CRIMED: Lower and Upper Bounds on Regret for Bandits with Unbounded Stochastic Corruption

Shubhada Agrawal SAGRAWAL362@GATECH.EDU
Georgia Institute of Technology

Timothée Mathieu TIMOTHEE.MATHIEU@INRIA.FR
Université de Lille, Inria, CNRS, Centrale Lille UMR 9189 – CRIStAL, F-59000 Lille, France

Debabrota Basu DEBABROTA.BASU@INRIA.FR
Université de Lille, Inria, CNRS, Centrale Lille UMR 9189 – CRIStAL, F-59000 Lille, France

Odalric-Ambrym Maillard ODALRIC.MAILLARD@INRIA.FR
Université de Lille, Inria, CNRS, Centrale Lille UMR 9189 – CRIStAL, F-59000 Lille, France

Abstract

We study the regret-minimisation problem in the multi-armed bandit setting with *unbounded stochastic corruption*. By unbounded stochastic corruption, we mean that the agent might not observe the reward generated from the selected arm, but another one generated from an arbitrary corruption distribution with potentially *unbounded support*. Amount of corruption is controlled by a corruption proportion $\varepsilon \in (0, \frac{1}{2})$. Unlike the bounded stochastic corruption setting, we do not have any existing generic lower bound on regret for the unbounded setting, and no existing algorithm that yields the same regret upper bound. In this paper, we address this gap. First, we propose a generic problem-dependent lower bound on regret that holds for any family of reward and corruption distributions. The proposed bound quantifies the hardness of this problem in terms of corrupted KL-inf, which is a generalisation of classical KL-inf known for uncorrupted bandits. Second, we study properties of corrupted KL-inf for Gaussian reward distributions with known variance, where it is numerically computable. We observe that the classical $\Omega(\log T)$ bound (for horizon T) is achievable for suboptimality gap greater than $2\Phi^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)$, where Φ is Gaussian CDF. This lower bound leads to a corruption robust extension of IMED, namely `CRIMED`, that uses median as a robust estimate of the central moment while accessing only the corrupted observations. Finally, we provide a regret analysis proving `CRIMED` to be asymptotically optimal. Due to a novel concentration result of medians, we show that `CRIMED` is the first algorithm that can handle corruption level up to $1/2$. We also provide numerical results confirming that `CRIMED` outperforms the existing algorithms.

Keywords: Multi-Armed bandit, Corruption neighbourhood, IMED, Robust estimation

1. Introduction

We study the problem of sequential decision-making under partial information, where the observation yielded by an action is arbitrarily but *stochastically corrupted*. Specifically, we consider a variant of the stochastic multi-armed bandits problem (Lattimore and Szepesvári, 2020) with *unbounded stochastic corruption* (Altschuler et al., 2019; Basu et al., 2022). In this setting, an agent (equivalently, an algorithm) is presented with a set of K *arms*, representing K *unknown* probability distributions, each coming from a family of distributions $\mathcal{L}$. We denote this set of K-distributions by $\mu \coloneqq (\mu_1, \dots, \mu_K)$, where $\forall a \in [K], \mu_a \in \mathcal{L}$. When the algorithm selects an arm A_n at time n, an independent sample, Y_n, is drawn from the corresponding distribution μ_{A_n}. This is the *reward* of

the algorithm for pulling the chosen arm. But unlike the classical setup, Y_n is not directly observed by the algorithm. Instead, the observations are corrupted with probability $\varepsilon \in (0, 0.5)$. More precisely, at time n, the algorithm observes $Y_n \sim \mu_{A_n}$ with probability $1-\varepsilon$, and with probability ε, it observes a sample from some other distribution, $H_{A_n} \in \mathcal{P}(\mathbb{R})$. We call $\mathbf{H} \coloneqq (H_1, \dots, H_K)$ the set of *corruption distributions* and each of them can have unbounded support. Following the classical regret-minimization setting, the goal of the algorithm is to *sequentially* sample these arms in order to maximize the expected cumulative reward.

Multi-armed bandits problem, in brief *bandits*, is the archetypal setting of Reinforcement Learning (RL). It serves as the theoretical founding stone of the modern RL theory and algorithmic basis of recommender systems. As bandits and RL are gradually reaching the realm of practical deployment, the question of robustness against corrupted (or externally perturbed) observations has gained significant interest. Specially, researchers have studied three types of settings: *bounded adversarial corruptions* (Auer et al., 2002; Hajiesmaili et al., 2020; Pogodin and Lattimore, 2020) (a.k.a adversarial bandits), *bounded stochastic corruptions* (Lykouris et al., 2018; Kapoor et al., 2019), in which an adversary shifts the rewards observed by the agent under constraint on the total shift budget, and *unbounded stochastic corruptions* (Altschuler et al., 2019; Basu et al., 2022). Unlike the first two settings, there exists to our knowledge no generic lower bound on regret for corrupted bandits under unbounded stochastic corruptions that holds for any family of reward or corruption distributions. There exists also no algorithm yielding a befitting regret upper bound, while being robust to the unbounded stochastic corruptions. Thus, in this paper, *we study bandits with unbounded stochastic corruptions, specifically Bandits Corrupted by Nature, and aim to fill up these two gaps.*

Regret. For $\mu \in \mathcal{L}^K$, let $m^*(\mu)$ denote the mean of the optimal arm in μ, and let $m(\mu_a)$ denote the mean of arm a. Then for any sub-optimal arm $a \notin \operatorname{argmax}_b\, m(\mu_b)$, $\Delta_a \coloneqq m^*(\mu) - m(\mu_a)$ denotes the instantaneous mean regret incurred by pulling the sub-optimal arm a. We refer to Δ_a as the *sub-optimality gap* of arm a. Recall that Y_n denotes the independent (uncorrupted) sample drawn from the arm A_n pulled at time n, and let $Y_{a,j}$ denote the j^{th} independent sample drawn from arm a. Using these notations and following (Kapoor et al., 2019; Basu et al., 2022), we define the *expected regret under corruption* ($\mathbb{E}[R_T]$) till time T as $\mathbb{E}\left[\sum_{n=1}^{T}(m^*(\mu) - Y_n)\right]$. Here, the expectation in the above expression is with respect to all the randomness present in the system including the impact of corruption on action selection. We further observe that $\mathbb{E}[R_T] \coloneqq \sum_{a=1}^{K}\mathbb{E}[N_a(T)]\,\Delta_a$, where $N_a(T)$ is the number of pulls of the arm a till time T. Since Δ_a's are constant for a given μ and for the optimal arm(s) $\Delta_a = 0$, minimizing the expected regret reduces to minimising the expected number of pulls of the sub-optimal arms $\mathbb{E}[N_a(T)]$.

Notation. We denote by $\mathbb{R}$ the set of real numbers and by $\mathcal{P}(\mathbb{R})$ the set of all probability distributions on $\mathbb{R}$. $\mathbb{R}^+$ denotes the set of non-negative real numbers. For any set S, we denote by 2^S the set of subsets of S, i.e. the power set of S. For $\mu, \mathbf{H} \in \mathcal{P}(\mathbb{R})^K$, let $\mu \odot_\varepsilon \mathbf{H} \coloneqq (1-\varepsilon)\mu + \varepsilon\mathbf{H}$ denote the vector of distributions in μ corrupted by the vector of corruption distributions $\mathbf{H}$ with corruption proportion ε. We use similar notation for the mixture of each component $\mu_a, H_a \in \mathcal{P}(\mathbb{R})$, i.e. $\mu_a \odot_\varepsilon H_a \coloneqq (1-\varepsilon)\mu_a + \varepsilon H_a$ for any $a \in [K]$. Finally, we denote by $\mathcal{G}$ the set of all the Gaussian distributions with variance 1, by φ the Gaussian pdf, and by Φ the Gaussian CDF.

1.1. Outline and contributions

Specifically, we investigate three questions in this paper:

1. Can we derive a problem-dependent lower bound on regret for a given set of reward distributions and the corresponding worst-case corruption distributions?
2. Can we leverage this lower bound to design an asymptotically optimal algorithm that is robust to unbounded stochastic corruptions?
3. Which robust statistics of observed rewards should we use for algorithm design and how does its concentration impacts the achieved regret?

Our study leads to the following results and observations.

1. A Generic Lower Bound on Regret: To the best of our knowledge, we derive the first generic lower bound for the unbounded stochastic corruption setting that holds for any family of reward distributions and corruption distributions (Section 2.1). Specifically, we show in Theorem 1 that in this setting a uniformly-good algorithm (Definition 1) has to pull any of the suboptimal arms $\Omega(\log T)$ number of times, in expectation. This resonates with the known $\Omega(\log T)$ problem-dependent lower bound for uncorrupted bandits (Lai and Robbins, 1985). In the corrupted setting, the lower bound depends on a novel quantity, $\mathrm{KL}^{\varepsilon}_{\inf}$, that we call the *corrupted KL-inf* (Equation (2.1)) that captures the complexity of the learning problem. Theorem 1 shows that inverse of the corrupted KL-inf is the cost that any algorithm need to pay to distinguish an optimal arm from a suboptimal one under the worst-case corruption. In the uncorrupted case, i.e. for corruption proportion $\varepsilon = 0$, the corrupted KL-inf reduces to the classical KL-inf yielding the known lower bound for classical uncorrupted bandits (Burnetas and Katehakis, 1996).

2. An Analytical Quantifier of Hardness: Though Lemma 1 provides a generic lower bound, we aim to derive an explicitly quantifiable formulation of corrupted KL-inf. For this reason, for the arm distributions, we focus on Gaussian family of reward distributions with known variance, while still allowing for unbounded and arbitrary corruptions (Chen et al., 2018). This leads to a more specific lower bound on the expected number of pulls of the suboptimal arms (Proposition 2):

$$\liminf_{T\to\infty} \frac{1}{\log T}\left(\sup_{\mathbf{H}\in\mathcal{P}(\mathbb{R})^K} \mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[N_a(T)\right]\right) \geq \frac{1}{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(m(\mu_a), m^*(\mu))},\ \forall a \text{ with } m(\mu_a) < m^*(\mu),$$

and an analytical formulation of KL-inf dictating the hardness of this setting

$$\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,y) \coloneqq \min_{\tilde{H}_1\in\mathcal{P}(\mathbb{R}),\tilde{H}_2\in\mathcal{P}(\mathbb{R})} \mathrm{KL}(\mathcal{N}(x,1)\odot_\varepsilon \tilde{H}_1, \mathcal{N}(y,1)\odot_\varepsilon \tilde{H}_2).$$

Formalising $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,y)$ allows us to *explicitly* study the most confusing pair of corruption distributions for a given family of reward distributions. Additionally, this allows us to show that in this setting, the minimum suboptimality gap $\Delta \coloneqq m^* - \max_{a\neq a^*} m(\mu_a)$, i.e. the difference of expected rewards of the optimal and second optimal arms, has to be greater than[1] $2\Phi^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)$ for any algorithm to achieve sublinear regret in a bandit problem. This comes in stark contrast with the uncorrupted stochastic bandit setting. We derive other interesting properties of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,y)$ in Section 2.3 and in Appendix C, which aid our algorithm design and corresponding regret analysis.

3. Algorithm Design: In Section 3, we leverage the formulation and properties of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,y)$ to propose an IMED-type index-based algorithm (Honda and Takemura, 2015), namely `CRIMED` (Corruption Robust IMED, Algorithm 1), for unbounded corruptions and Gaussian reward distributions with known variance. Designing `CRIMED` requires two main changes to the traditional IMED index.

1. Φ is the Gaussian Cumulative Distribution Function (CDF). Φ^{-1} is the inverse Gaussian CDF.

First, it *replaces the uncorrupted KL-divergence in IMED index with* $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$. Second, it uses *median* as the robust estimate of the mean, computes the empirical medians for each arm, and plugs them in $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(\cdot,\cdot)$ to compute the CRIMED index. In Section 3.2, we show that CRIMED *asymptotically achieves the regret lower bound* prescribed in Theorem 2. Thus, CRIMED is the first asymptotically optimal algorithm in the unbounded stochastic corruption setting with Gaussian rewards. Additionally, CRIMED is asymptotically optimal for any corruption level $\varepsilon < \frac{1}{2}$, whereas the algorithms of (Kapoor et al., 2019) and (Basu et al., 2022) could only tackle $\varepsilon < \Delta$ and $\frac{1}{5}$, respectively.

4. Median as the Robust Estimator and its Impact: Following the robust estimation literature, we choose median as the robust estimate of the mean. This choice is driven by the fact that median incurs the minimum bias among all the robust estimators of location parameter (Section B). In Theorem 3, we derive a novel concentration bound of the empirical median computed from the corrupted Gaussian rewards, which allows us to derive a regret upper bound for CRIMED. This concentration result is of parallel interest and may be extended to any *symmetric* unimodal distribution.

For brevity, we postpone technical derivations to the appendix.

1.2. Related work

Our work connects and is related with several research areas, which we now briefly summarize.

Multi-armed bandits. The problem of bandits was first introduced in the context of designing adaptive clinical trials by Thompson (1933), and later popularized under this name by Robbins (1952). Since then the variants of this problem have been widely studied and are used in practice. For the classical regret-minimization framework introduced earlier, asymptotic instance-dependent lower bounds on the regret suffered by an algorithm are well known (Lai and Robbins, 1985; Burnetas and Katehakis, 1996)). Algorithms matching the lower bounds are also developed. Cappé et al. (2013); Agrawal et al. (2021) propose index-based asymptotically optimal algorithms for parametric and heavy-tailed distributions, respectively. While these algorithms are statistically optimal, they can be computationally demanding. Honda and Takemura (2009, 2010, 2015) develop a different style of algorithms that have a lower computational cost and are also statistically optimal. Alternative optimal algorithms relying on Bayesian posteriors to sample arms have also been developed (Agrawal and Goyal, 2012, 2017; Kaufmann et al., 2012). In this paper, we follow a frequentist approach and *design an IMED-type algorithm due to its optimality and computational simplicity*.

