On Best-Arm Identification with a Fixed Budget in Non-Parametric Multi-Armed Bandits

Antoine Barrier, Aurélien Garivier, Gilles Stoltz

Preprint, pages 1–46, 2023

On Best-Arm Identification with a Fixed Budget in Non-Parametric Multi-Armed Bandits

Antoine Barrier ANTOINE.BARRIER@ENS-LYON.FR
Univ. Lyon, ENS de Lyon, UMPA UMR 5669, Lyon, France
Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France

Aurélien Garivier AURELIEN.GARIVIER@ENS-LYON.FR
Univ. Lyon, ENS de Lyon, UMPA UMR 5669, LIP UMR 5668, Lyon, France

Gilles Stoltz GILLES.STOLTZ@UNIVERSITE-PARIS-SACLAY.FR
Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France

Abstract

We lay the foundations of a non-parametric theory of best-arm identification in multi-armed bandits with a fixed budget T. We consider general, possibly non-parametric, models $\mathcal{D}$ for distributions over the arms; an overarching example is the model $\mathcal{D} = \mathcal{P}[0,1]$ of all probability distributions over $[0,1]$. We propose upper bounds on the average log-probability of misidentifying the optimal arm based on information-theoretic quantities that we name $\mathcal{L}^{<}_{\inf}(\cdot\,,\nu)$ and $\mathcal{L}^{>}_{\inf}(\cdot\,,\nu)$ and that correspond to infima over Kullback-Leibler divergences between some distributions in $\mathcal{D}$ and a given distribution ν. This is made possible by a refined analysis of the successive-rejects strategy of Audibert et al. (2010). We finally provide lower bounds on the same average log-probability, also in terms of the same new information-theoretic quantities; these lower bounds are larger when the (natural) assumptions on the considered strategies are stronger. All these new upper and lower bounds generalize existing bounds based, e.g., on gaps between distributions.

Keywords: Multi-armed bandits, best-arm identification, non-parametric models, Kullback-Leibler divergences, information-theoretic bounds

1. Introduction and brief literature review

We consider a class $\mathcal{D}$ of distributions over $\mathbb{R}$ with finite first moments, which we refer to as the model $\mathcal{D}$. A K–armed bandit problem in $\mathcal{D}$ is a K–tuple $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ of distributions in $\mathcal{D}$. We denote by $(\mu_1, \ldots, \mu_K)$ the K–tuple of their expectations. An agent sequentially interacts with $\underline{\nu}$: at each step $t \geqslant 1$, she selects an arm A_t and receives a reward Y_t drawn from the distribution ν_{A_t}. This is the only feedback that she obtains.

While regret minimization has been vastly studied (see Lattimore and Szepesvári, 2020), another relevant objective is *best-arm identification*, that is, identifying the distribution with highest expectation. In the fixed-confidence setting, this identification is performed under the constraint that a given confidence level $1 - \delta$ is respected, while minimizing the expected number of pulls of the arms. This setting is fairly well understood (see Lattimore and Szepesvári, 2020, Chapter 33 for a review). A turning point in this literature was achieved by Garivier and Kaufmann (2016), who provided matching upper and lower bounds on the expected number of pulls of the arms in the case of canonical one-parameter exponential families. Since then, improvements have been made in several directions, including for example non-asymptotic bounds (Degenne et al., 2019) and the

problem of ε–best-arm identification (Garivier and Kaufmann, 2021); however, no generalization to non-parametric models has been considered yet, to the best of our knowledge.

Best-arm identification with a fixed budget. The *fixed-budget setting* is much less understood. Therein, the total number T of pulls of the arms is fixed. After these T pulls, a strategy must issue a recommendation I_T. Assuming that $\underline{\nu}$ contains a unique optimal distribution $\nu^\star$ of index $a^\star(\underline{\nu})$, one aims at minimizing $\mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big)$. We are interested in (upper and lower) bounds that hold for all problems $\underline{\nu}$ in $\mathcal{D}$, possibly under the restriction that they only contain a unique optimal arm. It may be straightforwardly seen that the probability of error can decay exponentially fast—for instance, by uniformly exploring the arms (pulling each of them about T/K times) and recommending the one with the largest empirical average. This is why the literature focuses on upper and lower bound functions $\ell \leqslant U < 0$ of the typical form: *for all bandit problems* $\underline{\nu}$ in $\mathcal{D}$, with a unique optimal arm,

$$\ell(\underline{\nu}) \leqslant \liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant \limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant U(\underline{\nu}) < 0\,,$$

or, put differently, $$\exp\Big(\ell(\underline{\nu})\,T\big(1+o(1)\big)\Big) \leqslant \mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant \exp\Big(U(\underline{\nu})\,T\big(1+o(1)\big)\Big)\,.$$

This problem is generally considered more difficult than the fixed-confidence setting, and even for parametric models like canonical one-parameter exponential models, no strategy with matching upper and lower bounds (i.e., no optimal strategy) is known so far. Indeed, even the possibility of obtaining such bounds is disputed.

Earlier approaches. So far, four main approaches were considered for the problem of best-arm identification with a fixed budget. *First*, the early approach by Audibert et al. (2010) relies on gaps: we define the gap Δ_a of arm a as the difference $\mu^\star - \mu_a$ between the largest expectation $\mu^\star$ in $\underline{\nu}$ and the expectation of the distribution ν_a. They introduce a successive-rejects strategy and provide gap-based upper bounds for sub-Gaussian models, based on Hoeffding's inequality. They however propose a lower bound only in the case of a Bernoulli model, not for larger, non-parametric, models. This lower bound was further discussed by Carpentier and Locatelli (2016). *A second series of approaches* (see, e.g., Kaufman et al., 2016) focused on Gaussian bandits with fixed variances, but results do not seem to be easily generalized to other models as they often rely on specific facts (related, among others, to the symmetry of the Kullback-Leibler divergence). *A third approach*, led by Russo (2016, 2020), considered canonical one-parameter exponential families, but a for a different target probability. Namely, a Bayesian setting is considered and the quality of a strategy is measured as the posterior probability of identifying the best arm. An optimal non-gap-based complexity is exhibited, together with optimal strategies matching this complexity. However, Komiyama (2022) argue that such an approach is specific to the Bayesian case and is not suited to the frequentist case that we consider. *A fourth approach* is to focus on the case of $K = 2$ arms, see, e.g., Kaufman et al. (2016). The non-parametric bounds obtained therein do not enjoy any obvious generalization to the case of $K \geqslant 3$ arms. By considering very specific models, Kato et al. (2022) constructed a strategy that is optimal (only) in the regime where the gap between the 2 arms is small—yet, this gap-based approach does not, by nature, go in the direction of non-parametric bounds.

We will provide more details concerning some of these approaches while presenting and discussing our main results, in Section 2.2; see also Appendix E.

Content and outline of this article. We focus our attention on general upper and lower bounds, holding for all possible models $\mathcal{D}$, including non-parametric models, and valid for any number K of arms. Put differently, we target a high degree of generality. While admittedly not exhibiting matching upper and lower bounds, we show that the same information-theoretic quantities $\mathcal{L}_{\inf}^{<}$ and $\mathcal{L}_{\inf}^{>}$ are at stake in these upper and lower bounds. They are defined, in Section 2, as infima of Kullback-Leibler divergences and provide a quantification of the difficulty of the identification in terms of the geometry of information of the problem. We also provide in this section an overview of our results, which we carefully compare to existing bounds (restated therein, occasionally with some improvements). We study in Section 3 the classical successive-rejects strategy and provide an improved analysis, not relying on gaps through Hoeffding's lemma. Section 4 exhibits several possible lower bounds, which are inversely larger to the strength of the assumptions made on the strategies. These lower bounds generalize known lower bounds in the literature, like the lower bound for Bernoulli models by Audibert et al. (2010), but hold for arbitrary models. They share some similar flavor with the lower bounds by Lai and Robbins (1985) and Burnetas and Katehakis (1996) for the cumulative regret.

2. Overview of the results and more extended literature review

Before being able to actually provide a formal summary of our results, we introduce new quantifications of the difficulty of a bandit problem in terms of geometry of the information. These quantifications should be reminiscent of the $\mathcal{K}_{\inf}$ quantity that appears in the optimal bounds on the cumulative regret; it was introduced by Honda and Takemura (2015), see also Garivier et al. (2022). Denoting by ν the distribution of interest, this $\mathcal{K}_{\inf}$ is defined as an infimum over divergences of the form $\mathrm{KL}(\nu, \zeta)$. For our objectives, the arguments in the KL are in reverse order, and we are rather interested in infima over divergences of the form $\mathrm{KL}(\zeta, \nu)$.

Except for very specific models (e.g., the model $\mathcal{D}_{\sigma^2}$ of Gaussian distributions with a fixed variance $\sigma^2 > 0$), the Kullback-Leibler divergence is not symmetric, i.e., $\mathrm{KL}(\zeta, \nu)$ and $\mathrm{KL}(\nu, \zeta)$ differ in general. Specific best-arm-identification results were obtained by Kaufman et al. (2016) for the model $\mathcal{D}_{\sigma^2}$, based on the Bretagnolle-Huber inequality; they indicate that the sum of the inverse squared gaps would be driving both the lower bound and upper bound functions ℓ and U. However, a close look at the proof reveals that they heavily rely (among others) on the symmetry of KL for this model. There is therefore absolutely no hope to provide any generalization beyond the Gaussian case. Appendix E.2 provides furthers details and discussions on this matter.

2.1. The key new quantities: $\mathcal{L}_{\inf}^{<}$ and $\mathcal{L}_{\inf}^{\leqslant}$, as well as $\mathcal{L}_{\inf}^{>}$ and $\mathcal{L}_{\inf}^{\geqslant}$

In this article, we only consider models $\mathcal{D}$ whose distributions all admit an expectation. We denote by $\mathrm{E}(\zeta)$ the expectation of a distribution $\zeta \in \mathcal{D}$. For a distribution $\nu \in \mathcal{D}$ and a real number $x \in \mathbb{R}$, we then introduce

$$\mathcal{L}_{\inf}^{<}(x, \nu) = \inf\big\{\mathrm{KL}(\zeta, \nu) : \zeta \in \mathcal{D} \text{ s.t. } \mathrm{E}(\zeta) < x\big\}$$
$$\text{and} \qquad \mathcal{L}_{\inf}^{\leqslant}(x, \nu) = \inf\big\{\mathrm{KL}(\zeta, \nu) : \zeta \in \mathcal{D} \text{ s.t. } \mathrm{E}(\zeta) \leqslant x\big\}\,,$$

where KL denotes the Kullback-Leibler divergence and with the usual convention that the infimum of an empty set equals $+\infty$. Symmetrically, by considering rather distributions ζ with expectations larger than x, we define

$$\mathcal{L}^{>}_{\inf}(x,\nu) = \inf\big\{\mathrm{KL}(\zeta,\nu) : \zeta \in \mathcal{D} \text{ s.t. } \mathrm{E}(\zeta) > x\big\}$$
$$\text{and} \qquad \mathcal{L}^{\geqslant}_{\inf}(x,\nu) = \inf\big\{\mathrm{KL}(\zeta,\nu) : \zeta \in \mathcal{D} \text{ s.t. } \mathrm{E}(\zeta) \geqslant x\big\}\,.$$

We state some general properties on these quantities in Appendix A—among others, that $\mathcal{L}^{<}_{\inf}$ and $\mathcal{L}^{\leqslant}_{\inf}$, as well as $\mathcal{L}^{>}_{\inf}$ and $\mathcal{L}^{\geqslant}_{\inf}$, are almost identical for the model $\mathcal{P}[0,1]$. The same holds for canonical one-parameter exponential models, as discussed in Appendix C.3.

Lower bounds will be typically expressed with $\mathcal{L}^{<}_{\inf}$ and $\mathcal{L}^{>}_{\inf}$ quantities, while upper bounds will rely on $\mathcal{L}^{\leqslant}_{\inf}$ and $\mathcal{L}^{\geqslant}_{\inf}$ quantities.

2.2. Overview of the results

The paper provides new and more general (possibly non-parametric) bounds based on information-theoretic quantities, namely the quantities $\mathcal{L}^{<}_{\inf}$, $\mathcal{L}^{\leqslant}_{\inf}$, $\mathcal{L}^{>}_{\inf}$ and $\mathcal{L}^{\geqslant}_{\inf}$. In particular, we consider a version of Chernoff information defined, for ν, ν' in $\mathcal{D}$ with $\mathrm{E}(\nu') < \mathrm{E}(\nu)$, as

$$\mathcal{L}(\nu',\nu) = \inf_{x\in[\mathrm{E}(\nu'),\mathrm{E}(\nu)]} \Big\{\mathcal{L}^{\geqslant}_{\inf}(x,\nu') + \mathcal{L}^{\leqslant}_{\inf}(x,\nu)\Big\}\,.$$

Given a bandit problem $\underline{\nu}$ with a unique optimal distribution denoted by $\nu^\star$, we may rank the arms a in non-decreasing order of $\mathcal{L}\big(\nu_a,\nu^\star\big)$, i.e., consider the permutation σ such that

$$0 = \mathcal{L}\big(\nu_{\sigma_1},\nu^\star\big) < \mathcal{L}\big(\nu_{\sigma_2},\nu^\star\big) \leqslant \ldots \leqslant \mathcal{L}\big(\nu_{\sigma_{K-1}},\nu^\star\big) \leqslant \mathcal{L}\big(\nu_{\sigma_K},\nu^\star\big)\,.$$

Our first main result (Corollary 3 together with Lemma 4) considers models $\mathcal{D}$ like $\mathcal{D} = \mathcal{P}[0,1]$, the set of all probability distributions over $[0,1]$, or $\mathcal{D} = \mathcal{D}_{\text{exp}}$, any canonical one-parameter exponential family. It studies the successive-rejects strategy, introduced by Audibert et al. (2010), for which arms are rejected one by one at the end of phases of uniform exploration, and states that this strategy is such that for all bandit problems $\underline{\nu}$ in $\mathcal{D}$ with a unique optimal arm,

$$\limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant -\frac{1}{\overline{\ln} K} \min_{2\leqslant k\leqslant K} \frac{\mathcal{L}\big(\nu_{\sigma_k},\nu^\star\big)}{k}\,, \tag{1}$$

where $\overline{\ln} K \approx \ln K$. The key for this result (Lemma 1, of independent interest) is a grid-based application of the Cramér-Chernoff bound to control $\mathbb{P}\big(\overline{X}_N \leqslant \overline{Y}_N\big)$, where $\overline{X}_N$ and $\overline{Y}_N$ are averages of two independent N–samples. This approach can be used to analyze similar algorithms, like sequential halving (Karnin et al., 2013).

The corresponding lower bounds are stated rather in terms of $\mathcal{L}^{<}_{\inf}$ and $\mathcal{L}^{>}_{\inf}$ quantities, but Appendix A explains why, except in a single pathological case, $\mathcal{L}(\nu',\nu)$ could be alternatively defined with $\mathcal{L}^{<}_{\inf}$ and $\mathcal{L}^{>}_{\inf}$ instead of $\mathcal{L}^{\leqslant}_{\inf}$ and $\mathcal{L}^{\geqslant}_{\inf}$. We actually state several lower bounds in Section 4, that are larger as the assumptions on the strategies considered are more restrictive; as usual, there is a trade-off between the strength of a lower bound and its generality. However, all assumptions considered remain rather mild and are satisfied by successive-rejects-type strategies: e.g., one may restrict the attention to strategies such that for all bandit problems, the arm associated with the smallest expectation is pulled less than a fraction $1/K$ of the time. *Our second main result* (Theorem 12) gives the following alma matter lower bound: for all bandit problems $\underline{\nu}$ with no two same expectations,

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -\min_{2\leqslant k\leqslant K} \inf_{x\in[\mu_{(k)},\mu_{(k-1)})} \left\{\frac{\mathcal{L}^{>}_{\inf}\big(x,\nu_{(k)}\big)}{k-1} + \frac{\mathcal{L}^{<}_{\inf}\big(x,\nu^\star\big)}{k}\right\}, \tag{2}$$

where we considered the order statistics in reverse order, $\mu_{(1)} > \mu_{(2)} > \mu_{(3)} > \cdots > \mu_{(K)}$, and where $\nu_{(a)}$ denotes the distribution with expectation $\mu_{(a)}$.

This lower bound does not match the exhibited upper bound, mainly because the infima in (2) are only taken on restricted ranges $[\mu_{(k)}, \mu_{(k-1)})$ and not on the entire intervals $[\mu_{(k)}, \mu^\star]$ as in (1). Still, we argue that quantities defined as infima over x of $\mathcal{L}^{>}_{\inf}\big(x, \nu_{(k)}\big) + \mathcal{L}^{<}_{\inf}\big(x, \nu^\star\big)$ should measure how difficult a best-arm-identification problem is under a fixed budget. *This is the main insight of this article.*

We now compare our general bounds to existing bounds, for sub-Gaussian models and for exponential models. To do so, we will sometimes consider the following consequence of the lower bound (2), obtained by picking $x = \mu_{(k)}$:

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \min_{2\leqslant k\leqslant K} \frac{\mathcal{L}^{<}_{\inf}\big(\mu_{(k)}, \nu^\star\big)}{k} \,. \tag{3}$$

Comparison to the gap-based approaches. Audibert et al. (2010) propose an analysis of the successive-rejects strategy based on Hoeffding's inequality, stating that for all bandit problems in $\mathcal{P}[0,1]$ with a unique optimal arm,

$$\limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant - \frac{1}{\overline{\ln} K} \min_{2\leqslant k\leqslant K} \frac{\Delta^2_{(k)}}{k} \,, \tag{4}$$

where we recall the definition of the gaps $\Delta_{(k)} = \mu^\star - \mu_{(k)}$. This bound is a consequence of (Corollary 3, a slightly more general form of the) bound (1), given Pinsker's inequality (18):

$$\mathcal{L}\big(\nu_{(k)}, \nu^\star\big) \geqslant \inf_{x\in[\mu_{(k)},\mu^\star]} \Big\{ 2\big(x - \mu_{(k)}\big)^2 + 2(x - \mu^\star)^2 \Big\} = \big(\mu^\star - \mu_{(k)}\big)^2 = \Delta^2_{(k)} \,. \tag{5}$$

We remark that the bound (4) and the lower bound on $\mathcal{L}\big(\nu_{(k)}, \nu^\star\big)$ may actually be extended to the model of σ^2–sub-Gaussian distributions, up to dividing the bound (4) by a factor $4\sigma^2$. We do not discuss the UCB-E algorithm of Audibert et al. (2010), as its performance and analysis crucially depend on a tuning parameter set with some knowledge of the gaps.

Audibert et al. (2010) also propose a carefully constructed lower bound for the model $\mathcal{D}_p = \big\{\mathrm{Ber}(x) : x \in [p, 1-p]\big\}$ of Bernoulli distributions $\mathrm{Ber}(x)$ with parameters x in $[p, 1-p]$ for some $p \in (0, 1/2)$. A key inequality in their proof follows from the Kullback-Leibler – χ^2-divergence bound:

$$\forall x, y \in [p, 1-p], \qquad \mathrm{KL}\big(\mathrm{Ber}(x), \mathrm{Ber}(y)\big) \leqslant \frac{(x-y)^2}{2p(1-p)} \,.$$

Their construction may actually be generalized to models $\mathcal{D}$ with $C_{\mathcal{D}} > 0$ such that for all ν, ν' in $\mathcal{D}$, one has $\mathrm{KL}(\nu, \nu') \leqslant C_{\mathcal{D}}\big(\mathrm{E}(\nu) - \mathrm{E}(\nu')\big)^2$. This is a property that clearly holds for some exponential problems: on top of the restricted Bernoulli model discussed above, for which $C_{\mathcal{D}_p} = 1/\big(2p(1-p)\big)$, we may cite the model $\mathcal{D}_{\sigma^2}$ of Gaussian distributions with variance σ^2, for which $C_{\mathcal{D}_{\sigma^2}} = 1/(2\sigma^2)$. For models enjoying the existence of such a constant $C_{\mathcal{D}}$, (a straightforward modification of) the analysis by Audibert et al. (2010) entails that for any $\underline{\nu}$ in $\mathcal{D}$,

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -5 C_{\mathcal{D}} \min_{2\leqslant k\leqslant K} \frac{\Delta^2_{(k)}}{k} \,. \tag{6}$$

As by the very assumption on the model, $\mathcal{L}^{<}_{\inf}\big(\mu_{(k)}, \nu^\star\big) \leqslant C_{\mathcal{D}}\, \Delta^2_{(k)}$, the lower bound (3) implies the stated lower bound (6), with an improved constant factor.

The lower bound (6) and the upper bound (4) differ in particular by a factor of about $\approx \ln K$. Carpentier and Locatelli (2016) discuss this gap in the case of a restricted Bernoulli model and improve the lower bound (6) by a factor of $\ln K$, but not simultaneously for all bandit problems $\underline{\nu}$ (as we aim for); they obtain the improvement just for one bandit problem $\underline{\nu}$. Their lower bound result (formally stated and discussed in Appendix E.1) is therefore of a totally different nature. More results on how and when given lower bounds with a given complexity measure may, or may not, be improved were stated by Komiyama et al. (2022).

Discussion of the non-parametric bound for $K = 2$ arms of Kaufman et al. (2016). It turns out that the existing literature offered so far a non-parametric bound, in the case of $K = 2$ arms. Namely, in a general, possibly non-parametric model $\mathcal{D}$, Kaufman et al. (2016, Theorem 12) stated a lower bound for all 2–armed bandit problems $\underline{\nu} = (\nu_1, \nu_2)$:

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \inf_{\substack{\underline{\lambda} \text{ in } \mathcal{D}: \\ \mathrm{E}(\lambda_{a^\star(\underline{\nu})}) < \mathrm{E}(\lambda_{w_\star(\underline{\nu})})}} \max\Big\{ \mathrm{KL}\big(\lambda_{w_\star(\underline{\nu})}, \nu_{w_\star(\underline{\nu})}\big),\, \mathrm{KL}\big(\lambda_{a^\star(\underline{\nu})}, \nu_{a^\star(\underline{\nu})}\big) \Big\}, \tag{7}$$

where $w_\star(\underline{\nu})$ denotes the suboptimal arm in $\underline{\nu}$ and where the infimum is over all alternative bandit problems (λ_1, λ_2) in $\mathcal{D}$ with inverse order on the expectations compared to $\underline{\nu}$. We note (see the proof of Theorem 13) that we may actually rewrite this lower bound in a more readable way, in terms of $\mathcal{L}^<_{\inf}$ and $\mathcal{L}^>_{\inf}$ quantities, illustrating once again that these quantities are key in measuring the complexity of best-arm-identification under a fixed budget:

$$\begin{aligned} \inf_{\substack{\underline{\lambda} \text{ in } \mathcal{D}: \\ \mathrm{E}(\lambda_{a^\star(\underline{\nu})}) < \mathrm{E}(\lambda_{w_\star(\underline{\nu})})}} & \max\Big\{ \mathrm{KL}\big(\lambda_{w_\star(\underline{\nu})}, \nu_{w_\star(\underline{\nu})}\big),\, \mathrm{KL}\big(\lambda_{a^\star(\underline{\nu})}, \nu_{a^\star(\underline{\nu})}\big) \Big\} \\ & = \inf_{x \in [\mu_{w_\star(\underline{\nu})}, \mu^\star]} \Big\{ \max\Big\{ \mathcal{L}^>_{\inf}\big(x, \nu_{w_\star(\underline{\nu})}\big),\, \mathcal{L}^<_{\inf}\big(x, \nu^\star\big) \Big\} \Big\}. \end{aligned} \tag{8}$$

The proof technique of Kaufman et al. (2016) may be applied in a pairwise fashion to generalize the above lower bound for 2 arms into a lower bound for $K \geqslant 2$ arms, stated in Theorem 13: for all $\underline{\nu}$ in $\mathcal{D}$ with a unique optimal arm,

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \min_{k \neq a^\star(\underline{\nu})} \inf_{x \in [\mu_k, \mu^\star]} \Big\{ \max\big\{ \mathcal{L}^>_{\inf}(x, \nu_k),\, \mathcal{L}^<_{\inf}(x, \nu^\star) \big\} \Big\}.$$

We however do not claim that this is a deep and interesting bound, as it only involves pairwise comparisons with the best arm. In particular, we lack divisions by the ranks of the arms, as in (2).

