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

Antoine Barrier, Aurélien Garivier, Gilles Stoltz

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

Antoine Barrier ANTOINE.BARRIER@ENS-LYON.FR
ENS de Lyon, UMPA UMR 5669, 46 allée d'Italie, 69364 Lyon Cedex 07, France
Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France

Aurélien Garivier AURELIEN.GARIVIER@ENS-LYON.FR
ENS de Lyon, UMPA UMR 5669 et LIP UMR 5668, 46 allée d'Italie, 69364 Lyon Cedex 07, France

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

Editors: Shipra Agrawal and Francesco Orabona

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 (the expected sample complexity). 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). The first generalization to non-parametric models in this fixed-confidence setting was achieved by Jourdan et al. (2022), who worked in a concurrent and independent manner from us. Their upper and lower bounds differ by a multiplicative factor of 2 (only).

Best-arm identification with a fixed budget. The *fixed-budget setting* is much less understood in our opinion. 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 (see, for instance, Audibert et al., 2010 and Lattimore and Szepesvári, 2020, Chapter 33) 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 (see, e.g., Lattimore and Szepesvári, 2020, Chapter 33 and Jourdan et al., 2022, Section 6), 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.

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), in a minimax sense. *A second series of approaches* (see, e.g., Kaufmann et al., 2016) focused on Gaussian bandits with fixed variances, but their results do not seem to be easily generalized to other models as they rely on specific properties (even stronger than the symmetry of the Kullback-Leibler divergence, namely, that in this model, the Kullback-Leibler divergence only depends on the gap between the expectations of the distributions). *A third approach*, led by Russo (2016, 2020), considered canonical one-parameter exponential families, but 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., Kaufmann et al. (2016). The non-parametric bounds obtained therein do not enjoy any obvious generalization to the case of $K \geqslant 3$ arms beyond the one stated in Theorem 14 and criticized in Section 2.3 for only involving pairwise comparisons with the

best arm. 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 instance-dependent upper and lower bounds, holding for all problems of general 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 (new) information-theoretic quantities $\mathcal{L}^{<}_{\inf}$ and $\mathcal{L}^{>}_{\inf}$ are at stake in these upper and lower bounds. These information-theoretic quantities 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 present in this section an overview of our results, which we carefully compare to existing bounds (restated therein, occasionally with some improvements). We state upper bounds in Section 3 and to do so, we provide an improved analysis of the classical successive-rejects strategy, 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.

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

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

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

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.

Remark 1 *The key quantities for the non-parametric study of best-arm identification with fixed confidence by Jourdan et al. (2022) are defined based on Kullback-Leibler divergences with arguments in reverse order, namely,*

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

where the first quantity was referred to as simply $\mathcal{K}_{\inf}(\nu, x)$ by Honda and Takemura (2015) in the regret-minimization literature (see also Appendix C and Garivier et al., 2022). Optimal bounds for regret minimization only depend on $\mathcal{K}_{\inf}(\nu, x)$.

For best-arm identification with fixed budget, the arguments in the KL are in reverse order compared to the fixed-confidence setting. 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 Kaufmann et al. (2016) for the model $\mathcal{D}_{\sigma^2}$, based on the Bretagnolle-Huber inequality (Bretagnolle and Huber, 1979); 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 on a property even stronger than the symmetry of KL for this model: details and discussions on this matter are provided in Appendix E.2. In particular, generalizations beyond the Gaussian case appear to be infeasible.

2.2. Overview of the results

The paper provides new and more general (possibly non-parametric) bounds on the misidentification errors based on the information-theoretic quantities introduced above. In particular, we consider a version of Chernoff information defined, for ν, ν' in $\mathcal{D}$ with $\mathrm{E}(\nu') < \mathrm{E}(\nu)$, as

$$\mathcal{L}(\nu', \nu) = \inf_{x \in [\mathrm{E}(\nu'), \mathrm{E}(\nu)]} \Big\{ \mathcal{L}^{\geqslant}_{\inf}(x, \nu') + \mathcal{L}^{\leqslant}_{\inf}(x, \nu) \Big\}. \tag{1}$$

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

$$0 = \mathcal{L}\big(\nu_{\sigma_1}, \nu^\star\big) < \mathcal{L}\big(\nu_{\sigma_2}, \nu^\star\big) \leqslant \ldots \leqslant \mathcal{L}\big(\nu_{\sigma_{K-1}}, \nu^\star\big) \leqslant \mathcal{L}\big(\nu_{\sigma_K}, \nu^\star\big). \tag{2}$$

Our first main result (Corollary 4 together with Lemma 5) 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. We study 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 state that this strategy is such that for all bandit problems $\underline{\nu}$ in $\mathcal{D}$ with a unique optimal arm,

$$\limsup_{T \to +\infty} \frac{1}{T} \ln \mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant -\frac{1}{\overline{\ln} K} \min_{2 \leqslant k \leqslant K} \frac{\mathcal{L}\big(\nu_{\sigma_k}, \nu^\star\big)}{k}, \tag{3}$$

where $\overline{\ln}\, K$ is defined in (16) and is of order $\ln K$. The key for this result (Lemma 2, 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: for instance, Definition 8 restricts 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. Out of all lower bounds exhibited, *our second main result* (Theorem 13) holds, as indicated, under mild assumptions on the model and sequences of strategies considered, and reads: for all bandit problems $\underline{\nu}$ with no two same expectations,

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \min_{2\leqslant k\leqslant K} \inf_{x\in[\mu_{(k)},\mu_{(k-1)})} \left\{ \frac{\mathcal{L}^{>}_{\inf}\big(x, \nu_{(k)}\big)}{k-1} + \frac{\mathcal{L}^{<}_{\inf}\big(x, \nu^\star\big)}{k} \right\}, \tag{4}$$

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

This lower bound does not match the exhibited upper bound, as is further discussed in Section 2.4. 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.*

2.3. Re-derivation of existing bounds

We now survey the most important existing bounds and re-derive them from our general bounds. These existing bounds all hold only for sub-Gaussian models and for exponential models when $K \geqslant 3$, while a non-parametric bound was only available in the case of $K = 2$ arms.

To do so, we will sometimes consider the following weaker version of the lower bound (4), obtained by picking $x = \mu_{(k)}$:

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

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,

$$\limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant -\frac{1}{\overline{\ln}\, K} \min_{2\leqslant k\leqslant K} \frac{\Delta^2_{(k)}}{k}\,, \tag{6}$$

where we recall the definition of the gaps $\Delta_{(k)} = \mu^\star - \mu_{(k)}$. This bound is a consequence of (Corollary 4, a slightly more general form of) the bound (3), given Pinsker's inequality (22):

$$\mathcal{L}\big(\nu_{(k)}, \nu^\star\big) \geqslant \inf_{x\in[\mu_{(k)},\mu^\star]} \Big\{ 2\big(x-\mu_{(k)}\big)^2 + 2(x-\mu^\star)^2 \Big\} = \big(\mu^\star - \mu_{(k)}\big)^2 = \Delta^2_{(k)}\,. \tag{7}$$

We remark that the bound (6) 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 considering factors $1/(4\sigma^2)$. 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 $\mathcal{B}_{[p,\,1-p]} = \big\{\mathrm{Ber}(x) : x \in [p, 1-p]\big\}$ 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 families: on top of the restricted Bernoulli model discussed above, for which

$$C_{\mathcal{B}_{[p,\,1-p]}} = 1/\big(2p(1-p)\big)\,,$$

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 $\mathcal{D}$,

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -5\, C_{\mathcal{D}} \min_{2\leqslant k\leqslant K} \frac{\Delta_{(k)}^2}{k}\,. \tag{8}$$

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 (5) implies the stated lower bound (8), with an improved constant factor.

The lower bound (8) and the upper bound (6) differ in particular by a factor proportional to $\overline{\ln} K$. Carpentier and Locatelli (2016) discuss this gap in the case of the Bernoulli model $\mathcal{B}_{[1/4,\,3/4]}$ and improve the lower bound (8) 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 Kaufmann et al. (2016). It turns out that the existing literature for the fixed-budget setting offered so far a non-parametric bound, in the case of $K = 2$ arms. Namely, in a general, possibly non-parametric model $\mathcal{D}$, Kaufmann et al. (2016, Theorem 12) stated a lower bound for all 2–armed bandit problems $\underline{\nu} = (\nu_1, \nu_2)$:

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant - \inf_{\substack{\underline{\lambda} \text{ in } \mathcal{D}\,:\\ \mathrm{E}(\lambda_{a^\star(\underline{\nu})}) < \mathrm{E}(\lambda_{w_\star(\underline{\nu})})}} \max\Big\{\mathrm{KL}\big(\lambda_{w_\star(\underline{\nu})}, \nu_{w_\star(\underline{\nu})}\big),\, \mathrm{KL}\big(\lambda_{a^\star(\underline{\nu})}, \nu_{a^\star(\underline{\nu})}\big)\Big\}\,, \tag{9}$$

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 reverse order on the expectations compared to $\underline{\nu}$. We note (see the proof of Theorem 14) 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:

$$\inf_{\substack{\underline{\lambda} \text{ in } \mathcal{D}:\\ \mathrm{E}(\lambda_{a^\star(\underline{\nu})})<\mathrm{E}(\lambda_{w_\star(\underline{\nu})})}} \max\Big\{\mathrm{KL}\big(\lambda_{w_\star(\underline{\nu})},\nu_{w_\star(\underline{\nu})}\big),\ \mathrm{KL}\big(\lambda_{a^\star(\underline{\nu})},\nu_{a^\star(\underline{\nu})}\big)\Big\} = \inf_{x\in[\mu_{w_\star(\underline{\nu})},\mu^\star]}\Big\{\max\Big\{\mathcal{L}^{>}_{\inf}\big(x,\nu_{w_\star(\underline{\nu})}\big),\ \mathcal{L}^{<}_{\inf}\big(x,\nu^\star\big)\Big\}\Big\}. \quad (10)$$

The proof technique of Kaufmann et al. (2016) may be applied in a pairwise fashion to generalize the lower bound (10) for 2 arms into a lower bound for $K \geqslant 2$ arms, stated in Theorem 14: for all $\underline{\nu}$ in $\mathcal{D}$ with a unique optimal arm,

$$\liminf_{T\to+\infty}\frac{1}{T}\ln\mathbb{P}_{\underline{\nu}}\big(I_T\neq a^\star(\underline{\nu})\big) \geqslant -\min_{k\neq a^\star(\underline{\nu})}\inf_{x\in[\mu_k,\mu^\star]}\Big\{\max\big\{\mathcal{L}^{>}_{\inf}(x,\nu_k),\ \mathcal{L}^{<}_{\inf}(x,\nu^\star)\big\}\Big\}. \quad (11)$$

We however do not claim that (11) 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 (4). This is why we had not stated the result (11) of Theorem 14 in Section 2.2 and mention it only here.

That being said, given that the infima in (4) are over more restricted ranges than in (11), we can see no obvious ranking between the two bounds, which rather look incomparable.

Bounds for $K = 2$ arms and exponential families, cf. comments after Theorem 12 of Kaufmann et al. (2016). We denote by $\mathcal{D}_{\exp}$ 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 $\mathcal{D}_{\exp}$ and for all $x \in \mathcal{M}$,

$$\forall x \leqslant \mathrm{E}(\nu), \qquad \mathcal{L}^{<}_{\inf}(x,\nu) = \mathcal{L}^{\leqslant}_{\inf}(x,\nu) = d\big(x,\mathrm{E}(\nu)\big)\,, \quad (12)$$

$$\text{and} \qquad \forall x \geqslant \mathrm{E}(\nu), \qquad \mathcal{L}^{>}_{\inf}(x,\nu) = \mathcal{L}^{\geqslant}_{\inf}(x,\nu) = d\big(x,\mathrm{E}(\nu)\big)\,. \quad (13)$$

Note that all bounds stated in Section 2.2 then admit simple reformulations in terms of d. The Chernoff-information-type quantity $\mathcal{L}$ introduced in (1) may also be mean-parameterized as follows: for $\mu' < \mu$,

$$L(\mu',\mu) = \min_{x\in[\mu',\mu]}\big\{d(x,\mu') + d(x,\mu)\big\}\,. \quad (14)$$

We now explain why we called L (and therefore $\mathcal{L}$) a version of Chernoff information. The original definition of the Chernoff information $D(\mu',\mu)$ is the value $d(y,\mu)$ for $y \in [\mu',\mu]$ such that $d(y,\mu') = d(y,\mu)$. As mentioned in the comments after Theorem 12 of Kaufmann et al. (2016), D is the quantity at stake in (10) for a canonical one-parameter exponential family: given that $d(\,\cdot\,,\mu')$ and $d(\,\cdot\,,\mu)$ are respectively increasing and decreasing on $[\mu',\mu]$,

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

Therefore, $D(\mu',\mu) \leqslant L(\mu',\mu) \leqslant 2\,D(\mu',\mu)$, which shows that L is related to D, as claimed.