Bandits with bounded corruption. In the bounded adversarial setting, the rewards are assumed to be generated by an adaptive adversary from a bounded interval (e.g. $[0, 1]$). In the literature, this setting is also referred as adversarial multi-armed bandits (Pogodin and Lattimore, 2020). Researchers have aimed to design the best of the both worlds algorithms that perform almost optimally for this setting as well as the stochastic setting without any adversary, and are of parallel interest (Seldin and Slivkins, 2014; Pogodin and Lattimore, 2020). In the bounded stochastic setting (Lykouris et al., 2018; Kapoor et al., 2019), when an arm a is pulled, the reward r is stochastically generated from the corresponding reward distribution ν_a. But an adversary switches the rewards observed R' by the agent from the one generated from the distribution, i.e. R, such that $\sum_{n=1}^{T} |R'_n - R_n| \leq C$. Here, C is a non-negative upper bound on corruption. In this setting, we observe the corruption in observation appearing, and multiple variants of such bounded corruptions are studied (Lykouris et al., 2018; Gajane et al., 2018; Kapoor et al., 2019). But defining C plays a critical role in this setting, and existing regret bounds are linearly dependent on C (Lykouris et al., 2018). Thus, the existing analysis of regret bounds, and the proposed algorithms are unfit to handle large amounts

of corruptions. This propels the study of the third setting of the *unbounded stochastic corruption* described in the introduction.

Robust estimation under unbounded stochastic corruption. Robust estimation dates back (Box and Andersen, 1953). A robust estimator is an estimator that perform *well* even in the presence of anomalous (outlier) data. In Huber (1964), Huber developed a theory of asymptotic minimax optimality of estimators for distributions in a corruption neighbourhood of P, we use this same setting. Sampling from a corrupted distributions consist in sampling from the law P with probability $1-\varepsilon \geq 1/2$ and sampling according to an outlier distribution with probability ε, no assumption is made on the outlier distributions. Since then, several methods have been devised to assess the asymptotic robustness of estimators (Huber and Ronchetti, 2009; Hampel et al., 1986), in particular with the stability of the limit of an estimator when the samples come from a corrupted distribution. On the other hand, in recent years a non-asymptotic notion of robustness was created, the goal is to obtain estimators that concentrate fast even when the data are heavy-tailed (Catoni, 2012; Devroye et al., 2016; Lugosi and Mendelson, 2019; Agrawal, 2022), or corrupted (Wang and Ramdas, 2023; Chen et al., 2018). The two concepts of asymptotic and non-asymptotic robustness are closely linked and the estimators that perform well in asymptotic robustness have been shown to perform well in the non-asymptotic sense. Our article is situated between the two approaches: *we use robust concentrations inequality for the empirical median in order to prove asymptotic optimality of our algorithm in a corruption neighbourhood.*

Median. It is customary to estimate the mean regret when dealing with Multi-armed bandits, and in most cases (when the distributions are not symmetric) the median is not a good estimator for the mean. In the case of symmetric distributions, however, there is no bias and estimating the median is the same as estimating the mean, this corresponds to our setting. The median has also been used for the best-arm identification algorithms in which the goal is to find the arm with the largest median (Altschuler et al., 2019; Even-Dar et al., 2006; Nikolakakis et al., 2021), which is significantly different from *the regret-minimization setting considered in this paper*.

Bandits with unbounded stochastic corruption. To the best of our knowledge, unbounded stochastic corruption in bandits have only been studied in two papers Altschuler et al. (2019) and Basu et al. (2022). In Altschuler et al. (2019) the authors study the best arm identification problem in which one tries to find the arm with the largest median in a corruption neighbourhood setting, with inliers featuring only a finite second moment. They use a modification of the successive elimination algorithm with an expected stopping time smaller than a constant times the optimal expected stopping time. Though Basu et al. (2022) adhere to the same corruption model, they consider the problem of regret minimization. Since regret is best defined using the mean of rewards, Basu et al. (2022) use robust mean estimation to devise a UCB-type algorithm featuring a $\log(T)$ instance-dependent regret bound which is within a constant times the optimal regret attainable. Significantly improving on (Basu et al., 2022), *we devise* `CRIMED`*, an algorithm whose regret matches the lower bound asymptotically*, and we also experimentally demonstrate its performances.

2. Lower bound and KL-divergence in corrupted neighbourhoods

Given a class of probability distributions $\mathcal{L}$, we are interested in algorithms that perform uniformly well on all the K-armed bandit instances with the distributions of all the arms belonging to $\mathcal{L}$, when the observations are corrupted with a positive probability $\varepsilon > 0$. This requirement imposes a lower

bound on the number of samples that the algorithm needs to generate from each arm, which we now present. In the reminder of this paper, ε denotes a fixed constant in the range $(0, 0.5)$.

2.1. Problem-dependent lower bound

Definition 1 (Uniformly good algorithm) *An algorithm acting on a distribution in $\mathcal{L}$ is said* uniformly good *for a corruption level ϵ, if for all $\mu \in \mathcal{L}^K$ and the sub-optimal arms a in μ, it satisfies*

$$\sup_{\mathbf{H}\in\mathcal{P}(\mathbb{R})^K} \mathbb{E}_{\mu\odot_\varepsilon \mathbf{H}}\left[N_a(T)\right] = o\left(T^\alpha\right), \quad \text{for all } \alpha > 0.$$

Here the notation $\mathbb{E}_{\mu\odot_\varepsilon \mathbf{H}}[\cdot]$ denotes the expectation with respect to both the corrupted bandit process $\mu \odot_\varepsilon H$ and the possible randomness of the algorithm (omitted from notation).

Observe that unlike in the uncorrupted setting, for every instance, the algorithm needs to perform well with respect to the worst corruption distribution for each arm. The lower bound on the expected number of times a uniformly-good algorithm pulls a sub-optimal arm involves an optimisation problem, which we first present.

Corrupted KL-inf. We define corrupted KL-inf as the function $\mathrm{KL}^\varepsilon_{\inf} : \mathcal{P}(\mathbb{R}) \times \mathbb{R} \times 2^{\mathcal{P}(\mathbb{R})} \to \mathbb{R}^+$, that associates to each $\eta \in \mathcal{P}(\mathbb{R})$, $x \in \mathbb{R}$, and $\mathcal{L} \subset \mathcal{P}(\mathbb{R})$, the quantity

$$\mathrm{KL}^\varepsilon_{\inf}(\eta, x; \mathcal{L}) := \min_{H,H',\kappa} \left\{\mathrm{KL}\big(\eta \odot_\varepsilon H, \kappa \odot_\varepsilon H'\big) :\ \kappa\in\mathcal{L}, H, H' \in \mathcal{P}(\mathbb{R}), m(\kappa)\geq x\right\}. \tag{2.1}$$

This is equivalent to the optimisation problem that appears in the lower bound of the uncorrupted regret-minimisation setting leading to traditional $\mathrm{KL}_{\inf} := \min_\kappa \{\mathrm{KL}(\eta,\kappa) :\ \kappa\in\mathcal{L},\ m(\kappa) \geq x\}$. (c.f. Burnetas and Katehakis (1996), Lattimore and Szepesvári (2020, Chapter 16)). For $\varepsilon = 0$, $\mathrm{KL}^\varepsilon_{\inf}$ reduces to the traditional $\mathrm{KL}_{\inf}$ appearing in the lower bound of uncorrupted bandits. We also observe that the additional optimisation over the corruption distributions H and H' makes $\mathrm{KL}^\varepsilon_{\inf}$ smaller than that in the $\mathrm{KL}_{\inf}$ in the uncorrupted setting. Moreover, for $\eta \in \mathcal{L}$ and $x \leq m(\eta)$, $\mathrm{KL}^\varepsilon_{\inf}(\eta, x; \mathcal{L})$ is 0. This follows from the non-negativity of $\mathrm{KL}^\varepsilon_{\inf}$ and for any H, $\kappa = \eta$ and $H' = H$ are feasible for $\mathrm{KL}^\varepsilon_{\inf}$.

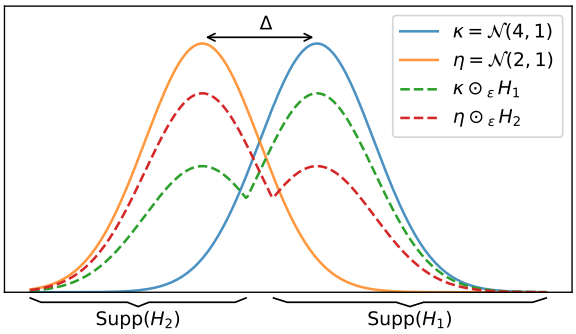
Theorem 1 (Corrupted lower bound) *For $\epsilon > 0$, $\mathcal{L} \subset \mathcal{P}(\mathbb{R})$, and a bandit instance $\mu \in \mathcal{L}^K$, for any sub-optimal arm a in μ, a uniformly good algorithm satisfies*

$$\liminf_{T\to\infty} \frac{1}{\log T}\left(\sup_{\mathbf{H}\in\mathcal{P}(\mathbb{R})^K} \mathbb{E}_{\mu\odot_\varepsilon \mathbf{H}}\left[N_a(T)\right]\right) \geq \frac{1}{\mathrm{KL}^\varepsilon_{\inf}(\mu_a, m^*(\mu); \mathcal{L})}.$$

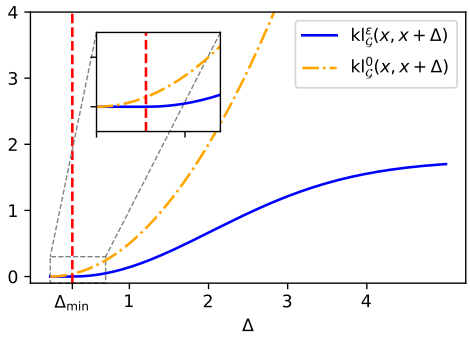
Since $\mathrm{KL}^\varepsilon_{\inf} \leq \mathrm{KL}_{\inf}$ for $\varepsilon \geq 0$, the lower bound above is higher than that in the uncorrupted setting. This quantifies that for $\varepsilon > 0$, the setting of stochastically-corrupted bandits is inherently harder than the classical uncorrupted setting. The central idea of our proof is to extend the classical change of measure lemma (Garivier et al., 2019) over the ε corruption neighbourhood of reward distributions. We refer the reader to Section A.1 for a complete proof of Theorem 1.

2.2. Huber's most confusing pair of distributions and KL-inf on corrupted neighbourhoods

For $\mathcal{L} \subset \mathcal{P}(\mathbb{R})$, $\eta \in \mathcal{L}$ and $x \in \mathbb{R}$, we now characterise the optimisers for $\mathrm{KL}^\varepsilon_{\inf}(\eta, x; \mathcal{L})$ (Equation (2.1)), i.e., the closest inlier distribution κ, and the most-confusing pair of corruption distributions H_1 and H_2. Let $\mathrm{Sp}(\eta)$ denote the support of η. First, we fix $\kappa \in \mathcal{L}$, and consider the optimisation problem over the two corruption distributions in $\mathrm{KL}^\varepsilon_{\inf}$. Consider



(*a*) Plot of corrupted distributions for $\varepsilon = 0.2$.



(*b*) Plot of $\mathrm{kl}^\varepsilon_{\mathcal{G}}$ for $\sigma = 1$ and $\varepsilon = 0.1$.

Figure 1: Illustration of the corrupted distributions from Lemma 1 and $\mathrm{kl}^\varepsilon_{\mathcal{G}}$.

$$\mathrm{d}(\eta \odot_\varepsilon H_1)(x) := \begin{cases} (1-\varepsilon)\mathrm{d}\eta(x), & \text{for } \frac{\mathrm{d}\eta}{\mathrm{d}\kappa}(x) \geq c_1 \\ c_1(1-\varepsilon)\mathrm{d}\kappa(x), & \text{otherwise,} \end{cases} \tag{2.2}$$

and

$$\mathrm{d}(\kappa \odot_\varepsilon H_2)(x) := \begin{cases} (1-\varepsilon)\mathrm{d}\kappa(x), & \text{for } \frac{\mathrm{d}\eta}{\mathrm{d}\kappa}(x) \leq \frac{1}{c_2} \\ c_2(1-\varepsilon)\mathrm{d}\eta(x), & \text{otherwise.} \end{cases} \tag{2.3}$$

Here, $\mathrm{d}f$ denotes the differential of a distribution function f, and $\frac{\mathrm{d}\eta}{\mathrm{d}\kappa}(x)$ denotes the Radon-Nikodym derivative of η with respect to κ. For $x \in \mathrm{Sp}(\eta) \cap \mathrm{Sp}(\kappa)^c$, $\frac{\mathrm{d}\eta}{\mathrm{d}\kappa}(x) := \infty$, and for $x \in \mathrm{Sp}(\eta)^c \cap \mathrm{Sp}(\kappa)$, $\frac{\mathrm{d}\eta}{\mathrm{d}\kappa}(x) := 0$. c_1 and c_2 are the normalisation constants ensuring that $\mathrm{d}(\eta \odot_\varepsilon H_1)$ and $\mathrm{d}(\kappa \odot_\varepsilon H_2)$ are probability measures, and also satisfying $0 \leq c_2 \leq \frac{1}{c_1} \leq \infty$. Observe that Equations (2.2) and (2.3) implicitly define H_1 and H_2. Now, we show that for a fixed η and κ, Equations (2.2) and (2.3) together yield the optimal corruption distributions in Equation (2.1).

Lemma 1 (Most confusing pairs) *Let $\eta, \kappa \in \mathcal{P}(\mathbb{R})$ be two probability distributions, and suppose that $\varepsilon < \frac{1}{2}$. Then, H_1 and H_2 defined above in Equations (2.2) and (2.3), respectively, satisfy*

$$(H_1, H_2) \in \operatorname{argmin} \left\{ \mathrm{KL}(\eta \odot_\varepsilon H, \kappa \odot_\varepsilon H') : H \in \mathcal{P}(\mathbb{R}),\ H' \in \mathcal{P}(\mathbb{R}) \right\}.$$

To prove the above result, we show that the directional derivative (for appropriate notion in the space of probability measures) of KL in every direction is non-negative at (H_1, H_2). We refer the reader to Section A.2 for a proof of the above result. Remark that these are the similar pair of distributions considered by Huber (1965) for a hypothesis testing setup.

Lemma 1 implies that the optimal corruption pair depends on the two input distributions. In particular, even though the corruption distributions are allowed to be arbitrary, corruption stays within the support of the considered pair of distributions. For illustration, we present the optimal-corruption pair for the Gaussian-inlier setting, i.e., when both η and κ are Gaussian distributions with unit variance, in Figure 1(*a*). We observe that the sets on which there is corruption are located in the right-tail (respectively left-tail) of the distribution on the left (respectively on the right). These and other interesting properties are formally proven in Lemma 3 and Appendix C.2.