Bounds for exponential models. We denote by $\mathcal{D}_{\mathrm{exp}}$ the model corresponding to a canonical one-parameter exponential family with expectations defined on an open interval $\mathcal{M}$ (see Appendix C.3 for a reminder on this matter). For such a model, we denote by d the mean-parameterized Kullback-Leibler divergence. By continuity of d, we have that for all ν in $\mathcal{D}_{\mathrm{exp}}$ and for all $x \in \mathcal{M}$,

$$\forall x \leqslant \mathrm{E}(\nu), \qquad \mathcal{L}^<_{\inf}(x, \nu) = \mathcal{L}^\leqslant_{\inf}(x, \nu) = d\big(x, \mathrm{E}(\nu)\big), \tag{9}$$

$$\text{and} \qquad \forall x \geqslant \mathrm{E}(\nu), \qquad \mathcal{L}^>_{\inf}(x, \nu) = \mathcal{L}^\geqslant_{\inf}(x, \nu) = d\big(x, \mathrm{E}(\nu)\big). \tag{10}$$

All bounds above then admit simple reformulations in terms of d. The version $\mathcal{L}$ of Chernoff information we introduced may also be mean-parameterized: for $\mu' < \mu$,

$$L(\mu', \mu) = \min_{x \in [\mu', \mu]} \big\{ d(x, \mu') + d(x, \mu) \big\}.$$

ALGORITHM: SUCCESSIVE-REJECTS STRATEGY

Parameters: K arms, budget T, lengths $\ell_1, \dots, \ell_{K-1} \geqslant 1$ with $\ell_1 + \dots + \ell_{K-1} = T$

Initialization: $S_0 = \{1, \dots, K\}$

For each regime: $r \in \{1, \dots, K-1\}$:

1. For each arm $a \in S_{r-1}$
   (a) Pull it $\lfloor \ell_r / (K - 1 + r) \rfloor$ times
   (b) Compute the empirical average $\overline{X}_a^r$ of the payoffs obtained in this regime and in the previous regimes
2. Drop the arm a_r with smallest average (ties broken arbitrarily):

$$S_r = S_{r-1} \setminus \{a_r\}\,, \qquad \text{where} \qquad a_r \in \operatorname*{argmin}_{a \in S_{r-1}} \overline{X}_a^r$$

Output: Recommend arm I_T, where $S_{K-1} = \{I_T\}$

The original definition of the Chernoff information $D(\mu', \mu)$ is the value $d(y, \mu)$ for $y \in [\mu', \mu]$ such that $d(y, \mu') = d(y, \mu)$. This is the quantity at stake in (8): given that $d(\,\cdot\,, \mu')$ and $d(\,\cdot\,, \mu)$ are respectively increasing and decreasing on $[\mu', \mu]$,

$$\min_{x \in [\mu', \mu]} \max\bigl\{d(x, \mu'),\, d(x, \mu)\bigr\} = D(\mu', \mu)\,.$$

Therefore, $D(\mu', \mu) \leqslant L(\mu', \mu) \leqslant 2\, D(\mu', \mu)$. This justifies that we called L (and therefore $\mathcal{L}$) a version of Chernoff information.

3. Upper bound: successive-rejects strategy, with an improved analysis

We consider the successive-rejects strategy introduced by Audibert et al. (2010), for K arms and a budget T. Regime lengths are set beforehand; they are denoted by $\ell_1, \dots, \ell_{K-1} \geqslant 1$ and satisfy $\ell_1 + \dots + \ell_{K-1} = T$. The strategy maintains a list of candidate arms, starting with all arms, i.e., $S_0 = \{1, \dots, K\}$. At the end of each regime $r \in \{1, \dots, K-1\}$, it drops an arm to get S_r, while during the regime r, it operates with the $K - r + 1$ arms in S_{r-1}.

More precisely, during regime $r \in \{1, \dots, K-1\}$, the strategy draws $\lfloor \ell_r / (K - r + 1) \rfloor$ times each arm in S_{r-1} (and does not use the few remaining time steps, if there are some). At the end of each regime r, the strategy computes the empirical averages $\overline{X}_a^r$ of the payoffs obtained by each arm $a \in S_{r-1}$ since the beginning; i.e., $\overline{X}_a^r$ is an average over

$$N_r = \lfloor \ell_1 / K \rfloor + \dots + \lfloor \ell_r / (K - r + 1) \rfloor$$

i.i.d. realizations of ν_a. It then drops the arm a_r with smallest empirical average (ties broken arbitrarily). This description is summarized in the algorithm box.

3.1. General analysis

The key quantities for the general analysis will be the logarithmic moment-generating function ϕ_ν of a distribution $\nu \in \mathcal{D}$, and its Fenchel-Legendre transform $\phi_\nu^\star$:

$$\forall \lambda \in \mathbb{R}, \quad \phi_\nu(\lambda) = \ln \int_{\mathbb{R}} \mathrm{e}^{\lambda x} \, \mathrm{d}\nu(x) \qquad \text{and} \qquad \forall x \in \mathbb{R}, \quad \phi_\nu^\star(x) = \sup_{\lambda \in \mathbb{R}} \big\{ \lambda x - \phi_\nu(\lambda) \big\} . \quad (11)$$

Based on them, we can now define, for all $\nu, \nu' \in \mathcal{D}$ with $\mathrm{E}(\nu') < \mathrm{E}(\nu)$,

$$\Phi(\nu', \nu) \stackrel{\text{def}}{=} \inf_{x \in [\mathrm{E}(\nu'), \mathrm{E}(\nu)]} \big\{ \phi_{\nu'}^\star(x) + \phi_\nu^\star(x) \big\} .$$

The following simple lemma shows that Φ plays a significant role for bounding the probability that two sample averages are in inverse order compared to the expectations of the underlying distributions. It supersedes the use of Hoeffding's inequality in Audibert et al. (2010).

Lemma 1 *Fix ν and ν' in $\mathcal{D}$, with respective expectations $\mu = \mathrm{E}(\nu) > \mu' = \mathrm{E}(\nu')$. For all $N \geqslant 1$, let $\overline{X}_N$ and $\overline{Y}_N$ be the averages of N–samples with respective distributions ν and ν'. Then,*

$$\limsup_{N \to +\infty} \frac{1}{N} \ln \mathbb{P}\big(\overline{X}_N \leqslant \overline{Y}_N \big) \leqslant - \inf_{x \in [\mu', \mu]} \big\{ \phi_{\nu'}^\star(x) + \phi_\nu^\star(x) \big\} \stackrel{\text{def}}{=} -\Phi(\nu', \nu) .$$

Proof sketch The fact $\overline{X}_N \leqslant \overline{Y}_N$ entails the existence of x such that $\overline{X}_N \leqslant x \leqslant \overline{Y}_N$. By independence, together with two applications of the Cramér-Chernoff bound (recalled in Appendix B.1),

$$\mathbb{P}\big(\overline{X}_N \leqslant x \leqslant \overline{Y}_N \big) = \mathbb{P}\big(\overline{X}_N \leqslant x \big) \, \mathbb{P}\big(x \leqslant \overline{Y}_N \big) \leqslant \exp\big(-N \, \phi_\nu^\star(x) \big) \ \exp\big(-N \, \phi_{\nu'}^\star(x) \big) .$$

The technical issue is then to deal with some union over x of the events $\big\{ \overline{X}_N \leqslant x \leqslant \overline{Y}_N \big\}$. We do so with a sequence of finite grids, with vanishing steps, and use lower-semi-continuity arguments to obtain an infimum over an interval based on a sequence of finite minima. A complete proof is to be found in Appendix B.2. ∎

The main performance upper bound is stated below in terms of Φ, that is, in terms of Fenchel-Legendre transforms of logarithmic moment-generating functions. Section 3.2 will later explain why and when the latter may be replaced by $\mathcal{L}_{\inf}^{\leqslant}$ and $\mathcal{L}_{\inf}^{\geqslant}$ quantities, leading to a rewriting $\Phi = \mathcal{L}$ and to the bound claimed in (1).

Theorem 2 *Fix $K \geqslant 2$ and a model $\mathcal{D}$. Consider a sequence of successive-rejects strategies, indexed by T, such that $N_r / T \to \gamma_r > 0$ as $T \to +\infty$ for all $r \in \{1, \dots, K-1\}$. Let $\underline{\nu}$ be a bandit problem in $\mathcal{D}$ with a unique optimal arm and, for each $r \in \{1, \dots, K-1\}$, let $\mathcal{A}_r$ be a subset of arms of cardinality r that does not contain $a^\star(\underline{\nu})$. Then*

$$\limsup_{T \to +\infty} \frac{1}{T} \ln \mathbb{P}\big(I_T \neq a^\star(\underline{\nu}) \big) \leqslant - \min_{1 \leqslant r \leqslant K-1} \Big\{ \gamma_r \min_{k \in \mathcal{A}_r} \Phi\big(\nu_k, \nu^\star \big) \Big\} .$$

Proof sketch A complete proof may be found in Appendix B.3; it mimics the analysis by Audibert et al. (2010), the main modification being the substitution of Hoeffding's inequality by the bound of Lemma 1. We have $I_T \neq a^\star(\underline{\nu})$ if and only if $a^\star(\underline{\nu})$ is rejected in some regime, i.e.,

$$\big\{ I_T \neq a^\star(\underline{\nu}) \big\} = \bigcup_{r=1}^{K-1} \big\{ a_r = a^\star(\underline{\nu}) \big\} \subseteq \bigcup_{r=1}^{K-1} \Big\{ a^\star(\underline{\nu}) \in S_{r-1} \ \text{ and } \ \forall k \in S_{r-1}, \ \ \overline{X}^r_{a^\star(\underline{\nu})} \leqslant \overline{X}^r_k \Big\} .$$

By optional skipping (see Doob, 1953, Chapter III, Theorem 5.2, p. 145) and by the fact that by the pigeonhole principle, the (random) set S_{r-1} necessarily contains one element of the deterministic set $\mathcal{A}_r$,

$$\mathbb{P}\Big(a^\star(\underline{\nu}) \in S_{r-1} \text{ and } \forall k \in S_{r-1},\ \overline{X}^r_{a^\star(\underline{\nu})} \leqslant \overline{X}^r_k\Big) \leqslant \sum_{k\in\mathcal{A}_r} \mathbb{P}\Big(\overline{Y}^r_{a^\star(\underline{\nu})} \leqslant \overline{Y}^r_k\Big)\,,$$

where, for all a, the $\overline{Y}^r_a$ are the averages of independent N_r–samples distributed according to ν_a. The proof is concluded by Lemma 1 and the fact that a sum of exponentially fast decaying quantities is driven by its largest term. ■

We conclude this subsection by stating the bound of Theorem 2 for the regime lengths suggested by Audibert et al. (2010), namely, $\ell_1 = T/\overline{\ln}\,K$ and for $r \in \{2,\dots,K-1\}$,

$$\ell_r = \frac{T}{(K-r+2)\,\overline{\ln}\,K}\,, \qquad \text{where} \qquad \overline{\ln}\,K = \frac{1}{2} + \sum_{k=2}^{K} \frac{1}{k}\,. \tag{12}$$

We also consider lower bounds $f\big(\nu_k,\nu^\star\big)$ on the $\Phi\big(\nu_k,\nu^\star\big)$. We may of course use $f = \Phi$ but sometimes, it is handy to rely on more readable lower bounds. For instance, in the case of the $\mathcal{P}[0,1]$ model, Hoeffding's inequality entails that

$$\phi^\star_\nu(x) \geqslant 2\big(x - \mathrm{E}(\nu)\big)^2\,, \qquad \text{so that} \qquad \Phi\big(\nu_k,\nu^\star\big) \geqslant \Delta_k^2 \stackrel{\text{def}}{=} f\big(\nu_k,\nu^\star\big)\,; \tag{13}$$

see more details in Appendix B.4. Such bounds hold more generally in models consisting of sub-Gaussian distributions.

We now order the arms into $\sigma_1,\dots,\sigma_K$ based on f, namely, we let $\sigma_1 = a^\star(\underline{\nu})$ and

$$0 = f\big(\nu_{\sigma_1},\nu^\star\big) < f\big(\nu_{\sigma_2},\nu^\star\big) \leqslant \ldots \leqslant f\big(\nu_{\sigma_{K-1}},\nu^\star\big) \leqslant f\big(\nu_{\sigma_K},\nu^\star\big)\,, \tag{14}$$

and we take $\mathcal{A}_r = \{\sigma_{K-r+1},\dots,\sigma_K\}$. We obtain immediately the following corollary, for which a detailed proof may be found, for the sake of completeness, in Appendix B.4.

Corollary 3 *Fix $K \geqslant 2$, a model $\mathcal{D}$, and consider a lower bound f on Φ. The sequence of successive-rejects strategies based on the regime lengths (12) ensures, that for all bandit problems $\underline{\nu}$ in $\mathcal{D}$ with a unique optimal arm,*

$$\limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant -\frac{1}{\overline{\ln}\,K} \min_{2\leqslant k\leqslant K} \frac{f\big(\nu_{\sigma_k},\nu^\star\big)}{k}\,,$$

where arms were reordered as in (14).

3.2. On links between Φ and the quantities $\mathcal{L}^<_{\inf}$, $\mathcal{L}^\leqslant_{\inf}$, $\mathcal{L}^>_{\inf}$ and $\mathcal{L}^\geqslant_{\inf}$

The Fenchel-Legendre transform $\phi^\star_\nu$ of the logarithmic moment-generating function of ν admits a classical (see, e.g., Boucheron et al., 2013, Exercice 4.13) dual formulation in terms of infima of Kullback-Leibler divergences. The following lemma, proved in Appendix C.2, reveals that these infima correspond to $\mathcal{L}^\leqslant_{\inf}$ and $\mathcal{L}^\geqslant_{\inf}$ for the model $\mathcal{P}[0,1]$ of distributions supported on $[0,1]$.

Lemma 4 *Consider the model $\mathcal{D} = \mathcal{P}[0,1]$. For all $\nu \in \mathcal{P}[0,1]$,*

$$\forall x \leqslant \mathrm{E}(\nu), \quad \phi_\nu^\star(x) = \mathcal{L}_{\inf}^{\leqslant}(x,\nu) \qquad \textit{and} \qquad \forall x \geqslant \mathrm{E}(\nu), \quad \phi_\nu^\star(x) = \mathcal{L}_{\inf}^{\geqslant}(x,\nu)\,.$$

Based on this lemma, we have the following rewriting, which is useful to reinterpret the quantities appearing in Theorem 2 and Corollary 3: $\Phi(\nu',\nu) = \mathcal{L}(\nu',\nu)$ for the model $\mathcal{P}[0,1]$, i.e.,

$$\inf_{x\in[\mathrm{E}(\nu'),\mathrm{E}(\nu)]} \bigl\{\phi_{\nu'}^\star(x) + \phi_\nu^\star(x)\bigr\} = \inf_{x\in[\mathrm{E}(\nu'),\mathrm{E}(\nu)]} \bigl\{\mathcal{L}_{\inf}^{\geqslant}(x,\nu') + \mathcal{L}_{\inf}^{\leqslant}(x,\nu)\bigr\}\,. \tag{15}$$

For canonical one-parameter exponential models $\mathcal{D}_{\mathrm{exp}}$, a slightly weaker version of Lemma 4, only holding for x corresponding to expectations in $\mathcal{D}_{\mathrm{exp}}$ and provided in Appendix C.3, similarly shows (15), i.e., $\Phi = \mathcal{L}$. Conditions on general models for $\Phi = \mathcal{L}$ to hold are discussed in Appendix C.4.

4. Lower bounds

In most of this section, we restrict our attention to generic K–armed bandit problems $\underline{\nu}$, that are such that $\mu_j \neq \mu_k$ for $j \neq k$. In particular, the best arm $a^\star(\underline{\nu})$ is unique.

Definition of a strategy, and of a (doubly-indexed) sequence of strategies. A strategy (ψ,φ) depends on the budget T and the number K of arms; it consists of a sampling scheme $\psi = (\psi_t)_{1\leqslant t\leqslant T}$ and a recommendation function φ. At each round $t \in \{1,\dots,T\}$, the strategy picks an arm A_t, possibly at random using an auxiliary randomization U_{t-1}. Given this choice A_t, the strategy observes a payoff Y_t drawn at random according to ν_{A_t}, independently from the past. For $t \geqslant 2$, the choice A_t is therefore a measurable function $A_t = \psi_t(H_t)$ of the history $H_t = (U_0,\, Y_1,\, \dots,\, Y_{t-1},\, U_{t-1})$, while $A_1 = \psi_1(H_0)$, where $H_0 = U_0$. At round T, the strategy recommends the arm $I_T = \varphi(H_T)$.

For our lower bounds, we will consider sequences of strategies, either only indexed by $T \geqslant 1$ given a value of $K \geqslant 2$, or doubly indexed by T and K. These sequences will also be assumed to be "reasonable" in the sense below.

Consistent (or exponentially consistent) sequences of strategies. The probability $\mathbb{P}\bigl(I_T \neq a^\star(\underline{\nu})\bigr)$ of misidentifying the unique optimal arm may vanish asymptotically (and even vanish exponentially fast) for all bandit problems—in not too large a model $\mathcal{D}$, as illustrated in Section 3. We will therefore only be interested in such sequences of strategies, called (exponentially) consistent. In the sequel and for extra clarity, we index the probabilities by the ambient bandit problem $\underline{\nu}$ considered.

Definition 5 *Fix $K \geqslant 2$. A sequence of strategies indexed by $T \geqslant 1$ is consistent, respectively, exponentially consistent, on a model $\mathcal{D}$ if for all generic problems $\underline{\nu}$ in $\mathcal{D}$,*

$$\mathbb{P}_{\underline{\nu}}\bigl(I_T \neq a^\star(\underline{\nu})\bigr) \underset{T\to+\infty}{\longrightarrow} 0\,, \qquad \textit{respectively,} \qquad \limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\bigl(I_T \neq a^\star(\underline{\nu})\bigr) < 0\,.$$

By extension, a doubly-indexed sequence of strategies is (exponentially) consistent if for all $K \geqslant 2$, the associated sequences of strategies are so.

The fundamental inequality. We denote by

$$N_a(T) = \sum_{t=1}^{T} \mathbb{I}_{\{A_t=a\}}$$

the number of times arm a was pulled in the T exploration rounds of a given strategy with budget $T \geqslant 1$. The fundamental inequality by Garivier et al. (2019), together with the very definition of consistency, yields in a straightforward manner our building block for lower bounds. Details of the derivation are provided in Appendix D.1, for the sake of completeness.

Lemma 6 *Fix $K \geqslant 2$ and a model $\mathcal{D}$. Consider a consistent sequence of strategies on $\mathcal{D}$, and two generic bandit problems $\underline{\nu}$ and $\underline{\lambda}$ in $\mathcal{D}$ such that $a^\star(\underline{\lambda}) \neq a^\star(\underline{\nu})$. Then*

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -\limsup_{T\to+\infty} \sum_{a=1}^{K} \frac{\mathbb{E}_{\underline{\lambda}}[N_a(T)]}{T} \mathrm{KL}(\lambda_a, \nu_a)\,.$$

4.1. A lower bound revisiting and extending the one by Audibert et al. (2010)

The focus of this subsection is to establish the lower bound (3), from which we derived the gap-based lower bound (4) by Audibert et al. (2010). The lower bound (3) is smaller than the lower bound to be exhibited in the next subsection, but it comes with less restrictive assumptions on the behaviors of the sequences of strategies considered.

Firstly, we only consider sequences of strategies—actually, sequences of sampling schemes—that do not pull too often the worst arm, and which we will refer to as being balanced against the worst arm. Successive-rejects-type strategies sample the worst arm less than other arms in expectations, and hence, are indeed balanced against the worst arm. To define this constraint formally, we denote by $w_\star(\underline{\nu})$ the index of the unique worst arm of a generic bandit problem $\underline{\nu}$.

Definition 7 *A doubly-indexed sequence of strategies is balanced against the worst arm on a model $\mathcal{D}$ if for all $K \geqslant 2$, for all generic K–armed bandit problems $\underline{\nu}$ in $\mathcal{D}$,*

$$\limsup_{T\to+\infty} \frac{1}{T} \mathbb{E}_{\underline{\nu}}\big[N_{w_\star(\underline{\nu})}(T)\big] \leqslant \frac{1}{K}\,.$$

A second constraint is related to bandit subproblems. We say that $\underline{\nu}'$ is a subproblem of a K–armed bandit problem $\underline{\nu}$ if $\underline{\nu}' = (\nu_a)_{a\in\mathcal{A}}$ for a subset $\mathcal{A} \subseteq \{1,\dots,K\}$ of cardinality greater than or equal to 2; we denote by $\underline{\nu}' \subseteq \underline{\nu}$ this fact. We say in addition that $\underline{\nu}'$ and $\underline{\nu}$ feature the same optimal arm if $\nu'_{a^\star(\underline{\nu}')} = \nu_{a^\star(\underline{\nu})}$. It should be easier to identify the best arm in $\underline{\nu}'$ than in $\underline{\nu}$, in the sense below, and this defines the fact that a strategy cleverly exploits pruning of suboptimal arms. Again, successive-rejects-type strategies naturally satisfy this constraint.

Definition 8 *A doubly-indexed sequence of strategies cleverly exploits pruning of suboptimal arms on a model $\mathcal{D}$ if for all generic bandit problems $\underline{\nu}$ in $\mathcal{D}$ with $K \geqslant 2$ arms, for all subproblems $\underline{\nu}' \subseteq \underline{\nu}$ featuring the same optimal arm,*

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant \liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}'}\big(I_T \neq a^\star(\underline{\nu}')\big)\,.$$

We use again the order statistics $\mu_{w_\star(\underline{\nu})} = \mu_{(K)} < \mu_{(K-1)} < \ldots < \mu_{(1)} = \mu_{a^\star(\underline{\nu})}$.

