entropy

MDPI

Article

On Gap-Based Lower Bounding Techniques for Best-Arm Identification

Lan V. Truong [1,*] **and Jonathan Scarlett** [2]

[1] Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK
[2] Department of Computer Science & Department of Mathematics, National University of Singapore, Singapore 117418, Singapore; scarlett@comp.nus.edu.sg
[*] Correspondence: lt407@cam.ac.uk

Received: 16 June 2020; Accepted: 17 July 2020; Published: 20 July 2020

Abstract: In this paper, we consider techniques for establishing lower bounds on the number of arm pulls for best-arm identification in the multi-armed bandit problem. While a recent divergence-based approach was shown to provide improvements over an older gap-based approach, we show that the latter can be refined to match the former (up to constant factors) in many cases of interest under Bernoulli rewards, including the case that the rewards are bounded away from zero and one. Together with existing upper bounds, this indicates that the divergence-based and gap-based approaches are both effective for establishing sample complexity lower bounds for best-arm identification.

Keywords: multi-armed bandits; best-arm identification; information-theoretic lower bounds; PAC learning

1. Introduction

The multi-armed bandit (MAB) problem [1] provides a versatile framework for sequentially searching for high-reward actions, with applications including clinical trials [2], online advertising [3], adaptive routing [4], and portfolio design [5]. The best-arm identification problem seeks to find the arm with the highest mean using as few arm pulls as possible, and dates back to the works of Bechhofer [6] and Paulson [7]. More recently, several algorithms have been proposed for best-arm identification, including successive elimination [8], lower-upper confidence bound algorithms [9,10], PRISM [11], and gap-based elimination [12]. The latter establishes a sample complexity that is known to be optimal in the two-arm case [13], and more generally near-optimal.

Complementary to these upper bounds is information-theoretic lower bounds on the performance of any algorithm. Such bounds serve as a means to assess the degree of optimality of practical algorithms, and identify where further improvements are possible, thus focusing research towards directions that can have the greatest practical impact. Lower bounds were given by Mannor and Tsitsiklis [14] for Bernoulli bandits, and by Kaufmann et al. [15] for more general reward distributions. Both of these works were based on the difficulty of distinguishing bandit instances that differ in only a single arm distribution, but the subsequent analysis techniques differed significantly, with [14] using a direct change-of-measure analysis and introducing gap-based quantities equaling the difference between two arm means, and [15] using a form of the data processing inequality for KL divergence. We refer to these as the gap-based and divergence-based approaches, respectively. Further works on best-arm identification lower bounds include [16–18].

The divergence-based approach was shown in [15] to attain a stronger result than that of [14] with a simpler proof, as we outline in Section 2.2. In this paper, we address the question of whether the gap-based approach is fundamentally limited, or can be refined to attain a similar results to [15]. We show that the correct answer is the latter in many cases of interest, by suitable refinements of the analysis of [14]. The existing results and our results are presented in Section 2, and our analysis is presented in Section 3.

2. Overview of Results

2.1. Problem Setup

We consider the following setup:

- There are M arms with Bernoulli rewards; the means are $\mathbf{p} = (p_1, p_2, \cdots, p_M)$, and this set of means is said to define the bandit instance. Our analysis will consider instances with arms sorted such that $p_1 \geq p_2 \cdots \geq p_M$, without loss of generality.
- The agent would like to find an arm whose arm mean is within ϵ of the highest arm mean for some $0 < \varepsilon < 1$, i.e., $p_l > p_1 - \varepsilon$. Even if there are multiple such arms, just identifying one of them is good enough.
- In each round, the agent can pull any arm $l \in [M]$ and observe an reward $X_l^{(s)} \sim \text{Bernoulli}(p_l)$, where s is the number of times the l-th arm has been pulled so far. We assume that the rewards are independent, both across arms and across times.
- In each round, the agent can alternatively choose to terminate and output an arm index $\hat{l}$ believed to be ϵ-optimal. The index at which this occurs is denoted by T, and is a random variable because it is allowed to depend on the rewards observed. We are interested in the expected number of arm pulls (also called the sample complexity) $\mathbb{E}_{\mathbf{p}}[T]$ for a given instance $\mathbf{p}$, which should ideally be as low as possible.
- An algorithm is said to be (ε, δ)-PAC (Probably Approximately Correct) if, for all bandit instances, it outputs an ε-optimal arm with probability at least $1 - \delta$ when it terminates at the stopping time T.

We will frequently make use of some fundamental quantities. First, the best arm mean and the gap to the best arm are denoted by

$$p_* := p_1, \tag{1}$$

$$\Delta_l := p_* - p_l. \tag{2}$$

The set of ϵ-optimal arms and the set of ϵ-suboptimal arms are respectively given by

$$\mathcal{M}(\mathbf{p}, \varepsilon) := \{l \in [M] : p_l > p_* - \varepsilon\}, \tag{3}$$

$$\mathcal{N}(\mathbf{p}, \varepsilon) := \{l \in [M] : p_l \leq p_* - \varepsilon\}, \tag{4}$$

and we make use of the binary KL divergence function

$$\mathrm{KL}(p, q) := p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}, \tag{5}$$

where here and subsequently, $\log(\cdot)$ denotes the natural logarithm.

2.2. Existing Lower Bounds

For any fixed $\underline{p} \in (0, 1/2)$, Mannor and Tsitsiklis [14] showed that if an algorithm is (ϵ, δ)-PAC with respect to all instances with $\min_l p_l \geq \underline{p} > 0$, and if $\epsilon \leq \frac{1-p^*}{4}$ and $\delta \leq e^{-8}/8$, then for any constant $\alpha \in (0, 2)$, there exists $c_1 = O(\underline{p}^2)$ (depending on α) such that

$$\mathbb{E}_{\mathbf{p}}[T] \geq c_1 \left[\frac{\left(|\tilde{\mathcal{M}}(\mathbf{p}, \varepsilon)| - 1\right)^+}{\varepsilon^2} + \sum_{l \in \tilde{\mathcal{N}}(\mathbf{p}, \varepsilon)} \frac{1}{\Delta_l^2} \right] \log \frac{1}{8\delta} \tag{6}$$

where

$$\tilde{\mathcal{M}}(\mathbf{p},\varepsilon) = \mathcal{M}(\mathbf{p},\varepsilon) \cap \left\{ l \in [M] : p_l \geq \frac{\varepsilon + p_*}{2-\alpha} \right\}, \tag{7}$$

$$\tilde{\mathcal{N}}(\mathbf{p},\varepsilon) = \mathcal{N}(\mathbf{p},\varepsilon) \cap \left\{ l \in [M] : p_l \geq \frac{\varepsilon + p_*}{2-\alpha} \right\}. \tag{8}$$

Note that the subsets $\tilde{\mathcal{M}}(\mathbf{p},\varepsilon)$ and $\tilde{\mathcal{N}}(\mathbf{p},\varepsilon)$ do not always form a partition of the arms, i.e., it may hold that $\tilde{\mathcal{M}}(\mathbf{p},\varepsilon) \cup \tilde{\mathcal{N}}(\mathbf{p},\varepsilon) \subsetneq [M]$. The sets increase in size as α decreases, but implicitly this leads to a lower value of c_1. In addition, as we will see below, the $\underline{p}^2$ dependence entering via c_1 is not necessary.

We also note that the lower bound in (6) depends on the instance-specific quantities $\tilde{\mathcal{M}}(\mathbf{p},\varepsilon)$, $\tilde{\mathcal{N}}(\mathbf{p},\varepsilon)$, and Δ_l, and is thus an instance-dependent bound. On the other hand, the lower bound is only stated for (ε,δ)-PAC algorithms, and the PAC guarantee requires the algorithm to eventually succeed on any instance (subject to the assumptions given on p_l, ϵ, and δ).