2.3. The case of Gaussian rewards with known variance

With the above minimisers for the corruption pair for fixed η and κ, we are now left with characterising the optimal κ in $\mathrm{KL}^\varepsilon_{\inf}(\eta, x; \mathcal{L})$. When $\mathcal{L} = \mathcal{G}$ and $\eta \in \mathcal{G}$, it follows from Lemma 3 (later

in the section) that the optimiser $\kappa \in \mathcal{G}$ is the one with mean equal to x. Using these results in Theorem 1, we get the simplified lower bound for the Gaussian bandit models.

Proposition 2 (Lower bound for Gaussian rewards) *For $\mu \in \mathcal{G}^K$, any uniformly good algorithm satisfies for any sub-optimal arm a*

$$\liminf_{T\to\infty} \frac{1}{\log(T)} \left(\sup_{\mathbf{H}\in\mathcal{P}(\mathbb{R})^K} \mathbb{E}_{\mu\odot_\varepsilon \mathbf{H}}\left[N_a(T)\right] \right) \geq \frac{1}{\mathrm{kl}^\varepsilon_{\mathcal{G}}(m(\mu_a), m^*(\mu))}, \quad \textit{where}$$

$$\forall x,y\in\mathbb{R}, x\leq y \quad \mathrm{kl}^\varepsilon_{\mathcal{G}}(x,y) := \min_{H,H'} \left\{ \mathrm{KL}(\mathcal{N}(x,1)\odot_\varepsilon H, \mathcal{N}(y,1)\odot_\varepsilon H') \colon H, H' \in \mathcal{P}(\mathbb{R}) \right\}. \quad (2.4)$$

Here, the optimal pair of corruption distributions are given by Lemma 1.

We now present various important properties of $\mathrm{kl}^\varepsilon_{\mathcal{G}}$. First, we state a necessary and sufficient condition to have a finite bound in Proposition 2, i.e., the conditions under which $\mathrm{kl}^\varepsilon_{\mathcal{G}}(x,y) > 0$.

Lemma 2 (Non-intersection of corrupted neighbourhoods) *Let κ, η be Gaussian distributions with variance 1, i.e. $\kappa \sim \mathcal{N}(m(\kappa), 1)$ and $\eta \sim \mathcal{N}(m(\eta), 1)$. Then the following are equivalent:*
• $\forall (H_1, H_2) \in \mathcal{P}(\mathbb{R})^2,\ \kappa \odot_\varepsilon H_1 \neq \eta \odot_\varepsilon H_2$, • $|m(\kappa) - m(\eta)| > 2\Phi^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)$.

Lemma 2 shows that when the gap between the means of the two Gaussian distributions is too small, there exists a pair of corruption distributions that make the corrupted Gaussians arbitrarily close in KL divergence and hence, not distinguishable. We postpone the proof of this lemma to Section B.1. This justifies formally introducing the minimum gap between two means necessary to be able to distinguish between two Gaussian distributions under corruption proportion ε.

Definition 3 (Minimum distinction gap under corruption) $\Delta_{\min} := 2\Phi^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)$.

For Gaussian distributions, the KL can be expressed using CDF Φ and PDF φ of a standard Gaussian. Moreover, the corrupted KL, viz. $\mathrm{kl}^\varepsilon_{\mathcal{G}}$, enjoys nice properties like an almost closed-form expression, shift invariance, differentiability, etc., described in the following lemma.

Lemma 3 (Properties of $\mathrm{kl}^\varepsilon_{\mathcal{G}}(x,y)$) *Let H_1, H_2 be minimisers in Equation (2.4).Then,*

(a) The normalisation constants c_1 and c_2 are unique and equal, i.e., $c_1 = c_2 := c$.

(b) $\mathrm{kl}^\varepsilon_{\mathcal{G}}$ has an almost closed-form expression given in Equation (2.5) below. For $\Delta \leq \Delta_{\min}$, $\mathrm{kl}^\varepsilon_{\mathcal{G}}(x, x+\Delta) = 0$. Moreover, it is invariant under shift of means, i.e., for $\Delta \leq y - x$, $\mathrm{kl}^\varepsilon_{\mathcal{G}}(x+\Delta, y) = \mathrm{kl}^\varepsilon_{\mathcal{G}}(x, y-\Delta)$.

(c) For $\Delta \geq 0$, the function $\Delta \mapsto \mathrm{kl}^\varepsilon_{\mathcal{G}}(x, x+\Delta)$ is continuous and differentiable with continuous derivatives. For $\varepsilon > 0$ and $\Delta \leq \Delta_{\min}$, c equals 1, and $\partial\, \mathrm{kl}^\varepsilon_{\mathcal{G}}(x, x+\Delta)/\partial\Delta = 0$. For $\Delta > \Delta_{\min}$, and the corresponding normalising c, $\partial\, \mathrm{kl}^\varepsilon_{\mathcal{G}}(x, x+\Delta)/\partial\, \Delta > 0$ and is given by

$$\frac{\partial\, \mathrm{kl}^\varepsilon_{\mathcal{G}}(x, x+\Delta)}{\partial\Delta} = (1-\varepsilon)\,\Delta \left(\Phi\left(\frac{\Delta_+}{2}\right) - \Phi\left(\frac{\Delta_-}{2}\right) \right),$$

with $\Delta_+ := \Delta + \frac{2}{\Delta}\log\frac{1}{c}$ and $\Delta_- := \Delta - \frac{2}{\Delta}\log\frac{1}{c}$.

It follows from Lemma 3(c) above that $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ is strictly increasing in the second argument for values larger than the first argument, implying for $\eta \in \mathcal{G}$, $\mathrm{KL}^{\varepsilon}_{\inf}(\eta, x; \mathcal{G}) = \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(m(\eta), x)$. We refer the reader to Appendix C for a complete proof of the above lemma, and for other interesting properties of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$. Our proposed algorithm in Section 3 relies on computing $\mathrm{KL}^{\varepsilon}_{\inf}(\mu_a, x; \mathcal{G})$ for each arm a at every step. We do this by using a root-finding algorithm on Equation (C.1) to compute c in practice, and then use this value in Equation (2.5), the expression for $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ from Lemma 3(b).

For any $x < y$, below we give the closed-form expression for $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, y)$, once we know the optiaml normalising constant from Lemma 3(a). Here, $\Delta = y - x$, and Δ_+ and Δ_- are as defined in the Lemma 3(c).

$$\frac{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,y)}{1-\varepsilon} = (1-c)\log\left(\frac{1}{c}\right)\Phi\left(\frac{\Delta_-}{2}\right) + \frac{\Delta^2}{2}\left(\Phi\left(\frac{\Delta_+}{2}\right) - \Phi\left(\frac{\Delta_-}{2}\right)\right) - \Delta\left(\varphi\left(\frac{\Delta_-}{2}\right) - \varphi\left(\frac{\Delta_+}{2}\right)\right). \quad (2.5)$$

When Δ increases to infinity, the first term is dominant and $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ goes to a constant that depends on the corruption. On the other hand, for Δ converging to $\Delta_{\min}$, c can be shown to converge to 1 with Δ_- and Δ_+ converging to $\Delta_{\min}$. This can be seen from the defining equation for c (Equation (C.1)). In this limit, from Lemma 3(c), it follows that the derivative of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ converges to 0. Figure 1(*b*) illustrates $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x+\Delta)$ as a function of $\Delta \geq 0$. In particular, we observe that contrary to the uncorrupted Gaussian case, KL is not convex. Moreover, it has a flat region with value being 0 for $\Delta \leq \Delta_{\min}$. We leverage these properties of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ in the regret analysis of the proposed algorithm.

3. `CRIMED`: Algorithm and analysis

In this section, we leverage the lower bound in Proposition 2 to propose a corruption robust IMED-like algorithm, namely `CRIMED`. Then, we state the regret upper bound for `CRIMED` that shows it is asymptotically optimal for bandits with unbounded stochastic corruption. Finally, we explicate the technical novelty of our regret analysis.

3.1. Algorithm design: An IMED-based algorithm with estimated medians

First, we explicate our algorithm design. For $n \in \mathbb{N}$ and $a \in [K]$, let $\hat{\mu}_a(n)$ denote the empirical distribution constituted by $N_a(n)$ samples generated from arm a. We use median of the corrupted observations as the estimator of the central moment of underlying reward distributions. The choice of empirical median is natural in the case of Gaussian distributions because it has been proven that the median has the smallest *bias due to corruption* among all location estimators in a corruption neighbourhood of the Gaussian (ref. Lemma 5 and corresponding discussion in Section B). The fact that we use the median is also closely linked to the symmetry of the Gaussian distribution which means that estimating the median is the same as estimating the mean of the distribution. Let $\mathrm{Med}(\cdot)$ denote the median of the input distribution and define the maximum estimated median of the arms using the samples till time n as

$$\mathrm{Med}_*(n) := \max_a \ \mathrm{Med}(\hat{\mu}_a(n))\,.$$

We present in Algorithm 1 the `CRIMED` (Corruption Robust IMED) algorithm, which is an extension of the `IMED` algorithm of Honda and Takemura (2015) adapted to unbounded corruptions.

Note that in Algorithm 1 we introduce a forced exploration for $N_{\min}$ steps, where for $T > 0$,

$$N_{\min} := \left\lceil \frac{2\log(T)\log(1+\log(1+\log(T)))^2 s_\varepsilon^2}{\log(1+\log(T)^{0.99})} \right\rceil, \text{ and} \quad (3.1)$$

Algorithm 1: CRIMED for unit variance Gaussian bandits

Input: Horizon T, Corruption level ε, K

Initialisation phase: Compute $N_{\min}$ using Equation (3.1) and pull every arm $N_{\min}$ times.

for $n \in \{KN_{\min}+1, \dots, T-1, T\}$ **do**

Set $\mathrm{Med}_*(n) \leftarrow \max_a \mathrm{Med}(\hat{\mu}_a(n))$, $A_n^* \in \mathrm{argmax}_a \mathrm{Med}(\hat{\mu}_a(n))$, $I_{A_n^*}(n) \leftarrow \log N_{A_n^*}(n)$.

Compute, for each arm a different from A_n^*,

$$I_a(n) \leftarrow N_a(n)\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_a(n)) - \Delta_{\min}, \mathrm{Med}_*(n)\right) + \log N_a(n).$$

end Pull the arm $A_n \in \mathrm{argmin}_a I_a(n)$.

$$s_\varepsilon := \frac{\frac{1}{1-\varepsilon}}{\varphi\left(\frac{\Delta_{\min}}{2}+1\right)}\left(\frac{\varepsilon^{\frac{1}{2}}}{\sqrt{2}\log^{\frac{1}{2}}\left(\frac{1}{1-2\varepsilon}\right)} + \frac{(1-2\varepsilon)^{\frac{1}{2}}}{2\log^{\frac{1}{2}}\left(\frac{1-\varepsilon}{\varepsilon}\right)}\right). \tag{3.2}$$

The amount of forced-exploration $N_{\min}$ is proportional on the constant s_ε^2, which is a proxy of the variance of the empirical median from Theorem 3. It converges to a constant, $\frac{1}{2\varphi(1)}$, as $\varepsilon \to 0$.

Since we use empirical median as an estimate for the true mean using the corrupted observations, the empirically-optimal arm, or the arm with the maximum estimated mean is defined as $a^*(n) := \mathrm{argmax}_b\ \mathrm{Med}(\hat{\mu}_b(n))$. Moreover, since for this arm $\mathrm{kl}_{\mathcal{G}}^{\varepsilon}(\mathrm{Med}(\hat{\mu}_{a^*(n)}(n)) - \Delta_{\min}, \mathrm{Med}_*(n)) = 0$, its index is trivial to compute (Lemma 3(b)). For other arms, we can use the explicit formulation from Lemma 3 to compute $\mathrm{kl}_{\mathcal{G}}^{\varepsilon}$, while we compute c using a root-finding algorithm to normalise the corruption probabilities as stated in Lemma 1.

3.2. Theoretical results: Regret upper bound and concentration of median

We now present the theoretical guarantees of the proposed algorithm (Theorem 2), as well as the refined concentration inequality for median (Theorem 3, Lemma 4) that play a key role in our analysis and is of independent interest.

Theorem 2 (Asymptotic optimality of CRIMED under corruption) *For $\varepsilon \in (0, \frac{1}{2})$ and $\mu \in \mathcal{G}^K$ such that for $a \neq 1$, $m(\mu_1) - m(\mu_a) > \Delta_{\min}$, Algorithm 1 is asymptotically optimal, i.e.,*

$$\lim_{T\to\infty} \frac{\mathbb{E}[N_a(T)]}{\log(T)} \leq \frac{1}{\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(m(\mu_a), m(\mu_1)\right)}.$$

The upper bound from Theorem 2 matches with the lower bound from Proposition 2 showing that CRIMED is asymptotically optimal. We believe that the forced exploration for $N_{\min}$ steps is an artefact of the proof, and is needed in our analysis to handle the difficulties due to corruption. In Section 4, we numerically compare CRIMED and an aggressive version with $N_{\min} = 0$, called CRIMED*. We observe that CRIMED* suffers smaller regret.

The proof of Theorem 2, which we elaborate in Appendix D, proceeds by *controlling the probability of selecting a sub-optimal arm a* at each step n. CRIMED pulls an arm a at time n if the index for that arm $I_a(n)$ is the smallest. Thus, the probability of pulling an arm can be bounded by controlling the deviations of $\mathrm{kl}_{\mathcal{G}}^{\varepsilon}$ evaluated on the empirical estimates. This in turn is related to the probability of deviation of the empirical estimates themselves. In Theorem 3, we prove a new concentration result for the empirical median, which we leverage to prove that the regret of CRIMED

is well controlled. Later, in Lemma 4, we present the concentration results for the $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ evaluated at the empirical estimates. Given n samples $X_1, \dots, X_n$, $\mathrm{Med}(X_1^n)$ denotes the empirical median[2].

