, $\mathcal{K}_{\inf}^+$ is nonnegative and strictly convex for $u > m(F)$, then $u \mapsto \mathcal{K}_{\inf}^+(F,u)$ is increasing on $(m(F), B]$. □

Lemma 52. *For all $F, G \in \mathcal{F}$ with means $m(F) \le m(G)$ and for all $w \in \mathbb{R}_+^2$.*

- *If $\max\{w_1, w_2\} > 0$, then $\mu \mapsto w_1\mathcal{K}_{\inf}^+(F,\mu) + w_2\mathcal{K}_{\inf}^-(G,\mu)$ is strictly convex on $[m(F), m(G)]$.*
- *If $\min\{w_1, w_2\} > 0$, then $\mu \mapsto w_1\mathcal{K}_{\inf}^+(F,\mu) + w_2\mathcal{K}_{\inf}^-(G,\mu)$ is strictly convex on $[0, B]$.*

Proof. For $\mu \le m(F)$, the function is equal to $w_2\mathcal{K}_{\inf}^-(G,\mu)$, which is strictly convex unless $w_2 = 0$. For $\mu \ge m(G)$, the function is equal to $w_1\mathcal{K}_{\inf}^+(F,\mu)$, which is strictly convex unless $w_1 = 0$. In the interval $(m(F), m(G))$, it is the sum of two convex functions, one of which is strictly convex. Furthermore, the function is continuous at $m(F)$ and $m(G)$. □

F.3.1 More continuity, using convexity

Lemma 53. *Let $F, G \in \mathcal{F}$ with means $m(F) \le m(G)$ in $(0,B)$ and $w \in \mathbb{R}_+^2$. Then*

$$\inf_{\mu\in[0,B]}(w_1\mathcal{K}_{\inf}^+(F,\mu) + w_2\mathcal{K}_{\inf}^-(G,\mu)) = \inf_{\mu\in[m(F),m(G)]}(w_1\mathcal{K}_{\inf}^+(F,\mu) + w_2\mathcal{K}_{\inf}^-(G,\mu))\,.$$

Proof. On $[0,m(F)]$, $\mathcal{K}_{\inf}^+(F,\mu)$ is constant equal to 0 and $\mathcal{K}_{\inf}^-(G,\mu)$ is non-increasing, hence the minimum over that interval is attained at $m(F)$. We argue similarly for the interval $[m(G), B]$. □

Lemma 54. *Let $F, G \in \mathcal{F}$ with means in $(0,B)$ and $w \in \mathbb{R}_+^2$. Then*

1. *$(F, G, w) \mapsto \inf_{\mu\in[0,B]}(w_1\mathcal{K}_{\inf}^+(F,\mu) + w_2\mathcal{K}_{\inf}^-(G,\mu))$ is continuous on $\mathcal{F}\times\mathcal{F}\times\mathbb{R}_+^2$.*

2. *If* $\max\{w_1, w_2\} > 0$, $\mu_\star(F, G, w) = \arg\min_{\mu\in[0,B]}(w_1\mathcal{K}^+_{\inf}(F, \mu) + w_2\mathcal{K}^-_{\inf}(G, \mu))$ *is unique and continuous on* $\mathcal{F} \times \mathcal{F} \times \mathbb{R}^2_+$.

Proof. We can restrict the inf to $[\mu_F, \mu_G]$ by Lemma 53.

We will apply Berge's Maximum theorem [9, page 116]. Let

$$\begin{aligned}
\varphi(\mu, F, G, w) &= -w_1\mathcal{K}^+_{\inf}(F, \mu) - w_2\mathcal{K}^-_{\inf}(G, \mu)\,,\\
\Gamma(F, G, w) &= [\mu_F, \mu_G]\,,\\
M(F, G, w) &= \max\{\varphi(\mu, F, G, w) \mid \mu \in \Gamma(F, G, w)\}\,,\\
\Phi(F, G, w) &= \arg\max\{\varphi(\mu, F, G, w) \mid \mu \in \Gamma(F, G, w)\}\,.
\end{aligned}$$

We verify the hypotheses of the theorem:

- φ is continuous on $[\mu_F, \mu_G] \times \mathcal{F} \times \mathcal{F} \times C$, by Theorem 4 since $\mu_F, \mu_G \in (0, B)$.
- Γ is nonempty, compact-valued and continuous (since constant).

We obtain that M is continuous on $\mathcal{F} \times \mathcal{F} \times \mathbb{R}^2_+$ and that Φ is upper hemicontinuous.

Now since φ is a strictly concave function of μ (by Lemma 52) and Γ is convex, we can argue as in [40, Theorem 9.17] to prove that Φ is a single-valued upper hemicontinuous correspondence, hence a continuous function. □

Lemma 55. *Let* $F, G \in \mathcal{F}$ *with means* $m(F) < m(G)$ *in* $(0, B)$ *and* $w \in \mathbb{R}^2_+$ *such that* $\min\{w_1, w_2\} > 0$. *The value* $\mu_\star(F, G, w) = \arg\min_{\mu\in[0,B]}(w_1\mathcal{K}^+_{\inf}(F, \mu) + w_2\mathcal{K}^-_{\inf}(G, \mu))$ *(unique by Lemma 54) belongs to the interval* $(m(F), m(G))$ *and is such that both* $\mathcal{K}_{\inf}$ *are positive.*

Proof. We know by Lemma 53 that the minimum is attained inside $[m(F), m(G)]$. We only need to exclude the boundaries.

The function $\mu \mapsto w_1\mathcal{K}^+_{\inf}(F, \mu) + w_2\mathcal{K}^-_{\inf}(G, \mu)$ is differentiable on $(m(F), m(G))$ with derivative $w_1\lambda^+_\star(F, \mu) - w_2\lambda^-_\star(G, \mu)$. If we show that this derivatives takes the value zero in the open interval, then we prove the result.

For $\mu > m(F)$ in a neighborhood of $m(F)$, $\lambda^+_\star(F, \mu)$ tends to 0 by continuity (Lemma 46) and $\lambda^-_\star(G, \mu) > \lambda^-_\star(G, \frac{m(F)+m(G)}{2}) > 0$. We get that close to $m(F)$, the derivative is negative. Similarly, we get that the derivative is positive close to $m(G)$. We conclude that the infimum is indeed attained inside the open interval. □

Lemma 56. *Let* $F, G \in \mathcal{F}$ *with means* $m(F) < m(G)$ *in* $(0, B)$ *and* $w_1 > 0$. *Then* $w_2 \mapsto \min_{\mu\in[0,B]}(w_1\mathcal{K}^+_{\inf}(F, \mu) + w_2\mathcal{K}^-_{\inf}(G, \mu))$ *is increasing on* $\mathbb{R}_+$.

Proof. Let $w_2' > w_2 \geq 0$. Then by Lemma 55, since $w_2' > 0$, there exists $\mu' \in [0, B]$ with $\mathcal{K}^-_{\inf}(G, \mu') > 0$ such that $\min_{\mu\in[0,B]}(w_1\mathcal{K}^+_{\inf}(F, \mu) + w_2'\mathcal{K}^-_{\inf}(G, \mu)) = w_1\mathcal{K}^+_{\inf}(F, \mu') + w_2'\mathcal{K}^-_{\inf}(G, \mu')$. Then we have

$$\begin{aligned}
\min_{\mu\in[0,B]}(w_1\mathcal{K}^+_{\inf}(F, \mu) + w_2'\mathcal{K}^-_{\inf}(G, \mu)) &= w_1\mathcal{K}^+_{\inf}(F, \mu') + w_2'\mathcal{K}^-_{\inf}(G, \mu')\\
&> w_1\mathcal{K}^+_{\inf}(F, \mu') + w_2\mathcal{K}^-_{\inf}(G, \mu')\\
&\geq \min_{\mu\in[0,B]}(w_1\mathcal{K}^+_{\inf}(F, \mu) + w_2\mathcal{K}^-_{\inf}(G, \mu))\,.
\end{aligned}$$

□

Lemma 57. *Let* $F, G, F_1, \ldots, F_K \in \mathcal{F}$ *with means in* $(0, B)$.

The function $w \mapsto \min_{\mu\in[0,B]}(w_1\mathcal{K}^+_{\inf}(F, \mu) + w_2\mathcal{K}^-_{\inf}(G, \mu))$ *is concave on* $\mathbb{R}^2_+$.

The function $w \mapsto \min_{j\neq 1}\min_{\mu\in[0,B]}(w_1\mathcal{K}^-_{\inf}(F_1, \mu) + w_j\mathcal{K}^+_{\inf}(F_j, \mu))$ *is concave on* $\mathbb{R}^K_+$.

Proof. These functions are minimums of linear functions, hence concave. □

F.4 Properties of the characteristic time

Let $\boldsymbol{F} \in \mathcal{F}^K$ and $i^\star = i^\star(\boldsymbol{F})$, supposed unique. Recall that

$$
\begin{aligned}
T^\star(\boldsymbol{F})^{-1} &= \sup_{w \in \triangle_K} \min_{i \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_i \mathcal{K}_{\inf}^+(F_i, u) \right\} , \\
w^\star(\boldsymbol{F}) &= \arg\max_{w \in \triangle_K} \min_{i \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_i \mathcal{K}_{\inf}^+(F_i, u) \right\} , \\
T_\beta^\star(\boldsymbol{F})^{-1} &= \sup_{\substack{w \in \triangle_K \\ w_{i^\star} = \beta}} \min_{i \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_i \mathcal{K}_{\inf}^+(F_i, u) \right\} , \\
w_\beta^\star(\boldsymbol{F}) &= \arg\max_{\substack{w \in \triangle_K \\ w_{i^\star} = \beta}} \min_{i \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_i \mathcal{K}_{\inf}^+(F_i, u) \right\} .
\end{aligned}
$$

Lemma 58. *The functions $T^{\star -1}$ and $T_\beta^{\star -1}$ are continuous on $\mathcal{F}^K$. The correspondences $w^\star$ and $w_\beta^\star$ are upper hemicontinuous on $\mathcal{F}^K$ with compact convex values.*

Proof. Let $\mathcal{F}_i^K = \left\{ \boldsymbol{F} \in \mathcal{F}^K \mid i \in i^\star(\boldsymbol{F}) \right\}$. Since $\bigcup_{i \in [K]} \mathcal{F}_i^K = \mathcal{F}^K$, it is enough to show the property for all $\mathcal{F}_i^K$ for $i \in [K]$. Let $i^\star \in [K]$.

First, the function $(w, \boldsymbol{F}) \mapsto \min_{i \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_i \mathcal{K}_{\inf}^+(F_{i^\star}, u) \right\}$ is continuous on $\triangle_K \times \mathcal{F}^K$ by Lemma 54 and the fact that a minimum of continuous functions is continuous. It is concave in w by Lemma 57.

The correspondence $(w, \boldsymbol{F}) \mapsto \triangle_K$ is nonempty compact-valued and continuous (since constant). By Berge's maximum theorem, we get that $T^\star(\boldsymbol{F})^{-1}$ is continuous on $\mathcal{F}_{i^\star}^K$ and that $w^\star(\boldsymbol{F})$ is upper hemicontinuous with compact values. By [40, Theorem 9.17], the concavity of the function being maximized implies that $w^\star(\boldsymbol{F})$ is convex-valued.

The correspondence $(w, \boldsymbol{F}) \mapsto \triangle_K \cap \{w_{i^\star} = \beta\}$ is nonempty compact-valued and continuous (since constant). By Berge's maximum theorem, we get that $T_\beta^\star(\boldsymbol{F})^{-1}$ is continuous on $\mathcal{F}_{i^\star}^K$ and that $w_\beta^\star(\boldsymbol{F})$ is upper hemicontinuous with compact values. By [40, Theorem 9.17], the concavity of the function being maximized implies that $w_\beta^\star(\boldsymbol{F})$ is convex-valued. □

Lemma 59. *If $i^\star(\boldsymbol{F})$ is a singleton and $\beta \in (0,1)$, then $T^\star(\boldsymbol{F})^{-1} > 0$ and $T_\beta^\star(\boldsymbol{F})^{-1} > 0$.*

Proof.

$$
\begin{aligned}
T^\star(\boldsymbol{F})^{-1} &= \sup_{w \in \triangle_K} \min_{i \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_i \mathcal{K}_{\inf}^+(F_i, u) \right\} \\
&\geq \min_{i \neq i^\star} \inf_{u \in [0,B]} \left\{ \frac{1}{K} \mathcal{K}_{\inf}^-(F_{i^\star}, u) + \frac{1}{K} \mathcal{K}_{\inf}^+(F_i, u) \right\} > 0 ,
\end{aligned}
$$

since we proved that the inner infimum is positive for nonzero coefficients and $\mu_i < \mu_{i^\star}$. The proof for $T_\beta^\star$ is similar. □

Lemma 60. *If $i^\star(\boldsymbol{F})$ is a singleton and $\beta \in (0,1)$, then for all $i \in [K]$, $w_i^\star(\boldsymbol{F}) > 0$ and $w_{\beta,i}^\star(\boldsymbol{F}) > 0$.*

Proof. We proceed by contradiction. If $i^\star(\boldsymbol{F})$ is unique and there exists j with $w_j^\star(\boldsymbol{F}) = 0$, then we show $T^\star(\boldsymbol{F})^{-1} = 0$, which is absurd by Lemma 59. If $j = i^\star$ we have

$$
T^\star(\boldsymbol{F})^{-1} = \min_{i \neq i^\star} \inf_{u \in [0,B]} w_i^\star \mathcal{K}_{\inf}^+(F_i, u) \leq \min_{i \neq i^\star} w_i^\star \mathcal{K}_{\inf}^+(F_i, F_i) = 0 .
$$

If $j \neq i^\star$,

$$
\begin{aligned}
T^\star(\boldsymbol{F})^{-1} &= \min_{i \neq i^\star} \inf_{u \in [0,B]} \left\{ w_{i^\star}^\star \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_i^\star \mathcal{K}_{\inf}^+(F_i, u) \right\} \\
&\leq \inf_{u \in [0,B]} \left\{ w_{i^\star}^\star \mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_j^\star \mathcal{K}_{\inf}^+(F_j, u) \right\} \\
&= \inf_{u \in [0,B]} w_{i^\star}^\star \mathcal{K}_{\inf}^-(F_{i^\star}, u) = 0 .
\end{aligned}
$$

A similar proof holds for $T_\beta^\star$. □

Lemma 61. *If $i^\star(\boldsymbol{F})$ is a singleton and $\beta \in (0,1)$, then*

- *for all $i \neq i^\star(\boldsymbol{F})$, $\inf_{u\in[0,B]} \left\{ w^\star_{i^\star}(\boldsymbol{F}) \mathcal{K}^-_{\inf}(F_{i^\star}, u) + w^\star_i(\boldsymbol{F}) \mathcal{K}^+_{\inf}(F_i, u) \right\} = T^\star(\boldsymbol{F})^{-1}$,*
- *for all $i \neq i^\star(\boldsymbol{F})$, $\inf_{u\in[0,B]} \left\{ \beta \mathcal{K}^-_{\inf}(F_{i^\star}, u) + w^\star_{\beta,i}(\boldsymbol{F}) \mathcal{K}^+_{\inf}(F_i, u) \right\} = T^\star_\beta(\boldsymbol{F})^{-1}$,*
- *$w^\star(\boldsymbol{F})$ and $w^\star_\beta(\boldsymbol{F})$ are singletons: the optimal allocations are unique.*

Proof. At the optimal allocations, all w_i are positive. Suppose w.l.o.g. that 1 is the best arm. By dividing by w_1 and defining

$$G_j(x) = \inf_{u\in[0,B]} \left(\mathcal{K}^-_{\inf}(F_{i^\star}, u) + x \mathcal{K}^+_{\inf}(F_j, u) \right) ,$$

we obtain directly that

$$T^\star(\boldsymbol{F})^{-1} = \max_{w\in\triangle_K, w_1>0} w_1 \min_{j\neq 1} G_j\left(\frac{w_j}{w_1}\right) .$$

Let $w^\star \in w^\star(\boldsymbol{F})$. Then, using the above result, we obtain

$$w^\star \in \arg\max_{w\in\triangle_K} w_1 \min_{j\neq 1} G_j\left(\frac{w_j}{w_1}\right)$$

Introducing $x^\star_j = \frac{w^\star_j}{w^\star_1}$ for all $j \neq 1$, using that $\sum_{j\in[K]} w^\star_j = 1$, one has

$$w^\star_1 = \frac{1}{1+\sum_{j=2}^K x^\star_j} \quad \text{and, for } j \geq 2, w^\star_j = \frac{x^\star_j}{1+\sum_{j=2}^K x^\star_j} .$$

If $x^\star$ is unique, then so is $w^\star$.