Example 1 *We state the lower bound (5) and the upper bound (3) for the model $\mathcal{B}_{[p,\,1-p]}$ of Bernoulli distributions with parameters in $[p,1-p]$, where $p \in (0,1/2)$. We denote by*

$$\mathrm{kl}(x,y) = x\ln\frac{x}{y} + (1-x)\ln\frac{1-x}{1-y}\,, \qquad \textit{where} \qquad x,y\in[p,1-p]$$

the mean-parameterized Kullback-Leibler divergence of this model. We consider a generic bandit problem $\underline{\nu} = \big(\mathrm{Ber}(p_1), \dots, \mathrm{Ber}(p_K)\big)$. *We rank the parameters as in* (4)*, i.e., introduce the notation* $p^\star = p_{(1)} > p_{(2)} > \ldots > p_{(K)}$. *Then, after noticing (see Lemma 21 in Appendix C.3) that this ranking is the same as the one considered in* (2)*, the upper bound* (3) *rewrites as*

$$\limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant -\frac{1}{\overline{\ln} K} \min_{2\leqslant k\leqslant K} \frac{\min\limits_{x\in[p_{(k)},p^\star]} \Big\{\mathrm{kl}\big(x,p_{(k)}\big) + \mathrm{kl}(x,p^\star)\Big\}}{k},$$

while the lower bound (5) *rewrites as*

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -\min_{2\leqslant k\leqslant K} \frac{\mathrm{kl}\big(p_{(k)},p^\star\big)}{k}.$$

They should be compared to the upper (6) *and lower* (8) *bounds of Audibert et al. (2010), respectively.*

2.4. Discussion of the (lack of) optimality of the new bounds exhibited

The lower bound (4) does not match the upper bound (3) because of two aspects. First, the infima in (4) are only taken on restricted ranges $[\mu_{(k)}, \mu_{(k-1)})$ and not on the entire intervals $[\mu_{(k)}, \mu^\star]$ as in (3). Second, the upper bound (3) involves a $1/\overline{\ln} K$ factor, while the lower bound (4) does not. A similar $1/\overline{\ln} K$ factor was missing between the upper (6) and lower (8) bounds of Audibert et al. (2010) for Bernoulli models, together with a numerical factor of $5\, C_{\mathcal{B}_{[p,\,1-p]}}$. The non-parametric bounds exhibited in this article mainly generalize and extend the known parametric bounds but do not refine the latter in the sense that gaps between upper and lower bounds would be closed.

That being said, we would like to illustrate below on one specific example to which extent the gap-based bounds can be looser.

Example of an extreme improvement: distributions with separated supports. For general non-parametric models, gaps are not enough at all to measure complexity as we may well have a finite gap between two distributions ν_1 and ν_2 with $\mu_1 > \mu_2$, but $\mathcal{L}(\nu_2, \nu_1) = +\infty$. This holds, for instance, as soon as ν_1 and ν_2 have closed supports separated by a threshold x_0, i.e., the closed supports of ν_1 and ν_2 are included in $(-\infty, x_0)$ and $(x_0, +\infty)$, respectively. Indeed, by mimicking the beginning of the proof of Lemma 16 of Appendix A.2, it may be seen that $\mathcal{L}^{\leqslant}_{\mathrm{inf}}(x, \nu_1) = +\infty$ for $x \leqslant x_0$ and $\mathcal{L}^{\geqslant}_{\mathrm{inf}}(x, \nu_2) = +\infty$ if $x \geqslant x_0$, so that in all cases, the sum $\mathcal{L}^{\geqslant}_{\mathrm{inf}}(x, \nu_2) + \mathcal{L}^{\leqslant}_{\mathrm{inf}}(x, \nu_1)$ equals $+\infty$, and thus, $\mathcal{L}(\nu_2, \nu_1) = +\infty$. In our bounds, e.g., the upper bound (3), the pair of distributions ν_1, ν_2 will therefore not contribute—as intuition commands: these two distributions are easy to distinguish—, while it does contribute in the earlier gap-based bounds.

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. The strategy works in phases, and the lengths of the phases are set beforehand; they are denoted by $\ell_1, \dots, \ell_{K-1} \geqslant 1$ and satisfy $\ell_1 + \ldots + \ell_{K-1} = T$. The strategy maintains a list of candidate arms, starting with all arms, i.e., $S_0 = \{1, \dots, K\}$. At the end of each phase $r \in \{1, \dots, K-1\}$, it drops an arm to get S_r, while during phase r, it operates with the $K - r + 1$ arms in S_{r-1}.

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 phase $r \in \{1, \dots, K-1\}$**:**

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 phase and in the previous phases
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 $S_{K-1} = \{I_T\}$

More precisely, during pahse $r \in \{1, \dots, K-1\}$, 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 phase 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 $\phi_\nu^\star$:

$$\forall \lambda \in \mathbb{R}, \quad \phi_\nu(\lambda) = \ln \int_{\mathbb{R}} \mathrm{e}^{\lambda x} \, \mathrm{d}\nu(x) \qquad \text{and} \qquad \forall x \in \mathbb{R}, \quad \phi_\nu^\star(x) = \sup_{\lambda \in \mathbb{R}} \big\{ \lambda x - \phi_\nu(\lambda) \big\} . \quad (15)$$

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 reverse order compared to the expectations of the underlying distributions. It supersedes the use of Hoeffding's inequality in Audibert et al. (2010).

Lemma 2 *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),

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

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 $\mathcal{L}^{\leqslant}_{\inf}$ and $\mathcal{L}^{\geqslant}_{\inf}$ quantities, leading to a rewriting $\Phi = \mathcal{L}$ and to the bound claimed in (3).

Theorem 3 *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 2. We have $I_T \neq a^\star(\underline{\nu})$ if and only if $a^\star(\underline{\nu})$ is rejected in some phase, 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 2 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 3 for the phase lengths suggested by Audibert et al. (2010), namely, $\ell_1 = T/\overline{\ln}\,K$ and for $r \in \{2,\dots,K-1\}$,

$$\ell_r = \frac{T}{(K-r+2)\,\overline{\ln}\,K}\,, \qquad \text{where} \qquad \overline{\ln}\,K = \frac{1}{2} + \sum_{k=2}^{K} \frac{1}{k}\,. \tag{16}$$

We also consider lower bounds $f(\nu_k, \nu^\star)$ on the $\Phi(\nu_k, \nu^\star)$. 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

$$\phi_\nu^\star(x) \geqslant 2\big(x - \mathrm{E}(\nu)\big)^2\,, \qquad \text{so that} \qquad \Phi(\nu_k, \nu^\star) \geqslant \Delta_k^2 \stackrel{\text{def}}{=} f(\nu_k, \nu^\star)\,; \tag{17}$$

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 $\sigma_1, \ldots, \sigma_K$ based on f, namely, we let $\sigma_1 = a^\star(\underline{\nu})$ and

$$0 = f(\nu_{\sigma_1}, \nu^\star) < f(\nu_{\sigma_2}, \nu^\star) \leqslant \ldots \leqslant f(\nu_{\sigma_{K-1}}, \nu^\star) \leqslant f(\nu_{\sigma_K}, \nu^\star)\,, \tag{18}$$

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

Corollary 4 *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 phase lengths (16) ensures, that for all bandit problems $\underline{\nu}$ in $\mathcal{D}$ with a unique optimal arm,*

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

where arms were reordered as in (18).

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

The Fenchel-Legendre transform $\phi_\nu^\star$ 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 $\mathcal{L}_{\inf}^{\leqslant}$ and $\mathcal{L}_{\inf}^{\geqslant}$ for the model $\mathcal{P}[0,1]$ of distributions supported on $[0,1]$.

Lemma 5 *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 3 and Corollary 4: $\Phi(\nu', \nu) = \mathcal{L}(\nu', \nu)$ for the model $\mathcal{P}[0,1]$, i.e.,

$$\inf_{x\in[\mathrm{E}(\nu'),\mathrm{E}(\nu)]} \big\{\phi_{\nu'}^\star(x) + \phi_\nu^\star(x)\big\} = \inf_{x\in[\mathrm{E}(\nu'),\mathrm{E}(\nu)]} \big\{\mathcal{L}_{\inf}^{\geqslant}(x, \nu') + \mathcal{L}_{\inf}^{\leqslant}(x, \nu)\big\}\,. \tag{19}$$

For canonical one-parameter exponential models $\mathcal{D}_{\text{exp}}$, a slightly weaker version of Lemma 5, only holding for x corresponding to expectations in $\mathcal{D}_{\text{exp}}$ and provided in Appendix C.3, similarly shows (19), 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. (This is probably a new terminology[1] for referring to bandit problems with no two same expectations for the distributions over the arms.)

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 $t \in \{1, \ldots, T\}$, 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 $H_t = (U_0, Y_1, \ldots, Y_{t-1}, U_{t-1})$, while $A_1 = \psi_1(H_0)$, where $H_0 = U_0$. At round T, the strategy recommends the arm $I_T = \varphi(H_T)$.

Outline of this section. As always in lower-bound results, there is a trade-off between how restrictive are the assumptions on the (doubly-indexed) sequences of strategies, and sometimes on the models, and how large the lower bounds are: the more restrictive the assumptions, the larger the lower bounds. We are interested in assumptions on strategies that are natural in the sense that they should be satisfied by successive-rejects-type strategies. For instance, Theorem 14 comes with the least assumptions but provides a bound where there are no divisions by the ranks k of the arms, which Theorems 10 and 13 do. We may see Theorem 10 as a warm-up result: its main aim is to generalize the lower bound by Audibert et al. (2010) to non-parametric models with a (non-constructive) proof that is only a few-line long. Our preferred result is Theorem 13, which provides the largest lower bound while putting the heaviest (though natural) constraints on the sequences of strategies.

4.1. Common restriction: consistence

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 $\mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big)$ 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 6 *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 $\mathcal{D}$,*

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

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

1. The terminology comes from measure theory: if expectations were drawn at random according to some diffuse distribution, e.g., a uniform distribution over an interval, or a Gaussian distribution, then, almost surely, no two expectations would be equal.

The fundamental inequality. 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 7 *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*

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

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

denotes the number of times arm a was pulled in the T exploration rounds of a given strategy with budget $T \geqslant 1$.

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

The focus of this subsection is to establish the lower bound (5), from which we derived the gap-based lower bound (6) by Audibert et al. (2010). The lower bound (5) 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 8 *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 9 *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 10 *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 7 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. ■

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

In this section, we derive a slightly stronger version of the lower bound (4). 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 8 to provide frequency constraints on each arm $a \in \{1, \dots, K\}$.

Definition 11 *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 11 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 12 *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 13 *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 7, we obtain

$$\liminf_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -\left(\frac{\mathrm{KL}\big(\lambda_{(k)}, \nu_{(k)}\big)}{j-1} + \frac{\mathrm{KL}\big(\lambda_{(1)}, \nu^\star\big)}{j}\right). \tag{20}$$

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 (20) 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. ■

4.4. 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 (9) exhibited by Kaufmann 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 14 *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

Aurélien Garivier and Antoine Barrier acknowledge the support of the Project IDEXLYON of the University of Lyon, in the framework of the Programme Investissements d'Avenir (ANR-16-IDEX-0005), and Chaire SeqALO (ANR-20-CHIA-0020-01). 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.