Kaufmann et al. [15] improved Mannor and Tsitsiklis's lower bound by using a form of data processing inequality for KL divergence, leading to the following whenever $\delta \leq 0.15$ and $0 < \varepsilon < \min\{p_*, 1-p_*\}$ [15] (Remark 5):

$$\mathbb{E}_{\mathbf{p}}[T] \geq \left[\frac{|\mathcal{M}(\mathbf{p},\varepsilon)| - 1}{\mathrm{KL}(p_* - \varepsilon, p_* + \varepsilon)} + \sum_{l \in \mathcal{N}(\mathbf{p},\varepsilon)} \frac{1}{\mathrm{KL}(p_l, p_* + \varepsilon)} \right] \log \frac{1}{2.4\delta}. \tag{9}$$

To directly compare this result with (6), it is useful to apply the following inequality [19](Eq. (2.8)):

$$2(p-q)^2 \leq \mathrm{KL}(p,q) \leq \frac{(p-q)^2}{q(1-q)}, \tag{10}$$

which yields

$$\mathbb{E}_{\mathbf{p}}[T] \geq (p^* + \epsilon)(1 - p^* - \epsilon) \left[\frac{|\mathcal{M}(\mathbf{p},\varepsilon)| - 1}{4\epsilon^2} + \sum_{l \in \mathcal{N}(\mathbf{p},\varepsilon)} \frac{1}{(\varepsilon + \Delta_l)^2} \right] \log \frac{1}{2.4\delta}. \tag{11}$$

Even this weakened bound can significantly improve on (6), since (i) $\mathcal{M}(\mathbf{p},\varepsilon) \supset \tilde{\mathcal{M}}(\mathbf{p},\varepsilon)$ and $\mathcal{N}(\mathbf{p},\varepsilon) \supset \tilde{\mathcal{N}}(\mathbf{p},\varepsilon)$, (ii) the $\underline{p}^2$ dependence is replaced by $(p^* + \epsilon)(1 - p^* - \epsilon)$, so the dependence on the smallest arm mean is avoided (The $1 - p^* - \epsilon$ term is potentially small when ϵ is close to $1 - p^*$, but since (6) assumes $\epsilon \leq \frac{1-p^*}{4}$, we can still say that (11) is at least as good as (6)), and (iii) the assumption $\epsilon \leq \frac{1-p^*}{4}$ is avoided.

2.3. Our Result and Discussion

Our lower bound, stated in the following theorem, is developed based on Mannor and Tsitsiklis's analysis for best-arm identification [14] (Theorem 1), but uses novel refinements of the techniques therein to further optimize the bound (see Appendix C for an overview of these refinements).

Theorem 1. *For any bandit instance $\mathbf{p} \in (0, p_*]^M$ with $p_* \in (0,1)$, and any (ε,δ)-PAC algorithm with $0 < \varepsilon < 1 - p_*$ and $0 < \delta < \delta_0$ for some $\delta_0 < 1/4$, we have*

$$\mathbb{E}_{\mathbf{p}}[T] \geq \frac{2\gamma_0 (p_* + \varepsilon)(1 - p_* - \varepsilon)}{7(\xi + 1)} \left[\frac{|\mathcal{M}(\mathbf{p},\varepsilon)| - 1}{4\varepsilon^2} + \sum_{l \in \mathcal{N}(\mathbf{p},\varepsilon)} \frac{1}{(\varepsilon + \Delta_l)^2} \right] \log \frac{1 + 4\delta_0}{4\delta}, \tag{12}$$

where

$$\gamma_0 = \frac{1-4\delta_0}{8}, \tag{13}$$

$$\theta = \frac{2\delta}{1-4\gamma_0} = \frac{4\delta}{1+4\delta_0}, \tag{14}$$

and $\xi > 0$ *is the unique positive solution of the following quadratic equation:*

$$7\gamma_0\xi^2 \log\frac{1}{\theta} = 3(\xi+1). \tag{15}$$

Observe that this result matches (11) (with modified constants), and therefore exhibits the above benefit of depending on the full sets $\mathcal{M}$ and $\mathcal{N}$ without the condition $p_l \geq \frac{\varepsilon+p_*}{2-\alpha}$ (see (7)–(8)), as well as avoiding the dependence on $\underline{p}$, and permitting the broadest range of ϵ and δ among the above results.

The result (11) in turn matches (9) whenever the right-hand inequality in (10) is tight (i.e., whenever $\mathrm{KL}(p,q) = \Theta(\frac{(p-q)^2}{q(1-q)})$). This is clearly true when p and q (representing the arm means) are bounded away from zero and one, and also in certain limiting cases approaching these endpoints (e.g., when p and q both tend to one, but $\frac{1-p}{1-q} = \Theta(1)$). However, there are also limiting cases where the upper bound in (10) is not tight (e.g., $p = 1-\sqrt{\eta}$ and $q = 1-\eta$ as $\eta \to 0$), and in such cases, the bound (9) remains tighter than that of Theorem 1.

3. Proof of Theorem 1

We follow the general steps of (Theorem 5 [14]), but with several refinements to improve the final bound. The main differences are outlined in Appendix C.

Step 1: Defining a Hypothesis Test

Let us denote the true (unknown) expected reward of each arm by Q_l for all $l \in [M]$. Similarly to [14,15], we consider M hypotheses as follows:

$$H_1 : Q_l = p_l, \quad \forall l \in [M], \tag{16}$$

and for each $l \neq 1$,

$$H_l : Q_l = p_* + \varepsilon, \quad Q_{l'} = p_{l'} \quad \forall l' \in [M] \setminus \{l\}. \tag{17}$$

If hypothesis H_l is true, the (ϵ,δ)-PAC algorithm must return arm l with probability at least $1-\delta$. We will bound sample complexity when the hypothesis H_l is true. We denote by $\mathbb{E}_l$ and $\mathbb{P}_l$ the expectation and probability, respectively, under hypothesis H_l.

Let B_l be the event that the algorithm returns arm l. Since $\sum_{l\in\mathcal{M}(\mathbf{p},\varepsilon)} \mathbb{P}_1(B_l) \leq 1$ and $|\mathcal{M}(\mathbf{p},\varepsilon)| \geq 1$, there is at most one arm $l_0 \in \mathcal{M}(\mathbf{p},\varepsilon)$ that satisfies $\mathbb{P}_1(B_{l_0}) > \frac{1}{2}$. Defining

$$\mathcal{M}_0(\mathbf{p},\varepsilon) := \left\{ l \in \mathcal{M}(\mathbf{p},\varepsilon) : \mathbb{P}_1[B_l] \leq \frac{1}{2} \right\} = \{ l \in \mathcal{M}(\mathbf{p},\varepsilon) : l \neq l_0 \}, \tag{18}$$

it follows that

$$|\mathcal{M}_0(\mathbf{p},\varepsilon)| \geq (|\mathcal{M}(\mathbf{p},\varepsilon)| - 1)^+. \tag{19}$$

Define

$$\mathcal{T}(\mathbf{p},\varepsilon) := \mathcal{M}_0(\mathbf{p},\varepsilon) \cup \mathcal{N}(\mathbf{p},\varepsilon), \tag{20}$$

as well as

$$B_{\mathcal{M}(\mathbf{p},\varepsilon)} := \bigcup_{l \in \mathcal{M}(\mathbf{p},\varepsilon)} B_l, \tag{21}$$

which is the event that the policy eventually select an arm in the ε-neighborhood of the best arm in $[M]$. Since the policy is (ε, δ)-correct with $\delta < \delta_0$, we must have

$$\mathbb{P}_1[B_{\mathcal{M}(\mathbf{p},\varepsilon)}] \geq 1 - \delta > 1 - \delta_0, \tag{22}$$

and it follows from (18) and (22) that

$$\begin{aligned} \mathbb{P}_1[B_l] &\leq \max\{\delta_0, 1/2\} & (23)\\ &= \frac{1}{2} & (24) \end{aligned}$$

for all $l \in \mathcal{T}(\mathbf{p}, \varepsilon)$.

Step 2: Bounding the Number of Pulls of Each Arm

Before proceeding, we make some additional definitions:

$$\begin{aligned} \alpha_l &:= \frac{\varepsilon + \Delta_l}{(1-p_l)p_l}, & (25)\\ \beta_l &:= \alpha_l \sqrt{\frac{1-p_l}{1-(p_* + \varepsilon)}}, & (26)\\ \tilde{\alpha}_l &:= \alpha_l - \frac{4}{3}\big(\alpha_l(1-p_l)\big)^2, & (27)\\ \tilde{\beta}_l &:= \beta_l - \frac{4}{3}\big(\alpha_l(1-p_l)\big)^2, & (28) \end{aligned}$$

The definitions (27) and (28) will only be used for arms with $\frac{\varepsilon+\Delta_l}{p_l} \leq \frac{1}{2}$, and for such arms, we will establish in the analysis that $\tilde{\alpha}_l \geq 0$ and $\tilde{\beta}_l \geq 0$.