Since it is optimal, $\{x^\star_j\}_{j=2}^K \in \mathbb{R}^{K-1}$ belongs to

$$\arg\max_{\{x_j\}_{j=2}^K \in \mathbb{R}^{K-1}} \frac{\min_{j\neq 1} G_j(x_j)}{1+\sum_{j=2}^K x_j} \tag{42}$$

Let's show that all the $G_j\left(x^\star_j\right)$ have to be equal. Let $\mathcal{O} = \left\{ i \in [K]\setminus\{1\} \mid G_i(x^\star_i) = \min_{j\neq 1} G_j\left(x^\star_j\right) \right\}$ and $\mathcal{A} = [K]\setminus(\{1\}\cup\mathcal{O})$. Assume that $\mathcal{A} \neq \emptyset$. For all $a \in \mathcal{A}$ and $b \in \mathcal{O}$, one has $G_j\left(x^\star_j\right) > G_i(x^\star_i)$. Using the continuity of the G_j functions and the fact that they are increasing (Lemma 56), there exists $\varepsilon > 0$ such that

$$\forall j \in \mathcal{A}, i \in \mathcal{O}, \quad G_j\left(x^\star_j - \varepsilon/|\mathcal{A}|\right) > G_i(x^\star_i + \varepsilon/|\mathcal{O}|) > G_i(x^\star_i) .$$

We introduce $\bar{x}_j = x^\star_j - \varepsilon/|\mathcal{A}|$ for all $j \in \mathcal{A}$ and $\bar{x}_i = x^\star_i + \varepsilon/|\mathcal{O}|$ for all $i \in \mathcal{O}$, hence $\sum_{j=2}^K \bar{x}_j = \sum_{j=2}^K x^\star_j$. There exists $i \in \mathcal{O}$ such that $\min_{j\neq 1} G_j(\bar{x}_j) = G_i(x^\star_i + \varepsilon/|\mathcal{O}|)$, hence

$$\frac{\min_{j\neq 1} G_j(\bar{x}_j)}{1+\bar{x}_2+\ldots\bar{x}_K} = \frac{G_i(x^\star_i+\varepsilon/|\mathcal{O}|)}{1+x^\star_2+\cdots+x^\star_K} > \frac{G_i(x^\star_i)}{1+x^\star_2+\cdots+x^\star_K} = \frac{\min_{j\neq 1} G_j\left(x^\star_j\right)}{1+x^\star_2+\cdots+x^\star_K} .$$

This is a contradiction with the fact that $x^\star$ belongs to (42). Therefore, we have $\mathcal{A} = \emptyset$.

We have proved that there is a unique value by $y^\star \in \mathbb{R}_+$, such that for all $j \neq 1$, $G_j\left(x^\star_j\right) = y^\star$. Now since G_j is increasing, this defines a unique value for $x^\star_j$, equal to $G_j^{-1}(y^\star)$.

For y in the intersection of the ranges of all G_j, let $x_j(y) = G_j^{-1}(y)$. $y^\star$ belongs to

$$\arg\max_{y\in[0,\min_{j\neq 1}\lim_{+\infty} G_j(x))} \frac{y}{1+\sum_{j\neq 1} x_j(y)} . \tag{43}$$

For $\beta \in (0,1)$, the same results (and proof) hold for $w^\star_\beta(\boldsymbol{F})$ by noting that

$$T^\star_\beta(\boldsymbol{F})^{-1} = \max_{w\in\triangle_K : w_1=\beta} \beta \min_{j\neq 1} G_j\left(\frac{w_j}{\beta}\right) .$$

Let $w^\beta \in w^\star_\beta(\boldsymbol{F})$, since we have equality at the equilibrium, we obtain for all $j \neq 1$,

$$\beta G_j\left(\frac{w_j^\beta}{\beta}\right) = T^\star_\beta(\boldsymbol{F})^{-1}\,,$$

Using the inverse mapping x_j, we obtain for all $j \neq 1$,

$$w_j^\beta = \beta x_j\left(\frac{1}{T^\star_\beta(\boldsymbol{F})\beta}\right)\,.$$

Therefore, we have shown that $w^\star_\beta(\boldsymbol{F}) = \{w^\beta\}$, where

$$w_i^\beta = \begin{cases} \beta x_i\left(\frac{1}{T^\star_\beta(\boldsymbol{F})\beta}\right) & \text{if } i \neq i^\star \\ \beta & \text{else} \end{cases}\,.$$

□

Lemma 62. *$T^\star_{1/2}(\boldsymbol{F}) \leq 2T^\star(\boldsymbol{F})$ and with $\beta^\star = w^\star_{i^\star}(\boldsymbol{F})$,*

$$\frac{T^\star(\boldsymbol{F})^{-1}}{T^\star_\beta(\boldsymbol{F})^{-1}} \leq \max\left\{\frac{\beta^\star}{\beta}, \frac{1-\beta^\star}{1-\beta}\right\}\,.$$

Proof. Define for each non-negative vector $\boldsymbol{\psi} \in \mathbb{R}^K_+$,

$$f(\boldsymbol{\psi}) := \min_{i\neq i^\star(\boldsymbol{F})} \inf_{u\in[0,B]} \left\{\psi_{i^\star}\mathcal{K}^-_{\inf}(F_{i^\star}, u) + \psi_i \mathcal{K}^+_{\inf}(F_i, u)\right\}\,.$$

$T^\star(\boldsymbol{F})^{-1}$ is the maximum of $f(\boldsymbol{\psi})$ over probability vectors $\boldsymbol{\psi}$. Here, we instead define f for all non-negative vectors, and proceed by varying the total budget of measurement effort available $\sum_{a\in[K]}\psi_a$. f is non-decreasing in ψ_i for all i. f is homogeneous of degree 1. That is $f(c\boldsymbol{\psi}) = cf(\boldsymbol{\psi})$ for all $c \geq 1$. For each $c_1, c_2 > 0$ define

$$g\left(c_1, c_2\right) = \max\left\{ f(\boldsymbol{\psi}) \mid \boldsymbol{\psi} \in \mathbb{R}^K_+,\ \psi_{i^\star(\boldsymbol{F})} = c_1,\ \sum_{i\neq i^\star(\boldsymbol{F})} \psi_i \leq c_2,\right\}$$

The function g inherits key properties of f; it is also non-decreasing and homogeneous of degree 1 . We have

$$\begin{aligned} T^\star_\beta(\boldsymbol{F})^{-1} &= \max\left\{ f(\boldsymbol{\psi}) \mid \boldsymbol{\psi} \in \mathbb{R}^K_+,\ \psi_{i^\star(\boldsymbol{F})} = \beta,\ \sum_{i\in[K]} \psi_i = 1\right\} \\ &= \max\left\{ f(\boldsymbol{\psi}) \mid \boldsymbol{\psi} \in \mathbb{R}^K_+,\ \psi_{i^\star(\boldsymbol{F})} = \beta,\ \sum_{i\neq i^\star(\boldsymbol{F})} \psi_i \leq 1-\beta\right\} \\ &= g(\beta, 1-\beta) \end{aligned}$$

where the second equality uses that f is non-decreasing. Similarly, $T^\star(\boldsymbol{F})^{-1} = g\left(\beta^\star, 1-\beta^\star\right)$ where $\beta^\star = w^\star_{i^\star(\boldsymbol{F})}(\boldsymbol{F})$. Setting

$$r := \max\left\{\frac{\beta^\star}{\beta}, \frac{1-\beta^\star}{1-\beta}\right\}$$

implies $r\beta \geq \beta^\star$ and $r(1-\beta) \geq 1-\beta^\star$. Therefore

$$rT^\star_\beta(\boldsymbol{F})^{-1} = rg(\beta, 1-\beta) = g(r\beta, r(1-\beta)) \geq g\left(\beta^\star, 1-\beta^\star\right) = T^\star(\boldsymbol{F})^{-1}\,.$$

Taking $\beta = \frac{1}{2}$, yields that $T^\star(\boldsymbol{F})^{-1} \leq 2\max\{\beta^\star, 1-\beta^\star\}T^\star_{1/2}(\boldsymbol{F})^{-1} \leq 2T^\star_{1/2}(\boldsymbol{F})^{-1}$. □

Lemma 63. *Let $G_i(x) = \inf_{u\in[0,B]}\left(\mathcal{K}^-_{\inf}(F_{i^\star}, u) + x\mathcal{K}^+_{\inf}(F_i, u)\right)$ for $x \in [0, +\infty)$ and $i \neq i^\star$. Then,*

$$\lim_{x\to+\infty} G_i(x) = \mathcal{K}^-_{\inf}(F_{i^\star}, m(F_i))\,.$$

Proof. Let $u_i(x) \in \arg\min_{u\in[0,B]} \left(\mathcal{K}_{\inf}^-(F_{i^\star}, u) + x\mathcal{K}_{\inf}^+(F_i, u)\right)$. It is easy to see that $G_i(0) = 0$ and $u_i(0) = m(F_1)$. Likewise, we have $u_i(x) =_{+\infty} m(F_i) + o(1)$ by considering $(w_{i^\star}, w_i)$ instead of $x_i = \frac{w_i}{w_{i^\star}}$. By continuity of $u \mapsto \mathcal{K}_{\inf}^-(F_{i^\star}, u)$ and using the definition of $u_i(x)$

$$\lim_{x\to+\infty} G_i(x) = \mathcal{K}_{\inf}^-(F_{i^\star}, m(F_i)) + \lim_{x\to+\infty} x\mathcal{K}_{\inf}^+(F_i, u_i(x)) \,.$$

Using the deviations bounds on the $u \mapsto \mathcal{K}_{\inf}^-(F, u)$ (e.g. Lemma 6 in [19]), we obtain that

$$0 < x\mathcal{K}_{\inf}^-(F_i, u_i(x)) \le x\frac{(m(F_i) - u_i(x))^2}{2} \,.$$

Therefore, a sufficient condition to obtain $\lim_{x\to+\infty} x\mathcal{K}_{\inf}^+(F_i, u_i(x)) = 0$ is to show that $u_i(x) =_{+\infty} m(F_i) + o\left(\frac{1}{\sqrt{x}}\right)$. The first order condition of optimality on $u_i(x)$ can be expressed as

$$x\frac{\partial\mathcal{K}_{\inf}^+(F_i, u_i(x))}{\partial u} = -\frac{\partial\mathcal{K}_{\inf}^-(F_{i^\star}, u_i(x))}{\partial u} \iff x\lambda_\star^+(F_i, u_i(x)) = \lambda_\star^-(F_{i^\star}, u_i(x)) \,,$$

where we used Lemma 47 for the equivalent formulation.

Using that $u \mapsto \lambda_\star^-(F, u)$ is decreasing for $u < m(F_{i^\star})$ (Lemma 50 and Lemma 47) yields

$$x\lambda_\star^+(F_i, u_i(x)) = \lambda_\star^-(F_{i^\star}, u_i(x)) \le \lambda_\star^-(F_{i^\star}, m(F_i)) \le \frac{1}{m(F_i)} \,.$$

Using that $\lambda_\star^+(F_i, u) \ge \frac{u-m(F_i)}{u(B-u)}$ (Lemma 12 in [18]) and denoting $y(x) = u_i(x) - m(F_i) =_{+\infty} o(1)$, we obtain

$$\frac{1}{m(F_i)} \ge x\lambda_\star^+(F_i, u_i(x)) \ge \frac{xy(x)}{(m(F_i) + y(x))(B - m(F_i) - y(x))} \,.$$

Suppose towards contradiction that $y(x) = \mathcal{O}(\frac{1}{x})$ doesn't hold, i.e. $\lim_{+\infty} xy(x) = +\infty$. Using that $y(x) =_{+\infty} o(1)$ and taking the limit in the above inequality yields

$$\frac{1}{m(F_i)} \ge \frac{\lim_{+\infty} xy(x)}{m(F_i)(B - m(F_i))} = +\infty \,,$$

which is a direct contradiction. Therefore, we have shown that $y(x) = \mathcal{O}(\frac{1}{x})$. We showed above that a sufficient condition to conclude was $y(x) =_{+\infty} o\left(\frac{1}{\sqrt{x}}\right)$. Therefore, we have obtained that $\lim_{+\infty} x\mathcal{K}_{\inf}^-(F_i, u_i(x)) = 0$, which concludes the proof. □

G Boundary crossing probability bounds

In order to analyze the algorithms presented in this paper, we need to quantify probabilities of the form $\mathbb{P}(\theta_1 \ge \theta_2)$ for θ_1 and θ_2 two independent real random variables. We first show how such bounds can be obtained by quantifying the individual deviations $\mathbb{P}(\theta_i \ge u)$ and $\mathbb{P}(\theta_i \le u)$ for all $u \in \mathbb{R}$ (the so-called Boundary Crossing Probabilities). Then we prove upper and lower bounds on those probabilities when θ_1 and θ_2 are obtained from a Dirichlet sampler.

G.1 From one arm to two

As remarked in Appendix D.2.1, studying BAI randomized algorithms require to control probability of the form $\mathbb{P}(\theta_1 \ge \theta_2)$ where θ_1 and θ_2 are two independent real random variables. Thanks to Lemma 64, it is possible to obtain those by using Boundary Crossing Probability (BCP) bounds, which are extensively studied in the regret minimization literature. Therefore, while it is based on simple calculations, Lemma 64 is a powerful result of independent interest.

Lemma 64. *Let θ_1 and θ_2 be two independent real random variables with cdf F_1 and F_2. Let $x \in \arg\max_{u\in\mathbb{R}} \mathbb{P}(\theta_2 \ge u)\mathbb{P}(\theta_1 \le u)$. Then*

$$\mathbb{P}(\theta_2 \ge x)\mathbb{P}(\theta_1 \le x) \le \mathbb{P}(\theta_2 \ge \theta_1) \le g\big(\mathbb{P}(\theta_2 \ge x)\mathbb{P}(\theta_1 \le x)\big) \,.$$

where $g(u) = u(1 - \log(u))$ for all $u \in [0, 1]$.

Proof. To ease the notation, we introduce the cdfs $F_1(u) = \mathbb{P}(\theta_1 \leq u)$ and $F_2(u) = \mathbb{P}(\theta_2 \leq u)$. We can suppose that there exists $u \in \mathbb{R}$ with $(1 - F_2(u))F_1(u) > 0$. Otherwise the probability of $\theta_2 \geq \theta_1$ is 0, and both bounds are 0 as well. We start by proving the upper bound.

$$\begin{aligned}
\mathbb{P}(\theta_2 \geq \theta_1) &= \int_u \int_v \mathbb{1}\{u \geq v\} dF_1(v) dF_2(u) \\
&= \int_{u \leq x} \int_{v \leq x} \mathbb{1}\{u \geq v\} dF_1(v) dF_2(u) + \int_{u \leq x} \int_{v > x} \mathbb{1}\{u \geq v\} dF_1(v) dF_2(u) \\
&\quad + \int_{u > x} \int_{v \leq x} \mathbb{1}\{u \geq v\} dF_1(v) dF_2(u) + \int_{u > x} \int_{v > x} \mathbb{1}\{u \geq v\} dF_1(v) dF_2(u)
\end{aligned}$$

The second of those four integrals is equal to zero. We now bound integrals 1, 3, and 4.