Theorem 9 *Fix a model $\mathcal{D}$. Consider a doubly-indexed sequence of strategies that is consistent, balanced against the worst arm on $\mathcal{D}$, and that cleverly exploits the pruning of suboptimal arms on $\mathcal{D}$. For all generic bandit problems $\underline{\nu}$ in $\mathcal{D}$ with $K \geqslant 2$ arms,*

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -\min_{2\leqslant k\leqslant K} \frac{\mathcal{L}^{<}_{\inf}\big(\mu_{(k)}, \nu^\star\big)}{k}\,.$$

Proof sketch The bound is proved for $k = K$ by considering alternative bandit problems $\underline{\lambda}$ differing from $\underline{\nu}$ only at arm $a^\star(\underline{\nu})$, where $\nu^\star$ is replaced by distributions $\zeta \in \mathcal{D}$ with $\mathrm{E}(\zeta) < \mu_{(K)}$. For $\underline{\lambda}$, the arm $a^\star(\underline{\nu})$ is the worst arm, and is therefore pulled less than a fraction $1/K$ of the time,

asymptotically and on average, as the strategy is balanced against the worst arm. An application of Lemma 6 concludes the case $k = K$. The extension to $k \leqslant K-1$ is obtained by clever exploitation of the pruning of suboptimal arms. A complete proof may be found in Appendix D.2. ■

4.2. A larger lower bound, for a more restrictive class of strategies

In this section, we derive a slightly stronger version of the lower bound (2). This lower bound is larger than the bound exhibited in the previous subsection but relies on stronger assumptions on the strategies considered. Namely, we introduce an assumption of monotonicity, which extends Definition 7 to provide frequency constraints on each arm $a \in \{1, \dots, K\}$.

Definition 10 *Fix $K \geqslant 2$. A sequence of strategies is monotonous on a model $\mathcal{D}$ if for all generic problems $\underline{\nu}$ in $\mathcal{D}$, for all arms $a \in \{1, \dots, K\}$,*

$$\limsup_{T \to +\infty} \frac{\mathbb{E}_{\underline{\nu}}\big[N_{(a)}(T)\big]}{T} \leqslant \frac{1}{a}\,,$$

where arms are ordered such that $\mu_{(1)} > \mu_{(2)} > \cdots > \mu_{(K)}$.

This condition is satisfied as soon as a given arm is not pulled more often, asymptotically and on average, than better-performing arms (note that Definition 10 is slightly weaker than this). Successive-rejects-type strategies naturally satisfy this requirement.

We also rely on the following assumption on the model $\mathcal{D}$, which essentially indicates that there is "no gap" in $\mathcal{D}$. Once again, the model $\mathcal{P}[0,1]$ and canonical one-parameter exponential models $\mathcal{D}_{\text{exp}}$ all satisfy this mild requirement (see Appendix D.3 for the immediate details).

Definition 11 *A model $\mathcal{D}$ is normal if for all $\nu \in \mathcal{D}$, for all $x \geqslant \mathrm{E}(\nu)$,*

$$\begin{aligned} \forall \varepsilon > 0, \qquad \mathcal{L}^{>}_{\inf}(x, \nu) &\stackrel{\text{def}}{=} \inf\big\{\mathrm{KL}(\zeta, \nu) : \zeta \in \mathcal{D} \text{ s.t. } \mathrm{E}(\zeta) > x\big\} \\ &= \inf\big\{\mathrm{KL}(\zeta, \nu) : \zeta \in \mathcal{D} \text{ s.t. } x + \varepsilon > \mathrm{E}(\zeta) > x\big\}\,. \end{aligned}$$

Theorem 12 *Fix $K \geqslant 2$ and a normal model $\mathcal{D}$. Consider a sequence of strategies which is consistent and monotonous on $\mathcal{D}$. For all generic bandit problems $\underline{\nu}$ in $\mathcal{D}$,*

$$\liminf_{T \to +\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \min_{2 \leqslant k \leqslant K} \min_{2 \leqslant j \leqslant k} \inf_{x \in [\mu_{(j)}, \mu_{(j-1)})} \left\{ \frac{\mathcal{L}^{>}_{\inf}\big(x, \nu_{(k)}\big)}{j-1} + \frac{\mathcal{L}^{<}_{\inf}\big(x, \nu^\star\big)}{j} \right\}.$$

Proof sketch A complete proof may be found in Appendix D.4. For triplets (k, j, x) satisfying the stated requirements, we consider an alternative problem $\underline{\lambda}$ differing from the original bandit problem $\underline{\nu}$ at the best arm (1) and at the k–th best arm (k), for which we pick distributions such that $\mathrm{E}\big(\lambda_{(1)}\big) < x < \mathrm{E}\big(\lambda_{(k)}\big) < \mu_{(j-1)}$. Then arm (1) is at best the j–th best arm of $\underline{\lambda}$, while arm (k) is exactly the $j-1$–th best arm of $\underline{\lambda}$. By monotonicity and Lemma 6, we obtain

$$\liminf_{T \to +\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \left(\frac{\mathrm{KL}\big(\lambda_{(k)}, \nu_{(k)}\big)}{j-1} + \frac{\mathrm{KL}\big(\lambda_{(1)}, \nu^\star\big)}{j} \right). \tag{16}$$

We get $-\mathcal{L}^{>}_{\inf}\big(x, \nu_{(k)}\big)/(j-1) - \mathcal{L}^{<}_{\inf}\big(x, \nu^\star\big)/j$ as a lower bound by taking (separate) suprema of the lower bound (16) over $\mathrm{E}\big(\lambda_{(1)}\big) < x$ and $x < \mathrm{E}\big(\lambda_{(k)}\big) < \mu_{(j-1)}$, where the $< \mu_{(j-1)}$ constraint disappears thanks to normality of the model. ■

4.3. A general lower bound, valid for any strategy

The previous subsections illustrated what may be achieved under restrictions—though natural restrictions—on the classes of strategies considered. For the sake of completeness, we also provide a lower bound relying on no other restriction than consistency; it extends the lower bound (7) exhibited by Kaufman et al. (2016) for $K = 2$ arms, and is formulated in terms of $\mathcal{L}^{<}_{\inf}$ and $\mathcal{L}^{>}_{\inf}$. A proof of the following theorem may be found in Appendix D.5.

Theorem 13 *Fix $K \geqslant 2$ and a model $\mathcal{D}$. Consider a consistent sequence of strategies on $\mathcal{D}$. For all generic bandit problems $\underline{\nu}$ in $\mathcal{D}$,*

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \min_{k\neq a^\star(\underline{\nu})} \inf_{x\in[\mu_k,\mu^\star]} \max\Big\{\mathcal{L}^{>}_{\inf}(x,\nu_k), \mathcal{L}^{<}_{\inf}(x,\nu^\star)\Big\}\,.$$

Acknowledgments

We thank Hédi Hadiji for pointers relative to the equality between $\phi^\star$ and d in the case of exponential models $\mathcal{D}_{\text{exp}}$.

References

J.-Y. Audibert, S. Bubeck, and R. Munos. Best arm identification in multi-armed bandits. In *Proceedings of the 23th Conference on Learning Theory (COLT 2010)*, 2010.

S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

J. Bretagnolle and C. Huber. Estimation des densités: risque minimax. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 47, 1979.

A.N. Burnetas and M.N. Katehakis. Optimal adaptive policies for sequential allocation problems. *Advances in Applied Mathematics*, 17(2):122–142, 1996.

A. Carpentier and A. Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Proceedings of the 29th Conference on Learning Theory (COLT 2016)*, volume 49, pages 590–604. PMLR, 2016.

Y. Chow and H. Teicher. *Probability Theory*. Springer, 1988.

R. Degenne, W. Koolen, and P. Ménard. Non-asymptotic pure exploration by solving games. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

J.L. Doob. *Stochastic Processes*. Wiley Publications in Statistics. John Wiley & Sons, 1953.

A. Garivier and E. Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference on Learning Theory (COLT 2016)*, volume 49, pages 998–1027. PMLR, 2016.

A. Garivier and E. Kaufmann. Nonasymptotic sequential tests for overlapping hypotheses applied to near-optimal arm identification in bandit models. *Sequential Analysis*, 40(1):61–96, 2021.

A. Garivier, P. Ménard, and G. Stoltz. Explore first, exploite next: the true shape of regret in bandit problems. *Mathematics of Operations Research*, 44(2):377–399, 2019.

A. Garivier, H. Hadiji, P. Ménard, and G. Stoltz. KL-UCB-switch: optimal regret bounds for stochastic bandits from both a distribution-dependent and a distribution-free viewpoints. *Journal of Machine Learning Research*, 23(179):1–66, 2022.

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.

Z. Karnin, T. Koren, and O. Somekh. Almost optimal exploration in multi-armed bandits. In *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)*, volume 28, pages 1238–1246. PMLR, 2013.

M. Kato, K. Ariu, M. Imaizumi, M. Nomura, and C. Qin. Optimal best arm identification in two-armed bandits with a fixed budget under a small gap, 2022. Preprint, arXiv:2201.04469.

E. Kaufman, O. Cappé, and A. Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016.

J. Komiyama. Suboptimal performance of the Bayes optimal algorithm in frequentist best arm identification, 2022. Preprint, arXiv:2202.05193.

J. Komiyama, T. Tsuchiya, and J. Honda. Globally optimal algorithms for fixed-budget best arm identification, 2022. Preprint, arXiv:2206.04646.

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

T. Lattimore and C. Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

E.L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer, 2nd edition, 1998.

D. Russo. Simple Bayesian algorithms for best arm identification. In *Proceedings of the 29th Conference on Learning Theory (COLT 2016)*, volume 49, pages 1417–1418. PMLR, 2016.

D. Russo. Simple Bayesian algorithms for best arm identification. *Operations Research*, 68(6): 1625–1647, 2020.

Supplementary material for "On Best-Arm Identification with a Fixed Budget in Non-Parametric Multi-Armed Bandits"

The appendices of this article contain the following elements.

- Appendix A states and proves some basic properties on quantities $\mathcal{L}^{<}_{\inf}$, $\mathcal{L}^{\leqslant}_{\inf}$, $\mathcal{L}^{>}_{\inf}$, and $\mathcal{L}^{\geqslant}_{\inf}$ that were introduced in Section 2.1.
- Appendix B provides the proofs for the first part of the analysis of the successive-rejects strategy, namely, the general analysis in terms of Φ, to be found in Section 3.1.
- Appendix C provides the proofs for the second part of the analysis of the successive-rejects strategy, namely, the rewriting of Φ as $\mathcal{L}$ that was the key contribution of Section 3.2.
- Appendix D is related to the lower bounds of Section 4, and provides detailed proofs thereof.
- Appendix E contains additional elements on the literature review of Sections 1 and 2; it states and discusses some important existing lower bounds.

Appendix A. Properties of the $\mathcal{L}^{<}_{\inf}$, $\mathcal{L}^{\leqslant}_{\inf}$, $\mathcal{L}^{>}_{\inf}$, and $\mathcal{L}^{\geqslant}_{\inf}$ quantities

We separate the list of properties in two categories: general properties, that hold for all models $\mathcal{D}$, in Appendix A.1; specific properties for the model $\mathcal{D} = \mathcal{P}[0,1]$, in Appendix A.2. It also worth noting that the $\mathcal{L}^{<}_{\inf}$, $\mathcal{L}^{\leqslant}_{\inf}$, $\mathcal{L}^{>}_{\inf}$, and $\mathcal{L}^{\geqslant}_{\inf}$ quantities admit a simple rewriting in the case of canonical one-parameter exponential models $\mathcal{D}_{\exp}$, as the mean-parameterized Kullback-Leibler divergence d, see Appendix C.3. Properties in this case thus follow from classical properties of d.

A.1. General properties

We state some properties for $\mathcal{L}^{<}_{\inf}$, that all also hold for $\mathcal{L}^{\leqslant}_{\inf}$; the corresponding properties for $\mathcal{L}^{>}_{\inf}$ and $\mathcal{L}^{\geqslant}_{\inf}$ are deduced by symmetry.

The function $\mathcal{L}^{<}_{\inf}(\,\cdot\,,\nu)$ is non-increasing and satisfies $\mathcal{L}^{<}_{\inf}(x,\nu) = 0$ for all $x > \mathrm{E}(\nu)$, as can be seen by taking $\zeta = \nu$. Also, whenever $\mathcal{D}$ is convex, the function $\mathcal{L}^{<}_{\inf}$ is jointly convex over $\mathbb{R} \times \mathcal{D}$, as indicated in the lemma below. In particular, $x \mapsto \mathcal{L}^{<}_{\inf}(x,\nu)$ is continuous on the interior of its domain (the set where it takes finite values).

Lemma 14 *When $\mathcal{D}$ is a convex model, all four functions $\mathcal{L}^{<}_{\inf}$, $\mathcal{L}^{\leqslant}_{\inf}$, $\mathcal{L}^{>}_{\inf}$, and $\mathcal{L}^{\geqslant}_{\inf}$ are jointly convex over $\mathbb{R} \times \mathcal{D}$.*

Proof We provide the proof for $\mathcal{L}^{<}_{\inf}$, and it may be adapted in a straightforward manner for the other functions.

We set two distributions ν and ν' of $\mathcal{D}$, two expectation levels μ and μ' in $\mathbb{R}$, and a weight $\lambda \in (0,1)$. We want to prove that

$$\mathcal{L}^{<}_{\inf}\big(\lambda\mu + (1-\lambda)\mu', \lambda\nu + (1-\lambda)\nu'\big) \leqslant \lambda\mathcal{L}^{<}_{\inf}(\mu,\nu) + (1-\lambda)\mathcal{L}^{<}_{\inf}(\mu',\nu')\,. \tag{17}$$

The desired inequality holds whenever $\mathcal{L}^{<}_{\inf}(\mu,\nu) = +\infty$ or $\mathcal{L}^{<}_{\inf}(\mu',\nu') = +\infty$. Otherwise, assuming that both $\mathcal{L}^{<}_{\inf}(\mu,\nu)$ and $\mathcal{L}^{<}_{\inf}(\mu',\nu')$ are finite, we set $\delta > 0$ (which we will ultimately let converge to 0) and pick ζ and ζ' in $\mathcal{D}$ such that $\mathrm{E}(\zeta) < \mu$ and $\mathrm{E}(\zeta) < \mu'$, as well as

$$\mathrm{KL}(\zeta,\nu) \leqslant \mathcal{L}^{<}_{\inf}(\mu,\nu) + \delta \quad \text{and} \quad \mathrm{KL}(\zeta',\nu') \leqslant \mathcal{L}^{<}_{\inf}(\mu',\nu') + \delta\,.$$

Then, by joint convexity of the Kullback-Leibler divergence:

$$\begin{aligned}
\lambda\mathcal{L}^{<}_{\inf}(\mu,\nu) + (1-\lambda)\mathcal{L}^{<}_{\inf}(\mu',\nu') + \delta &\geqslant \lambda\mathrm{KL}(\zeta,\nu) + (1-\lambda)\mathrm{KL}(\zeta',\nu') \\
&\geqslant \mathrm{KL}\big(\lambda\zeta + (1-\lambda)\zeta', \lambda\nu + (1-\lambda)\nu'\big) \\
&\geqslant \mathcal{L}^{<}_{\inf}\big(\lambda\mu + (1-\lambda)\mu', \lambda\nu + (1-\lambda)\nu'\big)\,,
\end{aligned}$$

where for the last inequality, we used the definition of $\mathcal{L}^{<}_{\inf}$ as an infimum and the fact that by convexity, the distribution $\lambda\zeta + (1-\lambda)\zeta'$ belongs to $\mathcal{D}$, with expectation larger than $\lambda\mu + (1-\lambda)\mu'$. The desired convexity inequality (17) follows by letting $\delta \to 0$. ∎

A.2. Specific properties for $\mathcal{D} = \mathcal{P}[0,1]$

We now consider only the model $\mathcal{P}[0,1]$ of all distributions over $[0,1]$.

Since we are considering distributions over $[0,1]$, the data-processing inequality for Kullback-Leibler divergences ensures (see, e.g., Garivier et al., 2019, Lemma 1) that for all $\zeta \in \mathcal{P}[0,1]$,

$$\mathrm{KL}(\zeta,\nu) \geqslant \mathrm{KL}\Big(\mathrm{Ber}\big(\mathrm{E}(\zeta)\big),\, \mathrm{Ber}\big(\mathrm{E}(\nu)\big)\Big) \geqslant 2\big(\mathrm{E}(\zeta) - \mathrm{E}(\nu)\big)^2 ,$$

where $\mathrm{Ber}(p)$ denotes the Bernoulli distribution with parameter p and where we applied Pinsker's inequality for Bernoulli distributions. Therefore, taking the infimum over distributions $\zeta \in \mathcal{P}[0,1]$ with $\mathrm{E}(\zeta) < x$,

$$\forall x \leqslant \mathrm{E}(\nu), \quad \mathcal{L}^{<}_{\inf}(x,\nu) \geqslant 2\big(\mathrm{E}(\nu) - x\big)^2 . \tag{18}$$

We denote by $m(\nu) = \min\big(\mathrm{Supp}(\nu)\big) \geqslant 0$ the minimum of the closed support $\mathrm{Supp}(\nu)$ of ν; that is, $m(\nu)$ is the largest value such that $\mathrm{Supp}(\nu) \subseteq \big[m(\nu),\, 1\big]$. We will refer to $m(\nu)$ as the lower end of the support of ν. Though we will not need it immediately, we also define the upper end of the support of ν as $M(\nu) = \max\big(\mathrm{Supp}(\nu)\big) \leqslant 1$; by symmetry, it will be considered when studying $\mathcal{L}^{>}_{\inf}$ and $\mathcal{L}^{\geqslant}_{\inf}$ instead of $\mathcal{L}^{<}_{\inf}$ and $\mathcal{L}^{\leqslant}_{\inf}$.

The lemma below states that the functions $\mathcal{L}^{<}_{\inf}(\,\cdot\,,\nu)$ and $\mathcal{L}^{\leqslant}_{\inf}(\,\cdot\,,\nu)$ coincide, except maybe at $m(\nu)$. One may wonder what happens at $x = m(\nu)$. We denote by $\nu\big\{m(\nu)\big\}$ the probability mass assigned by ν to the point $m(\nu)$. It follows from the second part the lemma below that $\mathcal{L}^{<}_{\inf}\big(m(\nu),\nu\big) = \mathcal{L}^{\leqslant}_{\inf}\big(m(\nu),\nu\big)$ if and only if $\big\{m(\nu)\big\}$ is not an atom of ν.

Lemma 15 *For the model $\mathcal{D} = \mathcal{P}[0,1]$, we have, on the one hand,*

$$\forall \mu \neq m(\nu), \qquad \mathcal{L}^{<}_{\inf}(\mu,\nu) = \mathcal{L}^{\leqslant}_{\inf}(\mu,\nu) , \tag{19}$$

and on the other hand, at $\mu = m(\nu)$,

$$\ln \frac{1}{\nu\big\{m(\nu)\big\}} = \mathcal{L}^{\leqslant}_{\inf}\big(m(\nu),\nu\big) \leqslant \mathcal{L}^{<}_{\inf}\big(m(\nu),\nu\big) = +\infty . \tag{20}$$

Proof To prove (19), we first identify the interior of the domain of $\mathcal{L}^{<}_{\inf}$.

Distributions ζ such that $\mathrm{E}(\zeta) < m(\nu)$ cannot be absolutely continuous with respect to ν; otherwise, they would also give a null probability to values strictly smaller than $m(\nu)$, which contradicts the assumption $\mathrm{E}(\zeta) < m(\nu)$. Hence $\mathrm{KL}(\zeta,\nu) = +\infty$ for these distributions. It follows that $\mathcal{L}^{<}_{\inf}(\mu,\nu) = \mathcal{L}^{\leqslant}_{\inf}(\mu,\nu) = +\infty$ for $\mu < m(\nu)$; we note in passing that we also have $\mathcal{L}^{<}_{\inf}\big(m(\nu),\nu\big) = +\infty$.

For $\mu > m(\nu)$, we take $\varepsilon > 0$ with $m(\nu) + \varepsilon < \mu$ and have, by definition of the support of a measure, that $\big[m(\nu), m(\nu)+\varepsilon\big]$ has a positive ν–measure denoted by κ. The distribution ζ given by ν conditioned to the interval $\big[m(\nu), m(\nu)+\varepsilon\big]$ is absolutely continuous with respect to ν, with density $\mathrm{d}\zeta/\mathrm{d}\nu = 1/\kappa$ on $\big[m(\nu), m(\nu)+\varepsilon\big]$, and 0 elsewhere; therefore, $\mathrm{KL}(\zeta,\nu) = \ln(1/\kappa) < +\infty$ and $\mathcal{L}^{<}_{\inf}(\mu,\nu) < +\infty$.

The interior of the domain of $\mu \mapsto \mathcal{L}^{<}_{\inf}(\mu,\nu)$ is therefore $\big(m(\nu),\, +\infty\big)$, and we recall that $\mathcal{L}^{<}_{\inf}(\,\cdot\,,\nu)$ is continuous on this interval. We fix some $\mu > m(\nu)$. For all $\varepsilon > 0$, by the very definitions of all quantities as infima of nested sets, we have

$$\mathcal{L}^{<}_{\inf}(\mu,\nu-\varepsilon) \leqslant \mathcal{L}^{\leqslant}_{\inf}(\mu,\nu) \leqslant \mathcal{L}^{<}_{\inf}(\mu,\nu) .$$

Letting $\varepsilon \to 0$, we get, by a sandwich argument, that $\mathcal{L}^{\leqslant}_{\inf}(\mu,\nu) = \mathcal{L}^{<}_{\inf}(\mu,\nu)$. This concludes the proof of (19).

We turn our attention to (20). We already showed above that $\mathcal{L}^{<}_{\inf}\big(m(\nu),\nu\big) = +\infty$. Now, to compute $\mathcal{L}^{\leqslant}_{\inf}(\mu,\nu)$, we wonder which are the distributions ζ that are absolutely continuous with respect to ν, and thus, give a null probability to values strictly smaller than $m(\nu)$, and are also such that $\mathrm{E}(\zeta) \leqslant m(\nu)$: at most one such distribution exists, the Dirac mass at $m(\nu)$, denoted by $\delta_{m(\nu)}$. We then distinguish the cases $\nu\big\{m(\nu)\big\} > 0$ and $\nu\big\{m(\nu)\big\} = 0$ to establish, respectively, the equalities

$$\mathcal{L}^{\leqslant}_{\inf}\big(m(\nu),\nu\big) = \mathrm{KL}(\delta_{m(\nu)},\nu) = \ln\frac{1}{\nu\big\{m(\nu)\big\}} \quad \text{and} \quad \mathcal{L}^{\leqslant}_{\inf}\big(m(\nu),\nu\big) = +\infty = \ln\frac{1}{\nu\big\{m(\nu)\big\}}\,.$$

In both cases, the first equality in (20) is proved, which concludes the proof. ■

We also have the following result, which is the most important and useful one, as it discussed the quantity that appears in the upper bounds on the average log-probability of misidentification of the optimal arm; see Corollary 3 together with Lemma 4.

Lemma 16 *Let $\nu,\nu' \in \mathcal{P}[0,1]$ with $\mu = \mathrm{E}(\nu) > \mathrm{E}(\nu') = \mu'$. Then*

$$\inf_{x\in[\mu',\mu]} \mathcal{L}^{\leqslant}_{\inf}(x,\nu) + \mathcal{L}^{\geqslant}_{\inf}(x,\nu') = \inf_{x\in[\mu',\mu]} \mathcal{L}^{<}_{\inf}(x,\nu) + \mathcal{L}^{>}_{\inf}(x,\nu')$$

if and only if either $m(\nu) \neq M(\nu')$ or $\nu\big\{m(\nu)\big\} \times \nu'\big\{M(\nu')\big\} = 0$.