R.G. Bartle and D.R. Sherbert. *Introduction to Real Analysis*. John Wiley & Sons, 3rd edition, 2000.

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.

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

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.

M. Jourdan, R. Degenne, D. Baudry, R. de Heide, and E. Kaufmann. Top two algorithms revisited. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

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. Kaufmann, 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.

Content of the appendices

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 15 *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

$$\mathcal{L}^{<}_{\inf}\big(\lambda\mu+(1-\lambda)\mu',\lambda\nu+(1-\lambda)\nu'\big)\leqslant\lambda\mathcal{L}^{<}_{\inf}(\mu,\nu)+(1-\lambda)\mathcal{L}^{<}_{\inf}(\mu',\nu')\,. \tag{21}$$

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 (21) follows by letting $\delta\to0$. ∎

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 $\zeta\in\mathcal{P}[0,1]$,

$$\mathrm{KL}(\zeta,\nu)\geqslant\mathrm{KL}\Big(\mathrm{Ber}\big(\mathrm{E}(\zeta)\big),\,\mathrm{Ber}\big(\mathrm{E}(\nu)\big)\Big)\geqslant2\big(\mathrm{E}(\zeta)-\mathrm{E}(\nu)\big)^2\,,$$

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 $\mathrm{E}(\zeta)<x$,

$$\forall x\leqslant\mathrm{E}(\nu),\quad\mathcal{L}^{<}_{\inf}(x,\nu)\geqslant2\big(\mathrm{E}(\nu)-x\big)^2\,. \tag{22}$$

We denote by $m(\nu) = \min\bigl(\mathrm{Supp}(\nu)\bigr) \geqslant 0$ the minimum of the closed support $\mathrm{Supp}(\nu)$ of ν; that is, $m(\nu)$ is the largest value such that $\mathrm{Supp}(\nu) \subseteq \bigl[m(\nu),\, 1\bigr]$. 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 $M(\nu) = \max\bigl(\mathrm{Supp}(\nu)\bigr) \leqslant 1$; by symmetry, it will be considered when studying $\mathcal{L}^{>}_{\mathrm{inf}}$ and $\mathcal{L}^{\geqslant}_{\mathrm{inf}}$ instead of $\mathcal{L}^{<}_{\mathrm{inf}}$ and $\mathcal{L}^{\leqslant}_{\mathrm{inf}}$.

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

Lemma 16 *We consider the model $\mathcal{D} = \mathcal{P}[0,1]$. The function $\mathcal{L}^{<}_{\mathrm{inf}}(\,\cdot\,,\nu)$ is continuous on the interval $\bigl(m(\nu),\,+\infty\bigr)$. We also have, on the one hand,*

$$\forall \mu \neq m(\nu), \qquad \mathcal{L}^{<}_{\mathrm{inf}}(\mu,\nu) = \mathcal{L}^{\leqslant}_{\mathrm{inf}}(\mu,\nu)\,, \tag{23}$$

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

$$\ln \frac{1}{\nu\bigl\{m(\nu)\bigr\}} = \mathcal{L}^{\leqslant}_{\mathrm{inf}}\bigl(m(\nu),\nu\bigr) \leqslant \mathcal{L}^{<}_{\mathrm{inf}}\bigl(m(\nu),\nu\bigr) = +\infty\,. \tag{24}$$

Analogous results hold for $\mathcal{L}^{>}_{\mathrm{inf}}(\,\cdot\,,\nu)$, $\mathcal{L}^{\geqslant}_{\mathrm{inf}}(\,\cdot\,,\nu)$, and $M(\nu)$.

Proof To prove (23), we first identify the interior of the domain of $\mathcal{L}^{<}_{\mathrm{inf}}$.

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 $\mathcal{L}^{<}_{\mathrm{inf}}(\mu,\nu) = \mathcal{L}^{\leqslant}_{\mathrm{inf}}(\mu,\nu) = +\infty$ for $\mu < m(\nu)$; we note in passing that we also have $\mathcal{L}^{<}_{\mathrm{inf}}\bigl(m(\nu),\nu\bigr) = +\infty$.

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

The interior of the domain of $\mu \mapsto \mathcal{L}^{<}_{\mathrm{inf}}(\mu,\nu)$ is therefore $\bigl(m(\nu),\,+\infty\bigr)$, and we recall that $\mathcal{L}^{<}_{\mathrm{inf}}(\,\cdot\,,\nu)$ 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

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

Letting $\varepsilon \to 0$, we get, by a sandwich argument, that $\mathcal{L}^{\leqslant}_{\mathrm{inf}}(\mu,\nu) = \mathcal{L}^{<}_{\mathrm{inf}}(\mu,\nu)$. This concludes the proof of (23).

We turn our attention to (24). We already showed above that $\mathcal{L}^{<}_{\mathrm{inf}}\bigl(m(\nu),\nu\bigr) = +\infty$. Now, to compute $\mathcal{L}^{\leqslant}_{\mathrm{inf}}(\mu,\nu)$, 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 $\nu\bigl\{m(\nu)\bigr\} > 0$ and $\nu\bigl\{m(\nu)\bigr\} = 0$ to establish, respectively, the equalities

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

In both cases, the first equality in (24) 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 4 together with Lemma 5.

Lemma 17 *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 $\nu\{m(\nu)\} \times \nu'\{M(\nu')\} = 0$.

Remark 18 *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 16 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 16, the functions $\mathcal{L}^{\leqslant}_{\inf}(\,\cdot\,,\nu)$ and $\mathcal{L}^{\leqslant}_{\inf}(\,\cdot\,,\nu)$ coincide on $\mathbb{R}\backslash\big\{m(\nu)\big\}$. It may be similarly shown that $\mathcal{L}^{>}_{\inf}(\,\cdot\,,\nu')$ and $\mathcal{L}^{\geqslant}_{\inf}(\,\cdot\,,\nu')$ coincide on $\mathbb{R}\backslash\big\{M(\nu')\big\}$. 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 $\mathcal{I} = \big\{m(\nu)\big\} = \big\{M(\nu')\big\}$, and both infima are equal to the values of the sums at $m(\nu) = M(\nu')$. By Lemma 16 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,

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

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

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 the statement of the highly classical Cramér-Chernoff bound: with the notation introduced in Section 3, for an N–sample $X_1,\dots,X_N$, distributed according to ν and of average denoted by $\overline{X}_N$,

$$\forall x\leqslant \mathrm{E}(\nu),\qquad \mathbb{P}\big(\overline{X}_N\leqslant x\big)\leqslant\exp\big(-N\,\phi_\nu^\star(x)\big)\,, \tag{25}$$

$$\text{and}\qquad \forall x\geqslant \mathrm{E}(\nu),\qquad \mathbb{P}\big(\overline{X}_N\geqslant x\big)\leqslant\exp\big(-N\,\phi_\nu^\star(x)\big)\,. \tag{26}$$

Such a classical result would in principle not require to be proved here. However, it turns out that we will re-use parts of this proof in later proofs, like the application 27 of Jensen's inequality or the variations of $\phi_\nu^\star$ discussed at the end of this section. This is why, despite all, we now prove (25)–(26).

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,

$$\forall\lambda\in\mathbb{R},\qquad \phi_\nu(\lambda)=\ln\mathbb{E}\big[\mathrm{e}^{\lambda X}\big]\geqslant\lambda\,\mathbb{E}[X]=\lambda\,\mathrm{E}(\nu)\,; \tag{27}$$

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,

$$0=-\phi_\nu(0)\leqslant\sup_{\lambda\leqslant 0}\big\{\lambda x-\phi_\nu(\lambda)\big\}=\sup_{\lambda\in\mathbb{R}}\big\{\lambda x-\phi_\nu(\lambda)\big\}\overset{\text{def}}{=}\phi_\nu^\star(x)\,. \tag{28}$$

This concludes the proof of (25). The bound (26) 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 (28) 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 2

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

Lemma 2 *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 The proof consists in two parts. We first show that for any finite grid $\mathcal{G} = \{g_2, \dots, g_{G-1}\}$ in (μ', μ), to which we add the points $g_1 = \mu'$ and $g_G = \mu$, we have

$$\limsup_{N\to+\infty} \frac{1}{N} \ln \mathbb{P}\big(\overline{X}_N \leqslant \overline{Y}_N\big) \leqslant - \min\Big\{\phi_\nu^\star(\mu'),\ \min_{2\leqslant j\leqslant G-1} \big\{\phi_{\nu'}^\star(g_{j-1}) + \phi_\nu^\star(g_j)\big\},\ \phi_{\nu'}^\star(\mu)\Big\}\,. \quad (29)$$

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 (25) and (26),

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

Second, again by the Cramér-Chernoff inequalities,

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

By a union bound,

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

The stated bound (29) 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 (29) holds for any finite grid in (μ', μ), and we consider a sequence

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

of such finite grids. In particular,

$$\limsup_{N\to+\infty}\frac{1}{N}\ln\mathbb{P}\big(\overline{X}_N\leqslant\overline{Y}_N\big)\leqslant-\min\Big\{\phi_\nu^\star(\mu'),\ \max_{n\geqslant1}S_n,\ \phi_{\nu'}^\star(\mu)\Big\}\,,$$

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

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

$$\phi_\nu^\star(\mu)=0=\phi_{\nu'}^\star(\mu')\,,$$

it suffices to show that

$$\max_{n\geqslant1}S_n\geqslant\inf_{x\in[\mu',\mu]}\big\{\phi_{\nu'}^\star(x)+\phi_\nu^\star(x)\big\}\,.$$

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

$$\varepsilon_n\overset{\text{def}}{=}\max_{2\leqslant j\leqslant G_n}\Big|g_j^{(n)}-g_{j-1}^{(n)}\Big|\,,$$

vanish asymptotically, i.e., $\varepsilon_n\to0$. For each grid $\mathcal{G}^{(n)}$, we denote by $x_n^\star\in(\mu',\mu)$ the argument of the minimum in the definition of S_n. As a consequence, for each $n\geqslant1$,

$$S_n=\phi_{\nu'}^\star(x_n^\star-\varepsilon_n^\star)+\phi_\nu^\star(x_n^\star)\,,$$

for some $0<\varepsilon_n^\star\leqslant\varepsilon_n$. The quantity $x_n^\star-\varepsilon_n^\star$ denotes the point in the grid that is right before $x_n^\star$, and it belongs to $[\mu',\mu)$. We note that we also have $\varepsilon_n^\star\to0$. In the compact interval $[\mu',\mu]$, the Bolzano-Weierstrass theorem (see, e.g., Bartle and Sherbert, 2000, Section 3.4) ensures the existence of a converging subsequence: there exists $x_\infty^\star\in[\mu',\mu]$ and a sequence $(n_k)_{k\geqslant1}$ of integers such that

$$x_{n_k}^\star\underset{k\to+\infty}{\longrightarrow}x_\infty^\star\,,\qquad\text{which also entails}\qquad x_{n_k}^\star-\varepsilon_{n_k}^\star\underset{k\to+\infty}{\longrightarrow}x_\infty^\star\,.$$

Now, the functions $\phi_\nu^\star$, respectively, $\phi_{\nu'}^\star$, 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,

$$\begin{aligned}\max_{n\geqslant1}S_n\geqslant\liminf_{k\to+\infty}\phi_{\nu'}^\star(x_{n_k}^\star-\varepsilon_{n_k}^\star)+\phi_\nu^\star(x_{n_k}^\star)&\geqslant\phi_{\nu'}^\star(x_\infty^\star)+\phi_\nu^\star(x_\infty^\star)\\&\geqslant\inf_{x\in[\mu',\mu]}\big\{\phi_{\nu'}^\star(x)+\phi_\nu^\star(x)\big\}\,.\end{aligned}$$

This concludes the proof. ■

B.3. Proof of Theorem 3

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 2.