1. For x such that $F_2(x) < 1$,

$$\begin{aligned}
\int_{u \leq x} \int_{v \leq x} \mathbb{1}\{u \geq v\} dF_1(v) dF_2(u) &= \int_{u \leq x} F_1(u) dF_2(u) \\
&= \int_{u \leq x} \frac{1}{1 - F_2(u)} (1 - F_2(u)) F_1(u) dF_2(u) \\
&\leq \left(\sup_{u \leq x} (1 - F_2(u)) F_1(u) \right) \int_{u \leq x} \frac{1}{1 - F_2(u)} dF_2(u) \\
&= -\log(1 - F_2(x)) \sup_{u \leq x} (1 - F_2(u)) F_1(u) \,.
\end{aligned}$$

3.

$$\int_{u > x} \int_{v \leq x} \mathbb{1}\{u \geq v\} dF_1(v) dF_2(u) = F_1(x)(1 - F_2(x)) \,.$$

4. For x such that $F_1(x) > 0$,

$$\begin{aligned}
\int_{u > x} \int_{v > x} \mathbb{1}\{u \geq v\} dF_1(v) dF_2(u) &= \int_{v > x} (1 - F_2(v)) dF_1(v) \\
&= \int_{v > x} F_1(v)(1 - F_2(v)) \frac{1}{F_1(v)} dF_1(v) \\
&\leq \left(\sup_{v > x} F_1(v)(1 - F_2(v)) \right) \int_{v > x} \frac{1}{F_1(v)} dF_1(v) \\
&= -\log(F_1(x)) \sup_{v > x} F_1(v)(1 - F_2(v)) \,.
\end{aligned}$$

Putting things together:

$$\begin{aligned}
\mathbb{P}(\theta_2 \geq \theta_1) \leq &-\log(1 - F_2(x)) \sup_{u \leq x} (1 - F_2(u)) F_1(u) + F_1(x)(1 - F_2(x)) \\
&- \log(F_1(x)) \sup_{v > x} F_1(v)(1 - F_2(v)) \,.
\end{aligned}$$

Taking for x the argmax over $\mathbb{R}$ (which verifies $F_1(x) > 0$ and $F_2(x) < 1$), we get

$$\mathbb{P}(\theta_2 \geq \theta_1) \leq (1 - F_2(x)) F_1(x) \left[1 - \log((1 - F_2(x)) F_1(x))\right]$$

We now prove the lower bound. For $x \in \mathbb{R}$, by independence of θ_1 and θ_2,

$$\mathbb{P}(\theta_2 \geq \theta_1) \geq \mathbb{P}(\theta_2 \geq x \geq \theta_1) = \mathbb{P}(\theta_2 \geq x)\mathbb{P}(\theta_1 \leq x) = (1 - F_2(x)) F_1(x) \,.$$

□

G.2 Upper bounds

Theorem 5 gives a tight upper bound on the BCP.

Theorem 5. *Let $X = (X_1, \dots, X_n) \in [0, B]^n$, let $\hat{F}_n$ be the corresponding empirical distribution and let $\mu \in \mathbb{R}$. Then*

$$\mathbb{P}_{L\sim \mathrm{Dir}(1^n)}(L^\top X \geq \mu) \leq \exp\left(-n\mathcal{K}^+_{\inf}(\hat{F}_n, \mu)\right) ,$$
$$\mathbb{P}_{L\sim \mathrm{Dir}(1^n)}(L^\top X \leq \mu) \leq \exp\left(-n\mathcal{K}^-_{\inf}(\hat{F}_n, \mu)\right) .$$

Proof. We first prove the bound involving $\mathcal{K}^+_{\inf}$. This proof is extracted from the proof of Lemma 15 of [36].

Let $R_1, \dots, R_n$ be independent exponential random variables with parameter 1.

$$\mathbb{P}_{L\sim \mathrm{Dir}(1^n)}(L^\top X \geq \mu) = \mathbb{P}(\sum_{i=1}^n \frac{R_i}{\sum_j R_j} X_i \geq \mu) = \mathbb{P}(\sum_{i=1}^n R_i(X_i - \mu) \geq 0) .$$

For $t \geq 0$, we can compose with exponentials and use Markov's inequality to obtain

$$\mathbb{P}_{L\sim \mathrm{Dir}(1^n)}(L^\top X \geq \mu) = \mathbb{P}(\exp t \sum_{i=1}^n R_i(X_i - \mu) \geq 1) \leq \mathbb{E}e^{t\sum_{i=1}^n R_i(X_i-\mu)} .$$

By independence, this last expression is equal to $\prod_{i=1}^n \mathbb{E}e^{t(X_i-\mu)R_i}$. By a simple computation (See [36]) we get, for $t \in [0, \frac{1}{X_i-\mu})$ if $X_i \geq \mu$ and for $t \geq 0$ otherwise,

$$\mathbb{E}e^{t(X_i-\mu)R_i} = \frac{1}{1 - t(X_i - \mu)} .$$

We have proved that for all $t \in [0, \frac{1}{B-\mu})$,

$$\mathbb{P}_{L\sim \mathrm{Dir}(1^n)}(L^\top X \geq \mu) \leq \exp\left(-n\frac{1}{n}\sum_{i=1}^n \log(1 - t(X_i - \mu))\right) .$$

This is then also true for t minimizing the right-hand side.

It remains to show that $\sup_{t\in[0,\frac{1}{B-\mu})} \frac{1}{n}\sum_{i=1}^n \log(1 - t(X_i - \mu)) = \mathcal{K}^+_{\inf}(\hat{F}_n, \mu)$.

From [18], Theorem 8, for any distribution F with support in $[0, B]$,

$$\mathcal{K}^+_{\inf}(F, \mu) = \sup_{t\in[0,\frac{1}{B-\mu}]} \mathbb{E}_{X\sim F}[\log(1 - t(X - \mu))] .$$

Applying this to $\hat{F}_n$ gives $\mathcal{K}^+_{\inf}(\hat{F}_n, \mu) = \sup_{t\in[0,\frac{1}{B-\mu}]} \frac{1}{n}\sum_{i=1}^n \log(1 - t(X_i - \mu))$. The only difference with our target is that the supremum is over the closed interval and not the right-open interval, but either the sup is the same by continuity if there is no X_i equal to B, or the value at $1/(B - \mu)$ is $-\infty$ and hence not equal to the sup.

We now prove the bound involving $\mathcal{K}^-_{\inf}$. Let $\hat{F}_n^{B-X}$ be the empirical distribution corresponding to $(B - X_1, \dots, B - X_n)$.

$$\begin{aligned}\mathbb{P}_{L\sim \mathrm{Dir}(1^n)}(L^\top X \leq \mu) &= \mathbb{P}_{L\sim \mathrm{Dir}(1^n)}(L^\top (B - X) \geq B - \mu) \\ &\leq \exp\left(-n\mathcal{K}^+_{\inf}(\hat{F}_n^{B-X}, B - \mu)\right) \\ &= \exp\left(-n\mathcal{K}^-_{\inf}(\hat{F}_n, \mu)\right) .\end{aligned}$$

The last equality follows from Lemma 38. □

Corollary 2. *Let $X = (X_1, \dots, X_n) \in [0, B]^n$, and let $Y = (Y_1, \dots, Y_m) \in [0, B]^m$. let $\hat{F}_{n,X}$ be empirical distribution corresponding to X (and define $\hat{F}_{m,Y}$ similarly). Then*

$$\mathbb{P}_{L_X\sim \mathrm{Dir}(1^n), L_Y\sim \mathrm{Dir}(1^m)}(L_X^\top X \geq L_Y^\top Y) \leq f\left(-\inf_{\mu\in[0,B]}\left(n\mathcal{K}^+_{\inf}(\hat{F}_{n,X}, \mu) + m\mathcal{K}^-_{\inf}(\hat{F}_{m,Y}, \mu)\right)\right) .$$

where $f(x) = (1 + x)e^{-x}$.

Proof. Combine the two bounds of Theorem 5 using Lemma 64. □

G.3 Lower bounds

Lemma 65 gives a first, coarse lower bound on the BCP under a Dirichlet sampler. This result crucially relies on the fact that $\{0, B\}$ have been added to the support.

Lemma 65. *Let $X = (B, 0, X_1 \dots, X_n) \in [0, B]^{n+2}$ and $u \in (0, B)$. Then,*

$$\mathbb{P}_{L\sim\mathrm{Dir}(1^{n+2})}[L^\top X \geq u] \geq \left(1 - \frac{u}{B}\right)^{n+1} \quad \textit{and} \quad \mathbb{P}_{L\sim\mathrm{Dir}(1^{n+2})}[L^\top X \leq u] \geq \left(\frac{u}{B}\right)^{n+1} .$$

Proof. We consider $\tilde{X} = (B, 0, 0 \dots, 0) \in [0, B]^{n+2}$, use that the marginals of Dirichlet are Beta distributions and the Beta-Binomial trick (e.g. [2]) to obtain

$$\begin{aligned}
\mathbb{P}_{L\sim\mathrm{Dir}(1^{n+2})}[L^\top X \geq u] \geq \mathbb{P}_{L\sim\mathrm{Dir}(1^{n+2})}[L^\top \tilde{X} \geq u] &= \mathbb{P}_{w\sim\mathrm{Beta}(1,n+1)}\left[w \geq \frac{u}{B}\right] \\
&= \mathbb{P}_{k\sim\mathrm{Bin}\left(n+1,\frac{u}{B}\right)}[k \leq 0] \\
&= \left(1 - \frac{u}{B}\right)^{n+1} .
\end{aligned}$$

Similarly, considering $\tilde{X} = (B, 0, B \dots, B) \in [0, B]^{n+2}$, we obtain

$$\begin{aligned}
\mathbb{P}_{L\sim\mathrm{Dir}(1^{n+2})}[L^\top X \leq u] \geq \mathbb{P}_{L\sim\mathrm{Dir}(1^{n+2})}[L^\top \tilde{X} \leq u] &= \mathbb{P}_{w\sim\mathrm{Beta}(1,n+1)}\left[w \geq 1 - \frac{u}{B}\right] \\
&= \mathbb{P}_{k\sim\mathrm{Bin}\left(n+1,1-\frac{u}{B}\right)}[k \leq 0] \\
&= \left(\frac{u}{B}\right)^{n+1} .
\end{aligned}$$

□

In the rest of this section, we derive a tighter lower bound on the BCP which leads to a tight lower bound on the probability that one Dirichlet sample exceeds another (Theorem 8). These result rely on a discretization argument and on deriving lower bounds for multinomial distributions.

G.3.1 Multinomial distributions

Theorem 6 gives a tight lower bound on the BCP for multinomial distributions.

Theorem 6. *Let $X_1, \dots, X_M \in [0, B]$ with $X_M = B$ and let $\beta \in \mathbb{N}^M$ with $\beta_i > 0$ for all i. Define $n = \sum_{i=1}^M \beta_i$. For all $\mu \in [0, B]$ and $q \in \triangle_M$ such that $q^\top X \geq \mu$,*

$$\mathbb{P}_{L\sim Dir(\beta)}(L^\top X \geq \mu) \geq \frac{M^{M/2}}{2(8\pi)^{\frac{M-1}{2}}} \frac{1}{n^{\frac{M+1}{2}}} \exp\left(-n\mathrm{KL}_{\mathcal{M}}\left(\frac{\beta}{n}, q\right)\right) .$$

where $\mathrm{KL}_{\mathcal{M}}(p, q)$ is the Kullback-Leibler divergence between multinomial distributions with probability vectors p and q.

Proof. The proof is strongly inspired by the works of [18] and [7] who use lower bound on the BCP for analyzing regret minimization algorithms.

Let $\mathcal{S}_q = \{p \in \triangle_M \mid \forall i \in [M-1], p_i \leq q_i\}$. For $p \in \mathcal{S}_q$, we necessarily have $p_M \geq q_M$ and $p^\top X \geq q^\top X \geq \mu$. From that inequality we get

$$\mathbb{P}_{L\sim\mathrm{Dir}(\beta)}(L^\top X \geq \mu) \geq \mathbb{P}_{L\sim\mathrm{Dir}(\beta)}(L \in \mathcal{S}_q) .$$

We now quantify that probability, using the pdf of a Dirichlet distribution,

$$\begin{aligned}
\mathbb{P}_{L\sim\mathrm{Dir}(\beta)}(L\in\mathcal{S}_q) &= \frac{\Gamma(n)}{\prod_{i=1}^M\Gamma(\beta_i)}\int_{x\in\mathcal{S}_q}\prod_{i=1}^M x_i^{\beta_i-1}dx \\
&\geq \frac{\Gamma(n)}{\prod_{i=1}^M\Gamma(\beta_i)}q_M^{\beta_M-1}\prod_{j=1}^{M-1}\int_{x_j=0}^{q_j}x_j^{\beta_j-1}dx_j \\
&= \frac{\Gamma(n)}{\prod_{i=1}^M\Gamma(\beta_i)}q_M^{\beta_M-1}\prod_{j=1}^{M-1}\frac{q_j^{\beta_j}}{\beta_j} \\
&= \frac{\Gamma(n)}{\prod_{i=1}^M\Gamma(\beta_i)}\frac{\beta_M}{q_M}\prod_{j=1}^{M}\frac{q_j^{\beta_j}}{\beta_j} \\
&\geq \frac{\Gamma(n)}{\prod_{i=1}^M\Gamma(\beta_i)}\prod_{j=1}^{M}\frac{q_j^{\beta_j}}{\beta_j}\,.
\end{aligned}$$

The last line uses that since $\beta_M\geq 1$ and $q_M\leq 1$, $\beta_M/q_M\geq 1$. We transform that last expression to exhibit the Kullback-Leibler divergence between multinomial distributions.

$$\begin{aligned}
\mathbb{P}_{L\sim\mathrm{Dir}(\beta)}(L\in\mathcal{S}_q) &\geq \frac{\Gamma(n)}{\prod_{i=1}^M\Gamma(\beta_i)}\prod_{j=1}^{M}\frac{q_j^{\beta_j}}{\beta_j} \\
&= \frac{1}{n^M}\frac{\Gamma(n)}{\prod_{i=1}^M\Gamma(\beta_i)}\prod_{j=1}^{M}\left(\frac{q_j}{\beta_j/n}\right)^{\beta_j}\prod_{j=1}^{M}\left(\frac{\beta_j}{n}\right)^{\beta_j-1} \\
&= \frac{1}{n^M}\frac{\Gamma(n)}{\prod_{i=1}^M\Gamma(\beta_i)}\prod_{j=1}^{M}\left(\frac{\beta_j}{n}\right)^{\beta_j-1}\exp\left(-n\mathrm{KL}(\frac{\beta}{n},q)\right)\,.
\end{aligned}$$

We now simplify the factor in front of the exponential.

$$\frac{1}{n^M}\frac{\Gamma(n)}{\prod_{i=1}^M\Gamma(\beta_i)}\prod_{j=1}^{M}\left(\frac{\beta_j}{n}\right)^{\beta_j-1} = \frac{1}{n}\frac{n!}{\prod_{i=1}^M\beta_i!}\prod_{j=1}^{M}\left(\frac{\beta_j}{n}\right)^{\beta_j} = \frac{1}{n}\frac{n!}{n^n}\prod_{j=1}^{M}\frac{\beta_j^{\beta_j}}{\beta_j!}\,.$$

We use the following bound on the Stirling approximation of the factorial: $\frac{n!}{\sqrt{2\pi n}(n/e)^n}\in[1,2]$ for all $n\geq 1$. A tighter approximation is possible, but this one is sufficient for our purpose.

$$\begin{aligned}
\frac{1}{n}\frac{n!}{n^n}\prod_{j=1}^{M}\frac{\beta_j^{\beta_j}}{\beta_j!} \geq \frac{1}{n}\sqrt{2\pi n}e^n\prod_{j=1}^{M}\frac{1}{2\sqrt{2\pi\beta_j}e^{\beta_j}} &= \frac{1}{n}\sqrt{2\pi n}\prod_{j=1}^{M}\frac{1}{2\sqrt{2\pi\beta_j}} \\
&\geq \frac{1}{n}\sqrt{2\pi n}\prod_{j=1}^{M}\frac{1}{2\sqrt{2\pi n/M}} \\
&= \frac{M^{M/2}}{2(8\pi)^{\frac{M-1}{2}}}\frac{1}{n^{\frac{M+1}{2}}}\,.
\end{aligned}$$

□

Lemma 66 links $\mathrm{KL}_{\mathcal{M}}$ with $\mathcal{K}_{\inf}^{\pm}$.