We prove the following lemma, characterizing the probability of a certain event in which (i) the number of pulls of some arm $l \in \mathcal{T}(\mathbf{p}, \varepsilon)$ falls below a suitable threshold (event A_l below), (ii) a deviation bound holds regarding the number of observed 1's from pulling arm l (event C_l below), and (iii) arm l is not returned (event B_l^c).

Lemma 1. *For each $l \in [M]$, let T_l be the total number of times that arm l is pulled under the (ε, δ)-correct policy. Let $K_l = X_l^{(1)} + X_l^{(2)} + \cdots + X_l^{(T_l)}$ be the total number of unit rewards obtained from pulling the arm l up to the T_l-th time. Let*

$$\begin{aligned} G_{1,l} &:= \frac{7p_l^2\alpha_l^2(1-p_l)^2}{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}} T_l, & (29)\\ G_{2,l} &:= \Big[\tilde{\beta}_l(p_lT_l - K_l)\mathbf{1}\{p_lT_l > K_l\} + \tilde{\alpha}_l(p_lT_l - K_l)\mathbf{1}\{p_lT_l \leq K_l\}\Big]\mathbf{1}\Big\{\frac{\varepsilon+\Delta_l}{p_l} \leq \frac{1}{2}\Big\}, & (30) \end{aligned}$$

where α_l, $\tilde{\alpha}_l$, and $\tilde{\beta}_l$ are defined in (25), (27), and (28), respectively. Let

$$\nu_l := (\xi + 1)\left(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\right), \tag{31}$$

where ξ is defined in (15)*. Define the following events:*

$$A_l := \left\{ G_{1,l} \le \frac{1}{\nu_l} \log \frac{1}{\theta} \right\}, \tag{32}$$

$$C_l = \left\{ G_{2,l} \le \frac{\xi}{\nu_l} \left(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}} \right) \log \frac{1}{\theta} \right\}, \tag{33}$$

$$S_l := A_l \cap B_l^c \cap C_l. \tag{34}$$

If $l \in \mathcal{T}(\mathbf{p}, \varepsilon)$ (see (20)*), then under the condition that*

$$\mathbb{E}_1\left[G_{1,l}\right] < \frac{\gamma_0}{\nu_l} \log \frac{1}{\theta}, \tag{35}$$

we have

$$\mathbb{P}_1\left[S_l\right] > \frac{1-4\gamma_0}{2}. \tag{36}$$

Proof. See Appendix A. □

Intuitively, A_l is the event that the total number of times that arm l is pulled is small, and C_l is the event that $|p_l T_l - K_l|$ is not too large (since pulling an arm T_l times should produce roughly $p_l T_l$ ones). The lemma indicates that if $\mathbb{E}[T_l]$ is not too large, then $\mathbb{P}[A_l \cap B_l^c \cap C_l]$ is lower bounded, and this will ultimately lead to a lower bound on $\mathbb{P}[B_l^c]$, the event of primary interest.

In Lemma 2 below, we will use Lemma 1 to deduce a lower bound on $\mathbb{E}_1[G_{1,l}]$, which amounts to a lower bound on the average number of arm pulls by the definition of $G_{1,l}$. Before doing so, we introduce a likelihood ratio that will be used in a change-of-measure argument [14].

For any given time $t \ge 1$ and $l \in [M]$, let $T_l(t)$ be the total number of times that arm l is pulled by time t. Define

$$X_l^{T_l(t)} := \{X_l^{(1)}, X_l^{(2)}, \cdots, X_l^{(T_l(t))}\}, \tag{37}$$

and let

$$\mathcal{F}_t := \sigma(X_1^{T_1(t)}, X_2^{T_2(t)}, \cdots, X_M^{T_M(t)}) \tag{38}$$

be the σ-algebra generated by $X_1^{T_1(t)}, X_2^{T_2(t)}, \ldots, X_M^{T_M(t)}$ for all $t = 1, 2, \ldots$.

Recall that T is the stopping time of the algorithm, and that $T_l := T_l(T)$ for all $l \in [M]$. Moreover, let $W = \mathcal{F}_T$ be the entire history up to the stopping time T. We define the following likelihood ratio:

$$L_l(w) = \frac{\mathbb{P}_l(W = w)}{\mathbb{P}_1(W = w)} \tag{39}$$

for every possible history w. Moreover, we let $L_l(W)$ denote the corresponding random variable. Given the history up to time $T-1$ (i.e., $\mathcal{F}_{T-1}$), the arm reward at time T has the same probability distribution under H_1 and H_l unless the chosen arm is arm l. Therefore, we have

$$L_l(W) = \frac{(p_* + \varepsilon)^{K_l}(1-p_*-\varepsilon)^{T_l-K_l}}{p_l^{K_l}(1-p_l)^{T_l-K_l}}, \tag{40}$$

where $K_l := X_l^{(1)} + X_l^{(2)} + \cdots + X_l^{(T_l)}$ (or the total number of 1's in the T_l pulls of the arm l).

The following proposition presents one of our key technical results towards establishing the lower bound. We use the definitions in (1)–(5), along with (25)–(28).

Proposition 1. *Fix the bandit instance* $\mathbf{p}$, *the parameter* $0 < \varepsilon < 1 - p_*$, *and the history* W *with corresponding values* K_l *and* T_l. *Recalling the definitions of* α_l, $\tilde{\alpha}_l$, *and* $\tilde{\beta}_l$ *in* (25), (27), *and* (28), *respectively, we have*

$$L_l(W) \geq \exp\big(-G_{1,l} - G_{2,l}\big), \tag{41}$$

where $G_{1,l}$ *and* $G_{2,l}$ *are defined in* (29)–(30).

Proof. See Appendix B. □

Based on Lemma 1 and Proposition 1, we obtain the following extension of [14] (Lemma 6) lower bounding the average of each $G_{1,l}$; this lower bound will later be translated to a lower bound on the number of arm pulls T_l.

Lemma 2. *For any arm* $l \in \mathcal{T}(\mathbf{p}, \varepsilon)$, *the following holds:*

$$\mathbb{E}_1[G_{1,l}] \geq \frac{\gamma_0}{\nu_l} \log \frac{1}{\theta}, \tag{42}$$

where θ *and* ν_l *are defined in* (14) *and* (31), *respectively.*

Proof. We use a proof by contradiction. Assume that

$$\mathbb{E}_1[G_{1,l}] < \frac{\gamma_0}{\nu_l} \log \frac{1}{\theta}, \tag{43}$$

then by Lemma 1, Equation (36) holds. Moreover, by Proposition 1, we have

$$L_l(W) \geq \exp\big(-G_{1,l} - G_{2,l}\big), \tag{44}$$

and recalling the definition of S_l in (34), it follows from (44) that

$$\begin{aligned} L_l(W)\mathbf{1}_{S_l} &\geq \exp\big(-G_{1,l} - G_{2,l}\big)\mathbf{1}_{S_l} && (45)\\ &\geq \exp\left(-\frac{1}{\nu_l}\left[\xi\left(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\right)+1\right]\log\frac{1}{\theta}\right)\mathbf{1}_{S_l} && (46)\\ &\geq \exp\left(-\frac{1}{\nu_l}\left[(\xi+1)\left(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\right)\right]\log\frac{1}{\theta}\right)\mathbf{1}_{S_l} && (47) \end{aligned}$$

where (46) follows from the definitions in (32)–(33), and (47) follows from the fact that $1 - p_l \geq 1 - p_* \geq 1 - p_* - \varepsilon$ for all $l \in [M]$.

By the choice of $\nu_l > 0$ given in (31), it holds that

$$(\xi+1)\left(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\right)\frac{1}{\nu_l} = 1. \tag{48}$$

Hence, from (47) and (48), we have

$$L_l(W)\mathbf{1}_{S_l} \geq \theta\mathbf{1}_{S_l} = \frac{2\delta}{1-4\gamma_0}\mathbf{1}_{S_l}, \tag{49}$$

for all $l \in \mathcal{T}(\mathbf{p}, \varepsilon)$, by the definition of θ in (14).

We are now ready to complete the proof:

$$\begin{aligned}
\mathbb{P}_l[B_l^c] &\geq \mathbb{P}_l[S_l] && (50)\\
&= \mathbb{E}_l[\mathbf{1}_{S_l}] && (51)\\
&= \mathbb{E}_1\left[L_l(W)\mathbf{1}_{S_l}\right] && (52)\\
&\geq \mathbb{E}_1\left[\frac{2\delta}{1-4\gamma_0}\mathbf{1}_{S_l}\right] && (53)\\
&= \frac{2\delta}{1-4\gamma_0}\mathbb{P}_1[S_l] && (54)\\
&> \frac{2\delta}{1-4\gamma_0}\left(\frac{1-4\gamma_0}{2}\right) && (55)\\
&= \delta, && (56)
\end{aligned}$$

where (50) follows from the definition of set S_l in (34), (52) follows by a standard change of measure [20], (53) follows from (49), and (55) follows from (36) of Lemma 1 (recall that we assumed (43)).