Theorem 3 (Concentration of median for corrupted Gaussians) *For $H \in \mathcal{P}(\mathbb{R})$, let $X_1, \dots, X_n$ be i.i.d. samples from $\mathcal{N}(m, 1) \odot_{\varepsilon} H$ and let $\varepsilon < \frac{1}{2}$. For $y \in [0, 1]$,*

$$\mathbb{P}\left(\mathrm{Med}(X_1^n) - m \geq \frac{\Delta_{\min}}{2} + y\right) \vee \mathbb{P}\left(\mathrm{Med}(X_1^n) - m \leq -\frac{\Delta_{\min}}{2} - y\right) \leq 2\exp\left(\frac{-ny^2}{s_{\varepsilon}^2}\right).$$

Observe from Equation (3.2) that when ε goes to 0, s_{ε} goes to $\frac{1}{2\varphi(1)}$. On the other hand, when ε goes to $\frac{1}{2}$, $\varphi\left(\frac{1}{2}\Delta_{\min} + 1\right)$ goes to 0, hence s_{ε} goes to ∞, and we get a trivial bound in the theorem.

Remark 4 (A refined concentration result) *Theorem 3 is an improvement over the concentration in (Altschuler et al., 2019, Lemma 7) in which the variance term in the concentration does not depend on ε. Moreover, Theorem 3, enables an ε arbitrarily close to $1/2$ which is in contrast with usual robust mean estimators that often feature a small upper limit on ε. For example, $\varepsilon \leq 1/7$ in (Wang and Ramdas, 2023, Theorem 2), and $\varepsilon \leq 1/15$ in (Altschuler et al., 2019, Theorem 18).*

Using Theorem 3 and properties of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$, we prove the following concentration for corrupted KL.

Lemma 4 (Concentration of corrupted KL) *Let $\delta > 0$, $x \in \mathbb{R}$. For $H \in \mathcal{P}(\mathbb{R})$, let $X_1, \dots, X_n$ be n i.i.d. samples generated from $\mathcal{N}(m_a, 1) \odot_{\varepsilon} H$.*

(a) For $y \in [0, 1]$, with probability at least $1 - 2\exp\left(-ny^2/s_{\varepsilon}^2\right)$,

$$\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(X_1^n) - \frac{\Delta_{\min}}{2}, m_a - \delta\right) \leq (y - \delta)_+ \left(|y - \delta| + \frac{\Delta_{\min}}{2}\right).$$

(b) For $m_b > m_a + \Delta_{\min}$ and $y \in [0, 1]$, with probability at least $1 - 2\exp\left(-ny^2/s_{\varepsilon}^2\right)$,

$$\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m_a, m_b\right) - \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(X_1^n) - \frac{\Delta_{\min}}{2}, m_b\right) \leq y(m_b - m_a + y + \Delta_{\min}).$$

For $y \in [0, \delta)$, the probability of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(\mathrm{Med}(X_1^n) - \frac{\Delta_{\min}}{2}, m_a - y)$ being 0 is strictly positive. We get this from Lemma 4(a). In contrast, in the uncorrupted-Gaussian setting, this is a zero probability event, except when $y = 0$. We extensively use this property of corrupted KL in the proof of Theorem 2, which follows from the thresholding property that $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x + \Delta) = 0$ holds for any $\Delta \leq \Delta_{\min}$. Specifically, the probability of being 0 coincides with that for $m_a - y - \mathrm{Med}(X_1^n) \leq \frac{\Delta_{\min}}{2}$.

Challenges in the regret analysis. To prove Theorem 2, we modify the proof for the regret bound of Honda and Takemura (2015). The major difference comes from the fact that Theorem 3 *does not allow us to reach arbitrarily large level of confidence*, i.e. with $y \leq 1$, the probability in Theorem 3 cannot be smaller than $\exp(-\Omega(n))$. This implies that very large deviation of the median do not imply very small probabilities. This is a known limitation of robust estimators (Devroye et al., 2016). As a consequence, we change the decomposition of the bad event $A_n = a$ to also include an event on which the deviation of the corrupted KL is large. We decompose the event $A_n = a$ as a union of three disjoint events. (i) $E_n(a)$: when the sub-optimal arm are not well estimated.

2. This can be alternatively seen as $\mathrm{Med}(\hat{\mu}_n)$ if $\hat{\mu}_n$ denotes their empirical distribution

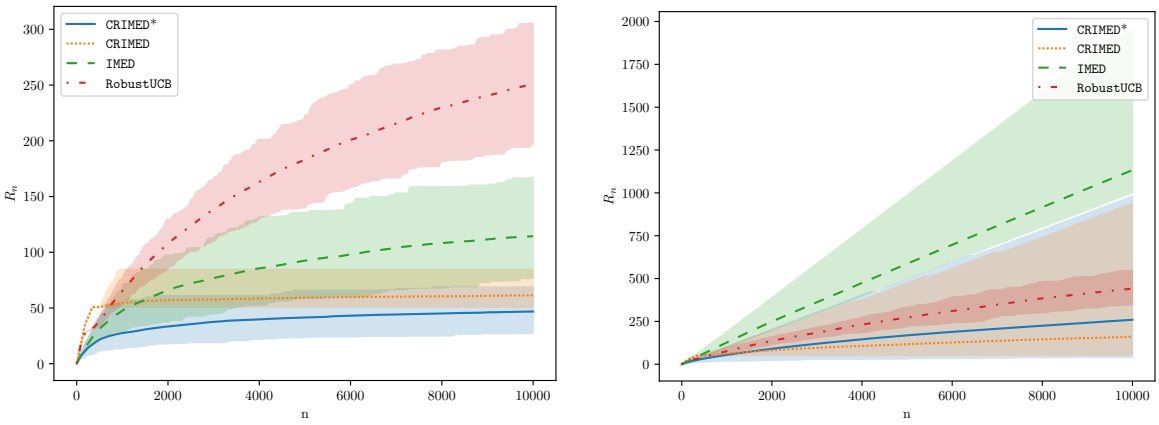


Figure 2: Cumulative regret for 100 repetitions of different algorithms on Setting 1 (left) and Setting 2 (right). Solid lines represent the means and shaded area are 90% percentile intervals.

(ii) $F_n(a)$: when the optimal arm is not well estimated and $G_n(a)$ when $\mathrm{kl}_{\mathcal{G}}^{\varepsilon}$ has large deviations (ref. Lemma 11 for formal definitions). We highlight that Lemma 4(a) with $y = \delta$, i.e. the fast concentration to 0, is specifically used to control the probabilities of events $F_n(a)$ and $G_n(a)$. (iii) Event $G_n(a)$ is further controlled thanks to the forced-exploration mechanism. Indeed, we observe that even refined concentration such as anytime concentration might only improve the lower order terms in the regret upper bound, but are not sufficient to control $G_n(a)$.

4. Experimental Analysis

In this section, we numerically illustrate the efficiency of our algorithm. The computation of indices in `CRIMED` depend on the threshold c that we compute using Equation (C.1) using the default *scipy* (Virtanen et al., 2020) root-finding algorithm. We consider distributions of the type $(1-\varepsilon)\mathcal{N}(m_a, \sigma_a^2) + \varepsilon\mathcal{N}(m_o, \sigma_o^2)$ with Gaussian inliers and Gaussian outliers in two different settings:

Parameters	Horizon	σ_a	means arms m_a	ε	means outliers m_b	σ_o outliers
Setting 1	5000	0.5	[0.8, 0.9, 1]	0.05	[1, 1, 0.8]	1
Setting 2	5000	0.5	[0.8, 0.9, 1]	0.05	[10, 10, -20]	1

Setting 1 corresponds to mild corruption in which the corrupted distribution of arm 2 still has the largest mean. In Setting 2, the corruption causes a change in the order of the arms and robustness is then needed to identify rightly the optimal arm. We compare four algorithms in each setting: `CRIMED` (Algorithm 1) with $N_{\min}$ set to Equation (3.1), `CRIMED*` is the aggressive version of `CRIMED`with $N_{\min}$ set to 1, `IMED` is the same as `CRIMED*` but in which there is no corruption ($\varepsilon = 0$) and the mean are estimated using the empirical mean, and finally `RobustUCB` is the algorithm from Basu et al. (2022). The results are illustrated in Figure 2.

Figure 2 illustrates that all the algorithms feature a logarithmic regret except `IMED` which features a linear regret in the case of large corruption (Setting 2). On the other hand, `CRIMED` and `CRIMED*` perform comparably and they are both performing better than `RobustUCB`.

5. Discussion and open questions

In this article, we studied a variant of stochastic multi-armed bandits corrupted by nature, with unbounded stochastic corruption. We studied the behaviour of KL divergences in a corrupted neighbourhood, and derived a tight lower bound on the achievable regret in the case of Gaussian arms with unbounded corruptions. This leads us to design of a corruption robust IMED-type algorithm, namely `CRIMED`, asymptotically achieves this lower bound. Additionally, we proposed a new concentration result for median that allows `CRIMED` to tackle corruption proportion up to $1/2$, which was not possible with existing algorithms.

To derive the upper bounds, we assumed that the observations were from corrupted Gaussian distributions. The Gaussian reward assumption does not seem to be essential and we believe it is possible to generalise these results to at least symmetric, unimodal distributions. Generalisation beyond symmetric, to non-symmetric and possibly non-parametric inliers is likely to be much more challenging. This is because the robust mean estimators are not consistent in this case, i.e., they do not converge to the true mean. This leads to the study of a non-trivial trade-off between robustness and asymptotic distance to the mean, which we leave as an open question.

References

Shipra Agrawal and Navin Goyal. Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on learning theory*, pages 39–1. JMLR Workshop and Conference Proceedings, 2012.

Shipra Agrawal and Navin Goyal. Near-optimal regret bounds for thompson sampling. *Journal of the ACM (JACM)*, 64(5):1–24, 2017.

Shubhada Agrawal. *Bandits with Heavy Tails: Algorithms, Analysis and Optimality*. PhD thesis, Tata Institute of Fundamental Research, Mumbai, 2022.

Shubhada Agrawal, Sandeep K Juneja, and Wouter M Koolen. Regret minimization in heavy-tailed bandits. In *Conference on Learning Theory*, pages 26–62. PMLR, 2021.

Jason Altschuler, Victor-Emmanuel Brunel, and Alan Malek. Best arm identification for contaminated bandits. *Journal of Machine Learning Research*, 20(91):1–39, 2019. URL http://jmlr.org/papers/v20/18-395.html.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multi-armed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.

Debabrota Basu, Odalric-Ambrym Maillard, and Timothée Mathieu. Bandits corrupted by nature: Lower bounds on regret and robust optimistic algorithm. *arXiv preprint arXiv:2203.03186*, 2022.

Hippolyte Bourel, Odalric Maillard, and Mohammad Sadegh Talebi. Tightening exploration in upper confidence reinforcement learning. In *International Conference on Machine Learning*, pages 1056–1066. PMLR, 2020.

GEP Box and SL Andersen. Preliminary results on a robust test for variances. Technical report, North Carolina State University. Dept. of Statistics, 1953.

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

Olivier Cappé, Aurélien Garivier, Odalric-Ambrym Maillard, Rémi Munos, and Gilles Stoltz. Kullback-leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, pages 1516–1541, 2013.

Olivier Catoni. Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. H. Poincaré Probab. Statist.*, 48(4):1148–1185, 11 2012. doi: 10.1214/11-AIHP454. URL https://doi.org/10.1214/11-AIHP454.

Mengjie Chen, Chao Gao, Zhao Ren, et al. Robust covariance and scatter matrix estimation under huber's contamination model. *The Annals of Statistics*, 46(5):1932–1960, 2018.

Luc Devroye, Matthieu Lerasle, Gabor Lugosi, and Roberto I. Oliveira. Sub-gaussian mean estimators. *The Annals of Statistics*, 44(6):2695–2725, 2016. ISSN 0090-5364. URL https://doi.org/10.1214/16-AOS1440.

Eyal Even-Dar, Shie Mannor, Yishay Mansour, and Sridhar Mahadevan. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(6), 2006.

Pratik Gajane, Tanguy Urvoy, and Emilie Kaufmann. Corrupt bandits for preserving local privacy. In *Algorithmic Learning Theory*, pages 387–412. PMLR, 2018.

Aurélien Garivier, Pierre Ménard, and Gilles Stoltz. Explore first, exploit next: The true shape of regret in bandit problems. *Mathematics of Operations Research*, 44(2):377–399, 2019.

Mohammad Hajiesmaili, Mohammad Sadegh Talebi, John Lui, Wing Shing Wong, et al. Adversarial bandits with corruptions: Regret lower bound and no-regret algorithm. *Advances in Neural Information Processing Systems*, 33:19943–19952, 2020.

Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. Wiley Series in Probability and Statistics. Wiley, 1st edition edition, January 1986. ISBN 0471829218. missing.

Junya Honda and Akimichi Takemura. An asymptotically optimal policy for finite support models in the multiarmed bandit problem. *arXiv preprint arXiv:0905.2776*, 2009.

Junya Honda and Akimichi Takemura. An asymptotically optimal bandit algorithm for bounded support models. In *COLT*, pages 67–79. Citeseer, 2010.

Junya Honda and Akimichi Takemura. Non-asymptotic analysis of a new bandit algorithm for semi-bounded rewards. *Journal of Machine Learning Research*, 16(113):3721–3756, 2015. URL http://jmlr.org/papers/v16/honda15a.html.

Peter J. Huber. Robust estimation of a location parameter. *Ann. Math. Statist.*, 35(1):73–101, 03 1964. ISSN 0003-4851. doi: 10.1214/aoms/1177703732. URL https://doi.org/10.1214/aoms/1177703732.

Peter J. Huber. A Robust Version of the Probability Ratio Test. *The Annals of Mathematical Statistics*, 36(6):1753 – 1758, 1965. doi: 10.1214/aoms/1177699803. URL https://doi.org/10.1214/aoms/1177699803.

Peter J Huber and Elvezio M Ronchetti. *Robust statistics; 2nd ed.* Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, 2009. URL https://cds.cern.ch/record/1254106.

Sayash Kapoor, Kumar Kshitij Patel, and Purushottam Kar. Corruption-tolerant bandit learning. *Machine Learning*, 108(4):687–715, 2019.

Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In *Algorithmic Learning Theory: 23rd International Conference, ALT 2012, Lyon, France, October 29-31, 2012. Proceedings 23*, pages 199–213. Springer, 2012.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Gábor Lugosi and Shahar Mendelson. Mean estimation and regression under heavy-tailed distributions: A survey. *Foundations of Computational Mathematics*, 19(5):1145–1190, 2019. doi: 10.1007/s10208-019-09427-x. URL https://doi.org/10.1007/s10208-019-09427-x.

Thodoris Lykouris, Vahab Mirrokni, and Renato Paes Leme. Stochastic bandits robust to adversarial corruptions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 114–122, 2018.

Konstantinos E Nikolakakis, Dionysios S Kalogerias, Or Sheffet, and Anand D Sarwate. Quantile multi-armed bandits: Optimal best-arm identification and a differentially private scheme. *IEEE Journal on Selected Areas in Information Theory*, 2(2):534–548, 2021.

Roman Pogodin and Tor Lattimore. On first-order bounds, variance and gap-dependent bounds for adversarial bandits. In *Uncertainty in Artificial Intelligence*, pages 894–904. PMLR, 2020.

Herbert Robbins. Some aspects of the sequential design of experiments. 1952.

Yevgeny Seldin and Aleksandrs Slivkins. One practical algorithm for both stochastic and adversarial bandits. In *International Conference on Machine Learning*, pages 1287–1295. PMLR, 2014.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der

Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Hongjian Wang and Aaditya Ramdas. Huber-robust confidence sequences. *arXiv preprint arXiv:2301.09573*, 2023.

Appendix

Appendix A. Proofs for results in Section 2.1 and Section 2.2: Lower bound and least informative pair

A.1. Proof of Theorem 1

Let the given bandit instance be denoted by $\mu \in \mathcal{L}^K$ and for $a \in [K]$ and let $\{X_{a,i}\}_{a,i}$, for $a \in [K], i \in \{1, \ldots, N_a(T)\}$ denote the $N_a(T)$ corrupted observations from arm a till time T. Under the instance μ with the corruption distributions $\mathbf{H} = \{H_1, \ldots, H_K\}$, the likelihood of observing the samples, denoted by $L_{\mu,\mathbf{H}}$, is

$$L_{\mu,\mathbf{H}} = \prod_{a=1}^{K} \prod_{i=1}^{N_a(T)} \left((1-\varepsilon)\mu_a(X_{a,i}) + \varepsilon H_a(X_{a,i})\right) = \prod_{a=1}^{K} \prod_{i=1}^{N_a(T)} \mu_a \odot_\varepsilon H_a(X_{a,i}).$$

Without loss of generality, we assume that arm 1 is the unique optimal arm in μ and establish the lower bound for the sub-optimal arm 2. To this end, consider an alternative bandit instance, $\nu = (\nu_1, \ldots, \nu_K)$, where, $\nu_b \in \mathcal{L}$ for each $b \in [K]$, for $b \neq 2$, $\nu_b = \mu_b$, and $m(\nu_2) \geq m^*(\mu)$. Clearly, $m^*(\nu) \geq m^*(\mu)$. The likelihood of observing samples under ν with corruption distributions $\mathbf{H}' = \{H'_1, \ldots, H'_K\}$, denoted by $L_{\nu,\mathbf{H}'}$ is given by

$$L_{\nu,\mathbf{H}'} = \prod_{a=1}^{K} \prod_{i=1}^{N_a(T)} \nu_a \odot_\varepsilon H'_a(X_{a,i}).$$

Writing the log-likelihood ratio,

$$LL_T = \sum_{a=1}^{K} \sum_{i=1}^{N_a(T)} \log\left(\frac{\mu_a \odot_\varepsilon H_a}{\nu_a \odot_\varepsilon H'_a}(X_{a,i})\right).$$

Taking average with respect to $\mu \odot_\varepsilon \mathbf{H}$, we get

$$\mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[LL_T\right] = \sum_{a=1}^{K} \mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[N_a(T)\right] \mathrm{KL}(\mu_a \odot_\varepsilon H_a, \nu_a \odot_\varepsilon H'_a).$$

Informally, let $\mathcal{F}_t$ be the $\sigma-$algebra generated by the randomness of the algorithm and the observations up to time t (see (Lattimore and Szepesvári, 2020, Chapter 4) for formal introduction to stochastic multi-armed bandits). An application of the data-processing inequality (see, Kaufmann et al. (2016); Garivier et al. (2019)) gives that for any $\mathcal{F}_T$ measurable event $\mathcal{E}_T$,

$$\sum_{a=1}^{K} \mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[N_a(T)\right] \mathrm{KL}(\mu_a \odot_\varepsilon H_a, \nu_a \odot_\varepsilon H'_a) \geq d\left(\mu_a \odot_\varepsilon H_a(\mathcal{E}_T), \nu_a \odot_\varepsilon H'_a(\mathcal{E}_T)\right),$$

where for $x \in (0,1)$ and $y \in (0,1)$, $d(x,y) := \mathrm{KL}(\mathrm{Ber}(x), \mathrm{Ber}(y))$ denotes the KL divergence between Bernoulli distributions with the given means. Since the r.h.s. above is true for all events $\mathcal{E}_T$ that are $\mathcal{F}_T$ measurable, optimizing over them we get

$$\sum_{a=1}^{K} \mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[N_a(T)\right] \mathrm{KL}(\mu_a \odot_\varepsilon H_a, \nu_a \odot_\varepsilon H'_a) \geq \sup_{\mathcal{E}_T \in \mathcal{F}_T} d\left(\mu_a \odot_\varepsilon H_a(\mathcal{E}_T), \nu_a \odot_\varepsilon H'_a(\mathcal{E}_T)\right).$$

Taking infimum over the corruptions $\mathbf{H} \in \mathcal{P}(\mathbb{R})^K$ and $\mathbf{H}' \in \mathcal{P}(\mathbb{R})^K$ on both sides, the above inequality implies

$$\inf_{\mathbf{H},\mathbf{H}'} \sum_{a=1}^{K} \mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[N_a(T)\right] \mathrm{KL}(\mu_a \odot_\varepsilon H_a, \nu_a \odot_\varepsilon H'_a) \\ \geq \inf_{\mathbf{H},\mathbf{H}'} \sup_{\mathcal{E}_T\in\mathcal{F}_T} d\left(\mu_a \odot_\varepsilon H_a(\mathcal{E}_T), \nu_a \odot_\varepsilon H'_a(\mathcal{E}_T)\right). \tag{A.1}$$

Since for every $\mathbf{H}$, $\mathbf{H}'$ with $H'_b = H_b$ for all $b \in [K]$ and $b \neq 2$ is a feasible candidate for $\mathbf{H}'$, the infimum in the l.h.s. above is at most

$$\left(\sup_{\mathbf{H}} \mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[N_2(T)\right]\right)\left(\inf_{H_2,H'_2} \mathrm{KL}(\mu_2 \odot_\varepsilon H_2, \nu_2 \odot_\varepsilon H'_2)\right). \tag{A.2}$$

Now, following along the arguments in Kaufmann et al. (2016) for the classical regret-minimization setting without corruption, we first choose

$$\mathcal{E}_T = \left\{N_1(T) \leq T - \sqrt{T}\right\}.$$

Then, we obtain by a simple application of Markov's inequality

$$\mathbb{P}_{\mu\odot_\varepsilon\mathbf{H}}\left(\mathcal{E}_T\right) = \mathbb{P}_{\mu\odot_\varepsilon\mathbf{H}}\left(T - N_1(T) \geq \sqrt{T}\right) \leq \frac{\sum_{a\neq 1} \mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[N_a(T)\right]}{\sqrt{T}} =: P_T^{\mathbf{H}},$$

and

$$\mathbb{P}_{\nu\odot_\varepsilon\mathbf{H}'}\left(\mathcal{E}_T^c\right) = \mathbb{P}_{\nu\odot_\varepsilon\mathbf{H}'}\left(N_1(T) \geq T - \sqrt{T}\right) \leq \frac{\sum_{a\neq 2} \mathbb{E}_{\nu\odot_\varepsilon\mathbf{H}'}\left[N_a(T)\right]}{T - \sqrt{T}} =: (Q_T^{\mathbf{H}'})^c.$$

Next, recall that the algorithm under consideration is uniformly-good (see Definition 1). This implies

$$\sup_{\mathbf{H}} \mathbb{P}_{\mu\odot_\varepsilon\mathbf{H}}\left(\mathcal{E}_T\right) \leq P_T := \sup_{H} P_T^H \xrightarrow{T\to\infty} 0,$$

and

$$\sup_{\mathbf{H}'} \mathbb{P}_{\nu\odot_\varepsilon\mathbf{H}'}\left(\mathcal{E}_T^c\right) \leq Q_T^c := \sup_{\mathbf{H}'} (Q_T^{\mathbf{H}'})^c \xrightarrow{T\to\infty} 0.$$

Clearly, for a fixed $\mathbf{H}$ and $\mathbf{H}'$, from the monotonicity of $d(\cdot,\cdot)$ in its arguments, it holds

$$d\left(\mathbb{P}_{\mu\odot_\varepsilon\mathbf{H}}\left(\mathcal{E}_T\right), \mathbb{P}_{\nu\odot_\varepsilon\mathbf{H}'}\left(\mathcal{E}_T\right)\right) \geq d\left(\sup_{\mathbf{H}} \mathbb{P}_{\mu\odot_\varepsilon\mathbf{H}}\left(\mathcal{E}_T\right), 1 - \sup_{\mathbf{H}'} \mathbb{P}_{\nu\odot_\varepsilon\mathbf{H}'}\left(\mathcal{E}_T^c\right)\right) \\ \geq d(P_T, Q_T). \tag{A.3}$$

Using (A.2) and (A.3) in (A.1), we get

$$\left(\sup_{\mathbf{H}} \mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[N_2(n)\right]\right)\left(\inf_{H_2,H'_2} \mathrm{KL}(\mu_2 \odot_\varepsilon H_2, \nu_2 \odot_\varepsilon H'_2)\right) \geq d(P_T, Q_T). \tag{A.4}$$

Next, we consider the following relation

$$\lim_{T\to\infty}\frac{d\left(P_T,Q_T\right)}{\log T}=\lim_{T\to\infty}\frac{1}{\log T}\log\frac{1}{Q_T^c}\geq\lim_{T\to\infty}\frac{1}{\log T}\log\frac{T-\sqrt{T}}{\sup_{\mathbf{H}'}\sum_{a\neq 2}\mathbb{E}_{\nu\odot_\varepsilon\mathbf{H}'}\left[N_a(T)\right]}.$$

We observe that the r.h.s. above equals

$$\lim_{T\to\infty}\left(1+\frac{\log\left(1-\frac{1}{\sqrt{T}}\right)}{\log T}-\frac{\log\left(\sup_{\mathbf{H}'}\sum_{a\neq 2}\mathbb{E}_{\nu\odot_\varepsilon\mathbf{H}'}\left[N_a(T)\right]\right)}{\log T}\right),$$

which in turns equals 1. Thus, we have obtained

$$\lim_{T\to\infty}\frac{d\left(P_T,Q_T\right)}{\log T}\geq 1.$$

Using this in Equation (A.4),

$$\liminf_{T\to\infty}\frac{\left(\sup_{\mathbf{H}}\mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[N_2(T)\right]\right)}{\log T}\geq\frac{1}{\inf_{H_2,H_2'}\mathrm{KL}_{\inf}^{\varepsilon}(\mu_2\odot_\varepsilon H_2,\nu_2\odot_\varepsilon H_2')}.$$

Since the above inequality is true for all the alternative bandit instances $\nu\in\mathcal{L}^K$ with $m(\nu_2)\geq m^*(\mu)$, we optimize over these to get

$$\liminf_{T\to\infty}\frac{\left(\sup_{\mathbf{H}}\mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[N_2(T)\right]\right)}{\log T}\geq\frac{1}{\mathrm{KL}_{\inf}(\mu_2,m^*(\mu);\mathcal{L})},$$

where $\mathrm{KL}_{\inf}^{\varepsilon}(\mu_2,m^*(\mu);\mathcal{L})$ equals

$$\inf\left\{\mathrm{KL}(\mu_2\odot_\varepsilon H_2,\nu_2\odot_\varepsilon H_2'):\ \nu_2\in\mathcal{L},\ m(\nu_2)\geq m^*(\mu),\ H_2\in\mathcal{P}(\mathbb{R}),\ H_2'\in\mathcal{P}(\mathbb{R})\right\}.$$

A.2. Proof of Lemma 1

Given $\eta\in\mathcal{L}$ and $\kappa\in\mathcal{L}$, we will show that H_1 and H_2 satisfying (2.2) and (2.3) are optimal for $\mathrm{KL}_{\inf}^{\varepsilon}$, for a fixed κ. To this end, consider any alternative corruption distributions, $H_1'\in\mathcal{P}(\mathbb{R})$ and $H_2'\in\mathcal{P}(\mathbb{R})$. For $t\in(0,1)$, define $H_{i,t}=(1-t)H_i+tH_i'$ for $i\in\{1,2\}$ and

$$J_{H_1',H_2'}(t)=\frac{1}{\varepsilon}\mathrm{KL}(\eta\odot_\varepsilon H_{1,t},\kappa\odot_\varepsilon H_{2,t}).$$