Lemma 66. *Let F be a multinomial distribution supported on points $X_1,\dots,X_M\in[0,B]$ and let $p\in\triangle_M$ be the corresponding probability vector.*

If there exists $i\in[M]$ with $X_i=B$ and $p_i>0$, then for all $\mu\in[0,B]$,

$$\mathcal{K}_{\inf}^{+}(F,\mu) = \inf_{q\in\triangle_M:q^\top X\geq\mu}\mathrm{KL}_{\mathcal{M}}(p,q)\,.$$

If there exists $i\in[M]$ with $X_i=0$ and $p_i>0$, then for all $\mu\in[0,B]$,

$$\mathcal{K}_{\inf}^{-}(F,\mu) = \inf_{q\in\triangle_M:q^\top X\leq\mu}\mathrm{KL}_{\mathcal{M}}(p,q)\,.$$

Proof. As remarked in [18], the probability measure that realizes the infimum (which exists by Lemma 41) in the $\mathcal{K}^+_{\inf}$ problem for distributions with finite support has mass on the same points and on B. Under the hypothesis that there exists $i \in [M]$ with $X_i = B$ and $p_i > 0$, we get that that infimum is also a multinomial with same support. Hence there exists q_F such that $\mathcal{K}^+_{\inf}(F, \mu) = \mathrm{KL}_{\mathcal{M}}(p, q_F)$ and we get $\inf_{q \in \triangle_M : q^\top X \geq \mu} \mathrm{KL}_{\mathcal{M}}(p, q) \leq \mathcal{K}^+_{\inf}(F, \mu)$. The reverse inequality comes from the definition of $\mathcal{K}^+_{\inf}$ as an infimum over all probability distributions (which is a larger set than the multinomial distributions).

The proof for $\mathcal{K}^-_{\inf}$ is similar. □

Theorem 7. *Let $X_1, \dots, X_M \in [0, B]$ with $X_M = B$ and let $\beta \in \mathbb{N}^M$ with $\beta_M > 0$. Define $n = \sum_{i=1}^M \beta_i$. Let F_n be the multinomial distributions over the X_i with weights β/n. For all $\mu \in [0, B]$,*

$$\mathbb{P}_{L \sim \mathrm{Dir}(\beta)}(L^\top X \geq \mu) \geq \frac{M^{M/2}}{2(8\pi)^{\frac{M-1}{2}}} \frac{1}{n^{\frac{M+1}{2}}} \exp\left(-n\mathcal{K}^+_{\inf}(F_n, \mu)\right) .$$

Proof. If $\beta_i > 0$ for all $i \in [M]$, this is the result of a supremum over the lower bounds of Theorem 6, to which we apply the equality of Lemma 66. Now if there are some i for which $\beta_i = 0$, we have for some $M_0 < M$,

$$\mathbb{P}_{L \sim \mathrm{Dir}(\beta)}(L^\top X \geq \mu) \geq \frac{M_0^{M_0/2}}{2(8\pi)^{\frac{M_0-1}{2}}} \frac{1}{n^{\frac{M_0+1}{2}}} \exp\left(-n\mathcal{K}^+_{\inf}(F_n, \mu)\right) .$$

But since the leading factor is non-increasing in M_0, we recover the result with M instead of M_0. □

Corollary 3. *Let $X_1, \dots, X_M \in [0, B]$ with $X_M = 0$ and let $\beta \in \mathbb{N}^M$ with $\beta_M > 0$. Define $n = \sum_{i=1}^M \beta_i$. Let F_n be the multinomial distributions over the X_i with weights β/n. For all $\mu \in [0, B]$,*

$$\mathbb{P}_{L \sim \mathrm{Dir}(\beta)}(L^\top X \leq \mu) \geq \frac{M^{M/2}}{2(8\pi)^{\frac{M-1}{2}}} \frac{1}{n^{\frac{M+1}{2}}} \exp\left(-n\mathcal{K}^-_{\inf}(F_n, \mu)\right) .$$

Proof. Remark that $\mathbb{P}_{L \sim \mathrm{Dir}(\beta)}(L^\top X \leq \mu) = \mathbb{P}_{L \sim \mathrm{Dir}(\beta)}(L^\top (B - X) \geq B - \mu)$ and apply Theorem 7 and Lemma 38. □

G.3.2 Lower bound for bounded distributions

Lemma 67. *Let $X_1, \dots, X_n \in [0, B]$. Let $\theta = L^\top X$, where L is a Dirichlet random variables, with $L \sim \mathrm{Dir}(1, \dots, 1)$ (n ones). Let $Y_1, \dots, Y_M \in [0, B]$, among which are the values 0 and B. For all i, let $X_i^+ = \min\{Y_k \mid k \in [M], Y_k \geq X_i\}$. Let F_n^+ be the empirical distribution corresponding to points X_i^+.*

If $0 \in \{X_1, \dots, X_n\}$, then for all $\mu \in [0, B]$,

$$\mathbb{P}(\theta \leq \mu) \geq \frac{M^{M/2}}{2(8\pi)^{\frac{M-1}{2}}} \frac{1}{n^{\frac{M+1}{2}}} \exp\left(-n\mathcal{K}^-_{\inf}(F_n^+, \mu)\right) .$$

If $B \in \{X_1, \dots, X_n\}$, then for all $\mu \in [0, B]$,

$$\mathbb{P}(\theta \geq \mu) \geq \frac{M^{M/2}}{2(8\pi)^{\frac{M-1}{2}}} \frac{1}{n^{\frac{M+1}{2}}} \exp\left(-n\mathcal{K}^+_{\inf}(F_n^-, \mu)\right) .$$

Proof. We have $\theta = L^\top X \leq L^\top X^+$ and $\theta = L^\top X \geq L^\top X^-$. Those scalar products can be written as scalar products of Dirichlet random variables with Y. We now apply Theorem 7 and its corollary. □

For a probability distribution with cdf F on $[0, B]$ and points $0 = x_0 < x_1 < \ldots < x_M < x_{M+1} = B$ we define two discretized distributions with cdf given by, for $x \in [x_m, x_{m+1})$,

$$F^-(x) = \lim_{y \to x_{m+1}, y \leq x_{m+1}} F(y) \,,$$
$$F^+(x) = F(x_m) \,.$$

Lemma 68. *For all $\varepsilon > 0$ and all probability distributions F supported on $[0, B]$, there exists a discretization over at most $2 + \lfloor 1/\varepsilon \rfloor$ points (counting points 0 and B) such that $\|F^- - F\|_\infty \leq \varepsilon$ and $\|F^+ - F\|_\infty \leq \varepsilon$.*

Proof. Let $M = \lfloor 1/\varepsilon \rfloor$. For $m \in \{0, \ldots, M\}$, let $x_m = \inf\{x \in [0, B] \mid F(x) \geq m\varepsilon\}$. Let $x_{M+1} = B$.

$$\begin{aligned}\|F^- - F\|_\infty &\leq \max_{0 \leq m \leq M} \Big| \lim_{y \to x_{m+1}, y \leq x_{m+1}} F(y) - F(x_m) \Big| \mathbb{1}\{x_{m+1} \neq x_m\} \\ &\leq \max_{0 \leq m \leq M} |(m+1)\varepsilon - m\varepsilon| \leq \varepsilon \,.\end{aligned}$$

The computation for F^+ is similar. □

Lemma 69. *For all $F, G \in \mathcal{P}(\mathbb{R})$ with support in $[0, B]$ and all finite discretizations,*

$$\|F^- - G^-\|_\infty \leq \|F - G\|_\infty \,,$$
$$\|F^+ - G^+\|_\infty \leq \|F - G\|_\infty \,.$$

Proof. For all $x \in [x_m, x_{m+1})$, $F^-(x) = \lim_{y \to x_{m+1}, y \leq x_{m+1}} F(y)$.

$$\|F^- - G^-\|_\infty \leq \max_{0 \leq m \leq M} \Big| \lim_{y \to x_{m+1}, y \leq x_{m+1}} F(y) - \lim_{y \to x_{m+1}, y \leq x_{m+1}} G(y) \Big| \leq \|F - G\|_\infty \,.$$

□

Lemma 70. *Let $F \in \mathcal{P}(\mathbb{R})$ with support in $[0, B]$. For all $\varepsilon > 0$, there exists a discretization of $[0, B]$ into $2 + \lfloor 2/\varepsilon \rfloor$ points such that for all G with $\|G - F\| \leq \varepsilon/2$, we have $\|G^- - F\|_\infty \leq \varepsilon$ and $\|G^+ - F\|_\infty \leq \varepsilon$.*

Proof. Let $\varepsilon > 0$. For a discretization of F verifying the result of Lemma 68 for $\varepsilon/2$,

$$\|G^- - F\|_\infty \leq \|G^- - F^-\|_\infty + \|F^- - F\|_\infty \leq \|G - F\|_\infty + \|F^- - F\|_\infty \leq \varepsilon \,.$$

Same computation for G^+. □

Lemma 71 gives a tight lower bound on the BCP for bounded distributions.

Lemma 71. *Let $a > 0$ and $b < B$. Let F be a probability distribution with support in $[0, B]$.*

For points $X, \ldots, X_n \in [0, B]$. let $\theta = L^\top X$, where L is a Dirichlet random variable, with $L \sim \mathrm{Dir}(1, \ldots, 1)$ (n ones). Let F_n be the empirical distribution corresponding to points $(X_i)_{i \in [n]}$.

For all $\varepsilon > 0$, there exists $\eta > 0$ such that for all such empirical distributions (and in particular for all n), if $\|F_n - F\|_\infty \leq \eta$ then for all $u \in [a, b]$,

$$\text{if } B \in \{X_1, \ldots, X_n\} \text{ then } \mathbb{P}(\theta \geq u) \geq \frac{M^{M/2}}{2(8\pi)^{\frac{M-1}{2}}} \frac{1}{n^{\frac{M+1}{2}}} \exp\left(-n\mathcal{K}_{\inf}^+(F, \mu) - n\varepsilon\right) \,,$$
$$\text{if } 0 \in \{X_1, \ldots, X_n\} \text{ then } \mathbb{P}(\theta \leq u) \geq \frac{M^{M/2}}{2(8\pi)^{\frac{M-1}{2}}} \frac{1}{n^{\frac{M+1}{2}}} \exp\left(-n\mathcal{K}_{\inf}^-(F, \mu) - n\varepsilon\right) \,.$$

with $M = 2 + \lfloor 2/\eta \rfloor$.

Proof. The function $(F, \mu) \mapsto \mathcal{K}_{\inf}^-(F, \mu)$ is continuous on $\mathcal{D} \times (0, B)$ by Theorem 4. The function $(F, \mu) \mapsto \mathcal{K}_{\inf}^-(F, \mu)$ is then uniformly continuous on $\mathcal{D} \times [a, b]$ (since $\mathcal{D}$ is compact). In particular, there exists $\eta > 0$ such that if $\|G - F\|_\infty \leq \eta$ then for all $\mu \in [a, b]$,

$$\mathcal{K}_{\inf}^-(G, \mu) \leq \mathcal{K}_{\inf}^-(F, \mu) + \varepsilon \,.$$

We have a similar property for $\mathcal{K}_{\inf}^{+}$.

We now build a discretization such that F_n^+ and F_n^- verify that condition under the hypothesis $\|F_n^{(1)} - F^{(1)}\|_\infty \leq \eta$ and $\|F_n - F\|_\infty \leq \eta$, using Lemma 70. The result is a combination of this continuity inequality and Lemma 67. □

Theorem 8 gives a tight lower bound on probabilities of the form $\mathbb{P}(\theta_2 \geq \theta_1)$ for bounded distributions.

Theorem 8. *Let $F^{(1)}$ and $F^{(2)}$ be two probability distributions with means in $(0, B)$.*

For points $X_1^{(1)}, \dots, X_{n_1}^{(1)}, X_1^{(2)}, \dots, X_{n_2}^{(2)} \in [0, B]$ such that $X_1^{(1)} = 0, X_{n_2}^{(2)} = B$, let $\theta_1 = (L^{(1)})^\top X^{(1)}$ and $\theta_2 = (L^{(2)})^\top X^{(2)}$, where $L^{(1)}$ and $L^{(2)}$ are independent Dirichlet random variables, with $L^{(1)} \sim \mathrm{Dir}(1, \dots, 1)$ (n_1 ones) and $L^{(2)} \sim \mathrm{Dir}(1, \dots, 1)$ (n_2 ones). Let $F_n^{(1)}$ be the empirical distribution corresponding to points $X_i^{(1)}$ and $F_n^{(2)}$ be the empirical distribution corresponding to points $X_i^{(2)}$.

For all $\varepsilon > 0$, there exists $\eta > 0$ such that for all such empirical distributions (and in particular for all n_1 and n_2), if $\|F_n^{(1)} - F^{(1)}\|_\infty \leq \eta$ and $\|F_n^{(2)} - F^{(2)}\|_\infty \leq \eta$ then

$$\mathbb{P}(\theta_2 \geq \theta_1) \geq \frac{M^M}{4(8\pi)^{M-1}} \frac{1}{(n_1 n_2)^{\frac{M+1}{2}}} \exp\left(- \inf_{\mu \in [0,B]} \left(n_1 \mathcal{K}_{\inf}^{-}(F^{(1)}, \mu) + n_2 \mathcal{K}_{\inf}^{+}(F^{(2)}, \mu) \right) - (n_1 + n_2)\varepsilon \right)$$

with $M = 2 + \lfloor 2/\eta \rfloor$.

Proof. By Lemma 64,

$$\mathbb{P}(\theta_2 \geq \theta_1) \geq \sup_{\mu \in [\mu_2, \mu_1]} \mathbb{P}(\theta_2 \geq \mu)\mathbb{P}(\theta_1 \leq \mu)$$

We now use Lemma 71 for both probabilities. This is valid since by hypothesis the interval $[\mu_2, \mu_1]$ is a subset of $(0, B)$. We get the wanted result, except that the infimum is over $\mu \in [\mu_2, \mu_1]$ instead of $\mu \in [0, B]$. But by the monotonicity of $\mathcal{K}_{\inf}^{\pm}$ in μ and the fact that $\mathcal{K}_{\inf}^{+}(F^{(2)}, \mu)$ (resp. $\mathcal{K}_{\inf}^{-}(F^{(1)}, \mu)$) is 0 for $\mu \leq \mu_2$ (resp. for $\mu \geq \mu_1$), we get that

$$\inf_{\mu \in [0,B]} \left(n_1 \mathcal{K}_{\inf}^{-}(F^{(1)}, \mu) + n_2 \mathcal{K}_{\inf}^{+}(F^{(2)}, \mu) \right) = \inf_{\mu \in [\mu_2, \mu_1]} \left(n_1 \mathcal{K}_{\inf}^{-}(F^{(1)}, \mu) + n_2 \mathcal{K}_{\inf}^{+}(F^{(2)}, \mu) \right)$$

and the theorem is proved. □

H Single parameter exponential families

In this section, we explain how our analysis can be used to analyze Top Two algorithms for Single Parameter Exponential Families (SPEF). More precisely, our results apply to SPEF of sub-exponential distributions. We recall below the definition of a sub-exponential distribution, which applies to several typical examples of SPEF: Bernoulli distribution, Gaussian distributions with known variances, exponential and Poissson distributions [42].

Definition 4. *A distribution X is sub-exponential with constant C if it satisfies $\mathbb{P}(|X| \geq x) \leq 2e^{-Cx}$.*

Preliminaries Let $\mathbb{P}^{(0)}$ be a sub-exponential probability distribution and let φ be the cumulant generating function of $\mathbb{P}^{(0)}$, defined by $\varphi(\lambda) = \log \mathbb{E}_{X \sim \mathbb{P}^{(0)}} e^{\lambda X}$. Let $\mathcal{I}^{(n)} \subseteq \mathbb{R}$ be the open interval on which it is defined (set of natural parameters).

The single parameter exponential family (SPEF) associated to a probability measure $\mathbb{P}^{(0)}$ is the set of probability measures $\{\mathbb{P}^{(\lambda)} \mid \lambda \in \mathcal{I}^{(n)}\}$, where $\mathbb{P}^{(\lambda)}$ is the probability measure absolutely continuous with respect to $\mathbb{P}^{(0)}$, with density $x \mapsto e^{\lambda x - \varphi(\lambda)}$ with respect to $\mathbb{P}^{(0)}$. That is, $\log \frac{d\mathbb{P}^{(\lambda)}}{d\mathbb{P}^{(0)}}(x) = \lambda x - \varphi(\lambda)$. Using that the reference probability measure $\mathbb{P}^{(0)}$ is assumed sub-exponential, we can verify that that for all $\lambda \in \mathcal{I}^{(n)}$, $\mathbb{P}^{(\lambda)}$ is also sub-exponential.