The inequality (56) shows a contradiction to the fact that under H_l, the (ε, δ)-correct bandit policy must return the arm l with probability at least $1 - \delta$, i.e., $\mathbb{P}_l(B_l^c) \leq \delta$. This concludes the proof. □

From Lemma 2 and the definition of $G_{1,l}$ in (29), it holds that

$$\mathbb{E}_1\left[\frac{7\alpha_l^2 p_l^2(1-p_l)^2}{2(p_* + \varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}T_l\right] \geq \frac{\gamma_0}{\nu_l}\log\frac{1}{\theta} \tag{57}$$

for all $l \in \mathcal{T}(\mathbf{p}, \varepsilon)$. Hence, and using the definition of ν_l in (31), we have

$$\begin{aligned}
\mathbb{E}_1\left[\alpha_l^2 p_l^2(1-p_l)^2 T_l\right] &\geq \frac{2\gamma_0(p_*+\varepsilon)\sqrt{(1-p_l)(1-p_*-\varepsilon)}}{7(1+\xi)}\left(\sqrt{\frac{1-p_*-\varepsilon}{1-p_l}}\right)\log\frac{1}{\theta} && (58)\\
&= \frac{2\gamma_0(p_*+\varepsilon)(1-p_*-\varepsilon)}{7(1+\xi)}\log\frac{1}{\theta}. && (59)
\end{aligned}$$

Step 3: Deducing a Lower Bound on the Sample Complexity

For any arm $l \in \mathcal{T}(\mathbf{p}, \varepsilon)$, by the definition of α_l in (25), we have

$$\alpha_l^2 p_l^2(1-p_l)^2 = (\varepsilon + \Delta_l)^2. \tag{60}$$

Note that $0 \leq \Delta_l < \varepsilon$ for all $l \in \mathcal{M}_0(\mathbf{p}, \varepsilon)$, since $\mathcal{M}_0 \subseteq \mathcal{M}$, the set of ϵ-optimal arms. Therefore, we can further simplify (60) to

$$\alpha_l^2 p_l^2(1-p_l)^2 \leq 4\epsilon^2 \tag{61}$$

for $l \in \mathcal{M}_0(\mathbf{p}, \varepsilon)$.

Substituting (60)–(61) into (59), we obtain

$$\mathbb{E}_{\mathbf{p}}[T] = \mathbb{E}_{\mathbf{p}}\left[\sum_{l=1}^{M} T_l\right] \tag{62}$$

$$\geq \mathbb{E}_{\mathbf{p}}\left[\sum_{l\in\mathcal{M}_0(\mathbf{p},\varepsilon)} T_l\right] + \mathbb{E}_{\mathbf{p}}\left[\sum_{l\in\mathcal{N}(\mathbf{p},\varepsilon)} T_l\right] \tag{63}$$

$$\geq \frac{2\gamma_0(p_*+\varepsilon)(1-p_*-\varepsilon)}{7(\xi+1)}\left[\frac{|\mathcal{M}_0(\mathbf{p},\varepsilon)|}{4\varepsilon^2} + \sum_{l\in\mathcal{N}(\mathbf{p},\varepsilon)} \frac{1}{(\varepsilon+\Delta_l)^2}\right]\log\frac{1}{\theta} \tag{64}$$

$$= \frac{2\gamma_0(p_*+\varepsilon)(1-p_*-\varepsilon)}{7(\xi+1)}\left[\frac{|\mathcal{M}_0(\mathbf{p},\varepsilon)|}{4\varepsilon^2} + \sum_{l\in\mathcal{N}(\mathbf{p},\varepsilon)} \frac{1}{(\varepsilon+\Delta_l)^2}\right]\log\frac{1+4\delta_0}{4\delta}, \tag{65}$$

where (65) uses the definition of θ in (14). Finally, we obtain (12) from (65) and (19).

4. Conclusion

We have presented a refined analysis of best-arm identification following the gap-based approach of [14], but incorporating refinements that circumvent some weaknesses, leading to a bound matching the divergence-based approach [15] in many cases. It would be of interest to determine whether further refinements could allow this approach to match [15] in all cases, or the extent to which the gap-based approach extends beyond Bernoulli rewards and/or beyond the standard best-arm identification problem (e.g., to ranking problems [21]).

Appendix A. Proof of Lemma 1 (Constant-Probability Event for Small Enough $\mathbb{E}_1[G_{1,l}]$)

As in [14], the proof of this lemma is based on Kolmogorov's maximum inequality [22].

Lemma A1. *(Kolmogorov's Theorem [22]) Let $Y_1, Y_2, \cdots, Y_n : \Omega \to \mathbb{R}$ be independent random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with expectation $\mathbb{E}[Y_k] = 0$ and variance $\mathsf{Var}[Y_k] < \infty$ for $k = 1, 2, \cdots, n$. Then, for each $\lambda > 0$,*

$$\mathbb{P}\left[\max_{1\leq k\leq n} |S_k| \geq \lambda\right] \leq \frac{1}{\lambda^2}\mathsf{Var}[S_n] = \frac{1}{\lambda^2}\sum_{k=1}^{n}\mathbb{E}[Y_k^2], \tag{A1}$$

where $S_k = Y_1 + Y_2 + \cdots + Y_k$.

We start by simplifying the main assumption of the lemma:

$$\frac{\gamma_0}{\nu_l}\log\frac{1}{\theta} > \mathbb{E}_1[G_{1,l}] \tag{A2}$$

$$\geq \left(\frac{1}{\nu_l}\log\frac{1}{\theta}\right)\mathbb{P}_1\left[G_{1,l} > \frac{1}{\nu_l}\log\frac{1}{\theta}\right] \tag{A3}$$

$$= \left(\frac{1}{\nu_l}\log\frac{1}{\theta}\right)\mathbb{P}_1[A_l^c], \tag{A4}$$

where (A3) follows from Markov's inequality, and (A4) follows from the definition of A_l in (32).

It follows from (A4) that

$$\mathbb{P}_1[A_l] \geq 1 - \gamma_0. \tag{A5}$$

We define

$$\mathcal{V} = \left\{ l \in [M] : \frac{\varepsilon + \Delta_l}{p_l} \leq \frac{1}{2} \right\} \tag{A6}$$

and we will find it convenient to treat the cases $l \in \mathcal{V}$ and $ł \notin \mathcal{V}$ separately. For $l \notin \mathcal{V}$, from (30) and (33), we have $A_l \cap C_l = A_l$ since $\theta \in (0,1), \xi > 0$, and $G_{2,l} = 0$, and it immediately follows from (A5) that

$$\mathbb{P}_1 \left[A_l \cap C_l \right] \geq 1 - \gamma_0. \tag{A7}$$

On the other hand, for $l \in \mathcal{V}$, we can simplify the definition of $\tilde{\alpha}_l$ in (27) as follows:

$$\begin{aligned}
\tilde{\alpha}_l &= \alpha_l - \frac{4}{3}\big(\alpha_l(1-p_l)\big)^2 && \text{(A8)}\\
&= \alpha_l - \frac{4}{3}\big(\alpha_l(1-p_l)\big)\big(\alpha_l(1-p_l)\big) && \text{(A9)}\\
&= \alpha_l - \frac{4}{3}\big(\alpha_l(1-p_l)\big)\left(\frac{\varepsilon + \Delta_l}{p_l}\right) && \text{(A10)}\\
&\geq \alpha_l - \frac{2}{3}\big(\alpha_l(1-p_l)\big) && \text{(A11)}\\
&\geq \alpha_l - \frac{2}{3}\alpha_l && \text{(A12)}\\
&= \frac{1}{3}\alpha_l && \text{(A13)}\\
&> 0, && \text{(A14)}
\end{aligned}$$

where (A10) follows from (25), and (A11) follows from the definition of the set $\mathcal{V}$ in (A6). It follows that

$$0 < \tilde{\alpha}_l \leq \alpha_l \leq \beta_l \tag{A15}$$

for all $l \in \mathcal{V}$, where the second inequality in (A15) follows from $p_l \leq p_* \leq p_* + \varepsilon$ and the definitions of α_l and β_l in (25) and (26), respectively.