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 phase r, i.e., belonging to S_{r-1}, was pulled in

phases 1 to r. For each arm a, we denote by $\overline{Y}_a^r$ 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 $r \in \{1, \dots, K-1\}$,

$$\text{on the event } \{a \in S_{r-1}\}, \qquad \overline{X}_a^r = \overline{Y}_a^r \,. \tag{30}$$

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 phases. This corresponds to the event

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

(We have an inclusion because ties are broken arbitrarily.) By optional skipping (30),

$$\begin{aligned}
&\bigcup_{r=1}^{K-1} \Big\{a^\star(\underline{\nu}) \in S_{r-1} \text{ and } \forall k \in S_{r-1},\ \overline{X}_{a^\star(\underline{\nu})}^r \leqslant \overline{X}_k^r\Big\} \\
&\qquad\qquad = \bigcup_{r=1}^{K-1} \Big\{a^\star(\underline{\nu}) \in S_{r-1} \text{ and } \forall k \in S_{r-1},\ \overline{Y}_{a^\star(\underline{\nu})}^r \leqslant \overline{Y}_k^r\Big\} \,.
\end{aligned}$$

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 phase $r \in \{1, \dots, K-1\}$,

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

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

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

or equivalently,

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

As $N_r/T \to \gamma_r > 0$ as $T \to +\infty$, we may apply Lemma 2, together with an exchange between the $\limsup$ and the maximum over a finite number of quantities. We obtain

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

This concludes the proof. ■

B.4. Proof of Corollary 4 and of the bound (17) on Φ

In this final subsection, we provide two series of proofs: first, a proof of Corollary 4; and then a proof of the bound $\Phi\big(\nu_k, \nu^\star\big) \geqslant \Delta_k^2$ stated as (17).

Proof of Corollary 4. To apply Theorem 3, we need only to show that the phase lengths of (16) 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 $r \in \{2, \dots, K-1\}$,

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

The bound of Theorem 3 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 $f(\nu_k, \nu^\star) \leqslant \Phi\big(\nu_k, \nu^\star\big)$,

$$\limsup_{T \to +\infty} \frac{1}{T} \ln \mathbb{P}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant -\frac{1}{\overline{\ln} K} \min_{1 \leqslant r \leqslant K-1} \left\{ \frac{1}{K-r+1} \min_{k \in \mathcal{A}_r} f\big(\nu_k, \nu^\star\big) \right\}. \tag{31}$$

The permutation σ in (18) 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 (31), 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 (17) on Φ, when the model is $\mathcal{D} = \mathcal{P}[0,1]$; we restate this bound 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 5) and then use Pinsker's inequality (7). We

provide here a slightly more direct but equivalent approach, based on Hoeffding's inequality.

Proof of the bound (17) 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 (17).

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

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

From the definition of Φ, it follows that

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

This corresponds to the second part of (17). ■

Appendix C. Proofs and details for Section 3.2: Rewriting 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 (19) holds. We do so for $\mathcal{D} = \mathcal{P}[0,1]$ in Appendix C.2 and for canonical one-parameter exponential families in Appendix C.3. Based on these two examples, we provide a set of conditions for general models, in Appendix C.4. A building block of these results is that for all these models $\mathcal{D}$, the functions $\mathcal{L}_{\inf}^{\leqslant}(\,\cdot\,, \nu)$ and $\mathcal{L}_{\inf}^{\geqslant}(\,\cdot\,, \nu)$ dominate the Fenchel-Legendre transform $\phi_\nu^\star$ defined in (15); we prove this in Appendix C.1.

All proofs of this section 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).

Remark 19 *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 5, 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, we recall (see Remark 1) that the $\mathcal{K}_{\inf}$ function is defined, for $\nu \in \mathcal{P}[0,1]$ and $x \in [0,1]$, as*

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

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

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

where X 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 Kullback-Leibler divergences as a supremum is not unexpected: a given Kullback-Leibler divergence can be formulated as a supremum, see (32), and equalities between $\inf \sup$ *and* $\sup \inf$ *holds under suitable assumptions (provided, e.g., by Sion's lemma).*

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

This domination is a consequence of a variational formula (32) for the Kullback-Leibler divergences.

Lemma 20 *For all models $\mathcal{D}$ containing distributions with finite first moments, for all distributions $\nu \in \mathcal{D}$,*

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

Proof We rely on a key variational formula for the Kullback-Leibler divergence, see Boucheron et al. (2013, Corollary 4.15): for all distributions ν, ν' over $\mathbb{R}$,

$$\begin{aligned}\mathrm{KL}(\nu',\nu) &= \sup\Big\{\mathbb{E}_{\nu'}[Y] - \ln \mathbb{E}_\nu\big[\mathrm{e}^Y\big] \,:\, \text{r.v. } Y \in \mathbb{L}^1(\nu') \ \text{ s.t. } \ \mathbb{E}_\nu\big[\mathrm{e}^Y\big] < +\infty\Big\}\,,\\ &= \sup\Big\{\mathbb{E}_{\nu'}[Y] - \ln \mathbb{E}_\nu\big[\mathrm{e}^Y\big] \,:\, \text{r.v. } Y \in \mathbb{L}^1(\nu')\Big\}\,, \end{aligned} \tag{32}$$

where the supremum is over random variables $Y : \mathbb{R} \to \mathbb{R}$ with a finite first moment with respect to ν', and 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 $\mathcal{D}$, they admit finite first moments, hence all random variables of the form $Y = \lambda\,\mathrm{id}_\mathbb{R}$ are ν'–integrable, where $\mathrm{id}_\mathbb{R}$ denotes the identity function over $\mathbb{R}$ and where $\lambda \in \mathbb{R}$. We have $\mathbb{E}_{\nu'}[Y] = \lambda\,\mathrm{E}(\nu')$. A consequence of (32) and of the definition (15) of $\phi^\star_\nu$ is therefore that

$$\mathrm{KL}(\nu',\nu) \geqslant \sup_{\lambda\in\mathbb{R}}\Big\{\lambda\,\mathrm{E}(\nu') - \ln \mathbb{E}_\nu\big[\mathrm{e}^{\lambda \mathrm{id}_\mathbb{R}}\big]\Big\} = \phi^\star_\nu\big(\mathrm{E}(\nu')\big)\,. \tag{33}$$

Using the variations of $\phi^\star_\nu$ indicated at the end of Appendix B.1, we see that

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

Therefore, taking an infimum in (33) yields, when $x \leqslant \mathrm{E}(\nu)$,

$$\mathcal{L}^{\leqslant}_{\inf}(x,\nu) = \inf\big\{\mathrm{KL}(\nu',\nu) : \mathrm{E}(\nu') \leqslant x\big\} \geqslant \phi^\star_\nu(x)\,,$$

and similarly for the other claimed inequality. ■

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 20 are in fact equalities, as claimed by Lemma 5, which we restate below. This yields, in particular, the target equality (19), as discussed after the statement of Lemma 5 in the main body of the article.

Lemma 5 *Consider the model $\mathcal{D} = \mathcal{P}[0,1]$. For all $\nu \in \mathcal{P}[0,1]$,*

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

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 $\phi_\nu^\star(x) = \mathcal{L}_{\inf}^{\leqslant}(x,\nu) = +\infty$. The counterpart statement $\phi_\nu^\star(x) = \mathcal{L}_{\inf}^{\geqslant}(x,\nu) = +\infty$ 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 20, it suffices 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 only deal with the first inequality, namely $\mathcal{L}_{\inf}^{\leqslant}(x,\nu) \leqslant \phi_\nu^\star(x)$ for $x \leqslant \mathrm{E}(\nu)$, as the other one may be obtained by symmetric arguments.

In the case $x = \mathrm{E}(\nu)$, we have $\phi_\nu^\star\big(\mathrm{E}(\nu)\big) = 0$, as stated at the end of Appendix B.1, and $\mathcal{L}_{\inf}^{\leqslant}\big(\mathrm{E}(\nu),\nu\big) = 0$, as can be seen by taking $\zeta = \nu$ in the infimum defining $\mathcal{L}_{\inf}^{\leqslant}$. 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,

$$\ell \stackrel{\text{def}}{=} \lim_{\lambda \to -\infty} \phi_\nu'(\lambda) = m(\nu)\,. \tag{34}$$

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 $-\infty$:

$$\forall \lambda \in \mathbb{R}, \qquad \phi_\nu'(\lambda) \geqslant \ell\,, \qquad \text{thus,} \qquad \mathbb{E}_\nu\Big[\big(\mathrm{id}_{[0,1]} - \ell\big)\,\mathrm{e}^{\lambda \mathrm{id}_{[0,1]}}\Big] \geqslant 0\,, \tag{35}$$

$$\text{or} \qquad \mathbb{E}_\nu\Big[\big(\mathrm{id}_{[0,1]} - \ell\big)\,\mathrm{e}^{\lambda (\mathrm{id}_{[0,1]} - \ell)}\Big] \geqslant 0\,. \tag{36}$$

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 (34).

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 $\mathcal{L}_{\inf}^{\leqslant}(x,\nu) \leqslant \phi_\nu^\star(x)$.

Case 1: $x > m(\nu)$. By Jensen's inequality (27) 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

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

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 $\mathcal{L}_{\inf}^{\leqslant}(x,\nu)$ and of the Kullback-Leibler divergence,

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

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,

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

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, Lemma 16 states that

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

We therefore have $\mathcal{L}_{\inf}^{\leqslant}(x,\nu) = \phi_\nu^\star(x)$ 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 $-\infty$, thus $\phi_\nu^\star(x) = \sup\Lambda = +\infty$. Now, 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; indeed, $x < m(\nu)$ imposes that ζ puts some probability mass to the left of the support of ν. Therefore, $\mathrm{KL}(\zeta,\nu) = +\infty$. All in all, $\mathcal{L}_{\inf}^{\leqslant}(x,\nu)$ appears as the infimum of either an empty set or of $+\infty$ values, so that $\mathcal{L}_{\inf}^{\leqslant}(x,\nu) = +\infty$. In this case as well, $\mathcal{L}_{\inf}^{\leqslant}(x,\nu) = \phi_\nu^\star(x)$, both being equal to $+\infty$. ■

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

In this section, we show that the target equality (19) is satisfied by so-called canonical one-parameter exponential families $\mathcal{D}_{\mathrm{exp}}$. Before we do so, we recall the definition and the properties of the latter.

Canonical one-parameter exponential families. We follow largely the exposition by Cappé et al. (2013, Section 4); more details, including the proofs of the stated properties may be found in the monograph by Lehmann and Casella (1998). A (regular) canonical one-parameter exponential family $\mathcal{D}_{\text{exp}}$ is a set of distributions ν_θ indexed by $\theta \in \Theta$, all absolutely continuous with respect to some measure ρ on $\mathbb{R}$, with densities given by

$$\frac{\mathrm{d}\nu_\theta}{\mathrm{d}\rho} = \exp\bigl(\theta\,\mathrm{id}_{\mathbb{R}} - b(\theta)\bigr)\,, \tag{37}$$

for some smooth enough normalization function b. More precisely, b is assumed to be twice differentiable. We also assume that Θ is the natural parameter space, i.e., that Θ contains all possible parameters for ρ:

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

and that Θ is an open interval (this latter fact is what regularity stands for). A closed-form expression of b is: for all $\theta \in \Theta$,

$$b(\theta) = \ln \int_{\mathbb{R}} e^{\theta y}\,\mathrm{d}\rho(y)\,. \tag{38}$$

The derivative b' of b is a continuous function, by assumption, and it may be shown that it is increasing, so that b' is a one-to-one mapping with a continuous inverse $(b')^{-1}$. In addition, it can be seen, by a differentiation under the integral sign, that $\mathrm{E}(\nu_\theta) = b'(\theta)$ for all $\theta \in \Theta$. Therefore, the distributions in $\mathcal{D}_{\text{exp}}$ may be rather parameterized by their expectations. We denote by $\mathcal{M} = b'(\Theta)$ the open interval of the expectations of distributions in $\mathcal{D}_{\text{exp}}$, and let μ_- and μ_+ be its lower and upper ends:

$$\mathcal{M} = (\mu_-, \mu_+)\,.$$

For each $x \in \mathcal{M}$, there exists a unique distribution in $\mathcal{D}_{\text{exp}}$ with expectation x, namely, $\nu_{(b')^{-1}(x)}$.