φ is an analytic, strictly convex function on $\mathcal{I}^{(n)}$. The distribution $\mathbb{P}^{(\lambda)}$ has mean $\varphi'(\lambda)$. Let $\mathcal{I} = \varphi'(\mathcal{I}^{(n)})$ be the open interval of means of the SPEF. Let φ^* be the convex conjugate of φ, which is also a strictly convex function. Recall that $(\varphi^*)' = (\varphi')^{-1}$. Let d_φ be the Bregman divergence associated to φ and d_{φ^*} be the Bregman divergence associated φ^*. We have, for $\lambda, \eta \in \mathcal{I}^{(n)}$,

$$d_\varphi(\lambda, \eta) = \varphi(\lambda) - \varphi(\eta) - (\lambda - \eta)^\top \varphi'(\eta) = d_{\varphi^*}(\varphi'(\eta), \varphi'(\lambda)) \, .$$

We write $\mathbb{P}_m$ for the unique member of the SPEF with mean m (if it exists, that is if $m \in \mathcal{I}$). It verifies $\mathbb{P}_m = \mathbb{P}^{(\varphi'^{-1}(m))}$. For two distributions in the SPEF with means m_1 and m_2, the Kullback-Leibler divergence between the corresponding distributions $\mathbb{P}_{m_1}$ and $\mathbb{P}_{m_2}$ is

$$\mathrm{KL}(\mathbb{P}_{m_1}, \mathbb{P}_{m_2}) = d_{\varphi^*}(m_1, m_2) \, .$$

In the following, we write simply $d(m_1, m_2)$ for $d_{\varphi^*}(m_1, m_2)$, the Kullback-Leibler divergence between the distributions in the SPEF with means m_1 and m_2.

The $\mathcal{K}_{\inf}$ minimization problem for exponential families In a SPEF, the quantity $\mathcal{K}^+_{\inf}$, infimum of the KL from a member of the SPEF to the subset of the family with mean larger that $\mu \in \mathcal{I}$ becomes

$$\begin{aligned}
\mathcal{K}^+_{\inf}(\mathbb{P}_m, \mu) &:= \inf\{\mathrm{KL}(\mathbb{P}_m, Q) \mid Q \in \{\mathbb{P}_{m'} \mid m' \in \mathcal{I}\}, \, \mathbb{E}_Q[X] \geq \mu\} \\
&= \inf\{\mathrm{KL}(\mathbb{P}_m, \mathbb{P}_{m'}) \mid m' \in \mathcal{I}, \, m' \geq \mu\} \\
&= \inf\{d(m, m') \mid m' \in \mathcal{I}, \, m' \geq \mu\} \\
&= d(m, \max\{m, \mu\}) \, .
\end{aligned}$$

Similarly for all $m, \mu \in \mathcal{I}$, $\mathcal{K}^-_{\inf}(\mathbb{P}_m, \mu) = d(m, \min\{m, \mu\})$. We abuse notations and write $\mathcal{K}^+_{\inf}(m, \mu) = \mathcal{K}^+_{\inf}(\mathbb{P}_m, \mu)$.

Properties of $\mathcal{K}_{\inf}$ in a SPEF The following properties are well-known in the bandit literature, see, e.g. [11]:

1. $\mu \mapsto \mathcal{K}^+_{\inf}(m, \mu)$ is differentiable on $\mathcal{I} \setminus \{m\}$.
2. $\mu \mapsto \mathcal{K}^+_{\inf}(m, \mu)$ is zero for $\mu \leq m$ and finite, increasing and strictly convex on $[m, +\infty) \cap \mathcal{I}$.
3. $\lim_{\mu \to \sup \mathcal{I}} \mathcal{K}^+_{\inf}(m, \mu) = +\infty$.
4. $(m, \mu) \mapsto \mathcal{K}^+_{\inf}(m, \mu)$ is jointly continuous on $\mathcal{I}^2$.

The transportation cost is

$$C_{i,j}(\mathcal{T}(\boldsymbol{F}), w) = \inf_{u \in \mathcal{I}} \left\{ w_i d(m_i, \min\{m_i, u\}) + w_j d(m_j, \max\{m_j, u\}) \right\} \, .$$

The infimum can equivalently be taken over $[m_j, m_i]$. That transportation cost is jointly continuous on $\mathcal{I}^K \times \triangle_K$ by Berge's Maximum theorem: Property 7 is verified.

Concentration results The following result is the counterpart of Lemma 35 for sub-exponential distributions.

Lemma 72. *Suppose that $(X_n)_{n \geq 1}$ are sub-exponential random variables with constants (C_n), such that $c := \inf_n C_n > 0$. Then $\sup_n (X_n / \log(e + n))$ is sub-exponential.*

Proof. This is due to a simple union bound:

$$\begin{aligned}
\mathbb{P}(|\sup_n X_n / \log(e+n)| \geq x) &\leq \sum_{n=1}^{+\infty} \mathbb{P}(|X_n| \geq x \log(e+n)) \\
&\leq 2 \sum_{n=1}^{+\infty} e^{-cx \log(e+n)} \, .
\end{aligned}$$

We now use that for $x \geq 4/c$, we have $cx/2 \geq 2$ and $2\log(e+n) \geq 2$, hence $cx\log(e+n) \geq cx/2 + 2\log(e+n)$: for $x \geq 4/c$,

$$\mathbb{P}(|\sup_n X_n/\log(e+n)| \geq x) \leq 2\left(\sum_{n=1}^{+\infty} \frac{1}{(e+n)^2}\right) e^{-cx/2} .$$

This shows that the random variable is sub-exponential. □

Lemma 73. *There exists a sub-exponential random variable W_d such that almost surely, for all $i \in [K]$ and all n such that $N_{n,i} \geq 1$,*

$$N_{n,i} d(\mu_{n,i}, \mu_i) \leq W_d \log(e + N_{n,i}) .$$

There exists a sub-exponential random variable W_μ such that almost surely, for all $i \in [K]$ and all n such that $N_{n,i} \geq 1$,

$$N_{n,i} |\mu_{n,i} - \mu_i| \leq W_\mu \log(e + N_{n,i}) .$$

In particular, any random variable which is polynomial in W_d or W_μ has a finite expectation.

Proof. We start with the proof of the concentration inequality on the divergence. Since the maximum of a finite number of sub-exponential random variables is sub-exponential, it suffices to show that $\sup_n \frac{N_{n,i} d(\mu_{n,i},\mu_i)}{\log N_{n,i}}$ is sub-exponential. Let then $i \in [K]$. Let $\hat{\mu}_{n,i}$ be the average of the first n samples from arm i. It suffices to show that $\sup_n \frac{n d(\hat{\mu}_{n,i},\mu_i)}{\log(e+n)}$ is sub-exponential. By [32, Lemma 4], we have that for any n,

$$\mathbb{P}(n d(\hat{\mu}_{n,i}, \mu_i) \geq \alpha) \leq 2e^{-\alpha} .$$

That is, for a fixed n, $n d(\hat{\mu}_{n,i}, \mu_i)$ is sub-exponential. We then apply Lemma 72 to obtain than $\sup_n \frac{n d(\hat{\mu}_{n,i},\mu_i)}{\log(e+n)}$ is sub-exponential.

By hypothesis, the distribution F_i is sub-exponential. Hence at any n, $n|\hat{\mu}_{n,i} - \mu_i|$ is as well. We then apply Lemma 72 to obtain than $\sup_n n|\hat{\mu}_{n,i} - \mu_i|/\log(e+n)$ is sub-exponential. We finally obtain that the maximum over the finitely many arms has the same property. □

H.1 Proof of the leader and challenger properties for EB, TC and TCI

We prove, in the case of a sub-exponential SPEF, that Properties 2 and 5 hold for the EB leader and that Properties 3 and 6 hold for the TC and TCI challengers.

For SPEF, Property 7 is a known result from the literature [38]. For sub-exponential SPEF, Property 8 is shown in Lemma 73. In [29], stopping threshold have been derived for general SPEF. Thanks to their dependency in (n, δ), those thresholds are also asymptotically tight. Therefore, applying Corollary 1 yields that β-EB-TC and β-EB-TCI are asymptotically β-optimal algorithms for sub-exponential SPEF with the corresponding threshold.

Rates for empirical transportation costs

Lemma 74. *Let $\boldsymbol{F} \in \mathcal{F}^K$ with $m(F_j) < m(F_i)$. There exists L with $\mathbb{E}[|L|^\alpha] < +\infty$ for all $\alpha > 0$ and $D_{\boldsymbol{F}} > 0$ such that for $N_{n,i} \geq L$ and $N_{n,j} \geq L$, $W_n(i,j) > L D_{\boldsymbol{F}}$.*

Proof. Suppose that $N_{n,i} \geq L$ and $N_{n,j} \geq L$, for some L to be determined. First we get

$$\begin{aligned} W_n(i,j) &= \inf_{u \in \mathcal{I}} \left\{ N_{n,i} \mathcal{K}_{\inf}^-(\mathcal{T}(F_{n,i}), u) + N_{n,j} \mathcal{K}_{\inf}^+(\mathcal{T}(F_{n,j}), u) \right\} \\ &\geq L \inf_{u \in \mathcal{I}} \left\{ \mathcal{K}_{\inf}^-(\mathcal{T}(F_{n,i}), u) + \mathcal{K}_{\inf}^+(\mathcal{T}(F_{n,j}), u) \right\} , \end{aligned}$$

where $\mathcal{T}(F_{n,i}) = \mu_{n,i}$ is simply the mean.

For any compact interval $\mathcal{I}_C \subseteq \mathcal{I}$, the function defined by $\mathcal{T}(\boldsymbol{F}) \mapsto \inf_{u \in \mathcal{I}_C} \left\{ \mathcal{K}_{\inf}^-(\mathcal{T}(F_i), u) + \mathcal{K}_{\inf}^+(\mathcal{T}(F_j), u) \right\}$ is continuous on $\mathcal{T}(\mathcal{F}^K)$.

For L greater than some L_1 with finite moments, the means $\mu_{n,i}$ and $\mu_{n,j}$ belong to $[\mu_j - \varepsilon, \mu_i + \varepsilon] \subseteq \mathcal{I}$ for some $\varepsilon > 0$. Furthermore, for L greater than some L_2 with finite moments, $\mathcal{T}(F_{n,i})$ is ε-close to

F_i (and same thing for F_j). The continuity then gives that there exists L with finite moments such that

$$\inf_{u\in\mathcal{I}} \left\{\mathcal{K}^-_{\inf}(\mathcal{T}(F_{n,i}),u) + \mathcal{K}^+_{\inf}(\mathcal{T}(F_{n,j}),u)\right\} \geq \frac{1}{2}\inf_{u\in\mathcal{I}_C} \left\{\mathcal{K}^-_{\inf}(\mathcal{T}(F_i),u) + \mathcal{K}^+_{\inf}(\mathcal{T}(F_j),u)\right\} .$$

This is positive since $m(F_j) < m(F_i)$ by an analogue of Lemma 55, which holds for exponential families due to the continuity and strict convexity properties detailed earlier. □

Lemma 75. *Let S_n^L and $\mathcal{I}_n^\star$ as in (19). There exists L_4 with $\mathbb{E}_{\boldsymbol{F}}[(L_4)^\alpha] < +\infty$ for all $\alpha > 0$ such that if $L \geq L_4$, for all n such that $S_n^L \neq \emptyset$,*

$$\forall (i,j) \in \mathcal{I}_n^\star \times \left(S_n^L \setminus \mathcal{I}_n^\star\right), \quad W_n(i,j) \geq L D_{\boldsymbol{F}} ,$$

where $D_{\boldsymbol{F}} > 0$ is a problem dependent constant.

Proof. Let S_n^L and $\mathcal{I}_n^\star$ as in (19). Assume that $S_n^L \neq \emptyset$. If $S_n^L \setminus \mathcal{I}_n^\star$ is empty, then the statement is not informative. Assume $S_n^L \setminus \mathcal{I}_n^\star$ is not empty. Let $(i,j) \in \mathcal{I}_n^\star \times \left(S_n^L \setminus \mathcal{I}_n^\star\right)$. We can now use $\{i,j\} \subseteq S_n^L$ and Lemma 74. □

Lemma 76 gives an upper bound on the transportation costs between a sampled enough arm and an under-sampled one.

Lemma 76. *Let S_n^L as in (19). There exists L_5 with $\mathbb{E}_{\boldsymbol{F}}[(L_5)^\alpha] < +\infty$ for all $\alpha > 0$ such that for all $L \geq L_5$ and all $n \in \mathbb{N}$,*

$$\forall (i,j) \in S_n^L \times \overline{S_n^L}, \quad W_n(i,j) \leq L(2W_d + D_1 + D_2 W_\mu) ,$$

where $D_1 > 0$ and $D_2 > 0$ are problem dependent constants and W_d, W_μ are the random variables defined in Lemma 73.

Proof. Let $(i,j) \in S_n^L \times \overline{S_n^L}$ (i is sampled more than L times, j is not). Taking $u = \mu_{n,i}$ yields

$$\begin{aligned} W_n(i,j) &= \inf_{u\in\mathbb{R}} \left\{N_{n,i}\mathcal{K}^-_{\inf}(\mathcal{T}(F_{n,i}),u) + N_{n,j}\mathcal{K}^+_{\inf}(\mathcal{T}(F_{n,j}),u)\right\} \\ &\leq N_{n,j}\mathcal{K}^+_{\inf}(\mathcal{T}(F_{n,j}),\mu_{n,i}) \leq L\mathcal{K}^+_{\inf}(\mathcal{T}(F_{n,j}),\mu_{n,i}) , \end{aligned}$$

where we used that $j \in \overline{S_n^L}$.

By definition of W_d and W_μ, we have

$$\begin{aligned} \mu_{n,j} &\leq \mu_j + W_\mu \log(e+N_{n,j})/N_{n,j} , \\ d(\mu_{n,j},\mu_j) &\leq W_d \log(e+N_{n,j})/N_{n,j} . \end{aligned}$$

The same is true for i. Then the $\mathcal{K}_{\inf}$ is bounded by

$$\begin{aligned} &\mathcal{K}^+_{\inf}(\mathcal{T}(F_{n,j}),\mu_{n,i}) = d(\mu_{n,j},\mu_{n,i}) \leq d\left(\mu_{n,j}, \mu_i + W_\mu \frac{\log(e+N_{n,i})}{N_{n,i}}\right) \\ &= d(\mu_{n,j},\mu_j) + d\left(\mu_j, \mu_i + W_\mu \frac{\log(e+N_{n,i})}{N_{n,i}}\right) \\ &\quad + \left|(\mu_{n,j}-\mu_j)\left((\varphi')^{-1}(\mu_j) - (\varphi')^{-1}\left(\mu_i + W_\mu \frac{\log(e+N_{n,i})}{N_{n,i}}\right)\right)\right| \\ &\leq W_d \frac{\log(e+N_{n,j})}{N_{n,j}} + d\left(\mu_j, \mu_i + W_\mu \frac{\log(e+N_{n,i})}{N_{n,i}}\right) \\ &\quad + W_\mu \frac{\log(e+N_{n,j})}{N_{n,j}} \left|(\varphi')^{-1}(\mu_j) - (\varphi')^{-1}\left(\mu_i + W_\mu \frac{\log(e+N_{n,i})}{N_{n,i}}\right)\right| . \end{aligned}$$

Since $x \mapsto \frac{\log(e+x)}{x}$ is decreasing on $\mathbb{R}_+^\star$, we have $\frac{\log(e+N_{n,j})}{N_{n,j}} \leq 2$ for $N_{n,j} \geq 1$ and $\frac{\log(e+N_{n,i})}{N_{n,i}} \leq \frac{\log(e+L)}{L}$ for $N_{n,i} \geq L$. For $\varepsilon > 0$ and $L \geq L_\varepsilon$ where $L_\varepsilon \geq W_\mu \log(e+L_\varepsilon)/\varepsilon$, we have

$$\mathcal{K}^+_{\inf}(\mathcal{T}(F_{n,j}),\mu_{n,i}) \leq 2W_d + d(\mu_j,\mu_i+\varepsilon) + 2W_\mu |(\varphi')^{-1}(\mu_j) - (\varphi')^{-1}(\mu_i+\varepsilon)| .$$

Since the means belong to the interior of the interval $\mathcal{I}$, there exists a $\varepsilon > 0$ such that $d(\mu_j,\mu_i+\varepsilon)$ and $|(\varphi')^{-1}(\mu_j) - (\varphi')^{-1}(\mu_i+\varepsilon)|$ are finite. We take the corresponding L, which is sub-exponential, and obtain the result. □

H.1.1 EB leader

The proof of Properties 2 and 5 is almost identical to that for bounded distributions. Only the lower bound on $W_n(i,j)$ is used, which has the same form for SPEFs and bounded distributions.