Remark 17 *In other words, the only case for which the two infima differ is when $m(\nu) = M(\nu')$, i.e., the upper end of the support of ν' equals the lower end of the support of ν, and both ν and ν' admit this common value as an atom.*

Proof The first lines of the proof of Lemma 15 show that $\mathcal{L}^{\leqslant}_{\inf}(x,\nu) = \mathcal{L}^{<}_{\inf}(x,\nu) = +\infty$ for $x < m(\nu)$. We can symmetrically show that $\mathcal{L}^{\geqslant}_{\inf}(x,\nu') = \mathcal{L}^{>}_{\inf}(x,\nu') = +\infty$ for $x > M(\nu')$. Therefore, $\mathcal{L}^{\leqslant}_{\inf}(x,\nu) + \mathcal{L}^{\geqslant}_{\inf}(x,\nu')$ and $\mathcal{L}^{<}_{\inf}(x,\nu) + \mathcal{L}^{>}_{\inf}(x,\nu')$ are infinite whenever x lies outside of $\big[m(\nu), M(\nu')\big]$. This implies that

$$\inf_{x\in[\mu',\mu]} \mathcal{L}^{\leqslant}_{\inf}(x,\nu) + \mathcal{L}^{\geqslant}_{\inf}(x,\nu') = \inf_{x\in[\mu',\mu]\cap[m(\nu),M(\nu')]} \mathcal{L}^{\leqslant}_{\inf}(x,\nu) + \mathcal{L}^{\geqslant}_{\inf}(x,\nu')$$

and

$$\inf_{x\in[\mu',\mu]} \mathcal{L}^{<}_{\inf}(x,\nu) + \mathcal{L}^{>}_{\inf}(x,\nu') = \inf_{x\in[\mu',\mu]\cap[m(\nu),M(\nu')]} \mathcal{L}^{<}_{\inf}(x,\nu) + \mathcal{L}^{>}_{\inf}(x,\nu')\,.$$

We now split the analysis according to how large the interval $\mathcal{I}$ is, where

$$\mathcal{I} = [\mu',\mu] \cap \big[m(\nu), M(\nu')\big] = \Big[\max\big\{\mu', m(\nu)\big\},\, \min\big\{\mu, M(\nu')\big\}\Big]\,.$$

Case 1: $\mathcal{I}$ is empty. In that case, the two infima are over an empty set and both equal $+\infty$.

Case 2: $\mathcal{I}$ has a non-empty interior. When $a \neq b$, the infimum of a convex function over a closed interval $[a,b]$ equals the infimum over (a,b), whether the function takes finite or infinite values at a and b. Now, the interior of $\mathcal{I} = [a,b]$ equals

$$(a,b) = \Big(\max\big\{\mu', m(\nu)\big\},\, \min\big\{\mu, M(\nu')\big\}\Big) = (\mu',\mu) \cap \big(m(\nu), M(\nu')\big)$$

and does not contain neither $m(\nu)$ nor $M(\nu')$. By Lemma 15, the functions $\mathcal{L}^{<}_{\inf}(\,\cdot\,,\nu)$ and $\mathcal{L}^{\leqslant}_{\inf}(\,\cdot\,,\nu)$ coincide on $\mathbb{R}\backslash\{m(\nu)\}$. It may be similarly shown that $\mathcal{L}^{>}_{\inf}(\,\cdot\,,\nu')$ and $\mathcal{L}^{\geqslant}_{\inf}(\,\cdot\,,\nu')$ coincide on $\mathbb{R}\backslash\{M(\nu')\}$. In particular, the functions $\mathcal{L}^{\leqslant}_{\inf}(\,\cdot\,,\nu)+\mathcal{L}^{\geqslant}_{\inf}(\,\cdot\,,\nu')$ and $\mathcal{L}^{<}_{\inf}(\,\cdot\,,\nu)+\mathcal{L}^{>}_{\inf}(\,\cdot\,,\nu')$ coincide on the interior of $\mathcal{I}$. Their infima over the interior of $\mathcal{I}$, which, by convexity, are equal to the infima over $\mathcal{I}$, are therefore equal.

Case 3: $\mathcal{I}$ is a singleton. This case arises if and only if $m(\nu)=M(\nu')$, as by definition, $m(\nu)\leqslant\mu$ and $M(\nu')\geqslant\mu'$. We then have $\mathcal{I}=\{m(\nu)\}=\{M(\nu')\}$, and both infima are equal to the values of the sums at $m(\nu)=M(\nu')$. By Lemma 15 and by symmetric results for $\mathcal{L}^{>}_{\inf}$ and $\mathcal{L}^{\geqslant}_{\inf}$, on the one hand,

$$\mathcal{L}^{<}_{\inf}\big(m(\nu),\nu\big)=\mathcal{L}^{>}_{\inf}\big(M(\nu'),\nu'\big)=+\infty\,,$$

and on the other hand,

$$\mathcal{L}^{\leqslant}_{\inf}\big(m(\nu),\nu\big)+\mathcal{L}^{\geqslant}_{\inf}\big(M(\nu'),\nu'\big)=\ln\frac{1}{\nu\{m(\nu)\}}+\ln\frac{1}{\nu'\{M(\nu')\}}\,.$$

We get the desired equality if and only if either $\nu\{m(\nu)\}=0$ or $\nu\{M(\nu')\}=0$. ■

Appendix B. General analysis of successive-rejects in terms of Φ

This appendix is devoted to the technical elements omitted in the general analysis of the successive-rejects strategy presented in Section 3.1.

B.1. The Cramér-Chernoff bound

In this section, we recall, for the sake of completeness, the highly classical Cramér-Chernoff bound. With the notation introduced in Section 3, it states that, for an N–sample $X_1, \dots, X_N$, distributed according to ν and of average denoted by $\overline{X}_N$,

$$\forall x \leqslant \mathrm{E}(\nu), \qquad \mathbb{P}\big(\overline{X}_N \leqslant x\big) \leqslant \exp\big(-N\,\phi_\nu^\star(x)\big)\,, \tag{21}$$

$$\text{and} \qquad \forall x \geqslant \mathrm{E}(\nu), \qquad \mathbb{P}\big(\overline{X}_N \geqslant x\big) \leqslant \exp\big(-N\,\phi_\nu^\star(x)\big)\,. \tag{22}$$

Proof For all $\lambda < 0$, by Markov's inequality first and then by independence,

$$\begin{aligned}\mathbb{P}\big(\overline{X}_N \leqslant x\big) = \mathbb{P}\Big(\mathrm{e}^{\lambda \overline{X}_N} \geqslant \mathrm{e}^{\lambda x}\Big) &\leqslant \mathrm{e}^{-\lambda x}\,\mathbb{E}\Big[\mathrm{e}^{\lambda \overline{X}_N}\Big] = \mathrm{e}^{-\lambda x}\,\Big(\mathbb{E}\big[\mathrm{e}^{\lambda X_1/N}\big]\Big)^N \\ &= \exp\big(-\lambda x + N\,\phi_\nu(\lambda/N)\big) = \exp\Big(-N\big(\lambda' x - \phi_\nu(\lambda')\big)\Big)\,,\end{aligned}$$

where $\lambda' = \lambda/N$. The bound also holds for $\lambda = \lambda' = 0$ given that $\phi_\nu(0) = 0$. Optimizing over $\lambda \leqslant 0$ (or, equivalently, over $\lambda' \leqslant 0$), we proved so far

$$\mathbb{P}\big(\overline{X}_N \leqslant x\big) \leqslant \exp\Big(-N \sup_{\lambda \leqslant 0}\big\{\lambda x - \phi_\nu(\lambda)\big\}\Big)\,.$$

Now, by Jensen's inequality,

$$\forall \lambda \in \mathbb{R}, \qquad \phi_\nu(\lambda) = \ln \mathbb{E}\big[\mathrm{e}^{\lambda X}\big] \geqslant \lambda\,\mathbb{E}[X] = \lambda\,\mathrm{E}(\nu)\,; \tag{23}$$

therefore, for $x \leqslant \mathrm{E}(\nu)$,

$$\forall \lambda \geqslant 0, \qquad \lambda x - \phi_\nu(\lambda) \leqslant \lambda\big(x - \mathrm{E}(\nu)\big) \leqslant 0\,.$$

In particular,

$$0 = -\phi_\nu(0) \leqslant \sup_{\lambda \leqslant 0}\big\{\lambda x - \phi_\nu(\lambda)\big\} = \sup_{\lambda \in \mathbb{R}}\big\{\lambda x - \phi_\nu(\lambda)\big\} \stackrel{\text{def}}{=} \phi_\nu^\star(x)\,. \tag{24}$$

This concludes the proof of (21). The bound (22) follows by symmetry. ■

We also note, in passing, that Jensen's inequality entails, for $x = \mathrm{E}(\nu)$, that

$$\forall \lambda \in \mathbb{R}, \qquad \lambda \mathrm{E}(\nu) - \phi_\nu(\lambda) \leqslant \lambda\big(\mathrm{E}(\nu) - \mathrm{E}(\nu)\big) = 0\,,$$

thus showing that $\phi_\nu^\star\big(\mathrm{E}(\nu)\big) = 0$. The property (24) and its counterpart for $x \geqslant \mathrm{E}(\nu)$ and $\lambda \geqslant 0$ actually show that $\phi_\nu^\star$ is non-increasing on $\big(-\infty, \mathrm{E}(\nu)\big]$ and non-decreasing on $\big[\mathrm{E}(\nu), +\infty\big)$.

B.2. Proof of Lemma 1

We first restate the lemma, for the convenience of the reader.

Lemma 1 *Fix ν and ν' in $\mathcal{D}$, with respective expectations $\mu = \mathrm{E}(\nu) > \mu' = \mathrm{E}(\nu')$. For all $N \geqslant 1$, let $\overline{X}_N$ and $\overline{Y}_N$ be the averages of N–samples with respective distributions ν and ν'. Then,*

$$\limsup_{N\to+\infty} \frac{1}{N} \ln \mathbb{P}\big(\overline{X}_N \leqslant \overline{Y}_N\big) \leqslant - \inf_{x\in[\mu',\mu]} \big\{\phi^\star_{\nu'}(x) + \phi^\star_\nu(x)\big\} \stackrel{\text{def}}{=} -\Phi(\nu',\nu)\,.$$

Proof The proof consists in two parts. We first show that for any finite grid $\mathcal{G} = \{g_2, \ldots, g_{G-1}\}$ in (μ', μ), to which we add the points $g_1 = \mu'$ and $g_G = \mu$, we have

$$\limsup_{N\to+\infty} \frac{1}{N} \ln \mathbb{P}\big(\overline{X}_N \leqslant \overline{Y}_N\big) \leqslant - \min\Big\{\phi^\star_\nu(\mu'),\ \min_{2\leqslant j\leqslant G-1} \big\{\phi^\star_{\nu'}(g_{j-1}) + \phi^\star_\nu(g_j)\big\},\ \phi^\star_{\nu'}(\mu)\Big\}\,. \quad (25)$$

Indeed, by identifying, when $\overline{X}_N$ and $\overline{Y}_N$ belong to $[\mu', \mu]$, in which interval $[g_{j-1}, g_j]$ lies $\overline{X}_N$, we note that

$$\big\{\overline{X}_N \leqslant \overline{Y}_N\big\} \quad \subseteq \quad \big\{\overline{X}_N \leqslant \mu'\big\} \,\cup\, \big\{\overline{Y}_N \geqslant \mu\big\} \,\cup\, \bigcup_{j=2}^{G-1} \big\{\overline{Y}_N \geqslant g_{j-1} \text{ and } \overline{X}_N \leqslant g_j\big\}\,.$$

First, by independence and by the Cramér-Chernoff inequalities (21) and (22),

$$\mathbb{P}\big(\overline{Y}_N \geqslant g_{j-1} \text{ and } \overline{X}_N \leqslant g_j\big) = \mathbb{P}\big(\overline{Y}_N \geqslant g_{j-1}\big)\ \mathbb{P}\big(\overline{X}_N \leqslant g_j\big) \leqslant \exp\Big(-N\big(\phi^\star_{\nu'}(g_{j-1}) + \phi^\star_\nu(g_j)\big)\Big)\,.$$

Second, again by the Cramér-Chernoff inequalities,

$$\mathbb{P}\big(\overline{X}_N \leqslant \mu'\big) \leqslant \exp\big(-N\,\phi^\star_\nu(\mu')\big) \qquad \text{and} \qquad \mathbb{P}\big(\overline{Y}_N \geqslant \mu\big) \leqslant \exp\big(-N\,\phi^\star_{\nu'}(\mu)\big)\,.$$

By a union bound,

$$\mathbb{P}\big(\overline{X}_N \leqslant \overline{Y}_N\big) \leqslant \exp\big(-N\,\phi^\star_\nu(\mu')\big) + \exp\big(-N\,\phi^\star_{\nu'}(\mu)\big) + \sum_{j=2}^{G-1} \exp\Big(-N\big(\phi^\star_{\nu'}(g_{j-1}) + \phi^\star_\nu(g_j)\big)\Big)\,.$$

The stated bound (25) follows by identifying the (finitely many) terms with the smallest rate in the exponent.

In the second part of the proof, we note that the bound (25) holds for any finite grid in (μ', μ), and we consider a sequence

$$\mathcal{G}^{(n)} = \Big\{g_2^{(n)}, \ldots, g_{G_n-1}^{(n)}\Big\}$$

of such finite grids. In particular,

$$\limsup_{N\to+\infty} \frac{1}{N} \ln \mathbb{P}\big(\overline{X}_N \leqslant \overline{Y}_N\big) \leqslant - \min\Big\{\phi^\star_\nu(\mu'),\ \max_{n\geqslant 1} S_n,\ \phi^\star_{\nu'}(\mu)\Big\}\,,$$

$$\text{where} \qquad S_n \stackrel{\text{def}}{=} \min_{2\leqslant j\leqslant G_n-1} \Big\{\phi^\star_{\nu'}\big(g_{j-1}^{(n)}\big) + \phi^\star_\nu\big(g_j^{(n)}\big)\Big\}\,.$$

To obtain the claimed bound, given that (see the end of Appendix B.1)

$$\phi^\star_\nu(\mu) = 0 = \phi^\star_{\nu'}(\mu')\,,$$

it suffices to show that

$$\max_{n\geqslant 1} S_n \geqslant \inf_{x\in[\mu',\mu]} \left\{\phi^\star_{\nu'}(x) + \phi^\star_{\nu}(x)\right\}.$$

To that end, we assume that the steps ε_n of the grids $\mathcal{G}^{(n)}$, which are defined as

$$\varepsilon_n \stackrel{\text{def}}{=} \max_{2\leqslant j\leqslant G_n} \left| g^{(n)}_j - g^{(n)}_{j-1} \right|,$$

vanish asymptotically, i.e., $\varepsilon_n \to 0$. For each grid $\mathcal{G}^{(n)}$, we denote by $x^\star_n \in (\mu',\mu)$ the argument of the minimum in the definition of S_n. As a consequence, for each $n \geqslant 1$,

$$S_n = \phi^\star_{\nu'}(x^\star_n - \varepsilon^\star_n) + \phi^\star_{\nu}(x^\star_n),$$

for some $0 < \varepsilon^\star_n \leqslant \varepsilon_n$. The quantity $x^\star_n - \varepsilon^\star_n$ denotes the point in the grid that is right before $x^\star_n$, and it belongs to $[\mu',\mu)$. We note that we also have $\varepsilon^\star_n \to 0$. In the compact interval $[\mu',\mu]$, the Bolzano-Weierstrass property ensures the existence of a converging subsequence: there exists $x^\star_\infty \in [\mu',\mu]$ and a sequence $(n_k)_{k\geqslant 1}$ of integers such that

$$x^\star_{n_k} \underset{k\to+\infty}{\longrightarrow} x^\star_\infty, \qquad \text{which also entails} \qquad x^\star_{n_k} - \varepsilon^\star_{n_k} \underset{k\to+\infty}{\longrightarrow} x^\star_\infty.$$

Now, the functions $\phi^\star_\nu$, respectively, $\phi^\star_{\nu'}$, are lower semi-continuous, as the suprema over $\lambda \in \mathbb{R}$ of the continuous functions $x \mapsto \lambda x - \varphi_\nu(\lambda)$, respectively, $x \mapsto \lambda x - \varphi_{\nu'}(\lambda)$. Therefore, by these lower semi-continuities,

$$\begin{aligned}\max_{n\geqslant 1} S_n \geqslant \liminf_{k\to+\infty} \phi^\star_{\nu'}(x^\star_{n_k} - \varepsilon^\star_{n_k}) + \phi^\star_{\nu}(x^\star_{n_k}) &\geqslant \phi^\star_{\nu'}(x^\star_\infty) + \phi^\star_{\nu}(x^\star_\infty) \\ &\geqslant \inf_{x\in[\mu',\mu]} \left\{\phi^\star_{\nu'}(x) + \phi^\star_{\nu}(x)\right\}.\end{aligned}$$

This concludes the proof. ■

B.3. Proof of Theorem 2

The proof mimics the analysis by Audibert et al. (2010), the main modification being the substitution of Hoeffding's inequality by the bound of Lemma 1.

Proof We recall that for $r \in \{1,\dots,K-1\}$, we denoted by $N_r = \lfloor \ell_1/K \rfloor + \ldots + \lfloor \ell_r/(K-r+1) \rfloor$ the total number of times an arm still considered in regime r, i.e., belonging to S_{r-1}, was pulled in regimes 1 to r. For each arm a, we denote by $\overline{Y}^r_a$ the average of a N_r–sample distributed according to ν_a. By optional skipping (see Doob, 1953, Chapter III, Theorem 5.2, p. 145, or Chow and Teicher, 1988, Section 5.3 for a more recent reference), we may assume, with no loss of generality, that for each $r \in \{1,\dots,K-1\}$,

$$\text{on the event } \{a \in S_{r-1}\}, \qquad \overline{X}^r_a = \overline{Y}^r_a. \tag{26}$$

We fix a bandit problem $\underline{\nu}$ with a unique optimal arm $a^\star(\underline{\nu})$. The successive-rejects strategy fails if (and only) if it rejects $a^\star(\underline{\nu})$ in ones of the regimes. This corresponds to the event

$$\left\{I_T \neq a^\star(\underline{\nu})\right\} = \bigcup_{r=1}^{K-1} \left\{a_r = a^\star(\underline{\nu})\right\} \subseteq \bigcup_{r=1}^{K-1} \left\{a^\star(\underline{\nu}) \in S_{r-1} \text{ and } \forall k \in S_{r-1},\ \overline{X}^r_{a^\star(\underline{\nu})} \leqslant \overline{X}^r_k\right\}.$$

(We have an inclusion because ties are broken arbitrarily.) By optional skipping (26),

$$\bigcup_{r=1}^{K-1}\Big\{a^\star(\underline{\nu})\in S_{r-1}\ \text{ and }\ \forall k\in S_{r-1},\ \ \overline{X}^r_{a^\star(\underline{\nu})}\leqslant\overline{X}^r_k\Big\}$$
$$=\bigcup_{r=1}^{K-1}\Big\{a^\star(\underline{\nu})\in S_{r-1}\ \text{ and }\ \forall k\in S_{r-1},\ \ \overline{Y}^r_{a^\star(\underline{\nu})}\leqslant\overline{Y}^r_k\Big\}.$$

Recall that the set S_{r-1} is a random set; dealing with it therefore requires some care. On the event of interest, S_{r-1} contains $K-r+1$ elements, among which $a^\star(\underline{\nu})$. The set $\mathcal{A}_r$ is of cardinality r and does not contain $a^\star(\underline{\nu})$. By the pigeonhole principle, S_{r-1} thus necessarily contains one arm in $\mathcal{A}_r$. As a consequence, for each regime $r\in\{1,\dots,K-1\}$,

$$\Big\{a^\star(\underline{\nu})\in S_{r-1}\ \text{ and }\ \forall k\in S_{r-1},\ \ \overline{Y}^r_{a^\star(\underline{\nu})}\leqslant\overline{Y}^r_k\Big\}\subseteq\bigcup_{k\in\mathcal{A}_r}\Big\{\overline{Y}^r_{a^\star(\underline{\nu})}\leqslant\overline{Y}^r_k\Big\}.$$

Summarizing the inclusions above, taking unions bounds, and upper bounding the obtained sum in a crude way, we proved so far

$$\mathbb{P}\big(I_T\neq a^\star(\underline{\nu})\big)\leqslant\sum_{r=1}^{K-1}\sum_{k\in\mathcal{A}_r}\mathbb{P}\Big(\overline{Y}^r_{a^\star(\underline{\nu})}\leqslant\overline{Y}^r_k\Big)\leqslant K^2\max_{1\leqslant r\leqslant K-1}\max_{k\in\mathcal{A}_r}\mathbb{P}\Big(\overline{Y}^r_{a^\star(\underline{\nu})}\leqslant\overline{Y}^r_k\Big),$$

or equivalently,

$$\frac{1}{T}\ln\mathbb{P}\big(I_T\neq a^\star(\underline{\nu})\big)\leqslant\frac{2}{T}\ln K+\max_{1\leqslant r\leqslant K-1}\max_{k\in\mathcal{A}_r}\frac{1}{T}\ln\mathbb{P}\Big(\overline{Y}^r_{a^\star(\underline{\nu})}\leqslant\overline{Y}^r_k\Big)$$
$$=\frac{2}{T}\ln K+\max_{1\leqslant r\leqslant K-1}\max_{k\in\mathcal{A}_r}\frac{N_r}{T}\frac{1}{N_r}\ln\mathbb{P}\Big(\overline{Y}^r_{a^\star(\underline{\nu})}\leqslant\overline{Y}^r_k\Big).$$

As $N_r/T\to\gamma_r>0$ as $T\to+\infty$, we may apply Lemma 1, together with an exchange between the $\limsup$ and the maximum over a finite number of quantities. We obtain

$$\limsup_{T\to+\infty}\frac{1}{T}\ln\mathbb{P}\big(I_T\neq a^\star(\underline{\nu})\big)\leqslant\max_{1\leqslant r\leqslant K-1}\max_{k\in\mathcal{A}_r}\Big\{\gamma_r\Big(-\Phi\big(\nu_k,\nu^\star\big)\Big)\Big\}$$
$$=-\min_{1\leqslant r\leqslant K-1}\Big\{\gamma_r\min_{k\in\mathcal{A}_r}\Phi\big(\nu_k,\nu^\star\big)\Big\}.$$

This concludes the proof. ■

B.4. Proof of Corollary 3 and of the bound (13) on Φ

In this final subsection, we provide two series of proofs: first, a proof of Corollary 3; and then a proof of the bound $\Phi\big(\nu_k,\nu^\star\big)\geqslant\Delta_k^2$ stated as (13).