Kullback-Leibler divergences for $\mathcal{D}_{\text{exp}}$. We may also parameterize the Kullback-Leibler divergence function by the expectations: we define, for all $\theta_1, \theta_2 \in \Theta$,

$$d\bigl(\mathrm{E}(\nu_{\theta_1}), \mathrm{E}(\nu_{\theta_2})\bigr) \overset{\text{def}}{=} \mathrm{KL}(\nu_{\theta_1}, \nu_{\theta_2})\,. \tag{39}$$

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 $\mathcal{M} \times \mathcal{M}$.

A direct application of the continuity and monotonicity properties of d is that all functions $\mathcal{L}^{<}_{\text{inf}}$, $\mathcal{L}^{\leqslant}_{\text{inf}}$, $\mathcal{L}^{>}_{\text{inf}}$, $\mathcal{L}^{\geqslant}_{\text{inf}}$ coincide with d in the sense of the stated equalities (12) and (13). Indeed and for instance, we have, for $\nu \in \mathcal{D}_{\text{exp}}$ and $x \leqslant \mathrm{E}(\nu)$ with $x \in \mathcal{M}$:

$$\mathcal{L}^{<}_{\text{inf}}(x, \nu) = \inf_{\mu < x} \bigl\{d(\mu, \nu)\bigr\} = \lim_{\substack{\mu \to x \\ \mu < x}} d(\mu, \nu) = d(x, \nu)\,.$$

When $x \notin \mathcal{M}$, by the convention on the infimum of an empty set, $\mathcal{L}^{<}_{\text{inf}}(x, \nu) = +\infty$, while by our definition of d outside $\mathcal{M} \times \mathcal{M}$, we also have $d(x, \nu) = +\infty$. But as Lemma 22 below illustrates, we will only be interested on the behaviors on $\mathcal{M} \times \mathcal{M}$.

We now state a monotonicity property of the Chernoff-information-type quantity L defined for exponential models in (14). This property was referred to in Example 1, when indicating that arms can be equivalently ranked in descending expectations or ascending values of $L(\,\cdot\,,\mu^\star)$.

Lemma 21 *Consider a canonical one-parameter exponential family $\mathcal{D}_{exp}$ and fix any $\mu \in \mathcal{M}$. Then $L(\,\cdot\,,\mu)$ is non-increasing on $(\mu_-,\mu]$.*

Proof Fix $\mu_- < \mu_2 \leqslant \mu_1 \leqslant \mu$. To get the desired inequality $L(\mu_2,\mu) \geqslant L(\mu_1,\mu)$, it suffices to show, by (14), that

$$\forall y \in [\mu_2,\mu], \qquad d(y,\mu_2) + d(y,\mu) \geqslant \min_{x\in[\mu_1,\mu]} d(x,\mu_1) + d(x,\mu) \stackrel{\text{def}}{=} L(\mu_1,\mu)\,. \tag{40}$$

We distinguish two cases. If $\mu_2 \leqslant \mu_1 \leqslant y \leqslant \mu$, then, since $d(y,\,\cdot\,)$ is increasing on $(\mu_-,y]$, we have $d(y,\mu_2) \geqslant d(y,\mu_1)$, from which the inequality (40) follows by considering $x = y$. If $\mu_2 \leqslant y \leqslant \mu_1 \leqslant \mu$, then similarly $d(y,\mu) \geqslant d(\mu_1,\mu)$, which yields

$$\underbrace{d(y,\mu_2)}_{\geqslant 0} + d(y,\mu) \geqslant d(\mu_1,\mu) = \underbrace{d(\mu_1,\mu_1)}_{=0} + d(\mu_1,\mu)\,,$$

from which the inequality (40) follows by considering $x = \mu_1$. ■

A slightly weaker version of Lemma 5, sufficient for our purposes. We may now come back to the proof of the target equality (19) for canonical one-parameter exponential families. The following slightly weaker version of Lemma 5 is enough to yield (19), given the rewritings (12) and (13).

Lemma 22 *Consider a canonical one-parameter exponential family $\mathcal{D} = \mathcal{D}_{exp}$. For all $\nu \in \mathcal{D}_{exp}$,*

$$\forall x \in \mathcal{M}, \qquad \phi_\nu^\star(x) = d\big(x, \mathrm{E}(\nu)\big)\,.$$

The result of the lemma holds, by conventions, for $x < \mu_-$ or $x > \mu_+$, but does not hold in general for $x \in \{\mu_-,\mu_+\}$.

Proof By Lemma 20, we only need to show that $\phi_\nu^\star(x) \geqslant d\big(x,\mathrm{E}(\nu)\big)$. Given the definition (15) of $\phi_\nu^\star$ as a supremum, it suffices to exhibit a $\lambda^\star \in \mathbb{R}$ such that

$$d\big(x,\mathrm{E}(\nu)\big) = \lambda^\star x - \phi_\nu(\lambda^\star)\,. \tag{41}$$

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 (41) with $\lambda^\star = \theta_2 - \theta_1$. Given the closed-form expression of the densities (37), the distribution ν_{θ_2} is absolutely continuous with respect to ν_{θ_1}, with density given by $(\theta_2-\theta_1)\mathrm{id}_\mathbb{R} - \big(b(\theta_2)-b(\theta_1)\big)$. Therefore, by definition of the Kullback-Leibler divergence,

$$\begin{aligned} d\big(x,\mathrm{E}(\nu)\big) = \mathrm{KL}(\nu_{\theta_2},\nu_{\theta_1}) &= \mathbb{E}_{\nu_{\theta_2}}\left[\ln\frac{\mathrm{d}\nu_{\theta_2}}{\mathrm{d}\nu_{\theta_1}}\right] = \mathbb{E}_{\nu_{\theta_2}}\Big[(\theta_2-\theta_1)\,\mathrm{id}_\mathbb{R} - \big(b(\theta_2)-b(\theta_1)\big)\Big] \\ &= (\theta_2-\theta_1)\,\mathrm{E}(\nu_{\theta_2}) - \big(b(\theta_2)-b(\theta_1)\big) = \lambda^\star x - \big(b(\theta_2)-b(\theta_1)\big)\,. \end{aligned} \tag{42}$$

To obtain (41), it only remains to show that $b(\theta_2) - b(\theta_1) = \phi_\nu(\lambda^\star)$. Using the closed-form expressions (38) of b at θ_2 and (37) of the density at θ_1, we obtain

$$\begin{aligned} b(\theta_2) &= \ln \int_{\mathbb{R}} e^{\theta_2 y} \,\mathrm{d}\rho(y) = b(\theta_1) + \ln \int_{\mathbb{R}} e^{(\theta_2-\theta_1)y} \overbrace{e^{\theta_1 y - b(\theta_1)} \,\mathrm{d}\rho(y)}^{=\mathrm{d}\nu_{\theta_1}(y)=\mathrm{d}\nu(y)} \\ &= b(\theta_1) + \ln \int_{\mathbb{R}} e^{\lambda^\star y} \,\mathrm{d}\nu(y) = b(\theta_1) + \phi_\nu(\lambda^\star)\,, \end{aligned} \tag{43}$$

which concludes the proof. ■

Remark 23 *A more direct approach bypassing Lemma 20 can be followed with $\mathcal{D}_{exp}$ models, along the following lines. The result (43) can be generalized into*

$$\forall \theta \in \Theta, \qquad \phi_\nu(\theta - \theta_1) = b(\theta) - b(\theta_1)\,. \tag{44}$$

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 closed-form calculation (42) and the rewriting (44) then lead to Lemma 22.

C.4. Conditions for general models

In this section, we extend Lemma 5, and thus the target equality (19), to more general models. We did so by mimicing the proof of Lemma 5: the result below can certainly be improved. We extend as follows the definitions of the lower and upper ends $m(\nu)$ and $M(\nu)$ of the closed support $\mathrm{Supp}(\nu)$ of a distribution ν over $\mathbb{R}$:

$$m(\nu) = \inf\big(\mathrm{Supp}(\nu)\big) \in \mathbb{R} \cup \{-\infty\} \qquad \text{and} \qquad M(\nu) = \sup\big(\mathrm{Supp}(\nu)\big) \in \mathbb{R} \cup \{+\infty\}\,.$$

Lemma 24 *Consider a model $\mathcal{D}$ containing distributions ν over $\mathbb{R}$ with finite first moments and with exponential moments: $\mathrm{e}^{\lambda \mathrm{id}_{\mathbb{R}}} \in \mathbb{L}^1(\nu)$ for all $\lambda \in \mathbb{R}$. Assume that the model $\mathcal{D}$ is stable by exponential reweighting of densities: for all $\nu \in \mathcal{D}$, for all $\lambda \in \mathbb{R}$, the distribution ν_λ with density*

$$\frac{\mathrm{d}\nu_\lambda}{\mathrm{d}\nu} = \frac{\mathrm{e}^{\lambda \mathrm{id}_{\mathbb{R}}}}{\mathbb{E}_\nu\big[\mathrm{e}^{\lambda \mathrm{id}_{\mathbb{R}}}\big]} \qquad \textit{with respect to } \nu \tag{45}$$

also 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 $x \in \big\{m(\nu), M(\nu)\big\} \cap \mathbb{R}$ and $\nu\{x\} > 0$; put differently, if a distribution $\nu \in \mathcal{D}$ puts some probability mass on an end x of its closed support, then the Dirac mass at x belongs to $\mathcal{D}$.

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 symmetry and by Lemma 20, we only need to prove that

$$\forall x \leqslant \mathrm{E}(\nu), \quad \phi_\nu^\star(x) \geqslant \mathcal{L}_{\inf}^{\leqslant}(x,\nu)\,. \tag{46}$$

For $x = \mathrm{E}(\nu)$, we have $\phi_\nu^\star\big(\mathrm{E}(\nu)\big) = 0 = \mathcal{L}_{\inf}^{\leqslant}\big(\mathrm{E}(\nu),\nu\big)$, as stated at the end of Appendix B.1 and by taking $\zeta = \nu$ in the infimum defining $\mathcal{L}_{\inf}^{\leqslant}$, respectively. Before moving to the case $x < \mathrm{E}(\nu)$, we establish a few properties of ϕ_ν based on the assumptions of Lemma 24. All random variables $\mathrm{e}^{\lambda \mathrm{id}_\mathbb{R}}$ are ν–integrable, for $\lambda \in \mathbb{R}$, which entails, by application of a standard theorem of differentiation under the integral sign together with local domination arguments of the form

$$\forall \lambda \in (\lambda_-, \lambda_+), \qquad \big|\mathrm{id}_\mathbb{R}\,\mathrm{e}^{\lambda \mathrm{id}_\mathbb{R}}\big| \leqslant \big|\mathrm{id}_\mathbb{R}\big|\,\big(\mathrm{e}^{\lambda_- \mathrm{id}_\mathbb{R}} + \mathrm{e}^{\lambda_+ \mathrm{id}_\mathbb{R}}\big) \leqslant \big(\mathrm{e}^{\mathrm{id}_\mathbb{R}} + \mathrm{e}^{-\mathrm{id}_\mathbb{R}}\big)\big(\mathrm{e}^{\lambda_- \mathrm{id}_\mathbb{R}} + \mathrm{e}^{\lambda_+ \mathrm{id}_\mathbb{R}}\big)\,,$$

that ϕ_ν is differentiable over $\mathbb{R}$, with derivative given by

$$\phi_\nu' : \lambda \in \mathbb{R} \longmapsto \frac{\mathbb{E}_\nu\big[\mathrm{id}_\mathbb{R}\,\mathrm{e}^{\lambda \mathrm{id}_\mathbb{R}}\big]}{\mathbb{E}_\nu\big[\mathrm{e}^{\lambda \mathrm{id}_\mathbb{R}}\big]}\,. \tag{47}$$

Hölder's inequality still entails that ϕ_ν is convex, thus its derivative ϕ_ν' is non-decreasing; therefore, ϕ_ν' admits a limit $\ell \in \{-\infty\} \cup \mathbb{R}$ at $-\infty$. Actually, we have $\ell = m(\nu)$, as can be seen by combining the following facts. First, by definition, $\mathrm{id}_\mathbb{R} \geqslant m(\nu)$ ν-a.s., thus $\phi_\nu' \geqslant m(\nu)$, hence $\ell \geqslant m(\nu)$. As a consequence, if $\ell = -\infty$, then we also have $m(\nu) = -\infty$. Otherwise, if $\ell \in \mathbb{R}$, the same arguments as in (35)–(36) show that $\mathrm{id}_\mathbb{R} - \ell \geqslant 0$ ν-a.s., i.e., $\ell \leqslant m(\nu)$.