H.1.2 TC challenger

Conditioned on $\mathcal{F}_n$ and given a leader B_{n+1}, the Transportation Cost (TC) challenger is defined in (26) as the arm with smallest transportation cost compared to the leader

$$C_{n+1}^{\mathrm{TC}} \in \operatorname*{arg\,min}_{j \neq B_{n+1}} W_n(B_{n+1}, j) \quad , \quad \mathbb{P}_{|n}[C_{n+1}^{\mathrm{TC}} = j | B_{n+1} = i] = \frac{\mathbb{1}\left(j \in \arg\min_{k \neq i} W_n(i,k)\right)}{|\arg\min_{k \neq i} W_n(i,k)|} ,$$

and $\widehat{C}_{n+1}^{\mathrm{TC}} \in \arg\min_{j \neq \widehat{B}_{n+1}} W_n(\widehat{B}_{n+1}, j)$.

Property 3 We prove Property 3 for C_{n+1}^{TC} in Lemma 77 by comparing the rates at which W_n increases.

Lemma 77. *Let B_{n+1} be a leader satisfying Property 2. Let $(C_{n+1}^{TC}, \widehat{C}_{n+1}^{TC})$ as in (26). Let U_n^L and V_n^L as in (20) and $\mathcal{J}_n^\star = \arg\max_{i \in \overline{V_n^L}} \mu_i$. There exists L_6 with $\mathbb{E}_{\boldsymbol{F}}[L_6] < +\infty$ such that if $L \geq L_6$, for all n such that $U_n^L \neq \emptyset$, $\widehat{B}_{n+1} \notin V_n^L$ implies $\widehat{C}_{n+1}^{TC} \in V_n^L \cup \left(\mathcal{J}_n^\star \setminus \left\{\widehat{B}_{n+1}\right\}\right)$.*

Proof. The proof proceeds similarly to the one of Lemma 19. The difference is the bounds on $W_n(i,j)$ that we get. For all L bigger than some random variable with finite expectation,

$$\begin{aligned} & \widehat{B}_{n+1} \in \mathcal{J}_n^\star \, , \\ \forall (i,j) \in \mathcal{J}_n^\star \times \left(\overline{V_n^L} \setminus \mathcal{J}_n^\star\right), \quad & W_n(i,j) \geq L^{3/4} D_{\boldsymbol{F}} \, , \\ \forall (i,j) \in \overline{U_n^L} \times U_n^L, \quad & W_n(i,j) \leq \sqrt{L}(2W_d + D_1 + D_2 W_\mu) \, . \end{aligned}$$

For all $L \geq L_7 := (2(2W_d + D_1 + D_2 W_\mu)/D_{\boldsymbol{F}})^4$,

$$L^{3/4} D_{\boldsymbol{F}} > \sqrt{L}(2W_d + D_1 + D_2 W_\mu) \, .$$

We now conclude that at least one under-sampled arm has transportation cost lower than all the ones that are much sampled, and proceed as in the proof of Lemma 19. □

Property 6 Lemma 78 shows that the Property 6 is satisfied by C_{n+1}^{TC}.

Lemma 78. *Assume Property 4 holds. Let $\varepsilon > 0$. Let B_{n+1} be a leader satisfying Property 5 and C_{n+1}^{TC} as in (26). There exists N_7 with $\mathbb{E}_{\boldsymbol{F}}[N_7] < +\infty$ such that for all $n \geq N_7$ and all $i \neq i^\star(\boldsymbol{F})$,*

$$\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon \quad \Longrightarrow \quad \mathbb{P}_{|n}[C_{n+1}^{TC} = i \mid B_{n+1} = i^\star(\boldsymbol{F})] = 0 \, . \tag{44}$$

Proof. This proof proceeds very similarly to the proof of lemma 20. Let $\varepsilon > 0$ and $i^\star = i^\star(\boldsymbol{F})$. Using the definition of C_{n+1}^{TC} in (26), we have

$$\mathbb{P}_{|n}[C_{n+1}^{\mathrm{TC}} = i \mid B_{n+1} = i^\star] = 0 \quad \Longleftrightarrow \quad \frac{1}{n}\left(W_n(i^\star, i) - \min_{j \neq i^\star} W_n(i^\star, j)\right) > 0 \, .$$

Let N_1 as in Property 4, then $N_{n,i} \geq \sqrt{\frac{n}{K}}$ for all $n \geq N_1$. Since $i^\star$ is unique, we have $\Delta' := \min_{j \neq i^\star} |\mu_{i^\star} - \mu_j| > 0$. Let u_0 be the minimal distance from any mean μ_i to an end of the interval of means $\mathcal{I}$ and let $\Delta = \min\{\Delta', u_0\}$. Lemma 73 yields that there exists $N_8 = Poly(W_\mu)$ such that for all $n \geq \max\{N_1, N_8\}$ and all $i \in [K]$, we have $|\mu_{n,i} - \mu_i| \leq \frac{\Delta}{4}$. Therefore, for all $n \geq \max\{N_1, N_8\}$, $\arg\max_{i \in [K]} \mu_{n,i} = \arg\max_{i \in [K]} \mu_i = i^\star$ and for all $i \in [K]$, $\mu_{n,i} \in \mathcal{I}$.

Let $\xi > 0$. Since Property 4 holds and B_{n+1} satisfies Property 5, we can use the results from Lemma 11. Let N_4 defined in Lemma 11. We have $\left|\frac{N_{n,i^\star}}{n} - \beta\right| \leq \xi$ for all $n \geq \max\{N_1, N_4\}$.

Let $i \neq i^\star$ such that $\frac{\Psi_{n,i}}{n} \geq w_i^\beta + \varepsilon$. Using Lemma 5, there exists $N_9 = Poly(W_1)$, such that for all $n \geq \max\{N_1, N_9\}$, we have $\frac{N_{n,i}}{n} \geq w_i^\beta + \frac{\varepsilon}{2}$. Therefore, for all $n \geq \max\{N_1, N_4, N_8, N_9\}$, as in the proof of Lemma 20,

$$\frac{1}{n}\left(W_n(i^\star, i) - \min_{j \neq i^\star} W_n(i^\star, j)\right) \geq \inf_{\tilde{\beta}:|\tilde{\beta}-\beta|\leq\xi} G_i(\mathcal{T}(\boldsymbol{F}_n), \tilde{\beta})$$

where

$$\begin{aligned} G_i(\boldsymbol{m}, \tilde{\beta}) = &\inf_{u\in[0,B]} \left\{\tilde{\beta}\mathcal{K}_{\inf}^-(m_{i^\star}, u) + \left(w_i^\beta + \frac{\varepsilon}{2}\right)\mathcal{K}_{\inf}^+(m_i, u)\right\} \\ &- \sup_{w\in\triangle_K : w_{i^\star}=\tilde{\beta}} \min_{j\neq i^\star} \inf_{u\in\mathcal{I}} \left\{w_{i^\star}\mathcal{K}_{\inf}^-(m_{i^\star}, u) + w_j\mathcal{K}_{\inf}^+(m_j, u)\right\} . \end{aligned}$$

Since all the means belong to a compact subset of $\mathcal{I}$ for $n \geq \max\{N_1, N_4, N_8, N_9\}$, we can prove continuity of the functions $(\boldsymbol{m}, \tilde{\beta}) \mapsto G_i(\boldsymbol{m}, \tilde{\beta})$ and $\boldsymbol{m} \mapsto \inf_{\tilde{\beta}:|\tilde{\beta}-\beta|\leq\xi} G_i(\boldsymbol{m}, \tilde{\beta})$ in a similar way as was done for bounded distributions in Lemma 31. Therefore, there exists $N_{10} = Poly(W_2)$ and ξ_0 such that for $n \geq N_7 := \{N_1, N_4, N_8, N_9, N_{10}\}$ and all $\xi \leq \xi_0$,

$$\inf_{\tilde{\beta}:|\tilde{\beta}-\beta|\leq\xi} G_i(\mathcal{T}(\boldsymbol{F}_n), \tilde{\beta}) \geq \frac{1}{2} \inf_{\tilde{\beta}:|\tilde{\beta}-\beta|\leq\xi} G_i(\mathcal{T}(\boldsymbol{F}), \tilde{\beta}) \geq \frac{1}{4} G_i(\mathcal{T}(\boldsymbol{F}), \beta) .$$

At the β-equilibrium all transportation costs are equal. Therefore, by definition of w^β,

$$\begin{aligned} &\sup_{w\in\triangle_K : w_{i^\star}=\beta} \min_{j\neq i^\star} \inf_{u\in[0,B]} \left\{w_{i^\star}\mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_j\mathcal{K}_{\inf}^+(F_j, u)\right\} \\ &= \min_{j\neq i^\star} \inf_{u\in[0,B]} \left\{\beta\mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_j^\beta\mathcal{K}_{\inf}^+(F_j, u)\right\} \\ &= \inf_{u\in[0,B]} \left\{\beta\mathcal{K}_{\inf}^-(F_{i^\star}, u) + w_i^\beta\mathcal{K}_{\inf}^+(F_i, u)\right\} \\ &< \inf_{u\in[0,B]} \left\{\beta\mathcal{K}_{\inf}^-(F_{i^\star}, u) + \left(w_i^\beta + \frac{\varepsilon}{2}\right)\mathcal{K}_{\inf}^+(F_i, u)\right\} \end{aligned}$$

where the strict inequality is obtained because the transportation costs are increasing in their allocation arguments (proved in a similar way as Lemma 56). Therefore, we have $G_i(\mathcal{T}(\boldsymbol{F}), \beta) > 0$. This yields that $W_n(i^\star, i) > \min_{j\neq i^\star} W_n(i^\star, j)$. As all moments of W_1 and λW_μ are finite, we have $\mathbb{E}_{\boldsymbol{F}}[N_i] < +\infty$ for $i \in \{8, 9, 10\}$. Hence this is also the case for N_7. □

H.1.3 TCI challenger

Showing Properties 3 and 6 for the TCI challenger uses the same arguments as for the proof of Lemma 77 and 78. Coping for the penalization term $\log N_{n,j}$ is done similarly as when we obtained Lemmas 21 and 22 by adapting the proof of Lemmas 19 and 20. Since there is no new arguments, we omit the proof.

I Implementation details and additional experiments

After presenting the implementations details in Appendix I.1, we display supplementary experiments in Appendix I.2.

I.1 Implementation details

In non-parametric settings, the algorithms are inherently more costly than their counterpart in parametric settings. First, the memory cost is linear in time as we need to maintain the whole history $\mathcal{F}_n$ in memory. This a direct consequence of the lack of sufficient statistics to summarize $\mathcal{F}_n$. In contrast, the memory cost is constant for single-parameter exponential families settings. Second, the computational cost per iteration of many algorithms is at least linear in time: a good algorithm should leverage all the observations to make a decision.

We detail below the most relevant implementation details regarding the sampling rules and discuss their computational cost. As mentioned above, the implemented algorithms for the bounded setting are computational expensive by nature. However, we aim at promoting the algorithm(s) achieving good empirical performance in terms of empirical stopping time at a reasonable computational cost.

Stopping-Recommendation pair The recommendation rule $\hat{\imath}_n \in \arg\max_{i\in[K]} \mu_{n,i}$ has a $\mathcal{O}(K)$ computational cost. This is achieved by simply maintaining the cumulative sum of the observations $\sum_{t\leq n} \mathbb{1}(i = I_t) X_{t,i}$ for each arm.

In contrast to the recommendation rule, the stopping rule defined in (2) is computationally expensive. At each time n, we need to compute $K-1$ transportation costs $W_n(\hat{\imath}_n, j)$ for $j \neq \hat{\imath}_n$. While each one can be evaluated efficiently for single-parameter exponential families (see below), this is not the case for bounded distributions where

$$W_n(\hat{\imath}_n, j) = \inf_{x\in[\mu_{n,j},\mu_{n,\hat{\imath}_n}]} g_n(\hat{\imath}_n, j, x)\,,$$
$$g_n(\hat{\imath}_n, j, x) = N_{n,\hat{\imath}_n}\mathcal{K}^-_{\inf}(F_{n,\hat{\imath}_n}, x) + N_{n,j}\mathcal{K}^+_{\inf}(F_{n,j}, x)\,.$$

Using Lemmas 52 and 55, the function $x \mapsto g_n(\hat{\imath}_n, j, x)$ is strictly convex and admit a unique minimizer in $[\mu_{n,j}, \mu_{n,\hat{\imath}_n}]$. Lemma 47 gives a formula for the derivatives of $x \mapsto \mathcal{K}^{\pm}_{\inf}(F, x)$. Unfortunately, $\lambda^{\pm}_\star(F, x)$ are often defined implicitly (Lemma 49), hence we can't leverage this knowledge and use first-order optimization methods. Therefore, in order to compute $W_n(\hat{\imath}_n, j)$, we rely on a zero-order optimization algorithm designed to minimize a univariate function on a bounded interval. In practice, we use Brent's method, which is implemented in the `Optim.jl` package under `Julia 1.7.2`.

To obtain $g_n(\hat{\imath}_n, j, x)$ for a given x, we need to compute $\mathcal{K}^-_{\inf}(F_{n,\hat{\imath}_n}, x)$ and $\mathcal{K}^+_{\inf}(F_{n,j}, x)$. This is made tractable thanks to their dual formulation. Taking $\mathcal{K}^+_{\inf}(F_{n,j}, x)$ as an example, Theorem 3 yields

$$N_{n,j}\mathcal{K}^+_{\inf}(F_{n,j}, x) = \sup_{\lambda\in[0,1]} \sum_{k\in[N_{n,i}]} \log\left(1 - \lambda\frac{X_{k,i} - x}{B - x}\right)\,,$$

where $(X_{k,i})_{k\in[N_{n,i}]}$ denotes the samples collected from arm i at time n. As the function is strictly concave (Lemma 44), we will also use Brent's method to compute it's maximum. Each computation requires to sum over the $N_{n,i}$ samples collected by arm i. Therefore, the computational cost is at least linear in time. Since we can compute the derivative, it would be possible to use first-order optimization algorithms. While this will improve the number of iterations required to reach convergence, it is not clear that the overall computational cost will be reduced since those gradient computations are also linear in time.

Top Two sampling rules We discuss the computational cost of the EB and TS leader, as well as the TC and RS challenger.

The EB leader has virtually no computational cost since it uses the candidate answer $\hat{\imath}_n$. Likewise, the TC challenger can also leverage the computations from the stopping rule (2). When $B_{n+1} = \hat{\imath}_n$, $C^{\mathrm{TC}}_{n+1} \in \arg\min_{j\neq\hat{\imath}_n} W_n(\hat{\imath}_n, j)$ was already computed in (2). When $B_{n+1} \neq \hat{\imath}_n$, we have $C^{\mathrm{TC}}_{n+1} \in \arg\min_{j\neq B_{n+1}} W_n(B_{n+1}, j) = \{j \neq B_{n+1} \mid \mu_{n,j} \geq \mu_{n,B_{n+1}}\}$, which contains at least $\hat{\imath}_n$. Therefore, the TC challenger has virtually no computational cost when paired with (2).