Similarly, for $l \in \mathcal{V}$, we can simplify $\tilde{\beta}_l$ from (28) as follows:

$$\begin{aligned}
\tilde{\beta}_l &= \beta_l - \frac{4}{3}\big(\alpha_l(1-p_l)\big)^2 && \text{(A16)}\\
&\geq \alpha_l - \frac{4}{3}\big(\alpha_l(1-p_l)\big)^2 && \text{(A17)}\\
&= \tilde{\alpha}_l && \text{(A18)}\\
&> 0, && \text{(A19)}
\end{aligned}$$

where (A17) follows from (A15), (A18) follows from (27), and (A19) again uses (A15). It follows that

$$0 < \tilde{\beta}_l \leq \beta_l \tag{A20}$$

for all $l \in \mathcal{V}$.

Now, let

$$Z_l^{(j)} := \beta_l (X_l^{(j)} - p_l), \quad j = 1, 2, \cdots \tag{A21}$$

Then, we have

$$\mathbb{E}_1[Z_l^{(j)}] = \mathbb{E}_1\left[\beta_l (X_l^{(j)} - p_l)\right] = \beta_l \mathbb{E}_1\left[X_l^{(j)} - p_l\right] = 0. \tag{A22}$$

In addition, we note that $Z_l^{(1)}, Z_l^{(2)}, \cdots$, are a i.i.d. sequence by the i.i.d. property of $X_l^{(1)}, X_l^{(2)}, \cdots$.

For each positive integer t_l, let $K_{l,t_l} := \sum_{j=1}^{t_l} X_l^{(j)}$, and define

$$U_l := \beta_l(K_l - p_l T_l), \tag{A23}$$
$$V_l(t_l) := \beta_l(K_{l,t_l} - p_l t_l) \tag{A24}$$
$$= \sum_{j=1}^{t_l} \beta_l(X_l^{(j)} - p_l) \tag{A25}$$
$$= \sum_{j=1}^{t_l} Z_l^{(j)}. \tag{A26}$$

Observe that

$$\sum_{j=1}^{t_l} \mathbb{E}_1\big[(Z_l^{(j)})^2\big] = \sum_{j=1}^{t_l} \beta_l^2 \mathbb{E}_1\big[(X_l^{(j)} - p_l)^2\big] \tag{A27}$$
$$= \sum_{j=1}^{t_l} \beta_l^2 p_l(1-p_l) \tag{A28}$$
$$= t_l \beta_l^2 p_l (1-p_l), \tag{A29}$$

where (A28) follows since a Bernoulli(ρ) variable has variance $\rho(1-\rho)$.

We are now ready to upper bound $\mathbb{P}_1[C_l^c \cap A_l]$ for $l \in \mathcal{V}$:

$$\mathbb{P}_1[C_l^c \cap A_l]$$
$$= \mathbb{P}_1\bigg[\bigg\{\tilde{\beta}_l(p_l T_l - K_l)\mathbf{1}\{p_l T_l > K_l\} + \tilde{\alpha}_l(p_l T_l - K_l)\mathbf{1}\{p_l T_l \le K_l\} > \frac{\xi}{\nu_l}\bigg(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\bigg)\log\frac{1}{\theta}\bigg\} \cap A_l\bigg] \tag{A30}$$
$$\le \mathbb{P}_1\bigg[\bigg\{\tilde{\beta}_l|p_l T_l - K_l|\mathbf{1}\{p_l T_l > K_l\} + \tilde{\alpha}_l|p_l T_l - K_l|\mathbf{1}\{p_l T_l \le K_l\} > \frac{\xi}{\nu_l}\bigg(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\bigg)\log\frac{1}{\theta}\bigg\} \cap A_l\bigg] \tag{A31}$$
$$\le \mathbb{P}_1\bigg[\bigg\{\beta_l|p_l T_l - K_l| > \frac{\xi}{\nu_l}\bigg(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\bigg)\log\frac{1}{\theta}\bigg\} \cap A_l\bigg] \tag{A32}$$
$$= \mathbb{P}_1\bigg[\bigg\{|U_l| > \frac{\xi}{\nu_l}\bigg(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\bigg)\log\frac{1}{\theta}\bigg\} \cap \bigg\{T_l < \frac{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}{7\nu_l\alpha_l^2 p_l^2(1-p_l)^2}\log\frac{1}{\theta}\bigg\}\bigg] \tag{A33}$$
$$\le \mathbb{P}_1\bigg[\max_{t_l \le \frac{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}\log\frac{1}{\theta}}{7\nu_l\alpha_l^2 p_l^2(1-p_l)^2}} |V_l(t_l)| > \frac{\xi}{\nu_l}\bigg(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\bigg)\log\frac{1}{\theta}\bigg] \tag{A34}$$
$$= \mathbb{P}_1\bigg[\max_{t_l \le \frac{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}\log\frac{1}{\theta}}{7\nu_l\alpha_l^2 p_l^2(1-p_l)^2}} \bigg|\sum_{j=1}^{t_l} Z_j\bigg| > \frac{\xi}{\nu_l}\bigg(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\bigg)\log\frac{1}{\theta}\bigg], \tag{A35}$$

where:

- (A30) uses the definitions of C_l and $G_{2,l}$;
- (A32) follows from (A15) and (A20), along with $\mathbf{1}\{p_l T_l > K_l\} + \mathbf{1}\{p_l T_l \le K_l\} = 1$;
- (A33) uses the definitions of U_l and A_l;
- (A34) follows from the definitions of U_l and $V_l(t_l)$ in (A23) and (A24) (which imply $U_l = V_l(T_l)$);
- (A35) follows from (A26);

Defining $n_l = \frac{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}\log\frac{1}{\theta}}{7\nu_l\alpha_l^2 p_l^2(1-p_l)^2}$ for brevity, we continue from (A35) as follows:

$$
\begin{aligned}
\mathbb{P}_1[C_l^c \cap A_l] &\le \max_{t_l \le n_l} \frac{\sum_{j=1}^{t_l} \mathbb{E}_1\left[Z_j^2\right]}{\left(\frac{\xi}{\nu_l}\left(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\right)\log\frac{1}{\theta}\right)^2} && \text{(A36)}\\
&= \max_{t_l \le n_l} \frac{p_l(1-p_l)t_l}{\left(\frac{\xi}{\nu_l}\left(\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\right)\log\frac{1}{\theta}\right)^2}\beta_l^2 && \text{(A37)}\\
&\le \frac{2\nu_l(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}{7\xi^2\left(\frac{1-p_l}{1-p_*-\varepsilon}\right)p_l(1-p_l)\log\frac{1}{\theta}}\left(\frac{\beta_l}{\alpha_l}\right)^2 && \text{(A38)}\\
&= \frac{2\nu_l(p_*+\varepsilon)\sqrt{\frac{1-(p_*+\varepsilon)}{1-p_l}}}{7\xi^2\left(\frac{1-p_l}{1-p_*-\varepsilon}\right)p_l\log\frac{1}{\theta}}\left(\frac{\beta_l}{\alpha_l}\right)^2 && \text{(A39)}\\
&\le \frac{3\nu_l\sqrt{\frac{1-(p_*+\varepsilon)}{1-p_l}}}{7\xi^2\left(\frac{1-p_l}{1-p_*-\varepsilon}\right)\log\frac{1}{\theta}}\left(\frac{\beta_l}{\alpha_l}\right)^2 && \text{(A40)}\\
&= \frac{3\nu_l\sqrt{\frac{1-(p_*+\varepsilon)}{1-p_l}}}{7\xi^2\left(\frac{1-p_l}{1-p_*-\varepsilon}\right)\log\frac{1}{\theta}}\left(\frac{1-p_l}{1-p_*-\varepsilon}\right) && \text{(A41)}\\
&= \frac{3(\xi+1)}{7\xi^2\log\frac{1}{\theta}} && \text{(A42)}\\
&= \gamma_0, && \text{(A43)}
\end{aligned}
$$

where:

- (A36) follows from Lemma A1 with $n = n_l$ and $k = t_l$;
- (A37) follows from (A29);
- (A38) follows from the definition of n_l;
- (A40) follows since the condition $\frac{\varepsilon+\Delta_l}{p_l} \le \frac{1}{2}$ in $\mathcal{V}$ yields $\frac{\varepsilon+p_*-p_l}{p_l} \le \frac{1}{2}$, which implies

$$p_l \ge \frac{2}{3}(p_*+\varepsilon). \tag{A44}$$

  for all $l \in \mathcal{V}$;
- (A41) follows from the definitions of α_l and β_l in (25)–(26);
- (A42) follows from the definition of ν_l in (31);
- (A43) follows from the definition of ξ in (15).