To prove the lemma, we show that $J_{H_1',H_2'}$ is a convex function that is minimized at $t=0$. Towards this, we derive

$$\begin{aligned}\frac{\mathrm{d}J_{H_1',H_2'}}{\mathrm{d}t}(t)&=\int\log\frac{\mathrm{d}\eta\odot_\varepsilon H_{1,t}}{\mathrm{d}\kappa\odot_\varepsilon H_{2,t}}(x)\left(\mathrm{d}H_1'-\mathrm{d}H_1\right)(x)\\&\qquad-\int\mathrm{d}\eta\odot_\varepsilon H_{1,t}(x)\left(\frac{(\mathrm{d}H_2'-\mathrm{d}H_2)}{\mathrm{d}\kappa\odot_\varepsilon H_{2,t}}(x)-\frac{(\mathrm{d}H_1'-\mathrm{d}H_1)}{\mathrm{d}\eta\odot_\varepsilon H_{1,t}}(x)\right)\\&=\int\log\frac{\mathrm{d}\eta\odot_\varepsilon H_{1,t}}{\mathrm{d}\kappa\odot_\varepsilon H_{2,t}}(x)\left(\mathrm{d}H_1'-\mathrm{d}H_1\right)(x)-\int\frac{\mathrm{d}\eta\odot_\varepsilon H_{1,t}}{\mathrm{d}\kappa\odot_\varepsilon H_{2,t}}(x)\left(\mathrm{d}H_2'-\mathrm{d}H_2\right)(x),\end{aligned}$$

where the last equality follows from the fact that H_1 and H_1' both integrate to 1. Differentiating again with respect to t, it comes

$$\begin{aligned}\frac{\mathrm{d}^2 J_{H_1',H_2'}}{\mathrm{d}t^2}(t) &= \int \left(\frac{\mathrm{d}H_1' - \mathrm{d}H_1}{\mathrm{d}\eta \odot_\varepsilon H_{1,t}} - \frac{\mathrm{d}H_2' - \mathrm{d}H_2}{\mathrm{d}\kappa \odot_\varepsilon H_{2,t}}\right)\left(\mathrm{d}H_1' - \mathrm{d}H_1\right)(x) \\ &\quad - \int \left(\frac{(\mathrm{d}H_1' - \mathrm{d}H_1)}{\mathrm{d}\kappa \odot_\varepsilon H_{2,t}}(x) - \frac{\mathrm{d}\eta \odot_\varepsilon H_{1,t}(\mathrm{d}H_2' - \mathrm{d}H_2)}{(\mathrm{d}\kappa \odot_\varepsilon H_{2,t})^2}(x)\right)\left(\mathrm{d}H_2' - \mathrm{d}H_2\right)(x) \\ &= \int \left(\frac{\mathrm{d}H_1' - \mathrm{d}H_1}{\sqrt{\mathrm{d}\eta \odot_\varepsilon H_{1,t}}} - \sqrt{\mathrm{d}\eta \odot_\varepsilon H_{1,t}}\frac{\mathrm{d}H_2' - \mathrm{d}H_2}{\mathrm{d}\kappa \odot_\varepsilon H_{2,t}}\right)^2 \geq 0,\end{aligned}$$

proving the convexity of $J_{H_1',H_2'}$ for any H_1', H_2'. Thus, it suffices to prove that its derivative is non-negative at $t = 0$. We now define the sets

$$A := \left\{x : \frac{d\eta}{\mathrm{d}\kappa}(x) < c_1\right\} \quad \text{and} \quad D := \left\{x : \frac{\mathrm{d}\eta}{\mathrm{d}\kappa}(x) > \frac{1}{c_2}\right\}. \tag{A.5}$$

Evaluating the term $\frac{\mathrm{d}J_{H_1',H_2'}}{\mathrm{d}t}(0)$, we get

$$\begin{aligned}\frac{\mathrm{d}J_{H_1',H_2'}}{\mathrm{d}t}(0) &= \int \log \frac{\mathrm{d}\eta \odot_\varepsilon H_1}{\mathrm{d}\kappa \odot_\varepsilon H_2}(x)\left(\mathrm{d}H_1' - \mathrm{d}H_1\right)(x) - \int \frac{\mathrm{d}\eta \odot_\varepsilon H_1}{\mathrm{d}\kappa \odot_\varepsilon H_2}(x)\left(\mathrm{d}H_2' - \mathrm{d}H_2\right)(x) \\ &= \int_A \log(c_2)\left(\mathrm{d}H_1' - \mathrm{d}H_1\right)(x) + \int_{A^c \cap D^c} \log \frac{\mathrm{d}\eta}{\mathrm{d}\kappa}(x)\mathrm{d}H_1'(x) + \int_D \log \frac{1}{c_1}\mathrm{d}H_1'(x) \\ &\qquad - \int_A c_2 \mathrm{d}H_2'(x) - \int_{A^c \cap D^c} \frac{\mathrm{d}\eta}{\mathrm{d}\kappa}(x)\mathrm{d}H_2'(x) - \int_D \frac{1}{c_1}\left(\mathrm{d}H_2' - \mathrm{d}H_2\right)(x) \\ &\geq \log c_2 \left(1 - H_1(A)\right) - \frac{1}{c_1}\left(1 - H_2(D)\right) \\ &= 0,\end{aligned}$$

where the last equality follows from the facts that H_1' and H_2' have supports equal to A and D, respectively, and integrate to 1.

Appendix B. Non-intersection of corruption neighbourhoods: Discussions and Proofs from Section 2.3

Let $\mathcal{T}$ be the set of all functionals $T : \mathcal{P} \to \mathbb{R}$ that are translation equivariant, meaning that for any $\tilde{\delta} > 0$, if κ is the law of X and η is the law of $X + \tilde{\delta}$, then $T(Q) = T(\kappa) + \tilde{\delta}$. For example, $\mathcal{T}$ includes common functionals like mean, median, mode, etc.

Now, when estimating the value of these functionals for a particular distribution (say κ) using samples that are generated from it, but are corrupted with probability $\varepsilon > 0$, one may not converge to the value for κ as the number of samples increases. For a given distribution κ and $\varepsilon > 0$, let κ_ε denote the collection of distributions in the corruption neighbourhood of κ, i.e.,

$$\kappa_\varepsilon := \{(1-\varepsilon)\kappa + \varepsilon H : \ H \in \mathcal{P}(\mathbb{R})\}.$$

The bias due to using the family of functionals $\mathcal{T}$ for distribution κ under corruption with probability ε is defined as

$$b_\kappa(\varepsilon) \coloneqq \inf_{T\in\mathcal{T}} \sup_{\kappa'\in\kappa_\varepsilon} |T(\kappa') - T(\kappa)|. \tag{B.1}$$

Lemma below is taken from (Huber and Ronchetti, 2009, Section 4.2) and is presented here for completeness. The proof of the lemma can be found in the same reference.

Lemma 5 (Optimality condition for median) *Let κ be a symmetric and unimodal distribution. The functional that achieves the infimum in $b_\kappa(\varepsilon)$ is median. Moreover,*

$$b_\kappa(\varepsilon) = F_\kappa^{-1}\left(\frac{1}{2(1-\varepsilon)}\right),$$

where F_κ is the c.d.f. of κ.

We now prove the equivalent conditions for $\mathrm{kl}^\varepsilon_{\mathcal{G}}$ to be non-zero.

B.1. Proof of Lemma 2

Without loss of generality, we assume that $m(\kappa) \le m(\eta)$. Define

$$b_0(\varepsilon) \coloneqq \Phi^{-1}\left(\frac{1}{2(1-\varepsilon)}\right).$$

We first prove that if for all $(H_1, H_2) \in \mathcal{P}(\mathbb{R})^2$, we have $\kappa \odot_\varepsilon H_1 \neq \eta \odot_\varepsilon H_2$, then $|m(\kappa)-m(\eta)| > 2\Phi^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)$ is true. We do this by showing the contrapositive. To this end, let us assume that $|m(\kappa) - m(\eta)| \le 2b_0(\varepsilon)$. Then

$$\exists\, \varepsilon' \le \varepsilon \quad \text{such that } |m(\kappa) - m(\eta)| = 2b_0(\varepsilon').$$

We construct a probability measure that belongs to the intersection of the corruption neighbourhoods of η and κ. Define

$$p'(x) \coloneqq \begin{cases} (1-\varepsilon')\varphi\left(x - m(\kappa)\right), & \text{for } (x - m(\kappa)) \le b_0(\varepsilon') \\ (1-\varepsilon')\varphi\left(x - m(\kappa) - 2b_0(\varepsilon')\right), & \text{for } (x - m(\kappa)) > b_0(\varepsilon'), \end{cases}.$$

We first show that $p' \in \kappa_\varepsilon$, i.e., it belongs to the corruption neighbourhood of κ. To this end, consider $(p' - (1-\varepsilon')\kappa)(x)$, which equals

$$\begin{cases} 0, & \text{for } (x - m(\kappa)) \le b_0(\varepsilon') \\ (1-\varepsilon')\left(\varphi\left(x - m(\kappa) - 2b_0(\varepsilon')\right) - \varphi\left(x - m(\kappa)\right)\right), & \text{for } (x - m(\kappa)) > b_0(\varepsilon'). \end{cases}$$

Now, if $x - m(\kappa) \in (b_0(\varepsilon'), 2b_0(\varepsilon')]$, then

$$0 \le 2b_0(\varepsilon') - (x - m(\kappa)) \le b_0(\varepsilon') \le x - m(\kappa).$$

Hence $p'(x) \ge 0$. Similarly, if $x - m(\kappa) \ge 2b_0(\varepsilon')$, then

$$0 \le x - m(\kappa) - 2b_0(\varepsilon') \le x - m(\kappa),$$

giving $p'(x) \geq 0$. Additionally,

$$\begin{aligned}
&\int_{\mathbb{R}} (p' - (1-\varepsilon')\kappa)(x)\mathrm{d}x \\
&= (1-\varepsilon') \int \left(\varphi\left(x - m(\kappa) - 2b_0(\varepsilon')\right) - \varphi\left(x - m(\kappa)\right)\right) \mathbb{1}\left\{x - m(\kappa) > b_0(\varepsilon')\right\} \\
&= (1-\varepsilon')\left(1 - \Phi\left(-b_0(\varepsilon')\right) - \left(1 - \Phi\left(b_0(\varepsilon')\right)\right)\right) \\
&= (1-\varepsilon')\left(\frac{1}{2(1-\varepsilon')} - 1 + \frac{1}{2(1-\varepsilon')}\right) = \varepsilon'.
\end{aligned}$$

Hence, $p' - (1-\varepsilon')\kappa$ is also a non-negative measure that sums to ε'. This implies that $p' \in \kappa_{\varepsilon'} \subset \kappa_\varepsilon$.

We next show that p' belongs to the corruption neighbourhood of η. Having that $|m(\kappa) - m(\eta)| = 2b_0(\varepsilon') = m(\eta) - m(\kappa)$, we can rewrite p' as

$$\begin{aligned}
p' &= \begin{cases} (1-\varepsilon')\varphi\left(x - m(\kappa)\right) & \text{for } x - m(\kappa) \leq b_0(\varepsilon') \\ (1-\varepsilon')\varphi\left(x - m(\kappa) - 2b_0(\varepsilon')\right) & \text{for } x - m(\kappa) > b_0(\varepsilon') \end{cases} \\
&= \begin{cases} (1-\varepsilon')\varphi\left(x - m(\eta) + 2b_0(\varepsilon')\right) & \text{for } x - m(\eta) \leq -b_0(\varepsilon') \\ (1-\varepsilon')\varphi\left(x - m(\eta)\right) & \text{for } x - m(\eta) > -b_0(\varepsilon'). \end{cases}
\end{aligned}$$

Then, $(p' - (1-\varepsilon')\eta)(x)$ equals

$$\begin{cases} (1-\varepsilon')\left(\varphi\left(x - m(\eta) + 2b_0(\varepsilon')\right) - \varphi\left(x - m(\eta)\right)\right) & \text{for } x - m(\eta) \leq -b_0(\varepsilon') \\ 0 & \text{for } x - m(\eta) > -b_0(\varepsilon'), \end{cases}$$

Again, if $x - m(\eta) \in (b_0(\varepsilon'), 2b_0(\varepsilon')]$, then

$$0 \leq 2b_0(\varepsilon') - (x - m(\eta)) \leq b_0(\varepsilon') \leq x - m(\eta),$$

implying that $p'(x) - (1-\varepsilon')\eta(x) \geq 0$. On the other hand, if $x - m(\eta) \geq 2b_0(\varepsilon')$, then

$$0 \leq x - m(\eta) - 2b_0(\varepsilon') \leq x - m(\eta),$$

and then $p'(x) - (1-\varepsilon')\eta(x) \geq 0$. Additionally, $p' - (1-\varepsilon')\eta$ sums to ε', as shown below:

$$\begin{aligned}
&\int_{\mathbb{R}} (p' - (1-\varepsilon')\eta)(x) \\
&= \int (1-\varepsilon')\left(\varphi\left(x - m(\eta) + 2b_0(\varepsilon')\right) - \varphi\left(x - m(\eta)\right)\right) \mathbb{1}\left\{x - m(\eta) \leq -b_0(\varepsilon')\right\} \\
&= (1-\varepsilon')\left(\Phi(b_0(\varepsilon')) - \Phi(-b_0(\varepsilon'))\right) \\
&= (1-\varepsilon')\left(\frac{1}{2(1-\varepsilon')} - 1 + \frac{1}{2(1-\varepsilon')}\right) = \varepsilon'.
\end{aligned}$$

Thus, p' also belongs to the corruption neighbourhood of η, i.e., $p' \in \eta_{\varepsilon'} \subset \eta_\varepsilon$, proving one direction.

We now prove that if $|m(\kappa) - m(\eta)| > 2b_0(\varepsilon)$, then $\kappa_\varepsilon \cap \eta_\varepsilon = \emptyset$, again by proving the contrapositive. Suppose $\exists \kappa' \in \kappa_\varepsilon \cap \eta_\varepsilon$. Since median is a minimax-bias functional, and $\kappa' \in \kappa_\varepsilon$,

$$|\mathrm{Med}(\kappa') - \mathrm{Med}(\kappa)| \leq b_0(\varepsilon).$$

Similarly, having $\kappa' \in \eta_\varepsilon$,

$$|\mathrm{Med}(\kappa') - \mathrm{Med}(\eta)| \le b_0(\varepsilon).$$

Hence, we obtain that

$$|m(\kappa) - m(\eta)| = |\mathrm{Med}(\kappa) - \mathrm{Med}(\eta)| \le 2b_0(\varepsilon),$$

proving the other direction.

Appendix C. Properties of $\mathrm{kl}^\varepsilon_{\mathcal{G}}$: a discussion and proofs

$\mathrm{kl}^\varepsilon_{\mathcal{G}}$ is crucial for our algorithm, both practically, and theoretically. We characterize its solutions and prove various nice properties that are useful in algorithmic implementation, as well as for its analysis. In this appendix, we discuss these properties of $\mathrm{kl}^\varepsilon_{\mathcal{G}}$, including those presented in the main text in Lemma 3.