The TS leader and the RS challenger use a sampler Π_n. In single-parameter exponential families, Π_n is a posterior distribution which can be computed in constant time by updating the posterior $\mathcal{O}(K)$ parameters. However, for bounded distributions, the Dirichlet sampler relies on the whole history $\mathcal{F}_n$. Therefore, for each arm i, we need to define and sample from a Dirichlet distribution with $N_{n,i} + 2$ parameters. This has linear computational cost per iteration. The TS leader requires only one Dirichlet observation per arm, hence it has constant cost once the Dirichlet distributions are defined (which is computationally expensive).

For the RS challenger, we re-sample until B_{n+1} is not an arm with highest mean in the corresponding vector of observations. The computational cost is proportional to the number of re-sampling steps which is on average $1/(1 - a_{n,B_{n+1}})$. Hence the computational cost can be very high when Π_n has

converged, that is when $a_{n,i} \approx 0$ for all $i \neq i^\star$. The analysis of the TS leader and the RS challenger reveals that this convergence is exponential, with a rate close to $T^\star_\beta(\boldsymbol{F})$. Therefore, when $B_{n+1} = i^\star$, it is highly unlikely to observe a vector θ for which $i^\star$ is not the best arm, and the average number of re-sampling steps is exponential.

Based on extensive experiments, we also have empirical evidence that the RS challenger has higher computational cost than the TC challenger. For Bernoulli instances, when using the stopping threshold defined in (4), the maximum number of re-sampling steps (set arbitrarily to 10^6) was always reached for large time n. As a direct consequence, the computational cost of the RS challenger explodes in those cases, e.g. 10^4 times slower than the TC challenger. As we use a uniform sampling when the maximum number of re-sampling steps is reached, the achieved empirical stopping time is also higher than for the variant using the TC challenger. In Appendix I.2.1, we perform experiments with the RS challenger for a heuristic stopping threshold defined in (46). This yields four times smaller empirical stopping time compared to using (4), see Appendix I.2.

Other sampling rules In LUCB-based sampling rules, we need to compute upper and lower confidence bounds based on the inversion of a distance function. For KL-LUCB, it requires inverting the KL divergence of Bernoulli distributions. This can be done efficiently by using a binary search algorithm. For $\mathcal{K}_{\inf}$-LUCB, we need to inverse $\mathcal{K}^{\pm}_{\inf}$, also by using a binary search. As explained above, computing $\mathcal{K}^{\pm}_{\inf}$ for the empirical cdfs yields a linear computation cost. Therefore, $\mathcal{K}_{\inf}$-LUCB will be significantly worse than KL-LUCB in terms of computational cost. However, $\mathcal{K}_{\inf}$-LUCB yields order of magnitude smaller empirical stopping time in the DSSAT instances compared to KL-LUCB.

The sampling rule $\mathcal{K}_{\inf}$-DKM is inspired by DKM [13], with only one learner on $\triangle_K$ instead of K learners. We replace the KL divergences by $\mathcal{K}^{\pm}_{\inf}$ as we are in the bounded setting, hence it will be more costly than DKM for single-parameter exponential families. Given the allocation w_n returned by the learner (e.g. AdaHedge), computing the most confusing alternative parameter has the same computational cost as evaluating the stopping rule (2). For bounded distributions, it is not clear how to define the optimistic bonuses. Therefore, we replace it by forced exploration, which yields worse empirical stopping times.

Adaptive choice of β Based on the theoretical lower bound, Top Two algorithms with a fixed allocation can be at best asymptotically β-optimal, not asymptotically optimal (meaning reaching $T^\star(\boldsymbol{F})$). To achieve asymptotic optimality, the fixed allocation should match the optimal allocation $\beta^\star = \arg\min_{\beta \in (0,1)} T^\star(\boldsymbol{F})$. As $\beta^\star$ is unknown, it should be learned from observations. Therefore, this desired adaptive Top Two algorithm should use an adaptive choice of β which converges towards $\beta^\star$. Proving optimality for adaptive Top Two algorithms is an interesting open problem, which is still unsolved even for Gaussian bandits. The very recent paper [43] proposes an update mechanism for β, but they study it only empirically and they don't provide any theoretical guaranty for that scheme.

Optimal allocation oracles In the following, we consider bounded distributions $\boldsymbol{F}$ having a discrete support. Computing the optimal allocation $w^\star(\boldsymbol{F})$ is computationally very expensive for bounded setting. Even for single-parameter exponential families, this can be very demanding. For more complex structure such as top-k identification or combinatorial bandits [26], advanced saddle-point algorithms are needed to obtain efficient implementation, even for Gaussian distributions. Therefore, Track-and-Stop algorithms computing $w^\star(\boldsymbol{F}_n)$ at each time n should not be used to tackle the bounded distributions setting. While it is costly, we can still compute $w^\star(\boldsymbol{F})$ for the true distribution $\boldsymbol{F}$ once. First, it allows to obtain a lower bound on the empirical stopping in $T^\star(\boldsymbol{F})\mathrm{kl}(\delta, 1-\delta)$. Second, we can implement the oracle algorithm (referred to as "fixed" in the experiments) which tracks the true optimal allocation $w^\star(\boldsymbol{F})$.

The strategy to compute $w^\star(\boldsymbol{F})$ is similar as done for single-parameter exponential families [16]. In the following, we describe the heuristic algorithm mimicking the behavior of the oracle in [16]. Let $i^\star = i^\star(\boldsymbol{F})$. Let $G_j(x)$ defined in Lemma 61. Recall that $x_j(y) = G_j^{-1}(y)$ and $u_j(x)$ is defined as the minimizer yielding $G_j(x)$. If x_j and u_j can be differentiated, we could directly derive (43) in Lemma 61. Additional manipulations [16] yield the reformulation of the optimization problem

defining $T^\star(\boldsymbol{F})$ as solving $F(y) = 1$ where

$$\forall y \in \left[0, \min_{j \neq i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, \mu_j)\right), \quad F(y) = \sum_{j \neq i^\star} \frac{\mathcal{K}_{\inf}^-(F_{i^\star}, u_j(x_j(y)))}{\mathcal{K}_{\inf}^+(F_j, u_j(x_j(y)))} \tag{45}$$

is a strictly increasing increasing function such that $F(0) = 0$ and $\lim_{y \to \min_{j \neq i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, \mu_j)} F(y) = +\infty$. We use nested binary searches to solve $F(y) = 1$. The outer binary search is done on $y \in \left[0, \min_{j \neq i^\star} \mathcal{K}_{\inf}^-(F_{i^\star}, \mu_j)\right)$ (see Lemma 63). The inner binary searches are done to compute $x_j(y)$ for all $j \neq i^\star$. To compute $u_j(x)$, we use the same procedure than described in the stopping rule. While proving that x_j and u_j are differentiable still eludes us, we conjecture it to be true. Therefore, this heuristic optimal allocation algorithm gives a good estimate of $T^\star(\boldsymbol{F})$ and $w^\star(\boldsymbol{F})$.

When $\boldsymbol{F}$ has a non-discrete support, computing numerically $\mathcal{K}_{\inf}^\pm$ by using the dual formulation would require having access to an oracle outputting $\mathbb{E}_{F_i}[\log(1 - \lambda(X - u))]$ (resp. $\mathbb{E}_{F_i}[\log(1 + \lambda(X - u))]$) for all $i \in [K]$, $u > \mu_i$ and $\lambda \in \left[0, \frac{1}{B-u}\right]$ (resp. $\lambda \in \left[0, \frac{1}{u}\right]$). For continuous distribution, those integrals could be computed by numerical integration. Instead we adopt a Monte-Carlo approach and use a discrete distribution $\widehat{\boldsymbol{F}}$ sampled from $\boldsymbol{F}$. By Lemma 14, taking a sufficiently large number of samples ensures that $\max_{i \in [K]} \|\widehat{F}_i - \boldsymbol{F}_i\|_\infty$ is small. Intuitively, this should ensures that $w^\star(\widehat{\boldsymbol{F}})$ and $T^\star(\widehat{\boldsymbol{F}})$ are a good approximation of $w^\star(\boldsymbol{F})$ and $T^\star(\boldsymbol{F})$. Formalizing this intuition theoretically requires having access to a Lipschitz constant for the $\|\cdot\|_\infty$. To our knowledge, proving that $\boldsymbol{F} \mapsto w^\star(\boldsymbol{F})$ and $\boldsymbol{F} \mapsto T^\star(\boldsymbol{F})$ are Lipschitz is still an open problem, even for Gaussian distributions.

Efficient implementation for Bernoulli For Bernoulli distributions, the computational cost is greatly reduced. In the following kl denotes the KL divergence for Bernoulli distributions.

First, the stopping rule can be computed in $\mathcal{O}(K)$ since we have have closed form formulas for $\mathcal{K}_{\inf}^\pm$, i.e. $\mathcal{K}_{\inf}^+(F_{n,i}, u) = \mathrm{kl}(\mu_{n,i}, \max\{\mu_{n,i}, u\})$ and $\mathcal{K}_{\inf}^-(F_{n,i}, u) = \mathrm{kl}(\mu_{n,i}, \min\{\mu_{n,i}, u\})$, and for the closest alternative parameter, i.e.

$$\operatorname*{arg\,min}_{x \in [\mu_{n,j}, \mu_{n,\hat{\imath}_n}]} g_n(\hat{\imath}_n, j, x) = \frac{N_{n,\hat{\imath}_n}\mu_{n,\hat{\imath}_n} + N_{n,j}\mu_{n,j}}{N_{n,\hat{\imath}_n} + N_{n,j}} \,.$$

By the same arguments, we have a more efficient computation of the optimal allocation $w^\star(\boldsymbol{F})$. As the differentiability of x_j and u_j holds in this setting, the optimal allocation oracle is theoretically validated.

Second, the sampler Π_n can be rewritten as a Beta distribution with parameters $(c_{n,i} + 1, N_{n,i} - c_{n,i} + 1)$, where $c_{n,i} = |\{t \in [n] \mid I_t = i, X_{t,i} = B\}|$. This leverages the fact that we can group the observations into two values $\{0, B\}$, and a classic results on marginals of Dirichlet distributions. While this reduces the cost of sampling observations from Π_n, the computational cost of the re-sampling procedure (discussed above) still remains.

For Bernoulli distributions, β-TS-TC coincide with T3C [39] and β-TS-RS coincide with TTTS [38]. While the algorithms were already known in this setting, we are the first to prove they achieve asymptotic β-optimality for Bernoulli distributions. In [39], the authors only provide a proof for Gaussian distributions.

Decision Support System for Agrotechnology Transfer (DSSAT) DSSAT[3] [22] is a crop modeling software that has been developed (mainly) to help agricultural production in developing countries. This simulator provides a standardized way to generate realistic crop yield for different plants and soil conditions, harnessing more than 30 years of historical field data on 42 different crops. Simulations are based on complex biophysical models, and take many parameters into account: local soil conditions, genetics, and crop management policy (e.g planting date, fertilization policy). Our experiment is inspired by the one proposed in [7]: we consider maize fields, and fixed challenging soil conditions (poor water retention and fertility), that are close to the conditions endured by small-holder farmers in Sub-Saharan Africa. As the biophysical models are fixed and the weather is sampled by the environment, in this example the learner can play on human decisions such as the planting date and

[3] DSSAT is an Open-Source project maintained by the DSSAT Foundation.

fertilization policy. To simplify, we only consider the planting date, which already provides a difficult problem as the distributions represented in Figure 3(b) and Figure 7 show. In those figures, each arm corresponds to a yield distribution with all parameters fixed, except for the plantings that are ~ 20 days apart from each other, ranging on two months. Our objective is to perform *in sillico* experiments to compare the performance of different Best Arm Identification bandit algorithms, to help a potential group of farmers to choose an algorithm to use for future real-world experiments.

As calling the simulator is computationally intensive and as we want to perform Monte-Carlo simulations we used the code provided in [7] to generate 10^6 empirical data from each distribution and store these points in a csv file that is provided with the code of this paper. We further re-scale the distributions in $[0, 1]$, which is equivalent as setting the known upper bound as the maximum value sampled by the simulator in our data collection process. Then, a call to an arm simply consists in sampling one of these points uniformly at random. We think this approximation is sufficient to reflect the difficulty of the problem, while being less demanding in terms of computation time.

Remark 1. *While this setting is simplified over a real-world experiment, it is an interesting and highly non-trivial first step to build more realistic algorithms taking in account contextual information, batch feedback, risk-aversion of farmers at a group and individual level (see [7]), . . . Furthermore, if the asymptotic guarantees of bandit algorithms make them non-realistic for a single farmer (one data point every 3-6 months), they may exhibit tremendous progress over uniform sampling for a group of farmers (typically a few hundreds of data points every 3-6 months) conducting an experiment for several years.*

Reproducibility Our code is implemented in `Julia 1.7.2`, and the plots are generated with the `StatsPlots.jl` package. Optimizations are performed based on the Brent's method available in the `Optim.jl` package. Other dependencies are listed in the `Readme.md`. The `Readme.md` file also provides detailed julia instructions to reproduce our experiments, and we provide a `script.sh` to run them all at once. The general structure of the code (and some functions) is taken from the tidnabbil library.[4]

I.2 Supplementary experiments

As in Section 5, we consider a moderate confidence regime in which $\delta = 0.1$ and Top Two algorithm with $\beta = 0.5$.

Heuristic GK16 threshold While the stopping threshold defined in (4) ensures δ-correctness of the stopping rule (2), it is conservative in practice. We denote it as TT (Theoretical Threshold). Aiming at running large scale experiments, we consider the GK16 heuristic threshold defined in [16],

$$\beta^{\mathrm{GK16}}(n, \delta) = \log\left(\frac{1 + \log(n)}{\delta}\right) . \tag{46}$$

Using GK16 yields an empirical error lower than δ, even if it has no δ-correctness guaranty. This threshold was extensively used in the BAI literature to conduct experiments. This idealized dependency in (n, δ) can be achieved for single-parameter exponential families [29]. In this work, we show that $\log n$ can be achieved for bounded distributions. Whether it is possible to achieve $\log \log n$ for bounded distributions is an interesting open research question. As it would require more sophisticated results on martingales to obtain such concentrations results for $\mathcal{K}_{\inf}$, we leave it for future work.

On simple Bernoulli instances, we compare the performance of the algorithms from Section 5 for the stopping rule (2) using GK16 or TT. We consider three instances: the *easy* instance with $\mu = (0.7, 0.5, 0.4, 0.3, 0.2)$, the *hard* instance with $\mu = (0.7, 0.6, 0.5, 0.4, 0.3)$ and the 3^{rd}*-equal* instance $\mu = (0.7, 0.6, 0.5, 0.5)$. While we average our results over 5000 runs for GK16, we only perform 1000 runs for TT.

In Figure 5, the empirical stopping time when using GK16 is on average four times lower than when using TT. As we discussed above, the computational cost of each iteration increases with the time n. Therefore, this speed-up in stopping time naturally yields a speed-up in the averaged computational

[4] This library was created by [13], see https://bitbucket.org/wmkoolen/tidnabbil. No license were available on the repository, but we obtained the authorization from the authors.

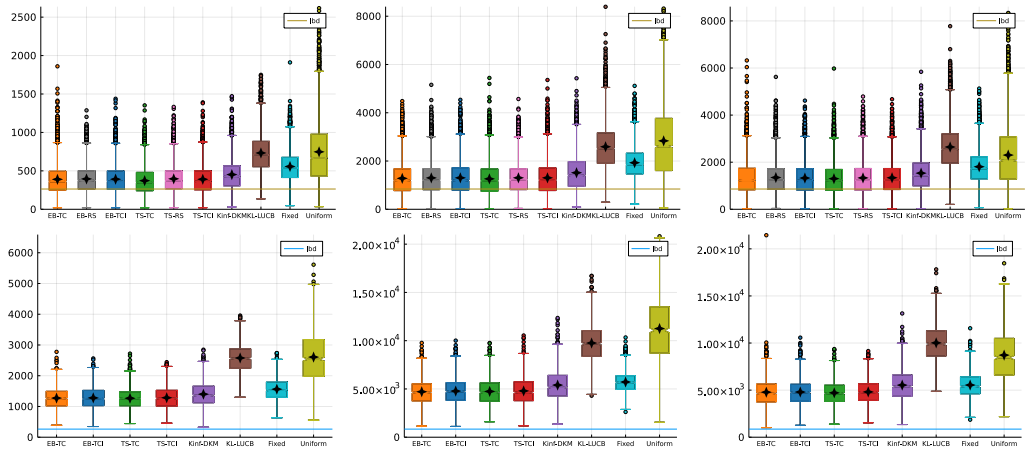

Figure 5: Empirical stopping time on the easy, hard and 3rd-equal instances (left to right) for GK16 (top) or TT (bottom) thresholds.

time per iteration. In some of our experiments, the averaged computational time per iteration was divided by 10.