We may now come back to establishing $\phi_\nu^\star(x) \geqslant \mathcal{L}_{\inf}^{\leqslant}(x,\nu)$ in the case $x < \mathrm{E}(\nu)$. We consider three sub-cases, depending on the respective positions of x and $m(\nu)$.

Case 1: $x > m(\nu)$. The properties of ϕ_ν ensure, exactly as in Case 1 of the proof of Lemma 5 (in Appendix C.2), the existence of $\lambda^\star$ such that $\phi_\nu'(\lambda^\star) = x$ and $\phi_\nu^\star(x) = \lambda^\star x - \phi_\nu(\lambda^\star)$. Given the assumption (45), we may consider the distribution $\nu_{\lambda^\star} \in \mathcal{D}$. We note, again exactly as in Case 1 of the proof of Lemma 5 and given the closed-form expression (47) for ϕ_ν', that $\mathrm{E}(\nu_{\lambda^\star}) = \phi_\nu'(\lambda^\star)$, thus $\mathrm{E}(\nu_{\lambda^\star}) = x$. Finally, an explicit computation yields

$$\mathrm{KL}(\nu_{\lambda^\star}, \nu) = \lambda^\star \mathrm{E}(\nu_{\lambda^\star}) - \ln \mathbb{E}_\nu\big[\mathrm{e}^{\lambda \mathrm{id}_\mathbb{R}}\big] = \lambda^\star x - \phi_\nu(\lambda^\star) = \phi_\nu^\star(x)\,.$$

By the defining infimum of $\mathcal{L}_{\inf}^{\leqslant}(x,\nu)$, we have indeed $\mathcal{L}_{\inf}^{\leqslant}(x,\nu) \leqslant \mathrm{KL}(\nu_{\lambda^\star}, \nu) = \phi_\nu^\star(x)$.

Case 2: $x = m(\nu)$. In particular, $m(\nu) \in \mathbb{R}$, which allows us to follow the monotone-convergence arguments of Case 2 of the proof of Lemma 5 (in Appendix C.2) and get the equality $\phi_\nu^\star\big(m(\nu)\big) = -\ln \nu\big\{m(\nu)\big\}$. Now, for the second part of this sub-case, we also adapt an argument of the second part of the proof of Lemma 16 (in Appendix A.2), namely, the fact that either there exists at most one distribution $\zeta \in \mathcal{D}$ absolutely continuous with respect to ν and satisfying $\mathrm{E}(\zeta) \leqslant m(\nu)$, namely, $\zeta = \delta_{m(\nu)}$, the Dirac mass at $m(\nu)$. The latter is indeed absolutely continuous with respect to ν if and only if $\nu\big\{m(\nu)\big\} > 0$. When $\nu\big\{m(\nu)\big\} > 0$, we have $\delta_{m(\nu)} \in \mathcal{D}$ by the Dirac assumption of the lemma, so that

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

Otherwise, when $\nu\big\{m(\nu)\big\} = 0$, the infimum defining $\mathcal{L}_{\inf}^{\leqslant}\big(m(\nu),\nu\big)$ is either over an empty set or of $+\infty$ values, and thus equals $+\infty = -\ln \nu\big\{m(\nu)\big\}$. In both situations, we obtained $\mathcal{L}_{\inf}^{\leqslant}\big(m(\nu),\nu\big) = \phi_\nu^\star\big(m(\nu)\big)$.

Case 3: $x < m(\nu)$. In particular, $m(\nu) \in \mathbb{R}$ in this sub-case as well, which allows us to repeat the exact same arguments as in Case 3 of the proof of Lemma 5 (in Appendix C.2): we may show that both $\mathcal{L}_{\inf}^{\leqslant}(x,\nu)$ and $\phi_\nu^\star(x)$ are equal to $+\infty$. ■

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 7

We restate the lemma for the convenience of the reader. The proof reveals that the inequality actually holds for limits taken along subsequences $(T_n)_{n\geqslant 1}$. Also, we may only relax the assumptions on the bandit models; e.g., they do not need to be generic and it suffices that they have different unique optimal arms. (The notion of a generic bandit problem is defined in the first lines of Section 4.)

Lemma 7 *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*

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

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

denotes the number of times arm a was pulled in the T exploration rounds of a given strategy with budget $T \geqslant 1$.

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

$$\sum_{a=1}^{K} \mathbb{E}_{\underline{\lambda}}[N_a(T)]\,\mathrm{KL}(\lambda_a,\nu_a) \geqslant \mathrm{KL}\big(\mathrm{Ber}(q_T),\mathrm{Ber}(p_T)\big)\,, \tag{48}$$

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 10

We restate the theorem for the convenience of the reader (and recall that the notion of a generic bandit problem is defined in the first lines of Section 4).

Theorem 10 *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,

$$\liminf_{T\to+\infty} \frac{1}{T}\ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -\frac{\mathcal{L}^{<}_{\inf}\big(\mu_{(K)}, \nu^\star\big)}{K}\,. \tag{49}$$

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 (49). 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 $\underline{\lambda} = (\lambda_1, \ldots, \lambda_K)$ 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 7 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 (49) 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 $K-k$ worst arms. Use the definition of clever exploitation of pruning of suboptimal arms and apply (49) 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. For the convenience of the reader, we first restate the definition of normality.

Definition 12 *A model $\mathcal{D}$ is normal if for all $\nu \in \mathcal{D}$, for all $x \geqslant \mathrm{E}(\nu)$,*

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

Proposition 25 *$\mathcal{P}[0,1]$ is a normal model.*

Proof We fix $\nu \in \mathcal{P}[0,1]$, a real $x \geqslant \mathrm{E}(\nu)$, and $\varepsilon > 0$. Recall the piece of notation $M(\nu)$ for the upper end of the support of ν, as introduced in Appendix A.2. As in Case 3 of the proof of Lemma 5 (in Appendix C.2), we note that when $x \geqslant M(\nu)$, there exists no distribution $\zeta \in \mathcal{P}[0,1]$ absolutely continuous with respect to ν and such that $\mathrm{E}(\zeta) > x$; hence, both infima in Definition 12 equal $+\infty$. We now tackle the case where $\mathrm{E}(\nu) \leqslant x < M(\nu)$. For all $\delta > 0$, we introduce

$$x'_\delta = \min\left\{x + \delta, \ \frac{x + M(\nu)}{2}\right\} < M(\nu)\,.$$

Case 1 of the proof of Lemma 5 and Lemma 20 reveal (by symmetry) that for each $\delta > 0$, there exists a distribution $\zeta_\delta \in \mathcal{P}[0,1]$ with expectation x'_δ and such that $\mathcal{L}^{\geqslant}_{\inf}(x'_\delta, \nu) = \phi^\star_\nu(x'_\delta) = \mathrm{KL}(\zeta_\delta, \nu)$. By Lemma 16, $\mathcal{L}^{\geqslant}_{\inf}(x'_\delta, \nu) = \mathcal{L}^{>}_{\inf}(x'_\delta, \nu)$ and $\mathcal{L}^{>}_{\inf}(\,\cdot\,, \nu)$ is continuous on $\big(-\infty, M(\nu)\big)$. Putting all these elements together, we obtain

$$\begin{aligned}\mathcal{L}^{>}_{\inf}(x, \nu) = \lim_{\delta\to 0} \mathcal{L}^{>}_{\inf}(x'_\delta, \nu) &= \liminf_{\delta\to 0} \mathrm{KL}(\zeta_\delta, \nu) \\ &\geqslant \inf\big\{\mathrm{KL}(\zeta_\delta, \nu) : \ \delta \in (0, \varepsilon)\big\} \\ &\geqslant \inf\big\{\mathrm{KL}(\zeta, \nu) : \ \zeta \in \mathcal{D} \ \text{ s.t. } \ x + \varepsilon > \mathrm{E}(\zeta) > x\big\}\,,\end{aligned}$$

where the first inequality is by the very definition of a $\liminf$. ■

Proposition 26 *All canonical one-parameter exponential models $\mathcal{D}_{exp}$ are normal.*

Proof The proof consists of rewriting $\mathcal{L}_{\inf}^{>}$ as d, as indicated by (13), and using the regularity properties for d exhibited in Appendix C.3. We fix $\nu \in \mathcal{D}_{\exp}$, a real $x \geqslant \mathrm{E}(\nu)$, and $\varepsilon > 0$. When $x \geqslant M(\nu)$, the same argument as in the previous proposition shows that both infima equal $+\infty$. For $x < M(\nu)$, we introduce $\delta \in (0, \mu_+ - x)$ and write

$$\begin{aligned}\mathcal{L}_{\inf}^{>}(x,\nu) = d\big(x, \mathrm{E}(\nu)\big) &= \lim_{\delta \to 0} d\big(x+\delta, \mathrm{E}(\nu)\big) \\ &= \inf\Big\{ d\big(x+\delta, \mathrm{E}(\nu)\big) : \delta \in (0,\varepsilon) \Big\} \\ &= \inf\big\{ \mathrm{KL}(\zeta,\nu) : \ \zeta \in \mathcal{D} \ \text{ s.t. } \ x+\varepsilon > \mathrm{E}(\zeta) > x \big\}\,,\end{aligned}$$

where the second and third equalities follow, respectively, by continuity of $d\big(\,\cdot\,, \mathrm{E}(\nu)\big)$ on $\mathcal{M}$ and by the fact that this function is non-decreasing on $(x, \mu_+) \subset [\mathrm{E}(\nu), \mu_+)$, and the final equality is by the rewriting (39). ■

D.4. Proof of Theorem 13

We restate the theorem for the convenience of the reader (and recall that the notion of a generic bandit problem is defined in the first lines of Section 4).