Proof of Corollary 3. To apply Theorem 2, we need only to show that the regime lengths of (12) are such that N_r/T converges to a positive value, and to identify this limit value γ_r. As $N_1=$

$\lfloor \ell_1/K \rfloor$, where $\ell_1 = T/\overline{\ln} K$, we immediately have $N_1/T \to \gamma_1 = 1/\big(K \overline{\ln} K\big) > 0$. For $r \in \{2, \dots, K-1\}$,

$$\frac{N_r}{T} = \sum_{p=1}^{r} \frac{1}{T} \left\lfloor \frac{\ell_p}{K} \right\rfloor = \frac{1}{T} \left(\left\lfloor \frac{T}{K \overline{\ln} K} \right\rfloor + \sum_{p=2}^{r} \left\lfloor \frac{T}{(K-p+1)(K-p+2) \overline{\ln} K} \right\rfloor \right)$$
$$\xrightarrow[T \to +\infty]{} \gamma_r \overset{\text{def}}{=} \frac{1}{\overline{\ln} K} \left(\frac{1}{K} + \sum_{p=2}^{r} \frac{1}{K-p+1} - \frac{1}{K-p+2} \right) = \frac{1}{(K-r+1) \overline{\ln} K} \,.$$

The bound of Theorem 2 reads:

$$\limsup_{T \to +\infty} \frac{1}{T} \ln \mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant -\frac{1}{\overline{\ln} K} \min_{1 \leqslant r \leqslant K-1} \left\{ \frac{1}{K-r+1} \min_{k \in \mathcal{A}_r} \Phi\big(\nu_k, \nu^\star\big) \right\}.$$

It implies, in terms of lower bounds $f(\nu_k, \nu^\star) \leqslant \Phi\big(\nu_k, \nu^\star\big)$,

$$\limsup_{T \to +\infty} \frac{1}{T} \ln \mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant -\frac{1}{\overline{\ln} K} \min_{1 \leqslant r \leqslant K-1} \left\{ \frac{1}{K-r+1} \min_{k \in \mathcal{A}_r} f\big(\nu_k, \nu^\star\big) \right\}. \tag{27}$$

The permutation σ in (14) and the sets $\mathcal{A}_r = \{\sigma_{K-r+1}, \dots, \sigma_K\}$ were exactly picked, for each $r \in \{1, \dots, K-1\}$, to minimize

$$\min_{k \in \mathcal{B}_r} f\big(\nu_k, \nu^\star\big)$$

over sets $\mathcal{B}_r$ abiding by the indicated constraints: being of cardinal r and not containing the optimal arm $a^\star(\underline{\nu}) = \sigma_1$. We get

$$\min_{k \in \mathcal{A}_r} f\big(\nu_k, \nu^\star\big) = \min_{K-r+1 \leqslant k \leqslant K} f\big(\nu_{\sigma_k}, \nu^\star\big) = f\big(\nu_{\sigma_{K-r+1}}, \nu^\star\big) \,,$$

which, together with (27), yields the stated bound, up to replacing $K-r+1$ with $r \in \{1, \dots, K-1\}$ by $k \in \{2, \dots, K\}$:

$$-\frac{1}{\overline{\ln} K} \min_{1 \leqslant r \leqslant K-1} \left\{ \frac{1}{K-r+1} f\big(\nu_{\sigma_{K-r+1}}, \nu^\star\big) \right\} = -\frac{1}{\overline{\ln} K} \min_{2 \leqslant k \leqslant K} \left\{ \frac{1}{k} f\big(\nu_{\sigma_k}, \nu^\star\big) \right\}. \qquad \blacksquare$$

We now move to the proof of the bound (13) on Φ, when the model is $\mathcal{D} = \mathcal{P}[0,1]$. We restate it here for the convenience of the reader:

$$\phi_\nu^\star(x) \geqslant 2\big(x - \mathrm{E}(\nu)\big)^2 \,, \qquad \text{so that} \qquad \Phi\big(\nu_k, \nu^\star\big) \geqslant \Delta_k^2 \overset{\text{def}}{=} f\big(\nu_k, \nu^\star\big) \,.$$

For the ease of exposition, the path followed in Section 2 to show that $\Phi\big(\nu_k, \nu^\star\big) \geqslant \Delta_k^2$ was to first note that $\Phi = \mathcal{L}$ when $\mathcal{D} = \mathcal{P}[0,1]$ (see Lemma 4) and then use Pinsker's inequality (5). We provide here a slightly more direct but equivalent approach, based on Hoeffding's inequality.

Proof of the bound (13) on Φ. When $\nu \in \mathcal{P}[0,1]$, Hoeffding's inequality exactly states that

$$\forall \lambda \in \mathbb{R}, \quad \phi_\nu(\lambda) \leqslant \lambda \mathrm{E}(\nu) + \frac{\lambda^2}{8} \,,$$
$$\text{so that} \qquad \forall x \in \mathbb{R}, \quad \phi_\nu^\star(x) \geqslant \sup_{\lambda \in \mathbb{R}} \left\{ \lambda\big(x - \mathrm{E}(\nu)\big) - \frac{\lambda^2}{8} \right\} = 2\big(x - \mathrm{E}(\nu)\big)^2 \,.$$

This corresponds to the first part of (13).

For its second part, we consider a pair ν, ν' of distributions in $\mathcal{P}[0,1]$, we set any $x \in [\mathrm{E}(\nu'), \mathrm{E}(\nu)]$, and we apply twice the bound of the first part to get

$$\phi^\star_{\nu'}(x) + \phi^\star_\nu(x) \geqslant 2\big(x - \mathrm{E}(\nu')\big)^2 + 2\big(x - \mathrm{E}(\nu)\big)^2 .$$

From the definition of Φ, it follows that

$$\Phi(\nu', \nu) \geqslant \inf_{x \in [\mathrm{E}(\nu'), \mathrm{E}(\nu)]} \Big\{ 2\big(x - \mathrm{E}(\nu')\big)^2 + 2\big(x - \mathrm{E}(\nu)\big)^2 \Big\} = \big(\mathrm{E}(\nu') - \mathrm{E}(\nu)\big)^2 .$$

This corresponds to the second part of (13). ■

Appendix C. Rewriting of $\phi_\nu^\star$ as some $\mathcal{L}_{\inf}^{\leqslant}$ or $\mathcal{L}_{\inf}^{\geqslant}$, i.e., of Φ as $\mathcal{L}$

We use the notation of Sections 2.1 and 3 and discuss conditions on models guaranteeing that $\Phi = \mathcal{L}$, i.e., that (15) holds. We do so for $\mathcal{D} = \mathcal{P}[0,1]$ in Section C.2 and for canonical one-parameter exponential families in Section C.3; based on these two examples, we provide a set of conditions for general models, in Section C.4. A building block of these results is that for all models $\mathcal{D}$, the functions $\mathcal{L}_{\inf}^{\leqslant}(\,\cdot\,,\nu)$ and $\mathcal{L}_{\inf}^{\geqslant}(\,\cdot\,,\nu)$ dominate $\phi_\nu^\star$ defined in (11); we prove this in Section C.1.

This rewriting of $\mathcal{L}_{\inf}^{\leqslant}(\,\cdot\,,\nu)$ or $\mathcal{L}_{\inf}^{\geqslant}(\,\cdot\,,\nu)$ as $\phi_\nu^\star$ claimed, e.g., by Lemma 4, can be seen as a counterpart to a similar rewriting of the $\mathcal{K}_{\inf}$ as the supremum of a function of $\lambda \in [0,1]$. More precisely, for $\nu \in \mathcal{P}[0,1]$ and $x \in [0,1]$,

$$\mathcal{K}_{\inf}(\nu, x) = \inf\big\{\mathrm{KL}(\nu,\zeta) :\ \zeta \in \mathcal{P}[0,1] \text{ s.t. } \mathrm{E}(\zeta) < x\big\}\,,$$

and Honda and Takemura (2015, Theorem 2)—see also Garivier et al., 2022, Lemma 18—show that

$$\mathcal{K}_{\inf}(\nu, x) = \sup_{0\leqslant\lambda\leqslant 1} \mathbb{E}\left[\ln\left(1 - \lambda\frac{Y-x}{1-x}\right)\right],$$

where Y is a random variable distributed according to ν. In both cases, for $\mathcal{L}_{\inf}^{\leqslant}(\,\cdot\,,\nu)$ or $\mathcal{L}_{\inf}^{\geqslant}(\,\cdot\,,\nu)$, and for $\mathcal{K}_{\inf}$, being able to rewrite the infimum of a set of Kullback-Leibler divergences as a supremum is not unexpected, as a Kullback-Leibler divergence can be formulated as a supremum, see (28), and exchanges between the infimum and the supremum may be justified (e.g., through Sion's lemma, when applicable).

C.1. $\mathcal{L}_{\inf}^{\leqslant}(\,\cdot\,,\nu)$ and $\mathcal{L}_{\inf}^{\geqslant}(\,\cdot\,,\nu)$ dominate $\phi_\nu^\star$

This domination is a consequence of a variational formula for the Kullback-Leibler divergence (28).

Lemma 18 *For all models $\mathcal{D}$, for all $\nu \in \mathcal{D}$,*

$$\forall x \leqslant \mathrm{E}(\nu),\quad \phi_\nu^\star(x) \leqslant \mathcal{L}_{\inf}^{\leqslant}(x,\nu) \qquad \textit{and} \qquad \forall x \geqslant \mathrm{E}(\nu),\quad \phi_\nu^\star(x) \leqslant \mathcal{L}_{\inf}^{\geqslant}(x,\nu)\,.$$

In the next sections we will obtain the converse inequalities for the specific models mentioned above. The proofs are immediate adaptations of a rather standard result, stated, among others, but in a slightly different form (and for the model $\mathcal{D}$ of all real-valued distributions with a first moment), by Boucheron et al. (2013, Exercise 4.13).

Proof We rely on a key variational formula for the Kullback-Leibler divergence, see Boucheron et al. (2013, Chapter 4): for all distributions ν, ν' over $\mathbb{R}$,

$$\begin{aligned}\mathrm{KL}(\nu',\nu) &= \sup\Big\{\mathbb{E}_{\nu'}[Y] - \ln \mathbb{E}_\nu\big[\mathrm{e}^Y\big] :\ \text{r.v. } Y \in \mathbb{L}^1(\nu') \text{ s.t. } \mathbb{E}_\nu\big[\mathrm{e}^Y\big] < +\infty\Big\}\,,\\ &= \sup\Big\{\mathbb{E}_{\nu'}[Y] - \ln \mathbb{E}_\nu\big[\mathrm{e}^Y\big] :\ \text{r.v. } Y \in \mathbb{L}^1(\nu')\Big\}\,, \qquad (28)\end{aligned}$$

where $\mathbb{E}_\nu$ and $\mathbb{E}_{\nu'}$ indicate that expectations are relative to ν and ν', respectively. In particular, when ν and ν' lie in $\mathcal{D}$, those two distributions admit a finite first moment, hence all random variables

$Y = \lambda\,\mathrm{id}_{\mathbb{R}}$ are ν'–integrable, where $\mathrm{id}_{\mathbb{R}}$ denotes the identity function on $\mathbb{R}$ and where $\lambda \in \mathbb{R}$. They satisfy $\mathbb{E}_{\nu'}[Y] = \lambda\,\mathrm{E}(\nu')$. A consequence of (28) is therefore

$$\mathrm{KL}(\nu', \nu) \geqslant \sup_{\lambda \in \mathbb{R}} \Big\{ \lambda\,\mathrm{E}(\nu') - \ln \mathbb{E}_{\nu}\big[\mathrm{e}^{\lambda \mathrm{id}_{\mathbb{R}}}\big] \Big\} = \phi_{\nu}^{\star}\big(\mathrm{E}(\nu')\big)\,. \tag{29}$$

Using the variations of $\phi_{\nu}^{\star}$ indicated at the end of Appendix B.1, we see that

$$\phi_{\nu}^{\star}\big(\mathrm{E}(\nu')\big) \geqslant \phi_{\nu}^{\star}(x) \quad \text{when} \quad \mathrm{E}(\nu') \leqslant x \leqslant \mathrm{E}(\nu) \quad \text{or} \quad \mathrm{E}(\nu') \geqslant x \geqslant \mathrm{E}(\nu)\,.$$

Therefore, by taking the infimum in (29) over $\nu' \in \mathcal{D}$ either with $\mathrm{E}(\nu') \leqslant x$ or $\mathrm{E}(\nu') \geqslant x$, we proved the claimed inequalities. ■

C.2. The case of $\mathcal{P}[0,1]$

In this section, we focus on the model $\mathcal{P}[0,1]$ and prove that the inequalities of Lemma 18 are in fact equalities, as claimed by Lemma 4.

Lemma 4 *Consider the model $\mathcal{D} = \mathcal{P}[0,1]$. For all $\nu \in \mathcal{P}[0,1]$,*

$$\forall x \leqslant \mathrm{E}(\nu), \quad \phi_{\nu}^{\star}(x) = \mathcal{L}_{\mathrm{inf}}^{\leqslant}(x, \nu) \qquad \textit{and} \qquad \forall x \geqslant \mathrm{E}(\nu), \quad \phi_{\nu}^{\star}(x) = \mathcal{L}_{\mathrm{inf}}^{\geqslant}(x, \nu)\,.$$

The lemma holds for all $x \in \mathbb{R}$, that is, even outside of the $[0,1]$ interval, though the proof reveals that when x is smaller than the lower end $m(\nu)$ of the support of ν, we actually have $\phi_{\nu}^{\star}(x) = \mathcal{L}_{\mathrm{inf}}^{\leqslant}(x, \nu) = +\infty$. The counterpart statement $\phi_{\nu}^{\star}(x) = \mathcal{L}_{\mathrm{inf}}^{\geqslant}(x, \nu) = +\infty$ holds for x larger than the upper end $M(\nu)$ of the support of ν. The pieces of notation $m(\nu)$ and $M(\nu)$ were formally defined in Appendix A.2.

Proof Note first that by Lemma 18, it suffices to prove that

$$\forall x \leqslant \mathrm{E}(\nu), \quad \phi_{\nu}^{\star}(x) \geqslant \mathcal{L}_{\mathrm{inf}}^{\leqslant}(x, \nu) \qquad \text{and} \qquad \forall x \geqslant \mathrm{E}(\nu), \quad \phi_{\nu}^{\star}(x) \geqslant \mathcal{L}_{\mathrm{inf}}^{\geqslant}(x, \nu)\,.$$

We only deal with the first inequality, namely $\mathcal{L}_{\mathrm{inf}}^{\leqslant}(x, \nu) \leqslant \phi_{\nu}^{\star}(x)$ for $x \leqslant \mathrm{E}(\nu)$, and obtain the other one by symmetry.

In the case $x = \mathrm{E}(\nu)$, we have $\phi_{\nu}^{\star}\big(\mathrm{E}(\nu)\big) = 0$, as stated at the end of Section B.1, and $\mathcal{L}_{\mathrm{inf}}^{\leqslant}\big(\mathrm{E}(\nu), \nu\big) = 0$ as can be seen by taking $\zeta = \nu$ in the defining infimum. We therefore only consider $x < \mathrm{E}(\nu)$ in the sequel. We will rely on the standard fact that, by Hölder's inequality, the logarithmic moment-generating function

$$\phi_{\nu} : \lambda \in \mathbb{R} \longmapsto \ln \mathbb{E}_{\nu}\big[\mathrm{e}^{\lambda \mathrm{id}_{[0,1]}}\big]\,,$$

is convex, where $\mathrm{id}_{[0,1]}$ denotes the identity function on $[0,1]$. Also, by two applications of a standard theorem of differentiation under the integral, given that ν is supported by $[0,1]$, we have that ϕ_{ν} is continuously differentiable over $\mathbb{R}$, with derivative

$$\phi_{\nu}' : \lambda \in \mathbb{R} \longmapsto \frac{\mathbb{E}_{\nu}\big[\mathrm{id}_{[0,1]}\,\mathrm{e}^{\lambda \mathrm{id}_{[0,1]}}\big]}{\mathbb{E}_{\nu}\big[\mathrm{e}^{\lambda \mathrm{id}_{[0,1]}}\big]}\,.$$

By convexity of ϕ_ν, this derivative is non-decreasing. Therefore, the limit of ϕ'_ν at $-\infty$ exists; we denote it by ℓ and have that a priori $\ell \in \{-\infty\} \cup \mathbb{R}$. We now prove that actually,

$$\lim_{\lambda \to -\infty} \phi'_\nu(\lambda) \stackrel{\text{def}}{=} \ell = m(\nu)\,. \tag{30}$$

On the one hand, by definition of $m(\nu)$, we have $\mathrm{id}_{[0,1]} \geqslant m(\nu)$ ν-a.s., which entails $\phi'_\nu(\lambda) \geqslant m(\nu)$ for all $\lambda \in \mathbb{R}$, and hence, $\ell \geqslant m(\nu)$. On the other hand, as ϕ'_ν is non-decreasing, it is always larger than its limit ℓ at $-\infty$:

$$\forall \lambda \in \mathbb{R}, \qquad \phi'_\nu(\lambda) \geqslant \ell\,, \qquad \text{thus,} \qquad \mathbb{E}_\nu\Big[\big(\mathrm{id}_{[0,1]} - \ell\big)\,\mathrm{e}^{\lambda \mathrm{id}_{[0,1]}}\Big] \geqslant 0\,, \tag{31}$$

$$\text{or} \qquad \mathbb{E}_\nu\Big[\big(\mathrm{id}_{[0,1]} - \ell\big)\,\mathrm{e}^{\lambda(\mathrm{id}_{[0,1]} - \ell)}\Big] \geqslant 0\,. \tag{32}$$

The last inequality and limit arguments as $\lambda \to -\infty$ impose that $\mathrm{id}_{[0,1]} - \ell \geqslant 0$ ν-a.s., which in turn entails that $\ell \leqslant m(\nu)$. This concludes the proof of (30).

The various properties exhibited above for ϕ_ν, including the fact that the derivative ϕ'_ν takes values in $\big[m(\nu), +\infty\big)$, entail that the function

$$\Lambda : \lambda \in \mathbb{R} \longmapsto \lambda x - \phi_\nu(\lambda)$$

is concave, continuously differentiable, with a non-increasing derivative Λ' taking values in the interval $\big(-\infty, x - m(\nu)\big]$ and with limit $x - m(\nu)$ at $-\infty$.

We split the analysis of the case $x < \mathrm{E}(\nu)$ into three sub-cases, depending on the respective positions of x and $m(\nu)$, and recall that we want to show that $\mathcal{L}^{\leqslant}_{\inf}(x, \nu) \leqslant \phi^\star_\nu(x)$.

Case 1: $x > m(\nu)$. By Jensen's inequality (23) and given that we consider $x < \mathrm{E}(\nu)$, the limit of Λ at $+\infty$ equals $-\infty$. The limit of Λ at $-\infty$ also equals $-\infty$, as the derivative Λ' has limit $x - m(\nu) > 0$ at $-\infty$. By concavity of Λ and the fact that Λ' is continuous, this implies the existence of some $\lambda^\star \in \mathbb{R}$ such that

$$\Lambda'(\lambda^\star) = x - \phi'_\nu(\lambda^\star) = 0 \qquad \text{and} \qquad \phi^\star_\nu(x) = \sup_{\lambda \in \mathbb{R}} \big\{\Lambda(\lambda)\big\} = \Lambda(\lambda^\star)\,.$$

Denoting by $\zeta_{\lambda^\star}$ the distribution absolutely continuous with respect to ν with density

$$\frac{\mathrm{d}\zeta_{\lambda^\star}}{\mathrm{d}\nu} = \frac{\mathrm{e}^{\lambda^\star \mathrm{id}_{[0,1]}}}{\mathbb{E}_\nu\big[\mathrm{e}^{\lambda^\star \mathrm{id}_{[0,1]}}\big]} = \mathrm{e}^{\lambda^\star \mathrm{id}_{[0,1]} - \phi_\nu(\lambda^\star)}\,,$$

we have $\mathbb{E}_{\zeta_{\lambda^\star}}\big[\mathrm{id}_{[0,1]}\big] = \mathrm{E}\big(\zeta_{\lambda^\star}\big) = \phi'_\nu(\lambda^\star) = x$. Therefore, by definition of $\mathcal{L}^{\leqslant}_{\inf}(x, \nu)$ and of the Kullback-Leibler divergence,

$$\mathcal{L}^{\leqslant}_{\inf}(x, \nu) \leqslant \mathrm{KL}\big(\zeta_{\lambda^\star}, \nu\big) = \mathbb{E}_{\zeta_{\lambda^\star}}\left[\ln \frac{\mathrm{d}\zeta_{\lambda^\star}}{\mathrm{d}\nu}\right] = \lambda^\star\, \mathbb{E}_{\zeta_{\lambda^\star}}\big[\mathrm{id}_{[0,1]}\big] - \phi_\nu(\lambda^\star) = \Lambda(\lambda^\star) = \phi^\star_\nu(x)\,.$$

Case 2: $x = m(\nu)$. In that case, $\Lambda' \to 0$ at $-\infty$ and Λ' is non-increasing, thus $\Lambda' \leqslant 0$ on $\mathbb{R}$ and Λ is non-increasing on $\mathbb{R}$. Thus,

$$\phi^\star_\nu\big(m(\nu)\big) = \sup_{\lambda \in \mathbb{R}} \big\{\Lambda(\lambda)\big\} = \lim_{\lambda \to -\infty} \Lambda(\lambda) = \lim_{\lambda \to -\infty} -\ln \mathbb{E}_\nu\Big[\mathrm{e}^{\lambda(\mathrm{id}_{[0,1]} - m(\nu))}\Big]\,.$$

By monotone convergence based on $\mathrm{id}_{[0,1]} - m(\nu) \geqslant 0$ ν-a.s.,

$$\lim_{\lambda \to -\infty} - \ln \mathbb{E}_\nu \Big[\mathrm{e}^{\lambda(\mathrm{id}_{[0,1]} - m(\nu))} \Big] = - \ln \nu\big\{ m(\nu) \big\} \,,$$

whether $\nu\big\{ m(\nu) \big\}$ is positive or null. Moreover, the very end of the proof of Lemma 15 shows that

$$\mathcal{L}_{\mathrm{inf}}^{\leqslant}\big(m(\nu), \nu \big) = - \ln \nu\big\{ m(\nu) \big\} \,.$$

We therefore have $\mathcal{L}_{\mathrm{inf}}^{\leqslant}(x, \nu) = \phi_\nu^\star(x)$ in this case.

Case 3: $x < m(\nu)$. In that case, as $\Lambda' \to x - m(\nu) < 0$ at $-\infty$, we get that $\Lambda \to +\infty$ at $-\infty$ and $\phi_\nu^\star(x) = +\infty$. No distribution $\zeta \in \mathcal{P}[0,1]$ with $\mathrm{E}(\zeta) \leqslant x$, if some exists, can be absolutely continuous with respect to ν, as $x < m(\nu)$ imposes that ζ puts some probability mass to the left of the support of ν. Therefore, $\mathrm{KL}(\zeta, \nu) = +\infty$. All in all, $\mathcal{L}_{\mathrm{inf}}^{\leqslant}(x, \nu)$ appears as the infimum of either an empty set or of $+\infty$ values, so that $\mathcal{L}_{\mathrm{inf}}^{\leqslant}(x, \nu) = +\infty$. In this case as well, $\mathcal{L}_{\mathrm{inf}}^{\leqslant}(x, \nu) = \phi_\nu^\star(x)$, both being equal to $+\infty$. ■

C.3. The case of canonical one-parameter exponential models $\mathcal{D}_{\mathbf{exp}}$

In this section, we consider a model $\mathcal{D}_{\mathrm{exp}}$ of distributions given by a canonical one-parameter exponential family and will prove that the target equality (15) is satisfied by $\mathcal{D}_{\mathrm{exp}}$. Before we do so, let us recall briefly some definitions and properties of exponential families; more detail (including proofs of the stated properties) may be found in Lehmann and Casella (1998).