In the following, all the experiments will be conducted with the GK16 heuristic threshold instead of the TT threshold.

I.2.1 RS challenger

In addition of the Top Two algorithms from Section 5, we assess the empirical performance of instances using the RS challenger. Moreover, we detail more on the lack of robustness of β-EB-TC (large outliers), explained in Appendix D.3. As explained in Appendix I.1, the RS challenger is computationally intractable for large n. The experiments with RS could be ran only because we used GK16 instead of TT, hence dividing the empirical stopping time by four on average. Due to their respective flaws, we do not recommend to use those algorithms in practice, even though they enjoy the same theoretical guarantees as β-EB-TCI, β-TS-TC and β-TS-TCI.

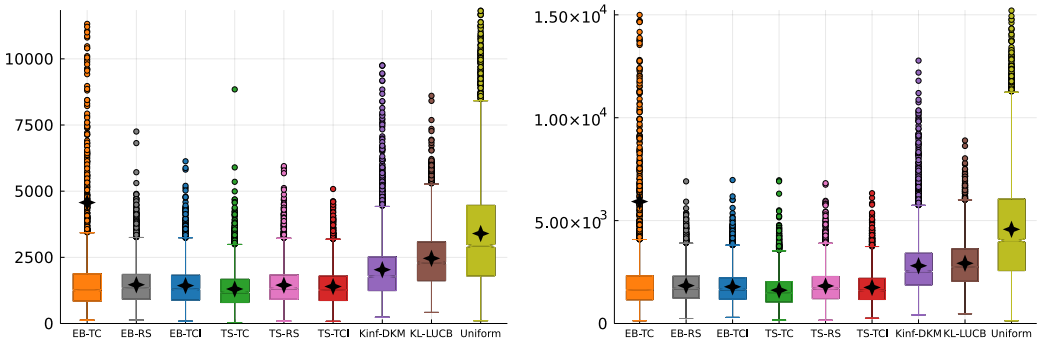

Figure 6: Empirical stopping time on random Bernoulli instances for $K \in (8, 10)$ (left to right).

Random Bernoulli instances For $K \in \{6, 8, 10\}$, we sample 5000 Bernoulli instances such that $\mu_1 = 0.6$ and $\mu_i \sim \mathcal{U}([0.2, 0.5])$ for all $i \neq 1$, where we enforce that $\Delta_{\min} \geq 0.01$. For all other algorithms, Figure 6 delivers the same messages as Figure 4. Therefore, we refer the reader to Section 5 for the corresponding comments.

Figure 6 confirms our theoretical intuition (Appendix D.3) hinting that β-EB-TC is not an empirically robust algorithm, even for $\Delta_{\min} > 0$. This is visible with the large number of outliers, which shift the

mean empirical stopping time away from the median empirical stopping time. Note that the y-axis was cut to provided visibility, hence it hides the largest outliers observed for β-EB-TC.

In Figure 6, we see that β-EB-RS and β-TS-RS perform on par with β-EB-TCI, β-TS-TC and β-TS-TCI, while having few outliers. This confirms our theoretical intuitions on the effect of randomization in the leader and/or challenger (see Appendix D.3).

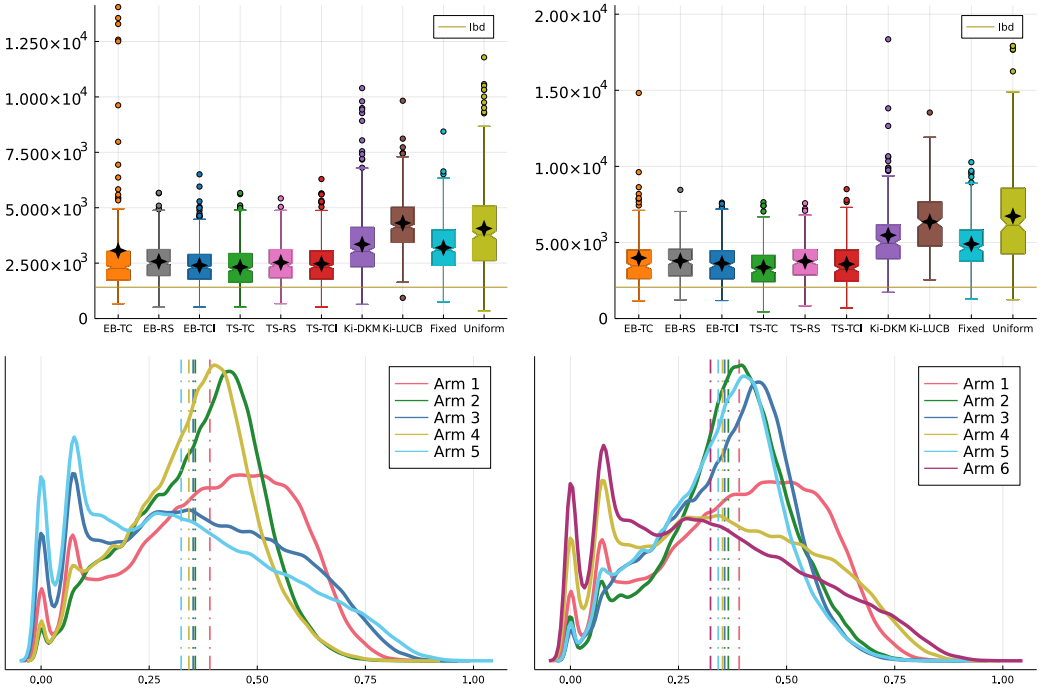

Figure 7: Empirical stopping time (top) on scaled DSSAT instances for $K \in (5, 6)$ (left to right) with their density (bottom), where "Ki" stands for "Kinf".

DSSAT instances We use the DSSAT real-world data for $K \in (5, 6)$, which we scale by the overall maximum so that $B = 1$. Their histograms define instances with bounded distributions, see the bottom plots of Figure 7, on which we can sample. We average our results over 500 runs for $K = 5$ and 250 runs for $K = 6$. For all other algorithms, Figure 7 delivers the same messages as Figure 3. Therefore, we refer the reader to Section 5 for the corresponding comments.

In Figure 7, we see that β-EB-RS and β-TS-RS perform on par with β-EB-TCI, β-TS-TC and β-TS-TCI. For $K = 5$, we observe large outliers on β-EB-TC. This is a symptom of its empirical lack of robustness, which would be more striking if more runs had been performed.

In Section 5, we mentioned that KL-LUCB was performing ten times worse than $\mathcal{K}_{\inf}$-LUCB on DSSAT instances. With the same setting at Figure 7, KL-LUCB used on average (standard deviation) a number of samples equal to 34635 (2860) for $K = 5$ and 57820 (5725) for $K = 6$.

I.2.2 On the distinct means assumption

In Figure 5, the considered Top Two algorithms have good empirical performance on the 3rd-equal instance. This instance violates Assumption 2, i.e. $\Delta_{\min} = \min_{i \neq j} |\mu_i - \mu_j| > 0$, under which we can prove sufficient exploration for the Top Two algorithms. We empirically study instances where $\Delta_{\min} = 0$ in order to confirm our intuition that some Top Two algorithms have good guarantees in this case (Appendix D.3).

The most difficult instances having $\Delta_{\min} = 0$ are the ones where the arms having the same mean are in second position. We consider four toy Bernoulli instances with three arms $\mu_i = (\mu_1, \mu_1 - \Delta_i, \mu_1 - \Delta_i)$, where $\Delta_i > 0$. The smaller Δ_i, the harder the identification problem is. Therefore,

with smaller values of Δ_i, it will be easier to see if Top Two algorithms are failing when $\Delta_{\min} = 0$. In the experiments below, we consider $\mu_1 = 0.5$ and $\Delta_i \in \{0.1, 0.075, 0.05\}$ and average our results over 5000 runs.

Since we aim at observing whether the algorithms get stuck (at least momentarily) without paying an infinite computational cost, we set a maximum number of iterations $T_{\max}$ for each run. In our plots, the quantity $T^\star(\mu)\log(1/\delta)$ acts as a lower bound. Therefore, we set $T_{\max} = 15T^\star(\mu)\log(1/\delta)$. Having an algorithm using more than $T_{\max}$ samples is a symptom of being stuck (at least momentarily).

We observe that Table 3(a) and Table 3(b) are very similar. Reaching $T_{\max}$ is a symptom of failing to identifying the best arm $i^\star$. As there is no δ-correctness guaranty when reaching $T_{\max}$, it is expected that the algorithm recommend an arm different from $i^\star$. In Table 3(a), we see that β-EB-TC is the only algorithm reaching $T_{\max}$ over the 5000 runs. Empirically, this is the only Top Two algorithms that seem to fail on instances with $\Delta_{\min} = 0$. In Table 3(b), all algorithms (except β-EB-TC) have an empirical error lower than $\delta = 1\%$.

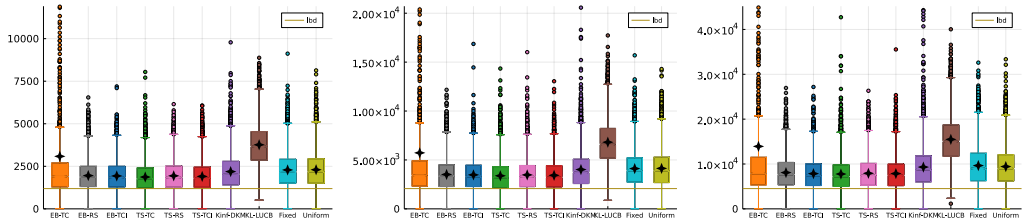

Figure 8: Empirical stopping time on Bernoulli with $\mu_i = (\mu_1, \mu_1 - \Delta_i, \mu_1 - \Delta_i)$ for $\mu_1 = 0.5$ and $\Delta_i \in (0.1, 0.075, 0.05)$ (left to right).

Table 3: Percentage (in %) of runs (a) achieving $T_{\max}$ and (b) failing at identifying $i^\star = 1$ on Bernoulli with $\mu_i = (\mu_1, \mu_1 - \Delta_i, \mu_1 - \Delta_i)$ for $\mu_1 = 0.5$ and $\Delta_i \in (0.1, 0.075, 0.05)$ for $\delta = 1\%$.

	Δ_1	Δ_2	Δ_3
β-EB-TC	6.66	7.62	9.56
β-EB-RS	0	0	0
β-EB-TCI	0	0	0
β-TS-TC	0	0	0
β-TS-RS	0	0	0
β-TS-TCI	0	0	0
$\mathcal{K}_{\inf}$-DKM	0	0	0
KL-LUCB	0	0	0
Fixed	0	0	0
Uniform	0	0	0

	Δ_1	Δ_2	Δ_3
EB-TC	6.74	7.66	9.54
EB-RS	0.02	0.02	0.02
EB-TCI	0.04	0.04	0.06
TS-TC	0.04	0.06	0.1
TS-RS	0.02	0.04	0.04
TS-TCI	0.04	0.08	0.08
$\mathcal{K}_{\inf}$-DKM	0.06	0.12	0.1
KL-LUCB	0.02	0.02	0.06
Fixed	0.02	0.02	0.06
Uniform	0.04	0.02	0.02

Lack of robustness of β-EB-TC Figure 8 confirms our theoretical intuition (Appendix D.3) hinting that β-EB-TC is not empirically robust and can fail when $\Delta_{\min} = 0$. This is visible with the large number of outliers, which shift the mean empirical stopping time away from the median empirical stopping time. In Table 3, we see observe that β-EB-TC is the only algorithm reaching $T_{\max}$ and it does it frequently, i.e. between 6% and 10% in our experiments.

TCI challenger Figure 8 confirms our theoretical intuition (Appendix D.3) that β-EB-TCI copes for the limitations of β-EB-TC, and that it should work when $\Delta_{\min} = 0$. Based on the comparison of β-EB-TC and β-EB-TCI in Figure 8 and Table 3, we see that adding the $\log(N_{n,j})$ term has a stabilization effect, hence reducing the number of outliers.

The difference between β-TS-TC and β-TS-TCI is milder. However, based on Figure 8, it seems that adding the $\log(N_{n,j})$ term slightly reduces the number of large outliers. This effect is less visible than when comparing β-EB-TC and β-EB-TCI due to the stabilization effect ensured by the randomization in the TS leader.

Randomized leader or challenger Figure 8 confirms our theoretical intuition (Appendix D.3) that randomized mechanisms have a stabilization effect, and that they should work when $\Delta_{\min} = 0$. Based on Figure 8 and Table 3, the TS leader or the RS challenger appear to prevent large outliers. For the TS leader, this effect is particularly striking when comparing the performance of β-EB-TC and β-TS-TC. For the RS challenger, the effect is striking when comparing β-EB-TC and β-EB-RS, and milder between β-TS-TC and β-TS-RS.

Symmetric instances When the two sub-optimal arms have the same mean, we have $w^\star(\mu)_2 = w^\star(\mu)_3 = (1 - w^\star(\mu)_1)/2$ by symmetry of the characteristic time $T^\star(\mu)$. Therefore, the optimal allocation is close to the uniform allocation $(1/3, 1/3, 1/3)$. Experimental results (Figure 4(b) and Figure 8) show that the uniform sampling performs on par with the fixed oracle algorithm tracking $w^\star(\mu)$. Therefore, it is not surprising that KL-LUCB performs worse than uniform sampling.

I.2.3 On larger sets of arms

Another interesting question that arises as regards our algorithms is to assess whether their performance scales with the number of arms. We consider the three problem scenarios from Jamieson and Nowak [24] with varying size of arms. The underlying distributions are Gaussian with mean $\mu \in \mathbb{R}^K$ and hardness $H_1 := \sum_{i \neq i^\star} \frac{1}{(\mu_{i^\star} - \mu_i)^2}$. The "1-sparse" scenario sets $\mu_1 = 1/4$ and $\mu_i = 0$ for all $i \in [K] \setminus \{1\}$, resulting in an hardness $H_1 \approx 4K$. The "$\alpha = 0.3$" and "$\alpha = 0.6$" scenarios consider $\mu_i = 1 - \left(\frac{i-1}{K-1}\right)^\alpha$ for all $i \in [K]$, with respective hardness $H_1 \approx 3K/2$ and $H_1 \approx 6K^{1.2}$. We only consider algorithms whose computational cost scales nicely with the number of arms, namely β-EB-TCI, β-TS-TC, LUCB and uniform sampling. We choose $\delta = 0.01$, $\beta = 1/2$ and average our results over 100 runs.

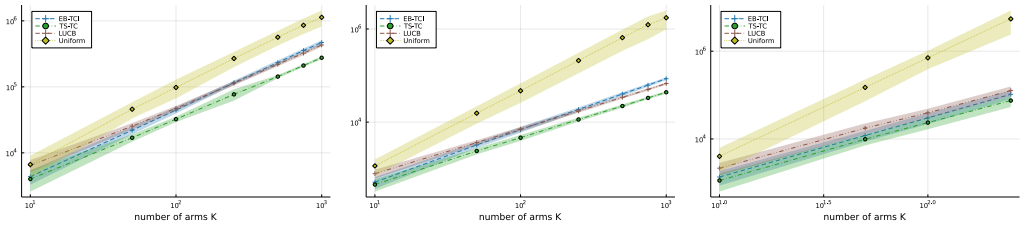

Figure 9: Empirical stopping time for the Gaussian benchmarks (left) "1-sparse", (middle) "$\alpha = 0.3$" and (right) "$\alpha = 0.6$".

In Figure 9, we observe that the performances of β-EB-TCI, β-TS-TC and LUCB scale proportionally to the hardness H_1 when the number of arms increase, while it is worsening for the uniform sampling. Surprisingly, the performance gap between β-EB-TCI and LUCB is diminishing with the number of arms. For larger experiments, LUCB seems to slightly outperform β-EB-TCI. Finally, β-TS-TC significantly outperforms all the other algorithms when the number of arms increase.