Recall that for $x \in \mathbb{R}$, $y \in \mathbb{R}$,

$$\mathrm{kl}^\varepsilon_{\mathcal{G}}(x, y) \coloneqq \inf_{H, H'} \left\{ \mathrm{KL}(\mathcal{N}(x,1) \odot_\varepsilon H, \mathcal{N}(y,1) \odot_\varepsilon H') : \; H \in \mathcal{P}(\mathbb{R}),\; H' \in \mathcal{P}(\mathbb{R}) \right\}.$$

Further, recall that Lemma 1 characterizes the optimal H and H' for this problem, and are defined by Equation (2.2) and Equation (2.3). The lemma below identifies the support sets for the optimal corruption pair (H_1, H_2) in the specific settig of $\eta = \mathcal{N}(x, 1)$ and $\kappa = \mathcal{N}(y, 1)$.

We now prove Lemma 3, before going on to developing additional properties, which will be handy in the analysis later.

C.1. Proof of Lemma 3

Let η represent the cdf of $\mathcal{N}(x, 1)$ and κ be that for $\mathcal{N}(y, 1)$. Recall that c_1 and c_2 are the normalization constants for the corrupted distributions in $\mathrm{kl}^\varepsilon_{\mathcal{G}}(x, y)$.

Proof of Lemma 3(a): Since $\mathrm{d}(\mathcal{N}(x,1) \odot_\varepsilon H_1)$ is a probability distribution, it sums to 1. Let $W \sim \mathcal{N}(0, 1)$ be a random variable distributed according to standard Gaussian. Using the explicit form of A_{c_1} and D_{c_1} from Lemma 6, we have

$$\begin{aligned}
1 &= \int_{\mathbb{R}} \mathrm{d}(\eta \odot_\varepsilon H_1) \\
&= (1-\varepsilon)\left(c_1 \kappa(A_{c_1}) + \eta(\mathbb{R} \setminus A_{c_1})\right) \\
&= (1-\varepsilon)\left(c_1 \mathbb{P}\left(W + y \ge \frac{x+y}{2} + \frac{\log(\frac{1}{c_1})}{y-x}\right) + \mathbb{P}\left(W + x < \frac{x+y}{2} + \frac{\log(\frac{1}{c_1})}{y-x}\right)\right) \\
&= (1-\varepsilon)\left(c_1\left(1 - \Phi\left(\frac{x+y}{2} + \frac{\log(\frac{1}{c_1})}{y-x} - y\right)\right) + \Phi\left(\frac{x+y}{2} + \frac{\log(\frac{1}{c_1})}{y-x} - x\right)\right) \\
&= (1-\varepsilon)\left(c_1\left(1 - \Phi\left(-\frac{\Delta}{2} + \frac{\log(\frac{1}{c_1})}{\Delta}\right)\right) + \Phi\left(\frac{\Delta}{2} + \frac{\log(\frac{1}{c_1})}{\Delta}\right)\right).
\end{aligned}$$

Similarly, $\mathrm{d}(\mathcal{N}(y,1) \odot_\varepsilon H_2)$ is a probability distribution, it sums to 1, giving

$$\begin{aligned}
1 &= \int \mathrm{d}(\mathcal{N}(y,1) \odot_\varepsilon H_2) \\
&= (1-\varepsilon)\left(c_2 \eta(D_{c_2}) + \kappa(\mathbb{R} \setminus D_{c_2})\right) \\
&= (1-\varepsilon)\left(c_2 \mathbb{P}\left(W + x \le \frac{x+y}{2} - \frac{\log(\frac{1}{c_2})}{y-x}\right) + \mathbb{P}\left(W + y > \frac{x+y}{2} - \frac{\log(\frac{1}{c_2})}{y-x}\right)\right) \\
&= (1-\varepsilon)\left(c_2 \Phi\left(\frac{x+y}{2} - \frac{\log(\frac{1}{c_2})}{y-x} - x\right) + 1 - \Phi\left(\frac{x+y}{2} - \frac{\log(\frac{1}{c_2})}{y-x} - y\right)\right) \\
&= (1-\varepsilon)\left(c_2 \Phi\left(\frac{\Delta}{2} - \frac{\log(\frac{1}{c_2})}{\Delta}\right) + 1 - \Phi\left(-\frac{\Delta}{2} - \frac{\log(\frac{1}{c_2})}{\Delta}\right)\right) \\
&= (1-\varepsilon)\left(c_2\left(1 - \Phi\left(-\frac{\Delta}{2} + \frac{\log(\frac{1}{c_2})}{\Delta}\right)\right) + \Phi\left(\frac{\Delta}{2} + \frac{\log(\frac{1}{c_2})}{\Delta}\right)\right).
\end{aligned}$$

From the above, observe that c_1 and c_2 solve the same equation. Hence they can be taken to be equal to a common value, say $c > 0$. We now prove uniqueness of this common value c.

From the discussion in the previous paragraph, c solves the following equation:

$$\frac{1}{1-\varepsilon} = c\Phi\left(\frac{\Delta_-}{2}\right) + \Phi\left(\frac{\Delta_+}{2}\right). \tag{C.1}$$

Observe that c is uniquely defined by Equation (C.1), indeed $c \mapsto c\Phi\left(\frac{\Delta_-}{2}\right) + \Phi\left(\frac{\Delta_+}{2}\right)$ is increasing because its derivative is

$$\Phi\left(\frac{\Delta_-}{2}\right) + \frac{1}{\Delta}\varphi\left(\frac{\Delta_-}{2}\right) - \frac{1}{c\Delta}\varphi\left(\frac{\Delta_+}{2}\right) = \Phi\left(\frac{\Delta_-}{2}\right) > 0. \quad \square$$

Proof for Lemma 3(b): From Lemma 1 and the using part (a) above in the definition of $\mathrm{kl}^\varepsilon_{\mathcal{G}}$, we have for any $x < y$,

$$\begin{aligned}
\mathrm{kl}^\varepsilon_{\mathcal{G}}(x,y) =& (1-\varepsilon)\Big(\int_{A_c} c\varphi(t-y)\log(c) + \int_{D_c} \varphi(t-x)\log(1/c) \\
&+ \int_{\mathbb{R}\setminus A_c \cup D_c} \varphi(t-x)\log\left(\frac{\varphi(t-x)}{\varphi(t-y)}\right)\mathrm{d}t\Big).
\end{aligned} \tag{C.2}$$

On simplifying, it is then equal to

$$(1-\varepsilon)\left(c\log(c)\,\Phi\left(\frac{\Delta_-}{2}\right) + \log(1/c)\Phi\left(\frac{\Delta_-}{2}\right) + \int_{\mathbb{R}\setminus A_c \cup D_c} \varphi(t-x)\log\left(\frac{\varphi(t-x)}{\varphi(t-y)}\right)\mathrm{d}t\right). \tag{C.3}$$

Let us now compute the integral on $\mathbb{R} \setminus A_c \cup D_c$. For this, let $a < b$. Then clearly,

$$\int_a^b \varphi(t-x)\log\left(\frac{\varphi(t-x)}{\varphi(t-y)}\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{(t-x)^2}{2}}\left(-\frac{(t-x)^2}{2}+\frac{(t-y)^2}{2}\right)\mathrm{d}t$$
$$= (x-y)\left(\frac{1}{\sqrt{2\pi}}\int_a^b te^{-\frac{(t-x)^2}{2}}\mathrm{d}t - \frac{x+y}{2}(\Phi(b-x)-\Phi(a-x))\right).$$

Using the mean of a truncated-Gaussian random variable, we get

$$\int_a^b \varphi(t-x)\log\left(\frac{\varphi(t-x)}{\varphi(t-y)}\right)$$
$$= (x-y)\left(x(\Phi(b-x)-\Phi(a-x)) + \varphi(a-x) - \varphi(b-x) - \frac{x+y}{2}(\Phi(b-x)-\Phi(a-x))\right)$$
$$= \frac{(x-y)^2}{2}(\Phi(b-x)-\Phi(a-x)) + (x-y)\left(\varphi(a-x)-\varphi(b-x)\right).$$

Now, substituting $\Delta = y-x$, $a = x+\frac{\Delta_-}{2}$ and $b = x+\frac{\Delta_+}{2}$, we have that the above integral equals

$$\frac{\Delta^2}{2}\left(\Phi\left(\frac{\Delta_+}{2}\right)-\Phi\left(\frac{\Delta_-}{2}\right)\right) - \Delta\left(\varphi\left(\frac{\Delta_-}{2}\right)-\varphi\left(\frac{\Delta_+}{2}\right)\right).$$

Substituting this in Equation (C.2) we have

$$\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,y) = (1-\varepsilon)\left(c\log(c)\,\Phi\left(\frac{\Delta_-}{2}\right) + \log(1/c)\Phi\left(\frac{\Delta_-}{2}\right) + \frac{\Delta^2}{2}\left(\Phi\left(\frac{\Delta_+}{2}\right)-\Phi\left(\frac{\Delta_-}{2}\right)\right)\right.$$
$$\left. -\Delta\left(\varphi\left(\frac{\Delta_-}{2}\right)-\varphi\left(\frac{\Delta_+}{2}\right)\right)\right). \tag{C.4}$$

Shift invariance now follows from the above expression for $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ only in terms of Δ. □

Proof for Lemma 3(c): Recall the defining equation for the normalizing constant c from Equation (C.1). Observe that c is a function of Δ. Then, by implicit function theorem, c is differentiable. Let c' denote the derivative of c with respect to Δ. Then, using the expressions for derivatives from Lemma 9,

$$\frac{\partial}{\partial\Delta}\varphi\left(\frac{\Delta_+}{2}\right) = \varphi\left(\frac{\Delta_+}{2}\right)\left(-\frac{\Delta_+\Delta_-}{4\Delta} - \frac{\Delta_+\varphi(\Delta_-/2)}{2\Delta\Phi(\Delta_-/2)}\right),$$

and similarly,

$$\frac{\partial}{\partial\Delta}\varphi\left(\frac{\Delta_-}{2}\right) = \varphi\left(\frac{\Delta_-}{2}\right)\left(-\frac{\Delta_+\Delta_-}{4\Delta} + \frac{\Delta_-\varphi(\Delta_-/2)}{2\Delta\Phi(\Delta_-/2)}\right).$$

Since $\Delta \mapsto c$ is differentiable with continuous derivative on $(0,\infty)$, from Equation (C.4), $\Delta \mapsto \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x+\Delta)$ is also differentiable with continuous derivative on $(0,\infty)$. Differentiating Equation (C.4) with respect to Δ (after setting $y = x+\Delta$), and substituting for c' from Lemma 9, we have that

$$
\begin{aligned}
&\frac{1}{(1-\varepsilon)}\frac{\partial \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x+\Delta)}{\partial \Delta} \\
&= -c\log c\phi\left(\frac{\Delta_-}{2}\right) + \frac{c\log c}{2}\varphi\left(\frac{\Delta_-}{2}\right)\frac{\Delta_+}{\Delta} - \frac{c\log c}{\Delta}\frac{\varphi^2(\Delta_-/2)}{\Phi(\Delta_-/2)} \\
&\quad - \log c\frac{\Delta_+}{2\Delta}\varphi\left(\frac{\Delta_-}{2}\right) + \frac{\log c}{\Delta}\frac{\varphi^2(\Delta_-/2)}{\Phi(\Delta_-/2)} \\
&\quad + \Delta\left(\Phi\left(\frac{\Delta_+}{2}\right) - \Phi\left(\frac{\Delta_-}{2}\right)\right) + \frac{\Delta\Delta_-}{4}\varphi\left(\frac{\Delta_+}{2}\right) - \frac{\Delta\Delta_+}{4}\varphi\left(\frac{\Delta_-}{2}\right) \\
&\quad + \frac{\Delta}{2}\frac{\varphi(\Delta_-/2)\varphi(\Delta_+/2)}{\Phi(\Delta_-/2)} + \frac{\Delta}{2}\frac{\varphi^2(\Delta_-/2)}{\Phi(\Delta_-/2)} \\
&\quad + \frac{\Delta_-\Delta_+}{4}\varphi\left(\frac{\Delta_-}{2}\right) - \frac{\Delta_+\Delta_-}{4}\varphi\left(\frac{\Delta_+}{2}\right) - \frac{\Delta_-\varphi^2(\Delta_-/2)}{2\Phi(\Delta_-/2)} - \frac{\Delta_+\varphi(\Delta_+/2)\varphi(\Delta_-/2)}{2\Phi(\Delta_-/2)}.
\end{aligned}
$$

Next, using that $c\varphi(\Delta_-/2) = \varphi(\Delta_+/2)$ and collecting the coefficients of like-terms, the required derivative scaled by $1-\varepsilon$ equals

$$
\begin{aligned}
&\Delta\left(\Phi\left(\frac{\Delta_+}{2}\right) - \Phi\left(\frac{\Delta_-}{2}\right)\right) \\
&\quad + \varphi\left(\frac{\Delta_+}{2}\right)\left(\log\frac{1}{c} - \frac{\Delta_+}{2\Delta}\log\frac{1}{c} + \frac{\Delta_+}{2\Delta c}\log\frac{1}{c} + \frac{\Delta\Delta_-}{4} - \frac{\Delta\Delta_+}{4c} + \frac{\Delta_+\Delta_-}{4c} - \frac{\Delta_+\Delta_-}{4}\right) \\
&\qquad + \frac{\varphi^2(\Delta_+/2)}{\Phi(\Delta_-/2)}\left(\frac{1}{c\Delta}\log\frac{1}{c} - \frac{1}{c^2\Delta}\log\frac{1}{c} + \frac{\Delta}{2c} + \frac{\Delta}{2c^2} - \frac{\Delta_-}{2c^2} - \frac{\Delta_+}{2c}\right).
\end{aligned}
$$

Substituting for Δ_+ and Δ_- in the above expression, one can see that the coefficients of $\varphi(\Delta_+/2)$ and $\frac{\varphi^2(\Delta_+/2)}{\Phi(\Delta_-/2)}$ are 0, giving

$$
\frac{1}{1-\varepsilon}\frac{\partial \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x+\Delta)}{\partial \Delta} = \Delta\left(\Phi(\Delta_+/2) - \Phi(\Delta_-/2)\right).
$$

For the inequality, observe that by definition of c, we have

$$
\Phi\left(\frac{\Delta_-}{2}\right) \geq c\Phi\left(\frac{\Delta_-}{2}\right) = \frac{1}{1-\varepsilon} - \Phi\left(\frac{\Delta_+}{2}\right).
$$

Using this inequality in the derivative $\frac{\partial \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x+\Delta)}{\partial \Delta}$ we get the result. □

C.2. Additional properties of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$

In this section, we state various properties of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ derived from the definitions of the optimal pair of corrupted distributions from Lemma 1.