Canonical one-parameter exponential families. We fix a reference real measure ρ and consider the set, called the natural parameter space,

$$\Theta = \left\{ \theta \in \mathbb{R} \,:\, \int_{\mathbb{R}} \exp(\theta\, y)\, \mathrm{d}\rho(y) < +\infty \right\} .$$

We assume that Θ is an open interval (the model is said to be regular) and we consider the model $\mathcal{D}_{\mathrm{exp}} = \{\nu_\theta : \theta \in \Theta\}$, where, for $\theta \in \Theta$, the distribution ν_θ is absolutely continuous with respect to ρ, with density

$$\frac{\mathrm{d}\nu_\theta}{\mathrm{d}\rho} = \exp\big(\theta\, \mathrm{id}_{\mathbb{R}} - b(\theta) \big) \,, \tag{33}$$

for a twice differentiable function $b : \Theta \to \mathbb{R}$. Due to density constraints, we get, for all $\theta \in \Theta$,

$$\int_{\mathbb{R}} \underbrace{\exp\big(\theta y - b(\theta)\big)\, \mathrm{d}\rho(y)}_{= \mathrm{d}\nu_\theta(y)} = \mathbb{E}_{\nu_\theta}[1] = 1 \quad \text{or, equivalently,} \quad b(\theta) = \ln \int_{\mathbb{R}} e^{\theta y}\, \mathrm{d}\rho(y) \,. \tag{34}$$

It can be seen that b is strictly convex and that $\mathrm{E}(\nu_\theta) = b'(\theta)$ for all $\theta \in \Theta$. Thus, b' is a one-to-one mapping between Θ and the set $\mathcal{M}$ of the expectations of the distributions of $\mathcal{D}_{\mathrm{exp}}$; the set $\mathcal{M}$ is an open interval of $\mathbb{R}$, by continuity of b', whose bounds are denoted by μ_- and μ^+. In particular, a distribution of $\mathcal{D}_{\mathrm{exp}}$ is characterised by its mean. We can also parameterize the Kullback-Leibler divergence function by the expectations: we set, for all $\theta_1, \theta_2 \in \Theta$,

$$d\big(\mathrm{E}(\nu_{\theta_1}), \mathrm{E}(\nu_{\theta_2})\big) \overset{\text{def}}{=} \mathrm{KL}(\nu_{\theta_1}, \nu_{\theta_2}) \,.$$

This defines a divergence d which is strictly convex and differentiable on the open set $\mathcal{M} \times \mathcal{M}$. In particular, d is continuous, is such that $d(\mu, \mu') = 0$ if and only if $\mu = \mu'$, and, for all $\mu \in \mathcal{M}$, both $d(\mu, \cdot\,)$ and $d(\,\cdot\,, \mu)$ are decreasing on $(\mu_-, \mu]$, and increasing on $[\mu, \mu^+)$.

In the following, we extend d to $\mathbb{R} \times \mathbb{R}$ by $+\infty$ values outside of $\mathcal{M} \times \mathcal{M}$.

Links between $\mathcal{L}_{\inf}^{<}$, $\mathcal{L}_{\inf}^{\leqslant}$, $\mathcal{L}_{\inf}^{>}$, $\mathcal{L}_{\inf}^{\geqslant}$ and d. A direct application of the continuity and monotonicity properties of d is that all functions $\mathcal{L}_{\inf}^{<}$, $\mathcal{L}_{\inf}^{\leqslant}$, $\mathcal{L}_{\inf}^{>}$, $\mathcal{L}_{\inf}^{\geqslant}$ easily rewrite as d, as stated in equalities (9) and (10). For instance, for $\nu \in \mathcal{D}_{\exp}$ and $x \leqslant \mathrm{E}(\nu)$, we have, when $x \in \mathcal{M}$:

$$\mathcal{L}_{\inf}^{<}(x, \nu) = \inf_{\mu < x} \big\{ d(\mu, \nu) \big\} = \lim_{\substack{\mu \to x \\ \mu < x}} d(\mu, \nu) = d(x, \nu)\,;$$

and when $x \notin \mathcal{M}$, by convention, $\mathcal{L}_{\inf}^{<}(x, \nu) = +\infty = d(x, \nu)$.

We are now able to prove that all canonical one-parameter exponential models satisfy Equation (15), which rewrites as follows, thanks to (9) and (10): for all $\nu, \nu' \in \mathcal{D}_{\exp}$ with $\mathrm{E}(\nu') < \mathrm{E}(\nu)$,

$$\inf_{x \in [\mathrm{E}(\nu'), \mathrm{E}(\nu)]} \Big\{ \phi_{\nu'}^\star(x) + \phi_\nu^\star(x) \Big\} = \inf_{x \in [\mathrm{E}(\nu'), \mathrm{E}(\nu)]} \Big\{ d\big(x, \mathrm{E}(\nu')\big) + d\big(x, \mathrm{E}(\nu)\big) \Big\}\,. \tag{35}$$

To obtain this result, we will not prove an exact counterpart of Lemma 4: we will rather only focus on points x belonging to $\mathcal{M}$, not to all $x \in \mathbb{R}$; see (36) below. In particular, we avoid the cases $x \in \{\mu_-, \mu^+\}$, the boundary points of $\mathcal{M}$, at which the equality of $d\big(\,\cdot\,, \mathrm{E}(\nu)\big)$ and $\phi_\nu^\star$ does not seem to hold in general.

Proof of (35). It suffices to show that, for all $\nu \in \mathcal{D}_{\exp}$ and $x \in \mathcal{M}$,

$$\phi_\nu^\star(x) \overset{\text{def}}{=} \sup_{\lambda \in \mathbb{R}} \big\{ \lambda x - \phi_\nu(\lambda) \big\} = d\big(x, \mathrm{E}(\nu)\big)\,. \tag{36}$$

Now, by Lemma 18, we only need to show that $\phi_\nu^\star(x) \geqslant d\big(x, \mathrm{E}(\nu)\big)$, and to prove this inequality, we will justify the existence of $\lambda^\star \in \mathbb{R}$ such that

$$d\big(x, \mathrm{E}(\nu)\big) = \lambda^\star x - \phi_\nu(\lambda^\star)\,. \tag{37}$$

Let $\theta_1 \in \Theta$ be such that $\nu = \nu_{\theta_1}$ and $\theta_2 = (b')^{-1}(x) \in \Theta$ be such that $\mathrm{E}(\nu_{\theta_2}) = x$. We will prove (37) with $\lambda^\star \overset{\text{def}}{=} \theta_2 - \theta_1$. Using the model density (33), we note that ν_{θ_2} is absolutely continuous with respect to ν_{θ_1} and compute, by definition of the Kullback-Leibler divergence,

$$\begin{aligned} d\big(x, \mathrm{E}(\nu)\big) = \mathrm{KL}(\nu_{\theta_2}, \nu_{\theta_1}) &= \mathbb{E}_{\nu_{\theta_2}}\left[\ln \frac{\mathrm{d}\nu_{\theta_2}}{\mathrm{d}\nu_{\theta_1}} \right] = \mathbb{E}_{\nu_{\theta_2}}\Big[(\theta_2 - \theta_1)\mathrm{id}_{\mathbb{R}} - \big(b(\theta_2) - b(\theta_1)\big) \Big] \\ &= (\theta_2 - \theta_1)\mathrm{E}(\nu_{\theta_2}) - \big(b(\theta_2) - b(\theta_1)\big) = \lambda^\star x - \big(b(\theta_2) - b(\theta_1)\big)\,. \end{aligned} \tag{38}$$

To obtain (37), it only remains to show that

$$b(\theta_2) - b(\theta_1) = \phi_\nu(\lambda^\star)\,. \tag{39}$$

Using (34) at θ_2 and again the density (33) at θ_1, we obtain

$$\begin{aligned} b(\theta_2) = \ln \int_{\mathbb{R}} e^{\theta_2 y}\, \mathrm{d}\rho(y) &= \ln \int_{\mathbb{R}} e^{(\theta_2 - \theta_1) y} \underbrace{e^{\theta_1 y - b(\theta_1)}\, \mathrm{d}\rho(y)}_{= \mathrm{d}\nu_{\theta_1}(y) = \mathrm{d}\nu(y)} + b(\theta_1) \\ &= \ln \int_{\mathbb{R}} e^{\lambda^\star y}\, \mathrm{d}\nu(y) + b(\theta_1) = \phi_\nu(\lambda^\star) + b(\theta_1)\,, \end{aligned} \tag{40}$$

which gives (39) and concludes the proof of (37).

Remark 19 *A more direct approach bypassing Lemma 18 can be followed with $\mathcal{D}_{exp}$ models, along the following lines. The result (40) can be generalized into*

$$\forall \theta \in \Theta, \quad \phi_\nu(\theta - \theta_1) = b(\theta) - b(\theta_1)\,.$$

As b is differentiable on Θ, the function ϕ_ν is also differentiable; at $\lambda^\star = \theta_2 - \theta_1$, we have

$$\phi_\nu'(\lambda^\star) = \phi_\nu'(\theta_2 - \theta_1) = b'(\theta_2) = x\,.$$

Thus, the derivative of the strictly concave function

$$\Lambda : \lambda \in \mathbb{R} \longmapsto \lambda x - \phi_\nu(\lambda)$$

vanishes at $\lambda^\star$, which is therefore the argument of its maximum: $\phi_\nu^\star(x) = \Lambda(\lambda^\star)$. The latter equality, together with the closed-form calculation (38), shows (36).

C.4. Condition for general models

In this section, we extend the equality (15) to more general models; we did so by analyzing which conditions were actually required in the proof of Lemma 4. We recall that the pieces of notation $m(\nu)$ and $M(\nu)$ for the lower and upper ends of the support of a distribution ν were introduced in Appendix A.2 and lie in $\mathbb{R} \cup \{-\infty\}$ and $\mathbb{R} \cup \{+\infty\}$ respectively in all generality. We also remind the reader that all considered models contain distributions with finite first moments.

Lemma 20 *Assume that a model $\mathcal{D}$ is such that for all distributions $\nu \in \mathcal{D}$ and all $\lambda \in \mathbb{R}$, on the one hand, the quantity $\phi_\nu(\lambda)$ is well-defined and finite, and, on the other hand, the distribution ν_λ with density*

$$\frac{\mathrm{d}\nu_\lambda}{\mathrm{d}\nu} = \frac{\mathrm{e}^{\lambda \mathrm{id}_\mathbb{R}}}{\mathbb{E}_\nu\left[\mathrm{e}^{\lambda \mathrm{id}_\mathbb{R}}\right]} \qquad \textit{with respect to } \nu$$

belongs to $\mathcal{D}$. Assume also that δ_x, the Dirac mass at x, belongs to $\mathcal{D}$ whenever there exists $\nu \in \mathcal{D}$ with $x \in \big\{m(\nu), M(\nu)\big\}$ and $\nu\{x\} > 0$.

Then, for all $\nu \in \mathcal{D}$,

$$\forall x \leqslant \mathrm{E}(\nu), \quad \phi_\nu^\star(x) = \mathcal{L}_{\inf}^{\leqslant}(x, \nu) \qquad \textit{and} \qquad \forall x \geqslant \mathrm{E}(\nu), \quad \phi_\nu^\star(x) = \mathcal{L}_{\inf}^{\geqslant}(x, \nu)\,.$$

Proof By Lemma 18, we only need to prove that

$$\forall x \leqslant \mathrm{E}(\nu), \quad \phi_\nu^\star(x) \geqslant \mathcal{L}_{\inf}^{\leqslant}(x, \nu) \qquad \text{and} \qquad \forall x \geqslant \mathrm{E}(\nu), \quad \phi_\nu^\star(x) \geqslant \mathcal{L}_{\inf}^{\geqslant}(x, \nu)\,.$$

We will follow the proof of Lemma 4 for the model $\mathcal{P}[0,1]$ (see Appendix C.2), and show that the same proof applies to $\mathcal{D}$, up to a few changes in the justifications. In particular, we only treat the first case, showing that

$$\forall x \leqslant \mathrm{E}(\nu), \quad \phi_\nu^\star(x) \geqslant \mathcal{L}_{\inf}^{\leqslant}(x, \nu)\,, \tag{41}$$

and obtain the other by symmetry. For $x = \mathrm{E}(\nu)$, we show, as in the $\mathcal{P}[0,1]$ case, that

$$\phi_\nu^\star\big(\mathrm{E}(\nu)\big) = 0 = \mathcal{L}_{\inf}^{\leqslant}\big(\mathrm{E}(\nu), \nu\big)\,.$$

Before proving (41) for $x < \mathrm{E}(\nu)$, we show that ϕ_ν satisfies the same required properties than in the $\mathcal{P}[0,1]$ case. As ϕ_ν takes finite values on $\mathbb{R}$, the moment-generating function of ν is a power series with infinite radius of convergence. In particular, ϕ_ν is differentiable over $\mathbb{R}$ and, as Hölder's inequality still entails that ϕ_ν is convex, the derivative ϕ_ν' is non-decreasing. We now prove that the limit of the latter at $-\infty$ (which we recall exists and belongs to $\{-\infty\} \cup \mathbb{R}$ by monotonicity) equals $m(\nu)$:

$$\lim_{\lambda \to -\infty} \phi_\nu'(\lambda) \overset{\text{def}}{=} \ell = m(\nu)\,. \tag{42}$$

If $m(\nu) \in \mathbb{R}$, the same proof as in the $\mathcal{P}[0,1]$ case applies. Otherwise, $m(\nu) = -\infty$ and we prove that ℓ cannot be finite. Indeed, if $\ell \in \mathbb{R}$, then using that ϕ_ν' is always larger than its limit ℓ by non-decreasing, and following the same analysis as in (31)–(32), we obtain $\mathrm{id}_\mathbb{R} - \ell \geqslant 0$ ν-a.s., which in turn entails that $\ell \leqslant m(\nu)$ and contradicts the fact that $m(\nu) = -\infty$. This concludes the proof of (42).

Finally, we prove that

$$\phi_\nu' : \lambda \in \mathbb{R} \longmapsto \frac{\mathbb{E}_\nu\left[\mathrm{id}_\mathbb{R}\, \mathrm{e}^{\lambda \mathrm{id}_\mathbb{R}}\right]}{\mathbb{E}_\nu\left[\mathrm{e}^{\lambda \mathrm{id}_\mathbb{R}}\right]}\,,$$

by applying a standard theorem of differentiation under the integral. To obtain the domination on a subset (λ_-, λ_+) of $\mathbb{R}$, we write that

$$\forall \lambda \in (\lambda_-, \lambda_+), \quad \left|\mathrm{id}_\mathbb{R}\, \mathrm{e}^{\lambda \mathrm{id}_\mathbb{R}}\right| \leqslant \left|\mathrm{id}_\mathbb{R}\right| \left(\mathrm{e}^{\lambda_- \mathrm{id}_\mathbb{R}} + \mathrm{e}^{\lambda_+ \mathrm{id}_\mathbb{R}}\right) \overset{\text{def}}{=} h$$

and get that h is ν–integrable whenever distributions ν_{λ_-} and ν_{λ_+} admit a finite first moment, which holds by assumption as those distributions belong to $\mathcal{D}$.

We now get all necessary properties to follow the proof of Lemma 4. Similarly to this proof, we split the analysis of the case $x < \mathrm{E}(\nu)$ into three sub-cases, depending on the respective positions of x and $m(\nu)$.

Case 1: $x > m(\nu)$. The properties of ϕ_ν ensure, as in the corresponding $\mathcal{P}[0,1]$ case, the existence of $\lambda^\star$ such that

$$\phi_\nu^\star(x) = \lambda^\star x - \phi_\nu(\lambda^\star)\,.$$

Besides, we get, again as in the proof of Lemma 4, that $\mathrm{E}(\nu_{\lambda^\star}) = \phi_\nu'(\lambda^\star) = x$ and compute that

$$\mathrm{KL}(\nu_{\lambda^\star}, \nu) = \lambda^\star x - \phi_\nu(\lambda^\star)\,.$$

This gives (41) noting that $\mathcal{L}_{\inf}^{\leqslant}(x, \nu) \leqslant \mathrm{KL}\big(\delta_{m(\nu)}, \nu\big)$, as by assumptions $\nu_{\lambda^\star} \in \mathcal{D}$.

Case 2: $x = m(\nu)$. As in the proof of this case in Lemma 4, we prove that

$$\phi_\nu^\star\big(m(\nu)\big) = -\ln \nu\big\{m(\nu)\big\}\,,$$

and get

$$\mathrm{KL}\big(\delta_{m(\nu)}, \nu\big) = -\ln \nu\big\{m(\nu)\big\}\,.$$

This also concludes this case, as $\delta_{m(\nu)} \in \mathcal{D}$ by assumption.

Case 3: $x < m(\nu)$. Similarly to the corresponding $\mathcal{P}[0,1]$ case, we show that both $\mathcal{L}_{\inf}^{\leqslant}(x, \nu)$ and $\phi_\nu^\star(x)$ are infinite. ∎

Appendix D. Proofs for lower bounds (Section 4)

This section provides the detailed proofs that were omitted when stating our various lower bounds in Section 4.

D.1. Proof of Lemma 6

We restate it for the convenience of the reader.

Lemma 6 *Fix $K \geqslant 2$ and a model $\mathcal{D}$. Consider a consistent sequence of strategies on $\mathcal{D}$, and two generic bandit problems $\underline{\nu}$ and $\underline{\lambda}$ in $\mathcal{D}$ such that $a^\star(\underline{\lambda}) \neq a^\star(\underline{\nu})$. Then*

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -\limsup_{T\to+\infty} \sum_{a=1}^{K} \frac{\mathbb{E}_{\underline{\lambda}}[N_a(T)]}{T} \mathrm{KL}(\lambda_a, \nu_a)\,.$$

Proof The considered sequence of strategies being consistent on $\mathcal{D}$, and as $a^\star(\underline{\lambda}) \neq a^\star(\underline{\nu})$,

$$q_T \overset{\text{def}}{=} \mathbb{P}_{\underline{\lambda}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant \mathbb{P}_{\underline{\lambda}}\big(I_T = a^\star(\underline{\lambda})\big) \underset{T\to+\infty}{\longrightarrow} 1\,,$$

$$\text{while} \qquad p_T \overset{\text{def}}{=} \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \underset{T\to+\infty}{\longrightarrow} 0\,.$$

Note that we introduced above short-hand notation p_T and q_T.

The fundamental inequality for lower bounds in bandit problems (which is a consequence of the chain rule and of the data-processing inequality for Kullback-Leibler divergences, see Garivier et al., 2019), applied for $Z = \mathbb{I}_{\{I_T \neq a^\star(\underline{\nu})\}}$, exactly states here that

$$\sum_{a=1}^{K} \mathbb{E}_{\underline{\lambda}}[N_a(T)]\, \mathrm{KL}(\lambda_a, \nu_a) \geqslant \mathrm{KL}\big(\mathrm{Ber}(q_T), \mathrm{Ber}(p_T)\big)\,, \tag{43}$$

where we recall that $\mathrm{Ber}(p)$ refers to the Bernoulli distribution with parameter p. Given the asymptotics of p_T and q_T,

$$\mathrm{KL}\big(\mathrm{Ber}(q_T), \mathrm{Ber}(p_T)\big) = q_T \ln \frac{q_T}{p_T} + (1-q_T) \ln \frac{1-q_T}{1-p_T} \sim -\ln p_T \qquad \text{as } T \to +\infty\,.$$

Put differently,

$$\frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \sim -\frac{\mathrm{KL}\big(\mathrm{Ber}(q_T), \mathrm{Ber}(p_T)\big)}{T}\,.$$

Combining this limit behavior with the previous inequality leads to the stated result, namely:

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -\limsup_{T\to+\infty} \sum_{a=1}^{K} \frac{\mathbb{E}_{\underline{\lambda}}[N_a(T)]}{T} \mathrm{KL}(\lambda_a, \nu_a)\,.$$

■

D.2. Proof of Theorem 9

We restate it for the convenience of the reader.

Theorem 9 *Fix a model $\mathcal{D}$. Consider a doubly-indexed sequence of strategies that is consistent, balanced against the worst arm on $\mathcal{D}$, and that cleverly exploits the pruning of suboptimal arms on $\mathcal{D}$. For all generic bandit problems $\underline{\nu}$ in $\mathcal{D}$ with $K \geqslant 2$ arms,*

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \min_{2\leqslant k\leqslant K} \frac{\mathcal{L}^<_{\inf}\big(\mu_{(k)}, \nu^\star\big)}{k} \,.$$

Proof The proof consists of two steps. The first step is to prove that for a generic bandit problem $\underline{\nu}$ in $\mathcal{D}$ with $K \geqslant 2$ arms, we have,

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \frac{\mathcal{L}^<_{\inf}\big(\mu_{(K)}, \nu^\star\big)}{K} \,. \tag{44}$$

In the second step, we use this lower bound and the very definition of the clever exploitation of the pruning of suboptimal arms to get the claimed bound.

Step 1: lower bound (44). We follow a well-established methodology and consider an alternative bandit problem only differing from $\underline{\nu}$ at one arm, namely, at the best arm. To do so, we set some distribution $\zeta \in \mathcal{D}$ with $\mathrm{E}(\zeta) < \mu_{(K)}$, if some exists, and define the bandit problem $\underline{\lambda} = (\lambda_1, \dots, \lambda_K)$ as

$$\lambda_a = \begin{cases} \zeta & \text{if } a = a^\star(\underline{\nu}), \\ \nu_a & \text{if } a \neq a^\star(\underline{\nu}). \end{cases}$$

Observe that $\underline{\lambda}$ is also a generic bandit problem in $\mathcal{D}$, that $a^\star(\underline{\nu})$ is the worst arm in $\underline{\lambda}$ (and also that the second best arm of $\underline{\nu}$ is the optimal arm in $\underline{\lambda}$, but we will not use this specific fact). Therefore, Lemma 6 yields, as $\underline{\lambda}$ and $\underline{\nu}$ only differ at arm $a^\star(\underline{\nu})$,

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \limsup_{T\to+\infty} \frac{\mathbb{E}_{\underline{\lambda}}[N_{a^\star(\underline{\nu})}(T)]}{T} \mathrm{KL}(\lambda_{a^\star(\underline{\nu})}, \nu^\star) \,,$$

where we recall that $\nu^\star = \nu_{a^\star(\underline{\nu})}$. Given that $a^\star(\underline{\nu})$ is the worst arm of $\underline{\lambda}$, and since by assumption, the sequence of strategies is balanced against the worst arm,

$$\limsup_{T\to+\infty} \frac{1}{T} \mathbb{E}_{\underline{\lambda}}\big[N_{a^\star(\underline{\nu})}(T)\big] \leqslant \frac{1}{K} \,,$$

proving that

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \frac{\mathrm{KL}(\zeta, \nu^\star)}{K} \,.$$

The claimed inequality (44) follows from taking the supremum in the right-hand side over distributions $\zeta \in \mathcal{D}$ with $\mathrm{E}(\zeta) < \mu_{(K)}$.