Combining (A5) and (A43), it follows that

$$
\begin{aligned}
\mathbb{P}_1[C_l \cap A_l] &= \mathbb{P}_1[A_l] - \mathbb{P}_1[C_l^c \cap A_l] && \text{(A45)}\\
&\ge 1 - 2\gamma_0 && \text{(A46)}
\end{aligned}
$$

for all $l \in \mathcal{V}$, and from (A7) and (A46), we obtain

$$\mathbb{P}_1[C_l \cap A_l] \ge 1 - 2\gamma_0 \tag{A47}$$

for all $l \in \mathcal{T}(\mathbf{p}, \varepsilon)$.

Finally, recall the definition of S_l in (34). From (24) and (A47), and using the union bound, we have

$$\mathbb{P}_1[S_l] > 1 - \left(2\gamma_0 + \frac{1}{2}\right) \tag{A48}$$
$$= \frac{1-4\gamma_0}{2} \tag{A49}$$

for all $l \in \mathcal{T}(\mathbf{p}, \varepsilon)$, as desired.

Appendix B. Proof of Proposition 1 (Bounding a Likelihood Ratio)

We first state the following lemma, which can easily be verified graphically, or proved using basic calculus.

Lemma A2. *For any $x \in [0,1)$, the following holds:*

$$1 - x \geq \exp\left(-\frac{x}{\sqrt{1-x}}\right). \tag{A50}$$

To prove Proposition 1, we consider two cases:

- **Case 1:** $\frac{\varepsilon+\Delta_l}{p_l} > \frac{1}{2}$. In this case, recalling that $\Delta_l = p_* - p_l$, we have

$$\frac{\varepsilon + p_*}{p_l} = \frac{\varepsilon+\Delta_l}{p_l} + 1 > \frac{3}{2} > 1. \tag{A51}$$

On the other hand, since $\varepsilon + p_l \leq \varepsilon + p_* < 1$, we have

$$0 < \frac{\varepsilon+\Delta_l}{1-p_l} = \frac{\varepsilon + p_* - p_l}{1-p_l} \tag{A52}$$
$$= 1 - \frac{1-(p_*+\varepsilon)}{1-p_l}, \tag{A53}$$

and applying Lemma A2 gives

$$\frac{1-\varepsilon-p_*}{1-p_l} = 1 - \frac{\varepsilon+\Delta_l}{1-p_l} \tag{A54}$$
$$\geq \exp\left[-\left(\sqrt{\frac{1-p_l}{1-(p_*+\varepsilon)}}\right)\left(\frac{\varepsilon+\Delta_l}{1-p_l}\right)\right] \tag{A55}$$
$$= \exp\left(-\frac{\varepsilon+\Delta_l}{\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}\right). \tag{A56}$$

Moreover, by the definition of α_l in (25), we have

$$\alpha_l = \frac{\varepsilon+\Delta_l}{(1-p_l)p_l} > \frac{1}{2(1-p_l)}, \tag{A57}$$

since $\frac{\varepsilon+\Delta_l}{p_l} > \frac{1}{2}$. It follows from (A57) that

$$\alpha_l < 2\alpha_l^2(1-p_l). \tag{A58}$$

In addition, again using $\frac{\varepsilon+\Delta_l}{p_l} > \frac{1}{2}$, we have

$$p_l < 2(\varepsilon + \Delta_l), \tag{A59}$$

and hence

$$p_* + \varepsilon = (\varepsilon + \Delta_l) + p_l \tag{A60}$$
$$< 3(\varepsilon + \Delta_l). \tag{A61}$$

We can now lower bound the likelihood ratio $L_l(W)$ as follows:

$$L_l(W) = \left(\frac{\varepsilon + p_*}{p_l}\right)^{K_l} \left(\frac{1-\varepsilon-p_*}{1-p_l}\right)^{T_l - K_l} \tag{A62}$$
$$\geq \exp\left(- \frac{\varepsilon + \Delta_l}{\sqrt{(1-p_l)(1-(p_*+\varepsilon))}} (T_l - K_l) \right) \tag{A63}$$
$$\geq \exp\left(- \frac{\varepsilon + \Delta_l}{\sqrt{(1-p_l)(1-(p_*+\varepsilon))}} T_l \right) \tag{A64}$$
$$= \exp\left(- \frac{(p_*+\varepsilon)(\varepsilon + \Delta_l)}{(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}} T_l \right) \tag{A65}$$
$$\geq \exp\left(- \frac{3(\varepsilon + \Delta_l)^2}{(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}} T_l \right) \tag{A66}$$
$$= \exp\left(- \frac{3\alpha_l^2 p_l^2 (1-p_l)^2}{(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}} T_l \right), \tag{A67}$$

where (A63) follows from (A51) and (A56), (A66) follows from (A61), and (A67) follows by the definition of α_l in (25). Hence, (41) holds for this case in which $G_{2,l} = 0$ (and also using $3 \leq \frac{7}{2}$).

- **Case 2:** $0 \leq \frac{\varepsilon+\Delta_l}{p_l} \leq \frac{1}{2}$**.** For this case, we have

$$L_l(W) = \left(\frac{\varepsilon + p_*}{p_l}\right)^{K_l} \left(\frac{1-\varepsilon-p_*}{1-p_l}\right)^{T_l - K_l} \tag{A68}$$
$$= \left(1 + \frac{\varepsilon + \Delta_l}{p_l}\right)^{K_l} \left(1 - \frac{\varepsilon + \Delta_l}{1-p_l}\right)^{T_l - K_l} \tag{A69}$$
$$= \left(1 - \left(\frac{\varepsilon + \Delta_l}{p_l}\right)^2\right)^{K_l} \left(1 - \frac{\varepsilon + \Delta_l}{p_l}\right)^{-K_l} \left(1 - \frac{\varepsilon + \Delta_l}{1-p_l}\right)^{T_l - K_l} \tag{A70}$$
$$= \left(1 - \left(\frac{\varepsilon + \Delta_l}{p_l}\right)^2\right)^{K_l} \left(1 - \frac{\varepsilon + \Delta_l}{p_l}\right)^{-K_l} \left(1 - \frac{\varepsilon + \Delta_l}{1-p_l}\right)^{K_l(1-p_l)/p_l} \left(1 - \frac{\varepsilon + \Delta_l}{1-p_l}\right)^{(p_l T_l - K_l)/p_l}, \tag{A71}$$

where (A69) follows from $\Delta_l = p_* - p_l$ along with (A53), and (A70) follows since $1 - a^2 = (1-a)(1+a)$.

From (A53), we have

$$0 < \left(\frac{\varepsilon + \Delta_l}{1-p_l}\right)^2 = \left(1 - \frac{1-(p_*+\varepsilon)}{1-p_l}\right)^2 \leq 1 - \frac{1-(p_*+\varepsilon)}{1-p_l} < 1. \tag{A72}$$

Hence, by Lemma A2, we have

$$1-\left(\frac{\varepsilon+\Delta_l}{1-p_l}\right)^2 \geq \exp\left[-\frac{1}{\sqrt{1-\left(\frac{\varepsilon+\Delta_l}{1-p_l}\right)^2}}\left(\frac{\varepsilon+\Delta_l}{1-p_l}\right)^2\right] \tag{A73}$$

$$\geq \exp\left[-\sqrt{\frac{1-p_l}{1-p_*-\varepsilon}}\left(\frac{\varepsilon+\Delta_l}{1-p_l}\right)^2\right] \tag{A74}$$

$$= \exp\left[-\left(\frac{1-p_l}{\sqrt{(1-p_l)(1-p_*-\varepsilon)}}\right)\left(\frac{\varepsilon+\Delta_l}{1-p_l}\right)^2\right], \tag{A75}$$

where (A74) follows from (A72).