Theorem 13 *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 7 may be applied; together with the two upper bounds above, it yields

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

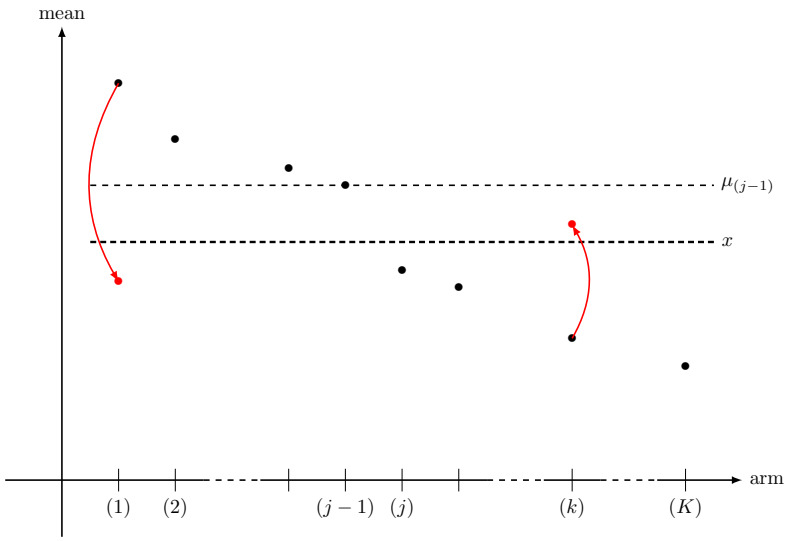

Figure 1: Original bandit problem $\underline{\nu}$ (in dark) and modifications made to arms (1) and (k) to obtain an alternative bandit problem $\underline{\lambda} \in \mathrm{Alt}_{k,j,x}(\underline{\nu})$ (in red): in $\underline{\lambda}$, arm (k) is the $j-1$–th best arm, while arm $(1) = a^\star(\underline{\nu})$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

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 $\mathcal{I}_{k,j,x}(\underline{\nu})$:

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

We prove below that

$$\mathcal{I}_{k,j,x}(\underline{\nu}) = \frac{\mathcal{L}^{>}_{\inf}\big(x, \nu_{(k)}\big)}{j-1} + \frac{\mathcal{L}^{<}_{\inf}\big(x, \nu^\star\big)}{j}\,, \tag{50}$$

from which the lower bound claimed in Theorem 13 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 (50). 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 obtained $\mathcal{L}^{<}_{\inf}\big(x,\nu^\star\big)$ by definition while we relied on the normality of the model (Definition 12) to obtain $\mathcal{L}^{>}_{\inf}\big(x,\nu_{(k)}\big)$. We did so with $\varepsilon = \mu_{(j-1)} - x$, which is indeed positive as we considered $x < \mu_{(j-1)}$. ■

D.5. Proof of Theorem 14

We restate the theorem for the convenience of the reader (and recall that the notion of a generic bandit problem is defined in the first lines of Section 4).

Theorem 14 *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 7 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 $\mathcal{I}_{k,x}(\underline{\nu})$:

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

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 13 (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 $\mathrm{Alt}_{k,x}(\underline{\nu})$,

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

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) in Appendix E.1 and the Bretagnolle-Huber technique in Appendix E.2.

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

Carpentier and Locatelli (2016, Theorem 1) proved (slightly stronger versions of) the following (non-asymptotic) minimax lower bound. Consider the model $\mathcal{B}_{[1/4,\,3/4]}$ of Bernoulli distributions $\mathrm{Ber}(p)$ with parameters $p \in [1/4,\,3/4]$. For all sequences of strategies that are consistent on $\mathcal{B}_{[1/4,\,3/4]}$, for all $T \geqslant 0.14\,K^4\ln(6KT)$,

$$\exists\,\underline{\nu}\text{ in }\mathcal{B}_{[1/4,\,3/4]},\qquad \frac{1}{T}\ln\mathbb{P}_{\underline{\nu}}\big(I_T\neq a^\star(\underline{\nu})\big)\geqslant -\frac{400}{\ln K}\left(\sum_{a\neq a^\star(\underline{\nu})}\frac{1}{\Delta_a^2}\right)^{-1}-\frac{\ln 6}{T}\,, \tag{51}$$

where, of course, we may rather use the weaker lower bound based on

$$-\left(\sum_{a\neq a^\star(\underline{\nu})}\frac{1}{\Delta_a^2}\right)^{-1}\geqslant -\min_{2\leqslant k\leqslant K}\frac{\Delta_{(k)}^2}{k}\,.$$

However, the bound (51) is different in nature from the lower bounds considered in this article, as first and foremost, it only guarantees a $1/\ln K$ improvement of the lower bound (8) of Audibert et al. (2010) for a single bandit problem $\underline{\nu}$ (actually belonging to a known collection of K bandit problems). This is in strong contrast with the uniform instance-dependent lower bounds presented in this article: bounds holding simultaneously for all bandit problems of a given model. Second, the proof of the result (see the simpler proof provided below for Proposition 27 stated next) is truly gap-based and does not seem to extend in any obvious way to non-parametric models.

As mentioned above, the proof of (51) in Carpentier and Locatelli (2016) uses only K different bandit problems in $\mathcal{B}_{[1/4,\,3/4]}$. We may therefore resort to the pigeonhole principle to exchange,

in some sense, the "for all $T \geqslant 0.14K^4 \ln(6KT)$" and "there exists $\underline{\nu}$ in $\mathcal{B}_{[1/4,\,3/4]}$" parts. More precisely, we obtain, from (51) the following proposition. For the sake of completeness, we provide a self-contained proof of this proposition closely following the original arguments by Carpentier and Locatelli (2016), except for the change-of-measure argument, for which we rather resort to Lemma 7. Doing so, we are able to improve the numerical factor 400 that would follow from (51) into a smaller factor of 30.

Proposition 27 *Fix $K \geqslant 3$ and consider the model $\mathcal{B}_{[1/4,\,3/4]}$ of Bernoulli distributions $\mathrm{Ber}(p)$ with parameters $p \in [1/4,\, 3/4]$. For all consistent sequences of strategies on $\mathcal{B}_{[1/4,\,3/4]}$, there exists an increasing sequence of budgets $(T_n)_{n\geqslant 1}$ such that*

$$\exists\, \underline{\nu} \text{ in } \mathcal{B}_{[1/4,\,3/4]}, \qquad \liminf_{n\to+\infty} \frac{1}{T_n} \ln \mathbb{P}_{\underline{\nu}}\big(I_{T_n} \neq a^\star(\underline{\nu})\big) \geqslant -\frac{30}{\ln K} \left(\sum_{a\neq a^\star(\underline{\nu})} \frac{1}{\Delta_a^2} \right)^{-1}. \tag{52}$$

Proof We consider some base Bernoulli bandit problem $\underline{\nu}^{\text{base}} = \big(\nu_1^{\text{base}}, \ldots, \nu_K^{\text{base}}\big)$, where

$$\nu_1^{\text{base}} = \mathrm{Ber}(1/2) \qquad \text{and} \qquad \forall j \in \{2,\ldots,K\}, \quad \nu_j^{\text{base}} = \mathrm{Ber}(p_j)\,,$$

for parameters $p_j \in [1/4,\, 1/2)$ to be specified later. For each $k \in \{2,\ldots,K\}$, we then define the alternative bandit problem $\underline{\nu}^{(k)} = \big(\nu_1^{(k)}, \ldots, \nu_K^{(k)}\big)$ as follows:

$$\nu_j^{(k)} = \begin{cases} \mathrm{Ber}(1-p_k) & \text{if } j = k, \\ \nu_j^{\text{base}} & \text{if } j \neq k. \end{cases}$$

Given the constraints on the p_j, the unique optimal arm of $\underline{\nu}^{\text{base}}$ is $a^\star\big(\underline{\nu}^{\text{base}}\big) = 1$, while the unique optimal arm of $\underline{\nu}^{(k)}$ is $a^\star\big(\underline{\nu}^{(k)}\big) = k$. We introduce, for a given bandit problem $\underline{\nu}$

$$H(\underline{\nu}) \overset{\text{def}}{=} \sum_{a\neq a^\star(\underline{\nu})} \frac{1}{\Delta_a^2}\,;$$

the right-hand side of (52) may be rewritten as $(34/\ln K)\, H(\underline{\nu})^{-1}$. The suboptimality gaps of the arms of $\underline{\nu}^{\text{base}}$ equal $\Delta_j^{\text{base}} = 1/2 - p_j$ for $j \neq 1$, while the ones of $\underline{\nu}^{(k)}$ equal

$$\begin{aligned} &\forall j \neq k, \qquad \Delta_j^{(k)} = 1 - p_k - p_j = (1/2 - p_k) + (1/2 - p_j) = \Delta_k^{\text{base}} + \Delta_j^{\text{base}}\,, \\ &\text{thus} \qquad H\big(\underline{\nu}^{(k)}\big) = \sum_{j\neq k} \frac{1}{\big(\Delta_k^{\text{base}} + \Delta_j^{\text{base}}\big)^2}\,. \end{aligned} \tag{53}$$

The proof is decomposed in two steps. First, we show that for all values of the p_j abiding by the constraints and for all weights $u_2, \ldots, u_K$ such that $u_j \geqslant 0$ for all j and $u_1 + \ldots + u_K = 1$, there exists $k^\star \in \{2,\ldots,K\}$ such that there exists an increasing sequence of budgets $(T_n)_{n\geqslant 1}$ with

$$\liminf_{n\to+\infty} \frac{1}{T_n} \ln \mathbb{P}_{\underline{\nu}^{(k^\star)}}\big(I_{T_n} \neq k^\star\big) \geqslant -9\, u_{k^\star} \big(\Delta_{k^\star}^{\text{base}}\big)^2. \tag{54}$$

Then, we set specific values of the u_j and p_j to get

$$\forall k \in \{2,\ldots,K\}, \qquad u_k \big(\Delta_k^{\text{base}}\big)^2 \leqslant \frac{10}{3\ln K}\, H\big(\underline{\nu}^{(k)}\big)^{-1}. \tag{55}$$

Proposition 27 follows by combining (54) and (55).

Part 1: Proof of (54). For all $T \geqslant 1$,

$$\sum_{k=2}^{K} \frac{\mathbb{E}_{\underline{\nu}^{\text{base}}}\big[N_k(T)\big]}{T} \leqslant 1 = \sum_{k=2}^{K} u_k \,;$$

therefore, for all $T \geqslant 1$, there exists $k_T \in \{2, \dots, K\}$ such that $\mathbb{E}_{\underline{\nu}^{\text{base}}}\big[N_{k_T}(T)\big]/T \leqslant u_{k_T}$. By the pigeonhole principle, there exists $k^\star \in \{2, \dots, K\}$ and an (infinite) increasing sequence $(T_n)_{n\geqslant 1}$ of integers such that $k_{T_n} = k^\star$ for all $n \geqslant 1$. In particular,

$$\limsup_{n\to+\infty} \frac{\mathbb{E}_{\underline{\nu}^{\text{base}}}\big[N_{k^\star}(T_n)\big]}{T_n} \leqslant u_{k^\star} \,.$$

Since $\underline{\nu}^{\text{base}}$ and $\underline{\nu}^{(k^\star)}$ only differ at arm $k^\star$, an application of Lemma 7 along subsequences (see the initial comments in Appendix D.1) guarantees that

$$\begin{aligned}\liminf_{n\to+\infty} \frac{1}{T_n} \ln \mathbb{P}_{\underline{\nu}^{(k^\star)}}\big(I_{T_n} \neq k^\star\big) &\geqslant -\left(\limsup_{n\to+\infty} \frac{\mathbb{E}_{\underline{\nu}^{\text{base}}}\big[N_{k^\star}(T_n)\big]}{T_n}\right) \mathrm{KL}\big(\mathrm{Ber}(1-p_{k^\star}), \mathrm{Ber}(p_{k^\star})\big) \\ &\geqslant -u_{k^\star} \times 9\,(1/2 - p_{k^\star})^2 = -9\,u_{k^\star}\,\big(\Delta_{k^\star}^{\text{base}}\big)^2 \,,\end{aligned}$$

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

$$\mathrm{KL}\big(\mathrm{Ber}(1-x), \mathrm{Ber}(x)\big) = (1-x)\ln\frac{1-x}{x} + x\ln\frac{x}{1-x} \leqslant 9\left(\frac{1}{2} - x\right)^2 .$$

Part 2: Proof of (55). We set, for $j \in \{2, \dots, K\}$,

$$u_j = \frac{U}{\big(\Delta_j^{\text{base}}\big)^2 H\big(\underline{\nu}^{(j)}\big)} \,, \qquad \text{where} \qquad U = \left(\sum_{k=2}^{K} \frac{1}{\big(\Delta_k^{\text{base}}\big)^2 H\big(\underline{\nu}^{(k)}\big)}\right)^{-1} .$$

Then, $u_k\big(\Delta_k^{\text{base}}\big)^2 = H\big(\underline{\nu}^{(k)}\big)^{-1} U$ for all $k \in \{2, \dots, K\}$. To get the desired result, it suffices to guarantee that $U \leqslant 10/(3\ln K)$. To do so, we consider the same values as in Carpentier and Locatelli (2016) for the p_j, i.e., we set, for $j \in \{2, \dots, K\}$,

$$p_j = \frac{1}{2} - \frac{j}{4K} \qquad \text{or, equivalently,} \qquad \Delta_j^{\text{base}} = \frac{j}{4K} \,.$$