Step 2: clever exploitation of pruning. For each $k \in \{2, \dots, K-1\}$, define $\underline{\nu}'_{1:k}$ as the sub-problem of $\underline{\nu}$ obtained by keeping the k best arms and dropping the $K-k$ worse arms. Use the definition of clever exploitation of pruning of suboptimal arms and apply (44) to $\underline{\nu}'_{1:k}$ to get

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant \liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}'_{1:k}}\Big(I_T \neq a^\star\big(\underline{\nu}'_{1:k}\big)\Big) \geqslant -\frac{\mathcal{L}^{<}_{\inf}\big(\mu_{(k)}, \nu^\star\big)}{k} \,.$$

Taking the maximum of all lower bounds exhibited as k varies between 2 and K, we proved the claimed result. ■

D.3. Proof of the normality of the models $\mathcal{P}[0,1]$ and $\mathcal{D}_{\mathbf{exp}}$

In this section, we show that $\mathcal{P}[0,1]$ and canonical one-parameter exponential models are normal. We focus first on $\mathcal{P}[0,1]$.

Proposition 21 *$\mathcal{P}[0,1]$ is a normal model.*

Proof We fix $\nu \in \mathcal{P}[0,1]$, a real $x \geqslant \mathrm{E}(\nu)$ and consider a positive ε. We recall that the pieces of notation $m(\nu)$ and $M(\nu)$ for the lower and upper ends of the support of a distribution ν were introduced in Appendix A.2. If $x \geqslant M(\nu)$, there exists no distribution ζ absolutely continuous with respect to ν and such that $\mathrm{E}(\zeta) > x$, hence the considered infima are infinite:

$$\mathcal{L}^{>}_{\inf}(x, \nu) = +\infty = \inf\big\{\mathrm{KL}(\zeta, \nu) : \ \zeta \in \mathcal{D} \ \text{ s.t. } \ x + \varepsilon > \mathrm{E}(\zeta) > x\big\} \,.$$

Assume now that $\mathrm{E}(\nu) \leqslant x < M(\nu)$ and note that this case only occurs when $\mathrm{E}(\nu) < M(\nu)$, i.e. when ν is not a Dirac distribution. It is clear that

$$\mathcal{L}^{>}_{\inf}(x, \nu) \leqslant \inf\big\{\mathrm{KL}(\zeta, \nu) : \ \zeta \in \mathcal{D} \ \text{ s.t. } \ x + \varepsilon > \mathrm{E}(\zeta) > x\big\} \,.$$

To prove the other inequality $\geqslant$, we will first remind why

$$\mathcal{L}^{>}_{\inf}(x, \nu) = \mathrm{KL}(\zeta_{\lambda^\star}, \nu) \,, \tag{45}$$

where $\zeta_{\lambda^\star}$ is a distribution of $\mathcal{D}$ of mean x introduced in the proof of Lemma 4. As the expectation of $\zeta_{\lambda^\star}$ does not belong to $(x, x+\varepsilon)$, we will then need to prove that

$$\inf\big\{\mathrm{KL}(\zeta, \nu) : \ \zeta \in \mathcal{D} \ \text{ s.t. } \ x + \varepsilon > \mathrm{E}(\zeta) > x\big\} \leqslant \mathrm{KL}(\zeta_{\lambda^\star}, \nu) \,. \tag{46}$$

In the remaining of the proof, we rely on notation and results proved in the first case (by symmetry) of the proof of Lemma 4, and for $\lambda \in \mathbb{R}$, we define ζ_λ the distribution absolutely continuous with respect to ν with density

$$\frac{\mathrm{d}\zeta_\lambda}{\mathrm{d}\nu} = \frac{\mathrm{e}^{\lambda \mathrm{id}_{[0,1]}}}{\mathbb{E}_\nu\big[\mathrm{e}^{\lambda \mathrm{id}_{[0,1]}}\big]} = \mathrm{e}^{\lambda \mathrm{id}_{[0,1]} - \phi_\nu(\lambda)} \,,$$

On the one hand, by Lemma 15 (or more precisely its formulation for $\mathcal{L}^{>}_{\inf}$ and $\mathcal{L}^{\geqslant}_{\inf}$ instead of $\mathcal{L}^{<}_{\inf}$ and $\mathcal{L}^{\leqslant}_{\inf}$), we know that

$$\mathcal{L}^{>}_{\inf}(x, \nu) = \mathcal{L}^{\geqslant}_{\inf}(x, \nu) \,,$$

while, on the other hand, the proof of the first case of Lemma 4 ensures the existence of $\lambda^\star \in \mathbb{R}$ such that $\mathrm{E}(\zeta_{\lambda^\star}) = \phi'_\nu(\lambda^\star) = x$ and

$$\mathcal{L}^{\geqslant}_{\inf}(x,\nu) = \mathrm{KL}(\zeta_{\lambda^\star},\nu)\,.$$

We thus obtained (45) and move to Equation (46).

We noticed that ν is not a Dirac distribution, as $\mathrm{E}(\nu) \leqslant x < M(\nu)$. This entails, by Hölder's inequality, that ϕ_ν is strictly convex, hence ϕ'_ν is an increasing function. Considering distributions ζ_λ for which $\lambda > \lambda^\star$, this implies that

$$\mathbb{E}_{\zeta_\lambda}\big[\mathrm{id}_{[0,1]}\big] = \mathrm{E}(\zeta_\lambda) = \phi'_\nu(\lambda) > x = \mathrm{E}(\zeta_{\lambda^\star})\,.$$

As ϕ'_ν is a continuous function (see, again, Lemma 4), we also get

$$\lim_{\substack{\lambda\to\lambda^\star\\ \lambda>\lambda^\star}} \mathrm{E}(\zeta_\lambda) = \mathrm{E}(\zeta_{\lambda^\star})\,,$$

so that there exists $\lambda_\varepsilon > \lambda^\star$ such that

$$\forall \lambda \in (\lambda^\star, \lambda_\varepsilon), \qquad \mathrm{E}(\zeta_\lambda) \in (x, x+\varepsilon)\,.$$

As a consequence,

$$\inf\big\{\mathrm{KL}(\zeta,\nu):\ \zeta\in\mathcal{D}\ \text{ s.t. }\ x+\varepsilon > \mathrm{E}(\zeta) > x\big\} \leqslant \inf_{\lambda\in(\lambda_\varepsilon,\lambda^\star)}\Big\{\mathrm{KL}(\zeta_\lambda,\nu)\Big\}\,, \tag{47}$$

and we compute that, for $\lambda \in (\lambda^\star, \lambda_\varepsilon)$

$$\mathrm{KL}(\zeta_\lambda,\nu) = \mathbb{E}_{\zeta_\lambda}\left[\ln\frac{\mathrm{d}\zeta_\lambda}{\mathrm{d}\nu}\right] = \lambda\,\mathbb{E}_{\zeta_\lambda}\big[\mathrm{id}_{[0,1]}\big] - \phi_\nu(\lambda) = \lambda\,\phi'_\nu(\lambda) - \phi_\nu(\lambda)$$
$$\underset{\lambda\to\lambda^\star}{\longrightarrow} \lambda^\star\,\phi'_\nu(\lambda^\star) - \phi_\nu(\lambda^\star) = \mathrm{KL}(\zeta_{\lambda^\star},\nu)\,,$$

by continuity of ϕ_ν and ϕ'_ν. Combining this limit behaviour with inequality (47) leads to (46). ■

We now consider canonical one-parameter exponential models, for which normality is easily obtained by the rewriting of $\mathcal{L}^{>}_{\inf}$ as d.

Proposition 22 *All canonical one-parameter exponential models $\mathcal{D}_{exp}$ are normal.*

Proof We fix $\nu \in \mathcal{D}_{\mathrm{exp}}$, a real $x \geqslant \mathrm{E}(\nu)$ and consider a positive ε. As in the $\mathcal{P}[0,1]$ case, the required equality holds (both terms are infinite) when $x \geqslant M(\nu)$.

Assume now that $x < M(\nu)$. By equality (10), we recall that

$$\mathcal{L}^{>}_{\inf}(x,\nu) = d\big(x, \mathrm{E}(\nu)\big)\,.$$

Using the basic properties of d (see Appendix C.3), namely that $d\big(\,\cdot\,, \mathrm{E}(\nu)\big)$ is continuous at x and is non-decreasing on $(x,+\infty) \subset [\mathrm{E}(\nu),+\infty)$, this gives

$$\begin{aligned}\mathcal{L}^{>}_{\inf}(x,\nu) &= \lim_{y\to x^+} d\big(y,\mathrm{E}(\nu)\big)\\ &= \inf_{y\in(x,x+\varepsilon)}\Big\{d\big(y,\mathrm{E}(\nu)\big)\Big\}\\ &= \inf\big\{\mathrm{KL}(\zeta,\nu):\ \zeta\in\mathcal{D}\ \text{ s.t. }\ x+\varepsilon > \mathrm{E}(\zeta) > x\big\}\,.\end{aligned}$$

■

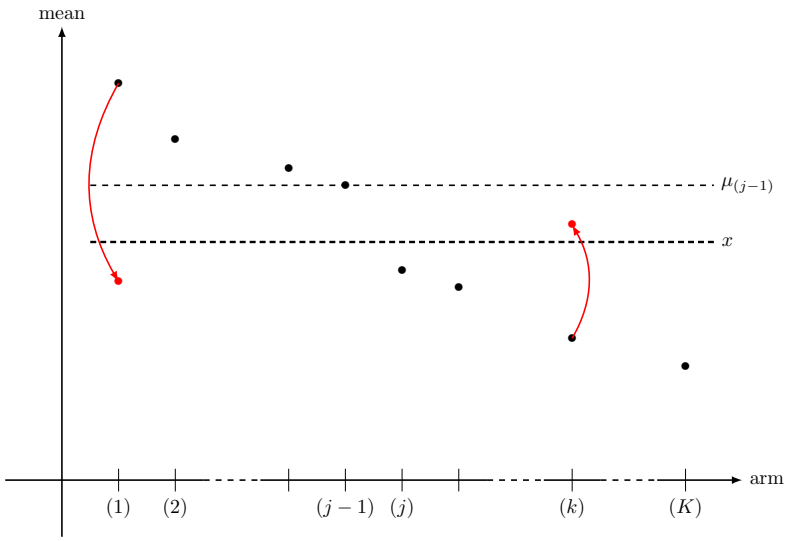

Figure 1: Original bandit problem $\underline{\nu}$ (in dark) and modifications made to arms (1) and (k) to obtain an alternative bandit problem $\underline{\lambda} \in \mathrm{Alt}_{k,j,x}(\underline{\nu})$ (in red): in $\underline{\lambda}$, arm (k) is the $j-1$–th best arm, while arm $(1) = a^\star(\underline{\nu})$ is at best the j–th best arm.

D.4. Proof of Theorem 12

We restate it for the convenience of the reader.

Theorem 12 *Fix $K \geqslant 2$ and a normal model $\mathcal{D}$. Consider a sequence of strategies which is consistent and monotonous on $\mathcal{D}$. For all generic bandit problems $\underline{\nu}$ in $\mathcal{D}$,*

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \min_{2\leqslant k\leqslant K} \min_{2\leqslant j\leqslant k} \inf_{x\in[\mu_{(j)},\mu_{(j-1)})} \left\{ \frac{\mathcal{L}_{\inf}^{>}\big(x,\nu_{(k)}\big)}{j-1} + \frac{\mathcal{L}_{\inf}^{<}\big(x,\nu^\star\big)}{j} \right\}.$$

Proof We fix a generic bandit $\underline{\nu}$ in $\mathcal{D}$ and consider the following sets of alternative bandit problems, indexed by triplets (k,j,x) satisfying $2 \leqslant k \leqslant K$ and $2 \leqslant j \leqslant k$, as well as $x \in [\mu_{(j)}, \mu_{(j-1)})$:

$$\mathrm{Alt}_{k,j,x}(\underline{\nu}) = \Big\{ \underline{\lambda} \text{ in } \mathcal{D} : \ \mathrm{E}\big(\lambda_{(1)}\big) < x < \mathrm{E}\big(\lambda_{(k)}\big) < \mu_{(j-1)} \ \text{ and } \ \lambda_a = \nu_a \ \text{ for } \ a \notin \big\{(1),(k)\big\} \Big\};$$

in particular, an alternative problem $\underline{\lambda}$ in $\mathrm{Alt}_{k,j,x}(\underline{\nu})$ only differ from the original bandit problem $\underline{\nu}$ at the best arm (1) and at the k–th best arm (k). Given $x \in \big[\mu_{(j)}, \mu_{(j-1)}\big)$ and $\mathrm{E}\big(\lambda_{(1)}\big) < x$, arm (1) is at best the j–th best arm of $\underline{\lambda}$, but it can be possibly worse. Similarly, the same condition on x and the fact that $x < \mathrm{E}\big(\lambda_{(k)}\big)$ implies that arm (k) is exactly the $j-1$–th best arm of $\underline{\lambda}$. Both facts are illustrated on Figure 1. Thus, by monotonicity of the strategy,

$$\limsup_{T\to+\infty} \frac{\mathbb{E}_{\underline{\lambda}}\big[N_{(k)}(T)\big]}{T} \leqslant \frac{1}{j-1} \qquad \text{and} \qquad \limsup_{T\to+\infty} \frac{\mathbb{E}_{\underline{\lambda}}\big[N_{(1)}(T)\big]}{T} \leqslant \frac{1}{j}\,.$$

Given that the optimal arm in $\underline{\lambda}$ is different from the optimal arm (1) of $\underline{\nu}$, Lemma 6 may be applied; together with the two upper bounds above, it yields

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -\left(\frac{\mathrm{KL}\big(\lambda_{(k)},\nu_{(k)}\big)}{j-1} + \frac{\mathrm{KL}\big(\lambda_{(1)},\nu^\star\big)}{j} \right).$$

We can now take the infimum over all bandit problems $\underline{\lambda} \in \mathrm{Alt}_{k,j,x}(\underline{\nu})$ and obtain the following lower bound, where we define a quantity $\mathcal{I}_{k,j,x}(\underline{\nu})$:

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \inf_{\underline{\lambda}\in\mathrm{Alt}_{k,j,x}(\underline{\nu})} \left\{ \frac{\mathrm{KL}\big(\lambda_{(k)},\nu_{(k)}\big)}{j-1} + \frac{\mathrm{KL}\big(\lambda_{(1)},\nu^\star\big)}{j} \right\} \stackrel{\text{def}}{=} -\mathcal{I}_{k,j,x}(\underline{\nu})\,.$$

We prove below that

$$\mathcal{I}_{k,j,x}(\underline{\nu}) = \frac{\mathcal{L}^{>}_{\inf}\big(x,\nu_{(k)}\big)}{j-1} + \frac{\mathcal{L}^{<}_{\inf}\big(x,\nu^\star\big)}{j}\,, \tag{48}$$

from which the lower bound claimed in Theorem 12 will follow, by taking the supremum of $-\mathcal{I}_{k,j,x}(\underline{\nu})$ first over $x \in \big[\mu_{(j)},\mu_{(j-1)}\big)$, then the maximum over $2 \leqslant j \leqslant k$, and finally, the maximum over $2 \leqslant k \leqslant K$.

We now prove (48). The infimum over $\underline{\lambda} \in \mathrm{Alt}_{k,j,x}(\underline{\nu})$ may be split into two separate infima, respectively over $\lambda_{(k)}$ and $\lambda_{(1)}$; given that each term of the sum of KL only depends either on $\lambda_{(k)}$, or on $\lambda_{(1)}$, but not on both, we may write

$$\begin{aligned}
\mathcal{I}_{k,j,x}(\underline{\nu}) &= \inf_{\substack{\lambda_{(1)},\lambda_{(k)}\in\mathcal{D}:\\ \mathrm{E}(\lambda_{(1)})<x\\ x<\mathrm{E}(\lambda_{(k)})<\mu_{(j-1)}}} \left\{ \frac{\mathrm{KL}\big(\lambda_{(k)},\nu_{(k)}\big)}{j-1} + \frac{\mathrm{KL}\big(\lambda_{(1)},\nu^\star\big)}{j} \right\} \\
&= \frac{1}{j-1} \underbrace{\inf_{\substack{\lambda_{(k)}\in\mathcal{D}:\\ x<\mathrm{E}(\lambda_{(k)})<\mu_{(j-1)}}} \mathrm{KL}\big(\lambda_{(k)},\nu_{(k)}\big)}_{=\mathcal{L}^{>}_{\inf}(x,\nu_{(k)})} + \frac{1}{j} \underbrace{\inf_{\substack{\lambda_{(1)}\in\mathcal{D}:\\ \mathrm{E}(\lambda_{(1)})<x}} \mathrm{KL}\big(\lambda_{(1)},\nu^\star\big)}_{=\mathcal{L}^{<}_{\inf}(x,\nu^\star)}\,,
\end{aligned}$$

where we obtain $\mathcal{L}^{<}_{\inf}\big(x,\nu^\star\big)$ by definition while we rely on the normality of the model to obtain $\mathcal{L}^{>}_{\inf}\big(x,\nu_{(k)}\big)$: we do so with $\varepsilon = \mu_{(j-1)} - x$, which is indeed positive as we considered $x < \mu_{(j-1)}$. ■

D.5. Proof of Theorem 13

We restate it for the convenience of the reader.

Theorem 13 *Fix $K \geqslant 2$ and a model $\mathcal{D}$. Consider a consistent sequence of strategies on $\mathcal{D}$. For all generic bandit problems $\underline{\nu}$ in $\mathcal{D}$,*

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \min_{k\neq a^\star(\underline{\nu})} \inf_{x\in[\mu_k,\mu^\star]} \max\big\{\mathcal{L}^{>}_{\inf}(x,\nu_k), \mathcal{L}^{<}_{\inf}(x,\nu^\star)\big\}\,.$$

Proof Let $\underline{\nu}$ be a generic bandit problem. We fix $k \neq a^\star(\underline{\nu})$ and $x \in [\mu_k,\mu^\star]$, and prove that

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \max\big\{\mathcal{L}^{>}_{\inf}(x,\nu_k), \mathcal{L}^{<}_{\inf}(x,\nu^\star)\big\}\,,$$

from which the stated lower bound follows, by taking suprema. To do so, we consider the set of alternative bandit problems

$$\mathrm{Alt}_{k,x}(\underline{\nu}) = \Big\{\underline{\lambda} \text{ in } \mathcal{D}:\ \mathrm{E}\big(\lambda_{a^\star(\underline{\nu})}\big) < x < \mathrm{E}(\lambda_k) \ \text{ and } \ \lambda_a = \nu_a \ \text{ for } \ a \notin \big\{a^\star(\underline{\nu}), k\big\}\Big\}\,;$$

it is composed of bandit problems, only differing from $\underline{\nu}$ at arms $a^\star(\underline{\nu})$ and k, and for which arm k is better than arm $a^\star(\underline{\nu})$, with associated expectations separated by x. In particular, the optimal arm in $\underline{\lambda}$ is different from the optimal arm $a^\star(\underline{\nu})$ of $\underline{\nu}$. Lemma 6 may therefore be applied; it states that

$$\begin{aligned}
&\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \\
&\geqslant -\limsup_{T\to+\infty} \frac{\mathbb{E}_{\underline{\lambda}}\big[N_k(T)\big]}{T} \mathrm{KL}(\lambda_k, \nu_k) + \frac{\mathbb{E}_{\underline{\lambda}}\big[N_{a^\star(\underline{\nu})}(T)\big]}{T} \mathrm{KL}\big(\lambda_{a^\star(\underline{\nu})}, \nu_{a^\star(\underline{\nu})}\big) \\
&\geqslant -\max\Big\{\mathrm{KL}(\lambda_k, \nu_k),\, \mathrm{KL}\big(\lambda_{a^\star(\underline{\nu})}, \nu_{a^\star(\underline{\nu})}\big)\Big\}\,,
\end{aligned}$$

where we used, for the second inequality, the crude upper bound $N_k(T) + N_{a^\star(\underline{\nu})}(T) \leqslant T$. Taking the supremum of the obtained lower bound over all $\underline{\lambda} \in \mathrm{Alt}_{k,x}(\underline{\nu})$ leads to the following inequality, where we define the short-hand notation $\mathcal{I}_{k,x}(\underline{\nu})$:

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}(I_T \neq a^\star(\underline{\nu})) \geqslant -\inf_{\underline{\lambda}\in\mathrm{Alt}_{k,x}(\underline{\nu})} \max\Big\{\mathrm{KL}(\lambda_k, \nu_k), \mathrm{KL}\big(\lambda_{a^\star(\underline{\nu})}, \nu_{a^\star(\underline{\nu})}\big)\Big\} \stackrel{\text{def}}{=} -\mathcal{I}_{k,x}(\underline{\nu})\,.$$

The proof is concluded below by showing that $\mathcal{I}_{k,x}(\underline{\nu}) = \max\big\{\mathcal{L}^{>}_{\inf}(x, \nu_k), \mathcal{L}^{<}_{\inf}(x, \nu^\star)\big\}$.

As in the proof of Theorem 12 (see Appendix D.4), we use a separation of the infima, in the abstract form, for two functions f and g,

$$\inf_{u,v}\ \max\big\{f(u),\, g(v)\big\} = \max\Big\{\inf_u f(u)\,,\, \inf_v g(v)\Big\}\,.$$

Here, by definition of $\mathrm{Alt}_{k,x}(\underline{\nu})$,

$$\begin{aligned}
\mathcal{I}_{k,x}(\underline{\nu}) &= \inf_{\substack{\lambda_{a^\star(\underline{\nu})},\lambda_k\in\mathcal{D} \\ \mathrm{E}(\lambda_{a^\star(\underline{\nu})})<x \\ \mathrm{E}(\lambda_k)>x}} \max\Big\{\mathrm{KL}(\lambda_k, \nu_k),\ \mathrm{KL}\big(\lambda_{a^\star(\underline{\nu})}, \nu_{a^\star(\underline{\nu})}\big)\Big\} \\
&= \max\left\{\inf_{\substack{\lambda_k\in\mathcal{D} \\ \mathrm{E}(\lambda_k)>x}} \mathrm{KL}(\lambda_k, \nu_k),\quad \inf_{\substack{\lambda_{a^\star(\underline{\nu})}\in\mathcal{D} \\ \mathrm{E}(\lambda_{a^\star(\underline{\nu})})<x}} \mathrm{KL}\big(\lambda_{a^\star(\underline{\nu})}, \nu_{a^\star(\underline{\nu})}\big)\right\} \\
&= \max\Big\{\mathcal{L}^{>}_{\inf}(x, \nu_k), \mathcal{L}^{<}_{\inf}(x, \nu^\star)\Big\}\,,
\end{aligned}$$

which concludes the proof. ■

Appendix E. Additional comments for the literature review

This appendix is devoted to additional discussions concerning the fixed-budget literature. More precisely, we discuss in detail two gap-based lower bounds that we believe are somewhat detached from the spirit of the article, namely, the minimax lower bound of Carpentier and Locatelli (2016) (Appendix E.1) and the Bretagnolle-Huber technique (Appendix E.2).

E.1. The minimax lower bound of Carpentier and Locatelli (2016)

Carpentier and Locatelli (2016) proved the following non-asymptotic minimax lower bound: for the model $\mathcal{D}_{1/4}$ of Bernoulli distributions with parameters in $[1/4,\, 3/4]$, for a given consistent strategy, and each large enough budget T,

$$\exists\, \underline{\nu} \text{ in } \mathcal{D}_{1/4}, \quad \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -\frac{400}{\ln K}\left(\sum_{a \neq a^\star(\underline{\nu})} \frac{1}{\Delta_a^2}\right)^{-1} + o(1)$$

$$\geqslant -\frac{400}{\ln K} \min_{2 \leqslant k \leqslant K} \frac{\Delta_{(k)}^2}{k} + o(1)\,, \quad (49)$$

where the $o(1)$ is with respect to $T \to +\infty$.