For the third term in (A71), we proceed as follows:

$$\left(1-\frac{\varepsilon+\Delta_l}{1-p_l}\right)^{K_l(1-p_l)/p_l}$$

$$= \left(1-\left(\frac{\varepsilon+\Delta_l}{1-p_l}\right)^2\right)^{K_l(1-p_l)/p_l}\left(1+\frac{\varepsilon+\Delta_l}{1-p_l}\right)^{-K_l(1-p_l)/p_l} \tag{A76}$$

$$\geq \exp\left[-\left(\frac{1-p_l}{\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}\right)\left(\frac{\varepsilon+\Delta_l}{1-p_l}\right)^2\left(\frac{K_l(1-p_l)}{p_l}\right)\right]\left(1+\frac{\varepsilon+\Delta_l}{1-p_l}\right)^{-K_l(1-p_l)/p_l} \tag{A77}$$

$$= \exp\left[-\frac{(1-p_l)^2}{p_l\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}\left(\frac{\varepsilon+\Delta_l}{1-p_l}\right)^2 K_l\right]\left(1+\frac{\varepsilon+\Delta_l}{1-p_l}\right)^{-K_l(1-p_l)/p_l} \tag{A78}$$

$$\geq \exp\left[-\frac{3(1-p_l)^2}{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}\left(\frac{\varepsilon+\Delta_l}{1-p_l}\right)^2 K_l\right]\left(1+\frac{\varepsilon+\Delta_l}{1-p_l}\right)^{-K_l(1-p_l)/p_l} \tag{A79}$$

$$= \exp\left[-\frac{3\alpha_l^2 p_l^2(1-p_l)^2}{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}K_l\right]\left(1+\frac{\varepsilon+\Delta_l}{1-p_l}\right)^{-K_l(1-p_l)/p_l} \tag{A80}$$

$$\geq \exp\left[-\frac{3\alpha_l^2 p_l^2(1-p_l)^2}{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}T_l\right]\left(1+\frac{\varepsilon+\Delta_l}{1-p_l}\right)^{-K_l(1-p_l)/p_l} \tag{A81}$$

$$\geq \exp\left[-\frac{3\alpha_l^2 p_l^2(1-p_l)^2}{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}T_l\right]\exp\left[-\left(\frac{\varepsilon+\Delta_l}{1-p_l}\right)\left(\frac{K_l(1-p_l)}{p_l}\right)\right] \tag{A82}$$

$$= \exp\left[-\frac{3\alpha_l^2 p_l^2(1-p_l)^2}{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}T_l\right]\exp\left[-\left(\frac{\varepsilon+\Delta_l}{p_l}\right)K_l\right], \tag{A83}$$

where (A76) uses $1-a^2 = (1-a)(1+a)$, (A77) follows from (A75), (A79) follows from (A44), (A80) follows by definition of α_l in (25), (A81) follows from the fact that $K_l \leq T_l$, and (A82) follows from the fact that $(1+x)^{-y} \geq \exp(-xy)$ for all $0 \leq x$ and $y \geq 0$.

On the other hand, observe that

$$\left(1-\frac{\varepsilon+\Delta_l}{p_l}\right)^{-K_l} \geq \exp\left[\left(\frac{\varepsilon+\Delta_l}{p_l}\right)K_l\right] \tag{A84}$$

since $(1-x)^{-y} \geq \exp(xy)$ for all $0 \leq x \leq 1$ and $y \geq 0$. It follows from (A83) and (A84) that

$$\left(1-\frac{\varepsilon+\Delta_l}{p_l}\right)^{-K_l}\left(1-\frac{\varepsilon+\Delta_l}{1-p_l}\right)^{K_l(1-p_l)/p_l} \geq \exp\left[-\frac{3\alpha_l^2 p_l^2(1-p_l)^2}{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}T_l\right], \tag{A85}$$

and it follows from (A71) and (A85) that

$$L_l(W) \geq \left(1 - \left(\frac{\varepsilon + \Delta_l}{p_l}\right)^2\right)^{K_l} \left(1 - \frac{\varepsilon + \Delta_l}{1 - p_l}\right)^{(p_l T_l - K_l)/p_l} \times \exp\left[- \frac{3}{2(p_* + \varepsilon)\sqrt{(1 - p_l)(1 - (p_* + \varepsilon))}} \alpha_l^2 p_l^2 (1 - p_l)^2 T_l \right]. \tag{A86}$$

Now, since $0 < (\frac{\varepsilon+\Delta_l}{p_l})^2 \leq \frac{1}{4} = 1 - \frac{3}{4}$ (since we are in the case $0 \leq \frac{\varepsilon+\Delta_l}{p_l} \leq \frac{1}{2}$), by Lemma A2, we have

$$\left(1 - \left(\frac{\varepsilon + \Delta_l}{p_l}\right)^2\right)^{K_l} \geq \exp\left[- \sqrt{\frac{4}{3}} \left(\frac{\varepsilon + \Delta_l}{p_l}\right)^2 K_l \right] \tag{A87}$$

$$\geq \exp\left[- \frac{4}{3} \left(\frac{\varepsilon + \Delta_l}{p_l}\right)^2 K_l \right] \tag{A88}$$

$$= \exp\left[- \frac{4}{3} (\alpha_l (1 - p_l))^2 K_l \right] \tag{A89}$$

$$= \exp\left[\frac{4}{3} (\alpha_l (1 - p_l))^2 (p_l T_l - K_l) \right] \exp\left[- \frac{4}{3} (\alpha_l (1 - p_l))^2 p_l T_l \right], \tag{A90}$$

where (A89) follows from the definition of α_l in (25).

We now consider two further sub-cases:

(i) If $p_l T_l > K_l$, then we have

$$\left(1 - \frac{\varepsilon + \Delta_l}{1 - p_l}\right)^{(p_l T_l - K_l)/p_l} \geq \exp\left(- \left(\sqrt{\frac{1 - p_l}{1 - (p_* + \varepsilon)}}\right) \left(\frac{\varepsilon + \Delta_l}{1 - p_l}\right) \left(\frac{p_l T_l - K_l}{p_l}\right) \right) \tag{A91}$$

$$= \exp\left(- \left(\sqrt{\frac{1 - p_l}{1 - (p_* + \varepsilon)}}\right) \alpha_l (p_l T_l - K_l) \right) \tag{A92}$$

$$= \exp\Big(- \beta_l (p_l T_l - K_l) \Big), \tag{A93}$$

where (A91) follows from Lemma A2 along with (A53), and (A93) follows from the definition of β_l in (26).

(ii) If $p_l T_l \leq K_l$, then we have

$$\left(1 - \frac{\varepsilon + \Delta_l}{1 - p_l}\right)^{(p_l T_l - K_l)/p_l} \geq \exp\left[- \left(\frac{\varepsilon + \Delta_l}{1 - p_l}\right) \left(\frac{p_l T_l - K_l}{p_l}\right) \right] \tag{A94}$$

$$= \exp\Big[- \alpha_l (p_l T_l - K_l) \Big], \tag{A95}$$

where (A94) follows from the fact that $(1 - x)^y \geq \exp(-xy)$ if $1 \geq x \geq 0$ and $y \leq 0$.

From (A93) and (A95), we obtain

$$\left(1 - \frac{\varepsilon + \Delta_l}{1 - p_l}\right)^{(p_l T_l - K_l)/p_l} \geq \exp\left[- \Big(\beta_l (p_l T_l - K_l) \mathbf{1}\{p_l T_l > K_l\} + \alpha_l (p_l T_l - K_l) \mathbf{1}\{p_l T_l \leq K_l\} \Big) \right]. \tag{A96}$$