We show first that $\big(\Delta_k^{\text{base}}\big)^2 H\big(\nu^{(k)}\big) \leqslant 2k$, for all $k \in \{2, \dots, K\}$. Indeed, by (53) and by lower bounding $\Delta_k^{\text{base}} + \Delta_j^{\text{base}}$ either by Δ_k^{base} or Δ_j^{base}, we get

$$\begin{aligned}\big(\Delta_k^{\text{base}}\big)^2 H\big(\underline{\nu}^{(k)}\big) &= \sum_{j<k} \frac{\big(\Delta_k^{\text{base}}\big)^2}{\big(\Delta_k^{\text{base}} + \Delta_j^{\text{base}}\big)^2} + \sum_{j>k} \frac{\big(\Delta_k^{\text{base}}\big)^2}{\big(\Delta_k^{\text{base}} + \Delta_j^{\text{base}}\big)^2} \\ &\leqslant k - 1 + \sum_{j>k} \frac{\big(\Delta_k^{\text{base}}\big)^2}{\big(\Delta_j^{\text{base}}\big)^2} = k - 1 + \sum_{j>k} \frac{k^2}{j^2} \leqslant k - 1 + k^2 \int_k^K \frac{1}{v^2}\mathrm{d}v \leqslant 2k \,.\end{aligned}$$

Finally,

$$U \leqslant \left(\sum_{k=2}^{K} \frac{1}{2k}\right)^{-1} \leqslant \left(\int_{2}^{K+1} \frac{1}{2v}\mathrm{d}v\right)^{-1} = 2\left(\ln(K+1) - \ln 2\right)^{-1} \leqslant \frac{10}{3\ln K}\,,$$

where the final inequality holds since $K \geqslant 3$. ■

E.2. The Bretagnolle-Huber technique by Kaufmann et al. (2016, Section 5.2)

Kaufmann et al. (2016, Section 5.2) provide an interesting series of results relying on the so-called Bretagnolle-Huber inequality recalled below in (57); we state one of their lower bounds in Corollary 29. But as we argue in this section, the methodology followed seems extremely specific to the case of parametric models where Kullback-Leibler divergences could be controlled (lower bounded and upper bounded) in terms of gaps, like the model $\mathcal{D}_{\sigma^2}$ of Gaussian distributions with a fixed variance $\sigma^2 > 0$. In particular, we state in Proposition 28 what would be the straightforward extension to non-parametric models of the Gaussian results of (Kaufmann et al., 2016, Section 5.2), and we immediately discuss after this statement why this extension lacks interpretability and interest. Proposition 28 considers any sequence of strategies (not necessarily consistent) and provides an asymptotic bound; however, it does not directly control the target probability of error $\mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big)$, but a larger quantity. A proof of Proposition 28 is provided at the end of this section.

Proposition 28 *Fix $K \geqslant 2$, a model $\mathcal{D}$, and any 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 $\mathrm{E}(\zeta_k) > \mu^\star$. For $k \neq a^\star(\underline{\nu})$, denote by $\underline{\nu}^{(k)}$ the bandit problem obtained from $\underline{\nu}$ by changing the distribution of arm k into ζ_k. For all $T \geqslant 1$,*

$$\frac{1}{T}\ln \max\left\{\mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big),\ \max_{k\neq a^\star(\underline{\nu})} \mathbb{P}_{\underline{\nu}^{(k)}}\big(I_T \neq k\big)\right\} \geqslant -\left(\sum_{a\neq a^\star(\underline{\nu})} \frac{1}{\mathrm{KL}(\nu_a,\zeta_a)}\right)^{-1} - \frac{\ln 4}{T}\,.$$

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 a large lower bound by choosing ζ_k as close as possible to ν_k in terms of Kullback-Leibler divergences, and controlling the maximum of the misidentification probabilities: when ζ_k gets closer to ν_k while abiding by the constraint $\mathrm{E}(\zeta_k) > \mu^\star$, the probability $\mathbb{P}_{\underline{\nu}^{(k)}}\big(I_T \neq k\big)$ becomes larger, and should even intuitively converge to $1/2$. In any case, the target error $\mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big)$ should get dominated by $\mathbb{P}_{\underline{\nu}^{(k)}}\big(I_T \neq k\big)$ and the obtained bound is likely to be uninformative on the target error, due to the maximum in the left-hand side. This tradeoff seems to be unsolvable in general, unless there exist some specific properties for the Kullback-Leibler divergence of the model, as we illustrate below for a Gaussian model, which was the setting considered by Kaufmann et al. (2016, Section 5.2).

Another intuitive issue with the bound of Proposition 28 is that it involves Kullback-Leibler divergences with arguments in reverse order compared to the lower bounds presented in Section 4. Indeed, taking the supremum 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

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

rather than in terms of the $\mathcal{L}^{>}_{\inf}(\mu^\star, \nu_k)$. Our intuition, given all bounds presented in this article, is that the $\mathcal{K}^{>}_{\inf}(\nu_k, \mu^\star)$ would not form the correct notion of complexity for the fixed-budget best-arm identification.

How Kaufmann et al. (2016, Section 5.2) could exploit Proposition 28 in the Gaussian case. Yet, in the case of the model $\mathcal{D}_{\sigma^2}$ of Gaussian distributions with a fixed variance $\sigma^2 > 0$, for which KL is symmetric, Proposition 28 admits an interesting corollary, corresponding[2] to Theorem 16 of Kaufmann et al. (2016, Section 5.2). The corollary actually relies on a strong property of KL in this model: not only is it symmetric, but it only depends on the expectation gaps between its arguments. Namely, for all pairs $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{N}(\mu', \sigma^2)$ of distributions in $\mathcal{D}_{\sigma^2}$, for all $\Delta \in \mathbb{R}$,

$$\mathrm{KL}\big(\mathcal{N}(\mu, \sigma^2), \mathcal{N}(\mu', \sigma^2)\big) = \frac{(\mu - \mu')^2}{2\sigma^2} = \mathrm{KL}\big(\mathcal{N}(\mu + \Delta, \sigma^2), \mathcal{N}(\mu' + \Delta, \sigma^2)\big)\,. \tag{56}$$

We introduce the following short-hand notation:

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

Corollary 29 *For all sequences of 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 $\mathcal{D}_{\sigma^2}$, where each $\underline{\nu}^{(k)}$ admits k as a best arm and satisfies $C\big(\underline{\nu}^{(k)}\big) \leqslant C(\underline{\nu})$, and for which*

$$\frac{1}{T} \ln \max\left\{ \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big),\ \max_{k \neq a^\star(\underline{\nu})} \mathbb{P}_{\underline{\nu}^{(k)}}\big(I_T \neq k\big) \right\} \geqslant -4\, C(\underline{\nu})^{-1} - \frac{\ln 4}{T}\,.$$

The proof provided below is highly specific to the Gaussian model and exploits the gap-based rewriting (56) of the Kullback-Leibler divergence. The calculations led would only extend to models for which such gap-based rewritings of (upper and lower bounds on) the Kullback-Leibler divergence would be available.

To compare the result of Corollary 29 with the bound (8) stemming from Audibert et al. (2010), note that

$$C(\underline{\nu})^{-1} \geqslant \frac{2}{\sigma^2} \min_{2 \leqslant k \leqslant K} \frac{\Delta_{(k)}^2}{k}\,.$$

Proof We apply Proposition 28 with the distributions $\zeta_k = \mathcal{N}(\mu^\star + \Delta_k, \sigma^2)$, for $k \neq a^\star(\underline{\nu})$. On the one hand, the bound of Proposition 28 involves

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

2. The maximum of the left-hand side of Corollary 29 is present, but somewhat discrete, in the Theorem 16 of Kaufmann et al. (2016, Section 5.2): it corresponds to the "There exists an alternative bandit problem" part of the statement of the latter.

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

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

These two observations conclude the proof of Corollary 29. ■

Proof of Proposition 28. We conclude this section with a proof of Proposition 28. It relies on the Bretagnolle-Huber inequality (Bretagnolle and Huber, 1979), which states that, for all $p, q \in [0,1]$,

$$p + 1 - q \geqslant \frac{1}{2} \exp\Big(-\mathrm{KL}\big(\mathrm{Ber}(p), \mathrm{Ber}(q)\big)\Big)\,. \tag{57}$$

Proof We fix distributions ζ_k abiding by the conditions of the proposition and also fix $T \geqslant 1$. We will prove below that, for all convex weights $(u_b)_{b \neq a^\star(\underline{\nu})}$, i.e., non-negative weights summing up to 1,

$$\frac{1}{T} \ln \max\Big\{ \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big),\ \max_{k \neq a^\star(\underline{\nu})} \mathbb{P}_{\underline{\nu}^{(k)}}\big(I_T \neq k\big) \Big\} \geqslant - \max_{b \neq a^\star(\underline{\nu})} \big\{ u_b\, \mathrm{KL}(\nu_b, \zeta_b) \big\} - \frac{\ln 4}{T}\,, \tag{58}$$

from which Proposition 28 follows, by optimizing the obtained lower bound, i.e., by 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 convex weights $(u_b)_{b \neq a^\star(\underline{\nu})}$ and prove (58). As $b \neq a^\star(\underline{\nu})$ and b is the unique optimal arm of $\underline{\nu}^{(b)}$, for the first inequality, and by the Bretagnolle-Huber inequality (57), for the second inequality,

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

where $p_T \overset{\text{def}}{=} \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big)$ and $q_T \overset{\text{def}}{=} \mathbb{P}_{\underline{\nu}^{(b)}}\big(I_T \neq a^\star(\underline{\nu})\big)$. Inequality (48) reads, in the present case, as $\underline{\nu}$ and $\underline{\nu}^{(b)}$ only differ at arm b,

$$\mathrm{KL}\big(\mathrm{Ber}(p_T), \mathrm{Ber}(q_T)\big) \leqslant \mathbb{E}_{\underline{\nu}}\big[N_b(T)\big]\, \mathrm{KL}(\nu_b, \zeta_b)\,.$$

Using $\max\{u, v\} \geqslant (u+v)/2$ after collecting all bounds obtained so far yields

$$\max\Big\{ \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big), \mathbb{P}_{\underline{\nu}^{(b)}}\big(I_T \neq b\big) \Big\} \geqslant \frac{1}{4} \exp\Big(-\mathbb{E}_{\underline{\nu}}\big[N_b(T)\big]\, \mathrm{KL}(\nu_b, \zeta_b)\Big)\,.$$

We take the maxima over $b \neq a^\star(\underline{\nu})$ in both sides, apply logarithms, and conclude the proof of (58) by showing that

$$\min_{b \neq a^\star(\underline{\nu})} \Big\{ \mathbb{E}_{\underline{\nu}}\big[N_b(T)\big]\, \mathrm{KL}(\nu_b, \zeta_b) \Big\} \leqslant \max_{b \neq a^\star(\underline{\nu})} \big\{ u_b\, \mathrm{KL}(\nu_b, \zeta_b) \big\}\,. \tag{59}$$

Indeed,

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

so that there exists $b^\star \neq a^\star(\underline{\nu})$ such that $\mathbb{E}_{\underline{\nu}}\big[N_{b^\star}(T)\big]/T \leqslant u_{b^\star}$. We then have

$$\min_{b \neq a^\star(\underline{\nu})} \Big\{ \mathbb{E}_{\underline{\nu}}\big[N_b(T)\big] \, \mathrm{KL}(\nu_b, \zeta_b) \Big\} \leqslant u_{b^\star} \, \mathrm{KL}(\nu_{b^\star}, \zeta_{b^\star}) \leqslant \max_{b \neq a^\star(\underline{\nu})} \big\{ u_b \, \mathrm{KL}(\nu_b, \zeta_b) \big\} \,,$$

as desired in (59). ■