Different nature: a non-uniform lower bound. This result is different in nature from the lower bounds considered in this article, as we now discuss. First and foremost, it improves the lower bound (6) of Audibert et al. (2010) for only one (unspecified) bandit problem $\underline{\nu}$ (belonging to a known collection of K bandit problems). This is in strong contrast with the instance-dependent lower bounds (bounds holding simultaneously for all bandit problems of the model) presented in this article.

Second, the result is stated for the restricted Bernoulli model $\mathcal{D}_{1/4}$ and does not seem to be easily generalized beyond this model or beyond similar models (e.g., the model $\mathcal{D}_{\sigma^2}$ of Gaussian distributions with a fixed variance $\sigma^2 > 0$). This is because of the truly gap-based arguments used (see the second part of the proof below).

Asymptotic statement and proof. The bound (49) may actually be stated for a sequence of strategies (as we do for all other bounds in this article) thanks to a straightforward adaptation of its proof (relying on the pigeonhole principle). More precisely, the counterpart of (49) would be that there exists an increasing sequence of budgets $(T_n)_{n\in\mathbb{N}}$ such that

$$\exists\, \underline{\nu} \text{ in } \mathcal{D}_{1/4}, \qquad \liminf_{n\to+\infty} \frac{1}{T_n} \ln \mathbb{P}_{\underline{\nu}}\big(I_{T_n} \neq a^\star(\underline{\nu})\big) \geqslant -\frac{400}{\ln K}\left(\sum_{a \neq a^\star(\underline{\nu})} \frac{1}{\Delta_a^2}\right)^{-1}. \quad (50)$$

We are actually able to slightly improve the numerical factor 400 into $100/3$ by using a sharper change-of-measure argument (namely, Lemma 6) than the original argument by Carpentier and Locatelli (2016).

Proposition 23 *Fix $K \geqslant 3$, consider the model $\mathcal{D}_{1/4}$ of Bernoulli distributions with parameters in $[1/4,\, 3/4]$, and a consistent sequence of strategies on $\mathcal{D}_{1/4}$. Then, there exists an increasing*

sequence of budgets $(T_n)_{n\in\mathbb{N}}$ *such that*

$$\exists\, \underline{\nu} \text{ in } \mathcal{D}_{1/4}, \qquad \liminf_{n\to+\infty} \frac{1}{T_n} \ln \mathbb{P}_{\underline{\nu}}\big(I_{T_n} \neq a^\star(\underline{\nu})\big) \geqslant -\frac{100}{3\ln K}\left(\sum_{a\neq a^\star(\underline{\nu})} \frac{1}{\Delta_a^2}\right)^{-1}.$$

Proof We begin the proof by introducing a collection of bandit problems $(\underline{\nu}^{(k)})_{1\leqslant k\leqslant K}$ in $\mathcal{D}_{1/4}$ and associated notation. Let $\underline{\nu}^{(1)}$ be a Bernoulli bandit problem such that the mean of arm 1 is $p_1 = 1/2$ and the mean of arm $k \in \{2,\dots,K\}$ is $p_k \in [1/4, 1/2)$. Let $\Delta_k = 1/2 - p_k$ denotes the gap of arm k in $\underline{\nu}^{(1)}$. For $k \in \{2,\dots,K\}$, we set $\underline{\nu}^{(k)}$ the bandit problem obtained by changing the arm k of $\underline{\nu}^{(1)}$ to a Bernoulli distribution with mean $1 - p_k$, hence $a^\star\big(\underline{\nu}^{(k)}\big) = k$, and we define

$$H\big(\underline{\nu}^{(k)}\big) = \sum_{a\neq k} \frac{1}{\big(\Delta_a^{(k)}\big)^2},$$

where $\Delta_a^{(k)}$ is the gap of arm a in $\underline{\nu}^{(k)}$. As for $a \neq k$, we notice that

$$\Delta_a^{(k)} = (1 - p_k) - p_a = (1/2 - p_k) + (1/2 - p_a) = \Delta_k + \Delta_a,$$

we have

$$H\big(\underline{\nu}^{(k)}\big) = \sum_{a\neq k} \frac{1}{\big(\Delta_k + \Delta_a\big)^2}. \tag{51}$$

Finally we define

$$H^\star = \sum_{k=2}^{K} \frac{1}{\Delta_k^2 H\big(\underline{\nu}^{(k)}\big)}. \tag{52}$$

The proof consists of two steps. Firstly, we show that there exists $k \in \{2,\dots,K\}$ and an increasing sequence of budgets $(T_n)_{n\in\mathbb{N}}$ such that

$$\liminf_{n\to+\infty} \frac{1}{T_n} \ln \mathbb{P}_{\underline{\nu}^{(k)}}\big(I_{T_n} \neq k\big) \geqslant -\frac{10}{H^\star} \frac{1}{H\big(\underline{\nu}^{(k)}\big)}. \tag{53}$$

Then, in a second step, we set particular values of the $(p_k)_{2\leqslant k\leqslant K}$ and show that

$$H^\star \geqslant \frac{3}{10} \ln K. \tag{54}$$

Step 1: lower bound (53) for general values of $(p_k)_k$. We observe that, for each $T \in \mathbb{N}^*$,

$$\sum_{k=2}^{K} \frac{\mathbb{E}_{\underline{\nu}^{(1)}}\big[N_k(T)\big]}{T} \leqslant 1 = \frac{1}{H^\star} \sum_{k=2}^{K} \frac{1}{\Delta_k^2 H\big(\underline{\nu}^{(k)}\big)} = \sum_{k=2}^{K} \frac{1}{H^\star \Delta_k^2 H\big(\underline{\nu}^{(k)}\big)},$$

hence there exists an arm $k_T \in \{2,\dots,K\}$ such that

$$\frac{\mathbb{E}_{\underline{\nu}^{(1)}}\big[N_{k_T}(T)\big]}{T} \leqslant \frac{1}{H^\star \Delta_{k_T}^2 H\big(\underline{\nu}^{(k_T)}\big)}. \tag{55}$$

As there is a finite number of arms, by the pigeonhole principle, there exists an increasing sequence $(T_n)_{n\in\mathbb{N}}$ of budgets such that all arms $(k_{T_n})_{n\in\mathbb{N}}$ are the same. Letting k denotes this arm, (55) gives

$$\limsup_{n\to+\infty} \frac{\mathbb{E}_{\underline{\nu}^{(1)}}\big[N_k(T_n)\big]}{T_n} \leqslant \frac{1}{H^\star \Delta_k^2 H\big(\underline{\nu}^{(k)}\big)}\,,$$

and applying a slightly modified version of Lemma 6 (where we only consider the sequence of budgets $(T_n)_{n\in\mathbb{N}}$) with bandit problems $\underline{\nu}^{(k)}$ and $\underline{\nu}^{(1)}$ respectively, we obtain

$$\begin{aligned}
\liminf_{n\to+\infty} \frac{1}{T_n} \ln \mathbb{P}_{\underline{\nu}^{(k)}}\big(I_{T_n} \neq k\big) &\geqslant -\limsup_{n\to+\infty} \frac{\mathbb{E}_{\underline{\nu}^{(1)}}\big[N_k(T_n)\big]}{T_n} \times \mathrm{KL}\big(\mathrm{Ber}(1-p_k), \mathrm{Ber}(p_k)\big) \\
&\geqslant -\frac{\mathrm{KL}\big(\mathrm{Ber}(1-p_k), \mathrm{Ber}(p_k)\big)}{H^\star \Delta_k^2 H\big(\underline{\nu}^{(k)}\big)} \\
&\geqslant -\frac{10}{H^\star} \frac{1}{H\big(\underline{\nu}^{(k)}\big)}\,,
\end{aligned}$$

where, in the last inequality, we used that for all $x \in [1/4, 1/2)$,

$$\mathrm{KL}\big(\mathrm{Ber}(1-x), \mathrm{Ber}(x)\big) \leqslant 10\Big(\frac{1}{2} - x\Big)^2\,,$$

which can be checked analytically. This concludes the proof of (53).

Step 2: control (54) of $H^\star$ for specific values of $(p_k)_{2\leqslant k\leqslant K}$. We proceed as Carpentier and Locatelli (2016). We set, for $k \in \{2, \dots, K\}$,

$$p_k = \frac{1}{2} - \frac{1}{4}\frac{k}{K} \qquad \text{or, equivalently,} \qquad \Delta_k = \frac{1}{4}\frac{k}{K}\,,$$

and show first that $\Delta_k^2 H\big(\nu^{(k)}\big) \leqslant 2k$. Indeed, by (51):

$$\Delta_k^2 H\big(\nu^{(k)}\big) = \Delta_k^2 \sum_{a\neq k} \frac{1}{\big(\Delta_k + \Delta_a\big)^2} = \sum_{a<k} \frac{\Delta_k^2}{\big(\Delta_k + \Delta_a\big)^2} + \sum_{a>k} \frac{\Delta_k^2}{\big(\Delta_k + \Delta_a\big)^2}\,,$$

and, lower bounding $\Delta_k + \Delta_a$ by Δ_k in the first sum, and by Δ_a in the second, we obtain

$$\Delta_k^2 H\big(\nu^{(k)}\big) \leqslant k - 1 + \sum_{a>k} \Big(\frac{\Delta_k}{\Delta_a}\Big)^2 = k - 1 + \sum_{a>k} \frac{k^2}{a^2} = k - 1 + k^2\Big(\frac{1}{k} - \frac{1}{K}\Big) \leqslant 2k\,.$$

We finally get (54) by plugging this inequality into the definition (52) of $H^\star$:

$$H^\star = \sum_{k=2}^{K} \frac{1}{\Delta_k^2 H\big(\underline{\nu}^{(k)}\big)} \geqslant \sum_{k=2}^{K} \frac{1}{2k} = \frac{1}{2}\big(\ln(K+1) - \ln 2\big) \geqslant \frac{3}{10}\ln(K)\,,$$

the latter inequality being easily verified for $K \geqslant 3$. ∎

E.2. The Bretagnolle-Huber technique

An alternative (non-asymptotic) method to obtain lower bounds consists in using, together with the data-processing inequality and the chain rule for the Kullback-Leibler divergence, the Bretagnolle-Huber inequality (Bretagnolle and Huber, 1979), which states that, for $p, q \in [0,1]$,

$$p + 1 - q \geqslant \frac{1}{2} \exp\Big(-\mathrm{KL}\big(\mathrm{Ber}(p), \mathrm{Ber}(q)\big)\Big) . \tag{56}$$

Lower bound of Kaufman et al. (2016). The method was used by Kaufman et al. (2016) on the model $\mathcal{D}_{\sigma^2}$ of Gaussian distributions with a fixed variance $\sigma^2 > 0$. We state here an asymptotic version of their result, that will be generalized to all models, possibly non-parametric, in Proposition 24. The statement relies on a measure of complexity

$$C(\underline{\nu}) = \sum_{a \neq a^\star(\underline{\nu})} \frac{2\sigma^2}{\Delta_a^2} .$$

It reads: for all strategies and for all bandit problems $\underline{\nu}$ in $\mathcal{D}_{\sigma^2}$ with a unique optimal arm, there exists a set of alternative bandit instances $(\underline{\nu}^{(k)})_{k \neq a^\star(\underline{\nu})}$ in $\mathcal{D}_{\sigma^2}$ such that the best arm of $\underline{\nu}^{(k)}$ is arm k and, denoting $\underline{\nu}^{a^\star(\underline{\nu})} = \underline{\nu}$,

$$\liminf_{T \to +\infty} \frac{1}{T} \ln\left(\max_{1 \leqslant k \leqslant K} \Big\{ \mathbb{P}_{\underline{\nu}^{(k)}}\big(I_T \neq k\big) \Big\} \right) \geqslant -4\, C(\underline{\nu})^{-1} \geqslant -\frac{2}{\sigma^2} \min_{2 \leqslant k \leqslant K} \frac{\Delta_{(k)}^2}{k} , \tag{57}$$

$$\text{and} \qquad \forall k \neq a^\star(\underline{\nu}), \qquad C\big(\underline{\nu}^{(k)}\big) \leqslant C(\underline{\nu}) .$$

(The adaptation to a bound for a sequence of strategies comes at a small cost: the original bound proposed by Kaufman et al. (2016) for a given budget T only involved two bandit problems, $\underline{\nu}$ and a single alternative problem $\underline{\lambda}$.)

Generalization to possibly non-parametric models. The Bretagnolle-Huber methodology readily extends to general, possibly non-parametric, models. It leads to the following bound. However, the issue is the lack of interpretability of that bound, as we discuss after the statement of the proposition.

Proposition 24 *Fix $K \geqslant 2$, a model $\mathcal{D}$, and a sequence of strategies. Let $\underline{\nu}$ be a bandit problem in $\mathcal{D}$ with a unique optimal arm. Consider, for each $k \neq a^\star(\underline{\nu})$, a distribution $\zeta_k \in \mathcal{D}$ such that $\mathrm{E}(\zeta_k) > \mu^\star$. Denoting by $\underline{\nu}^{(k)}$, for $k \neq a^\star(\underline{\nu})$, the bandit problem obtained from $\underline{\nu}$ by changing arm k to distribution ζ_k, and by $\underline{\nu}^{(a^\star(\underline{\nu}))}$ the original bandit problem $\underline{\nu}$, we have*

$$\liminf_{T \to +\infty} \frac{1}{T} \ln\left(\max_{1 \leqslant k \leqslant K} \Big\{ \mathbb{P}_{\underline{\nu}^{(k)}}\big(I_T \neq k\big) \Big\} \right) \geqslant -\left(\sum_{a \neq a^\star(\underline{\nu})} \frac{1}{\mathrm{KL}(\nu_a, \zeta_a)} \right)^{-1} .$$

Before proving this result, we provide a few comments concerning the (lack of) interpretability of this result and explain how to derive the bound (57) for Gaussian models.

Lack of interpretability of the bound for general models. To derive an interesting and interpretable bound from this result, one needs to choose carefully the distributions ζ_k: there is a tradeoff between obtaining the best possible bound by choosing ζ_k close to ν_k in terms of Kullback-Leibler

divergences, and controlling the maximum of the misidentification probabilities. In particular, when $\mathrm{E}(\zeta_k)$ is close to $\mu^\star$, the probability of misidentification under $\underline{\nu}^{(k)}$ will be larger, so that

$$\max_{1\leqslant k\leqslant K}\Big\{\mathbb{P}_{\underline{\nu}^{(k)}}\big(I_T\neq k\big)\Big\}\gg\mathbb{P}_{\underline{\nu}}\big(I_T\neq a^\star(\underline{\nu})\big)\,.$$

That is, the bound will become uninformative on the targeted quantity $\mathbb{P}_{\underline{\nu}}\big(I_T\neq a^\star(\underline{\nu})\big)$. This trade-off seems to be unsolvable in general, unless there exist some specific properties for the Kullback-Leibler divergence of the model, as we illustrate now.

The case of Gaussian models $\mathcal{D}_{\sigma^2}$. For the model $\mathcal{D}_{\sigma^2}$, we obtain (57) and the desired inequalities on complexities by considering $\zeta_k=\mathcal{N}(\mu^\star+\Delta_k,\sigma^2)$, the Gaussian distribution of mean $\mu^\star+\Delta_k$ and variance σ^2. Indeed, we recall that

$$\forall\nu,\zeta\in\mathcal{D}_{\sigma^2},\qquad \mathrm{KL}(\zeta,\nu)=\frac{\big(\mathrm{E}(\nu)-\mathrm{E}(\zeta)\big)^2}{2\sigma^2}=\mathrm{KL}(\nu,\zeta)\,.$$

Note a key property (actually stronger than symmetry), which will turn out to be useful in the calculations below: the Kullback-Leibler divergence only depends on the expectation gaps between the two distributions. Now, with the distributions ζ_k defined above, on the one hand, the bound of Proposition 24 rewrites

$$\sum_{a\neq a^\star(\underline{\nu})}\frac{1}{\mathrm{KL}(\nu_a,\zeta_a)}=\sum_{a\neq a^\star(\underline{\nu})}\frac{2\sigma^2}{\big(\underbrace{\mathrm{E}(\nu_a)}_{\mu^\star-\Delta_a}-\underbrace{\mathrm{E}(\zeta_a)}_{\mu^\star+\Delta_a}\big)^2}=\frac{C(\underline{\nu})}{4}\,,$$

and on the other hand, for $k\neq a^\star(\underline{\nu})$, as the best arm of $\underline{\nu}^{(k)}$ is k, with associated expectation $\mu^\star+\Delta_k$,

$$\begin{aligned}C\big(\underline{\nu}^{(k)}\big)=\sum_{a\neq k}\frac{2\sigma^2}{(\mu^\star+\Delta_k-\mu_a)^2}&=\frac{2\sigma^2}{\Delta_k^2}+\sum_{a\notin\{k,a^\star(\underline{\nu})\}}\frac{2\sigma^2}{(\mu^\star+\Delta_k-\mu_a)^2}\\&\leqslant\frac{2\sigma^2}{\Delta_k^2}+\sum_{a\notin\{k,a^\star(\underline{\nu})\}}\frac{2\sigma^2}{(\mu^\star-\mu_a)^2}=\sum_{a\neq a^\star(\underline{\nu})}\frac{2\sigma^2}{\Delta_a^2}=C(\underline{\nu})\,.\end{aligned}$$

As underlined above, the calculations led are highly specific to the Gaussian model and exploit the gap-based rewriting of the Kullback-Leibler divergence. They would only extend to models for which gap-based rewritings of (or upper and lower bounds on) the Kullback-Leibler divergence would be available.

Reverse order of the arguments in the KL. We observe that the bound of Proposition 24 involves Kullback-Leibler divergences with arguments in reverse order compared to the lower bounds presented in Section 4. Indeed, taking the infimum of the lower bound over distributions ζ_k such that $\mathrm{E}(\zeta_k)>\mu^\star$ would lead to a complexity in terms of the $\mathcal{K}_{\inf}^{>}\big(\nu_k,\mu^\star\big)$, where

$$\mathcal{K}_{\inf}^{>}(\nu,x)\overset{\text{def}}{=}\inf\big\{\mathrm{KL}(\nu,\zeta):\zeta\in\mathcal{D}\text{ s.t. }\mathrm{E}(\zeta)>x\big\}\,,$$

rather than in terms of the $\mathcal{L}_{\inf}^{>}\big(\mu^\star,\nu_k\big)$. Given all bounds presented in this article, it does not seem that this would be the correct notion of complexity for the fixed-budget best-arm identification.

Proof of Proposition (24). We will prove that for all $(u_b)_{b\neq a^\star(\underline{\nu})} \in \Sigma_{K-1}$, where Σ_{K-1} is the simplex of $\mathbb{R}^{K-1}$, we get

$$\liminf_{T\to+\infty} \frac{1}{T} \ln\left(\max_{1\leqslant k\leqslant K} \left\{ \mathbb{P}_{\underline{\nu}^{(k)}}\left(I_T \neq k\right) \right\} \right) \geqslant - \max_{b\neq a^\star(\underline{\nu})} \left\{ u_b \mathrm{KL}(\nu_b, \zeta_b) \right\}. \tag{58}$$

The result will follow by choosing $(u_b)_{b\neq a^\star(\underline{\nu})}$ so as to maximize this lower bound, that is taking

$$u_b = \left(\sum_{a\neq a^\star(\underline{\nu})} \frac{1}{\mathrm{KL}(\nu_a, \zeta_a)} \right)^{-1} \times \frac{1}{\mathrm{KL}(\nu_b, \zeta_b)}\,.$$

We now fix $(u_b)_{b\neq a^\star(\underline{\nu})} \in \Sigma_{K-1}$ and prove (58). Consider a given budget T and let $b \neq a^\star(\underline{\nu})$. We get, as $a^\star(\underline{\nu}) \neq b$ first, and by the Bretagnolle-Huber inequality (56) then,

$$\begin{aligned}\mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) + \mathbb{P}_{\underline{\nu}^{(b)}}\left(I_T \neq b\right) &\geqslant \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) + \mathbb{P}_{\underline{\nu}^{(b)}}\left(I_T = a^\star(\underline{\nu})\right) \\ &\geqslant \frac{1}{2}\exp\Big(-\mathrm{KL}\big(\mathrm{Ber}(p_T), \mathrm{Ber}(q_T)\big)\Big)\,,\end{aligned}$$

where $p_T \overset{\text{def}}{=} \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big)$ and $q_T \overset{\text{def}}{=} \mathbb{P}_{\underline{\nu}^{(b)}}\left(I_T \neq a^\star(\underline{\nu})\right)$. Applying the combination (43) of the data-compressing inequality and the chain rule for Kullback-Leibler divergences leads to, as $\underline{\nu}$ and $\underline{\nu}^{(b)}$ only differ at arm b,

$$\mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) + \mathbb{P}_{\underline{\nu}^{(b)}}\left(I_T \neq b\right) \geqslant \frac{1}{2}\exp\Big(-\mathbb{E}_{\underline{\nu}}\big[N_b(T)\big]\mathrm{KL}(\nu_b, \zeta_b)\Big)\,.$$

As $\max(u,v) \geqslant (u+v)/2$, we obtained so far

$$\begin{aligned}\max_{1\leqslant k\leqslant K} \left\{ \mathbb{P}_{\underline{\nu}^{(k)}}\left(I_T \neq k\right) \right\} &\geqslant \max\left\{ \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big), \mathbb{P}_{\underline{\nu}^{(b)}}\left(I_T \neq b\right) \right\} \\ &\geqslant \frac{1}{4}\exp\Big(-\mathbb{E}_{\underline{\nu}}\big[N_b(T)\big]\mathrm{KL}(\nu_b, \zeta_b)\Big)\,. \end{aligned} \tag{59}$$

This bound holds for any $b \neq a^\star(\underline{\nu})$. In particular, as

$$\sum_{b\neq a^\star(\underline{\nu})} \frac{\mathbb{E}_{\underline{\nu}}\big[N_b(T)\big]}{T} \leqslant 1 = \sum_{b\neq a^\star(\underline{\nu})} u_b\,,$$

we know that there exists $b^\star \neq a^\star(\underline{\nu})$ such that $\mathbb{E}_{\underline{\nu}}\big[N_{b^\star}(T)\big] \leqslant T u_{b^\star}$, and, applying (59) with $b^\star$,

$$\max_{1\leqslant k\leqslant K} \left\{ \mathbb{P}_{\underline{\nu}^{(k)}}\left(I_T \neq k\right) \right\} \geqslant \frac{1}{4}\exp\Big(-T u_{b^\star}\mathrm{KL}(\nu_{b^\star}, \zeta_{b^\star})\Big) \geqslant \frac{1}{4}\exp\Big(-T \max_{b\neq a^\star(\underline{\nu})} \left\{ u_b \mathrm{KL}(\nu_b, \zeta_b) \right\}\Big)\,,$$

or, to put it differently,

$$\frac{1}{T}\ln\left(\max_{1\leqslant k\leqslant K} \left\{ \mathbb{P}_{\underline{\nu}^{(k)}}\left(I_T \neq k\right) \right\} \right) \geqslant \frac{1}{T}\ln\frac{1}{4} - \max_{b\neq a^\star(\underline{\nu})} \left\{ u_b \mathrm{KL}(\nu_b, \zeta_b) \right\}.$$

We obtain (58) by taking the $\liminf$ in T on that inequality. ■