Now, from (A86), (A90), and (A96), we have

$$\begin{aligned}
L_l(W) &\geq \exp\left[\frac{4}{3}\big(\alpha_l(1-p_l)\big)^2(p_lT_l-K_l)\right]\exp\left[-\frac{4}{3}\alpha_l^2p_l(1-p_l)^2T_l\right]\\
&\quad\times\exp\left[-\frac{3\alpha_l^2p_l^2(1-p_l)^2}{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}T_l\right]\\
&\quad\times\exp\left[-\Big(\beta_l(p_lT_l-K_l)\mathbf{1}\{p_lT_l>K_l\}+\alpha_l(p_lT_l-K_l)\mathbf{1}\{p_lT_l\leq K_l\}\Big)\right] &\text{(A97)}\\
&=\exp\left[-\frac{4}{3}\alpha_l^2p_l(1-p_l)^2T_l\right]\exp\left[-\frac{3\alpha_l^2p_l^2(1-p_l)^2}{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}T_l\right]\\
&\quad\times\exp\left[-\Big(\tilde{\beta}_l(p_lT_l-K_l)\mathbf{1}\{p_lT_l>K_l\}+\tilde{\alpha}_l(p_lT_l-K_l)\mathbf{1}\{p_lT_l\leq K_l\}\Big)\right] &\text{(A98)}\\
&\geq\exp\left[-\frac{2}{(p_*+\varepsilon)}\alpha_l^2p_l^2(1-p_l)^2T_l\right]\exp\left[-\frac{3\alpha_l^2p_l^2(1-p_l)^2}{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}T_l\right]\\
&\quad\times\exp\left[-\Big(\tilde{\beta}_l(p_lT_l-K_l)\mathbf{1}\{p_lT_l>K_l\}+\tilde{\alpha}_l(p_lT_l-K_l)\mathbf{1}\{p_lT_l\leq K_l\}\Big)\right] &\text{(A99)}\\
&\geq\exp\left[-\frac{7\alpha_l^2p_l^2(1-p_l)^2}{2(p_*+\varepsilon)\sqrt{(1-p_l)(1-(p_*+\varepsilon))}}T_l\right]\\
&\quad\times\exp\left[-\Big(\tilde{\beta}_l(p_lT_l-K_l)\mathbf{1}\{p_lT_l>K_l\}+\tilde{\alpha}_l(p_lT_l-K_l)\mathbf{1}\{p_lT_l\leq K_l\}\Big)\right], &\text{(A100)}
\end{aligned}$$

where (A98) follows from the definitions of $\tilde{\alpha}_l$ and $\tilde{\beta}_l$ in (27)–(28), (A99) follows by writing $\frac{4}{3}\alpha_l^2p_l(1-p_l^2)=\frac{4}{3p_l}\alpha_l^2p_l^2(1-p_l^2)$ and applying (A44), and (A100) uses $\sqrt{(1-p_l)(1-p^*+\epsilon)}\leq 1$. Hence, (41) also holds in this case, and the proof is complete.

Appendix C. Differences in Analysis Techniques

Here we briefly overview some of the main differences in our analysis techniques compared to [14], leading to the improvements highlighted in Section 2:

- We remove the restriction $p_l\geq\frac{\varepsilon+p_*}{1+\sqrt{\frac{1}{2}}}$ (or $\frac{\varepsilon+\Delta_l}{p_l}\leq\frac{1}{\sqrt{2}}$) used in the subsets $\mathcal{M}(\mathbf{p},\varepsilon)$ and $\mathcal{N}(\mathbf{p},\varepsilon)$ in (Equations (4) and (5) [14]), so that our lower bound depends on all of the arms. To achieve this, our analysis frequently needs to handle the cases $\frac{\varepsilon+\Delta_l}{p_l}>\frac{1}{2}$ and $\frac{\varepsilon+\Delta_l}{p_l}\leq\frac{1}{2}$ separately (e.g., see the proof of Proposition 1).
- The preceding separation into two cases also introduces further difficulties. For example, our definition of $G_{2,l}$ in (30) is modified to contain different constants for the cases $p_lT_l>K_l$ and $p_lT_l\leq K_l$, which is not the case in (Lemma 2 [14]). Accordingly, the quantities $\tilde{\alpha}_l$ in (27) and $\tilde{\beta}_l$ in (28) appear in our proof but not in [14].
- We replace the inequality $(1-x)^y\geq e^{-1.78xy}$ (for $x\in\big(0,\frac{1}{\sqrt{2}}\big)$ and $y\geq 0$) (Lemma 3 [14]) by Lemma A2. By using this stronger inequality, we can improve the constant term c_1 from $O(\underline{p}^2)$ to $(p^*+\varepsilon)^2$. In addition, Lemma A2 does not require the assumption $x\leq\frac{1}{\sqrt{2}}$ as in (Lemma 3 [14]), so we can use it for the case $p_*>\frac{1}{2}$, which required a separate analysis in [14].
- To further reduce the constant term from $(p^*+\varepsilon)^2$ to $(p^*+\varepsilon)$ (see Theorem 1), we also need to use other mathematical tricks to sharpen certain inequalities, such as (A83).

Author Contributions: L.V.T. conceptualized the problem, and established and wrote the main results and proofs. J.S. provided ongoing supervision and corrections, and assistance with the writing.

Funding: This research was funded by the Singapore National Research Foundation (NRF) under grant number R-252-000-A74-281.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lattimore, T.; Szepesvári, C. *Bandit Algorithms*; Cambridge University Press: Cambridge, UK, to appear.
2. Villar, S.S.; Bowden, J.; Wason, J. Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Stat. Sci.* **2015**, *30*, 199–215. [CrossRef] [PubMed]
3. Li, L.; Chu, W.; Langford, J.; Schapire, R.E. A contextual-bandit approach to personalized news article recommendation. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010.
4. Awerbuch, B.; Kleinberg, R.D. Adaptive routing with end-to-end feedback: Distributed learning and geometric approaches. In Proceedings of the Symposium of Theory of Computing (STOC04), Chicago, IL, USA, 5–8 June 2004.
5. Shen, W.; Wang, J.; Jiang, Y.G.; Zha, H. Portfolio Choices with Orthogonal Bandit Learning. In Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI-15), Bengaluru, India, 25–31 July 2015.
6. Bechhofer, R.E. A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs. *Biometrics* **1958**, *14*, 408–429. [CrossRef]
7. Paulson, E. A sequential procedure for selecting the population with the largest mean from k normal populations. *Ann. Math. Stat.* **1964**, *35*, 174–180. [CrossRef]
8. Even-Dar, E.; Mannor, S.; Mansour, Y. PAC bounds for multi-armed bandit and Markov decision processes. In Proceedings of the Fifteenth Annual Conference on Computational Learning Theory, Sydney, Australia, 8–10 July 2002.
9. Kalyanakrishnan, S.; Tewari, A.; Auer, P.; Stone, P. PAC subset selection in stochastic multi-armed bandits. In Proceedings of the International Conference on Machine Learning, Edinburgh, UK, 26 June–1 July 2012.
10. Gabillon, V.; Ghavamzadeh, M.; Lazaric, A. Best arm identification: A unified approach to fixed budget and fixed confidence. In Proceedings of the 26th Annual Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012.
11. Jamieson, K.; Malloy, M.; Nowak, R.; Bubeck, S. On finding the largest mean among many. *arXiv* **2013**, arxiv:1306.3917.
12. Karnin, Z.; Koren, T.; Somekh, O. Almost optimal exploration in multi-armed bandits. In Proceedings of the 30th International Conference on Machine Learning (ICML 2013), Atlanta, GA, USA, 16–21 June 2013.
13. Jamieson, K.; Malloy, M.; Nowak, R.; Bubeck, S. lil'UCB: An Optimal Exploration Algorithm for Multi-Armed Bandits. *arXiv* **2013**, arxiv:1312.7308.
14. Mannor, S.; Tsitsiklis, J.N. The Sample Complexity of Exploration in the Multi-Armed Bandit Problem. *J. Mach. Learn. Res.* **2004**, *5*, 623–648.
15. Kaufmann, E.; Cappé, O.; Garivier, A. On the Complexity of Best-arm Identification in Multi-armed Bandit Models. *J. Mach. Learn. Res.* **2016**, *17*, 1–42.
16. Carpentier, A.; Locatelli, A. Tight (Lower) Bounds for the Fixed Budget Best Arm Identification Bandit Problem. In Proceedings of the Conference On Learning Theory, New York, NY, USA, 23–26 June 2016.
17. Chen, L.; Li, J.; Qiao, M. Nearly Instance Optimal Sample Complexity Bounds for Top-k Arm Selection. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017), Fort Lauderdale, FL, USA, 20–22 April 2017.
18. Simchowitz, M.; Jamieson, K.G.; Recht, B. The Simulator: Understanding Adaptive Sampling in the Moderate-Confidence Regime. *arXiv* **2013**, arxiv:abs/1702.05186.
19. Bubeck, S.; Bianchi, N.C. Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. In *Foundations and Trends in Machine Learning*; Now Publishers Inc.: Hanover, MA, USA, 2012; Volume 5.
20. Royden, H.; Fitzpatrick, P. *Real Analysis*, 4th ed.; Pearson: New York, NY, USA, 2010.
21. Katariya, S.; Jain, L.; Sengupta, N.; Evans, J.; Nowak, R. Adaptive Sampling for Coarse Ranking. In Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS 2018), Lanzarote, Spain, 9–11 April 2018.
22. Billingsley, P. *Probability and Measure*, 3rd ed.; Wiley-Interscience: Hoboken, NJ, USA, 1995.