Lemma 6 *Let $y > x + \Delta_{\min}$. Let H_1 and H_2 be the pair of distributions from Lemma 1 for $\eta = \mathcal{N}(x, 1)$ and $\kappa = \mathcal{N}(y, 1)$. Then, $\mathrm{Sp}(H_1) = A_{c_1}$ and $\mathrm{Sp}(H_2) = D_{c_2}$, where*

$$
A_{c_1} = \left\{t \in \mathbb{R} : \quad t \geq \frac{y+x}{2} + \frac{\log(1/c_1)}{y-x}\right\} \quad \text{and} \quad D_{c_2} = \left\{t \in \mathbb{R} : \quad t \leq \frac{x+y}{2} - \frac{\log(1/c_1)}{y-x}\right\}.
$$

Proof First, by the definitions of H_1 and H_2, we have that H_1 is supported on

$$\begin{aligned} A_{c_1} = \left\{ \frac{\mathrm{d}\mathcal{N}(x,1)}{\mathrm{d}\mathcal{N}(y,1)}(t) \le c_1 \right\} &= \left\{ \log\left(\frac{\mathrm{d}\mathcal{N}(x,1)}{\mathrm{d}\mathcal{N}(y,1)}(t) \right) \le -\log(\frac{1}{c_1}) \right\} \\ &= \left\{ \frac{(t-y)^2}{2} - \frac{(t-x)^2}{2} \le -\log(\frac{1}{c_1}) \right\} \\ &= \left\{ t(x-y) + \frac{y^2-x^2}{2} \le -\log(\frac{1}{c_1}) \right\} \\ &= \left\{ t \ge \frac{x+y}{2} + \frac{\log(\frac{1}{c_1})}{y-x} \right\}. \end{aligned}$$

Similarly, we have the rewriting

$$\begin{aligned} D_{c_2} = \left\{ \frac{\mathrm{d}\mathcal{N}(x,1)}{\mathrm{d}\mathcal{N}(y,1)}(t) \ge \frac{1}{c_2} \right\} &= \left\{ \log\left(\frac{\mathrm{d}\mathcal{N}(x,1)}{\mathrm{d}\mathcal{N}(y,1)}(t) \right) \ge \log(\frac{1}{c_2}) \right\} \\ &= \left\{ t(x-y) + \frac{y^2-x^2}{2} \ge \log(\frac{1}{c_2}) \right\} \\ &= \left\{ t \le \frac{x+y}{2} - \frac{\log(\frac{1}{c_2})}{y-x} \right\}. \end{aligned}$$

■

Lemma 7 *For $y > x$, define $\Delta = y - x$,*

$$\Delta_+ := \Delta + 2\log\left(\frac{1}{c}\right)\frac{1}{\Delta} \quad \textit{and} \quad \Delta_- := \Delta - 2\log\left(\frac{1}{c}\right)\frac{1}{\Delta},$$

where c is the normalization constant. $\mathrm{Sp}\,(H_1) = A_c$ *and* $\mathrm{Sp}\,(H_2) = D_c$*, where*

$$A_c = \left\{ x \ge \frac{\Delta_+}{2} + m(\eta) \right\} \quad \textit{and} \quad D_c = \left\{ x \le \frac{\Delta_-}{2} + m(\eta) \right\}.$$

Proof This follows from Lemma 7 with $c_1 = c_2 = c$. ■

Lemma 8 *For any $x < y$, we have*

$$\begin{aligned} \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,y) = (1-\varepsilon)\Bigg((1-c)\log\left(\frac{1}{c}\right)\Phi\left(\frac{\Delta_-}{2}\right) + \frac{\Delta^2}{2}\left(\Phi\left(\frac{\Delta_+}{2}\right) - \Phi\left(\frac{\Delta_-}{2}\right)\right) \\ -\Delta\left(\varphi\left(\frac{\Delta_-}{2}\right) - \varphi\left(\frac{\Delta_+}{2}\right)\right)\Bigg). \end{aligned} \tag{C.5}$$

The above result was proven in the proof of Lemma 3(b).

Lemma 9 *We have that c is a continuous function of Δ with continuous derivative on $(0,\infty)$. Moreover, for any $\Delta > 0$,*

$$c' = \frac{-c\varphi(\Delta_-/2)}{\Phi(\Delta_-/2)}, \quad c\varphi\left(\frac{\Delta_-}{2}\right) = \varphi\left(\frac{\Delta_+}{2}\right), \quad \frac{\partial\Delta_+}{\partial\Delta} = \frac{\Delta_-}{\Delta} - \frac{2c'}{\Delta c}, \quad \frac{\partial\Delta_-}{\partial\Delta} = \frac{\Delta_+}{\Delta} + \frac{2c'}{\Delta c}.$$

Proof c is defined by the following equation:

$$\frac{1}{1-\varepsilon} = c\Phi\left(\frac{\Delta_-}{2}\right) + \Phi\left(\frac{\Delta_+}{2}\right).$$

Because $\Delta \mapsto \Delta_+$, $\Delta \mapsto \Delta_-$ and Φ are all differentiable with continuous derivative on $(0,\infty)$, we have by implicit function theorem, c is a differentiable function of Δ with continuous derivative, let us denote c' this derivative. We have on the one hand

$$\Delta'_+ := \frac{\mathrm{d}}{\mathrm{d}\Delta}\Delta_+ = 1 - 2\frac{c'(\Delta)}{\Delta c(\Delta)} - 2\frac{\log(1/c(\Delta))}{\Delta^2} = \frac{\Delta_-}{\Delta} - 2\frac{c'(\Delta)}{\Delta c(\Delta)},$$

and no the other hand

$$\Delta'_- := \frac{\mathrm{d}}{\mathrm{d}\Delta}\Delta_- = \frac{\Delta_+}{\Delta} + 2\frac{c'(\Delta)}{\Delta c(\Delta)}.$$

Then, taking the derivative with respect to Δ in Equation (C.1),

$$\begin{aligned}
0 =& c'(\Delta)\Phi\left(\frac{\Delta_-}{2}\right) + c(\Delta)\left(\frac{\Delta'_-}{2}\right)\varphi\left(\frac{\Delta_-}{2}\right) + \left(\frac{\Delta'_+}{2}\right)\varphi\left(\frac{\Delta_+}{2}\right) \\
=& c'(\Delta)\left(\Phi\left(\frac{\Delta_-}{2}\right) + \frac{1}{\Delta}\varphi\left(\frac{\Delta_-}{2}\right) - \frac{1}{c(\Delta)\Delta}\varphi\left(\frac{\Delta_+}{2}\right)\right) \\
&+ c(\Delta)\frac{\Delta_+}{2\Delta}\varphi\left(\frac{\Delta_-}{2}\right) + \frac{\Delta_-}{2\Delta}\varphi\left(\frac{\Delta_+}{2}\right).
\end{aligned} \tag{C.6}$$

Now, remark that $\Delta_+^2 = \Delta_-^2 + 8\log(1/c(\Delta))$, hence

$$\varphi\left(\frac{\Delta_+}{2}\right) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\Delta_+^2}{8}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{\Delta_-^2}{8}+\log(c)} = c(\Delta)\varphi\left(\frac{\Delta_-}{2}\right). \tag{C.7}$$

Plugging this in Equation (C.6), we have

$$0 = c'(\Delta)\Phi\left(\frac{\Delta_-}{2}\right) + c(\Delta)\varphi\left(\frac{\Delta_-}{2}\right).$$

Hence, we deduce that

$$c'(\Delta) = -\frac{c(\Delta)\varphi\left(\frac{\Delta_-}{2}\right)}{\Phi\left(\frac{\Delta_-}{2}\right)}. \tag{C.8}$$

■

As a direct consequence of the above properties of $\mathrm{kl}_{\mathcal{G}}^{\varepsilon}$ and Taylor's inequality, we also have the following mean-value theorem for $\mathrm{kl}_{\mathcal{G}}^{\varepsilon}$.

Lemma 10 (Mean-value theorem for $\mathrm{kl}^\varepsilon_{\mathcal{G}}$) *Suppose that $\mu_a \sim \mathcal{N}(m_a, 1)$, $\mu_b \sim \mathcal{N}(m_b, 1)$ and $m_* \in \mathbb{R}$ with both $\Delta_a := m_* - m_a > \Delta_{\min}$ and $\Delta_b := m_* - m_b > \Delta_{\min}$. Then,*

$$\mathrm{kl}^\varepsilon_{\mathcal{G}}(\mu_a, m_*) - \mathrm{kl}^\varepsilon_{\mathcal{G}}(\mu_b, m_*) \le (1-\varepsilon)(m_b - m_a)_+ \left(\Delta_a \vee \Delta_b\right).$$

Proof By Lemma 3, we have $\mathrm{KL}^\varepsilon_{\inf}(\nu_a, m_*) = \mathrm{kl}^\varepsilon_{\mathcal{G}}(m_a, m_*)$ and similarly for $\mathrm{KL}^\varepsilon_{\inf}(\nu_b, m_*)$. Using this and the shift invariance from Lemma 3(b),

$$\begin{aligned}\mathrm{KL}^\varepsilon_{\inf}(\nu_a, m_*) - \mathrm{KL}^\varepsilon_{\inf}(\nu_b, m_*) &= \mathrm{kl}^\varepsilon_{\mathcal{G}}(m_a, m_*) - \mathrm{kl}^\varepsilon_{\mathcal{G}}(m_b, m_*) \\ &= \mathrm{kl}^\varepsilon_{\mathcal{G}}(m_*, 2m_* - m_a) - \mathrm{kl}^\varepsilon_{\mathcal{G}}(m_*, 2m_* - m_b),\end{aligned}$$

and then, denoting $\Delta_a = m_* - m_a =$ and $\Delta_b = m_* - m_b$, if $m_a < m_b$ then from Taylor's inequality and Lemma 3,

$$\begin{aligned}&\mathrm{KL}^\varepsilon_{\inf}(\nu_a, m_*) - \mathrm{KL}^\varepsilon_{\inf}(\nu_b, m_*) \\ &\le (m_b - m_a) \sup_{t\in(0,1)} \left| \frac{\partial \mathrm{kl}^\varepsilon_{\mathcal{G}}(x, x+\Delta)}{\partial \Delta}\Big|_{\Delta=(1-t)(m_*-m_a)+t(m_*-m_b)} \right| \\ &\le (m_b - m_a)(1-\varepsilon) \sup_{\Delta=(1-t)(m_*-m_a)+t(m_*-m_b), t\in(0,1)} \Delta\left(2\Phi\left(\frac{\Delta_+}{2}\right) - \frac{1}{1-\varepsilon}\right) \\ &\le (1-\varepsilon)(m_b - m_a)\left(\Delta_a \vee \Delta_b\right).\end{aligned}$$

On the other hand, if $m_a \ge m_b$, then $\mathrm{KL}^\varepsilon_{\inf}(\nu_a, m_*) - \mathrm{KL}^\varepsilon_{\inf}(\nu_b, m_*) \le 0$. ■

Observe that Lemma 10 gives a bound very similar to that in the Gaussian setting without corruptions. Indeed, in the latter case,

$$\mathrm{KL}(\mu_a, \mathcal{N}(m_*, 1)) - \mathrm{KL}(\mu_b, \mathcal{N}(m_*, 1)) = \frac{(\Delta_a^2 - \Delta_b^2)}{2} \le (\Delta_a - \Delta_b)(\Delta_a \vee \Delta_b).$$

Lemma 10 is tight for Δ_a and Δ_b around $\Delta_{\min}$ but not when Δ_a and Δ_b are large, this is due to having bounded the derivative of $\mathrm{kl}^\varepsilon_{\mathcal{G}}(x, x+\Delta)$ by Δ in the proof, for simplicity because handling $\Phi(\Delta_+/2) - \Phi(\Delta_-/2)$ require knowledge on c which is defined implicitely.

Appendix D. Proofs of results from Section 3

D.1. Proof of Theorem 2: regret upper bound

To prove Theorem 2, we use that $N_a(T) = \sum_{n=1}^T \mathbb{1}\{A_n = a\}$ and decompose $\{A_n = a\}$ using Lemma 11 below.

Lemma 11 (Decomposition of bad event) *Let $M > 0$, we have $\{A_n = a\} \subset E_n(a) \cup F_n(a) \cup G_n(a)$ where $E_n(a)$, $F_n(a)$ and $G_n(a)$ are disjoint events defined by*

$$E_n(a) = \left\{ A_n = a, N_a(n)\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_a(n)) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq \log n \right\},$$

$$F_n(a) = \bigcup_{t=N_{\min}}^{n} \Big\{ A_n = a,\ \mathrm{Med}(\hat{\mu}_{1,t}) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2},$$
$$I_*(n) \leq t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) + \log t \leq tM + \log t \Big\},$$

$$G_n(a) = \bigcup_{t=N_{\min}}^{n} \left\{ A_n = a, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \geq M, N_1(n) = t \right\},$$

where $\hat{\mu}_{a,t}$ denotes the empirical distribution obtained using t samples generated from arm a.

Using Lemma 11, observe that for $T \geq KN_{\min}$,

$$N_a(T) \leq N_{\min} + \sum_{n=KN_{\min}}^{T} \mathbb{1}\left(E_n(a)\right) + \sum_{n=KN_{\min}}^{T} \mathbb{1}\left(F_n(a)\right) + \sum_{n=KN_{\min}}^{T} \mathbb{1}\left(G_n(a)\right).$$

Thus, to bound the average number of pulls of sub-optimal arm a, it suffices to bound the summation of the probabilities of the above indicator functions.

$$\mathbb{E}\left(N_a(T)\right) \leq N_{\min} + \sum_{n=KN_{\min}}^{T} \mathbb{P}\left(E_n(a)\right) + \sum_{n=1}^{T} \mathbb{P}\left(F_n(a)\right) + \sum_{n=1}^{T} \mathbb{P}\left(G_n(a)\right), \tag{D.1}$$

where $\sum_n \mathbb{P}(E_n(a))$ is at most

$$\sum_{n=1}^{T} \mathbb{P}\left(A_n = a, N_a(n)\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(\mathrm{Med}(\hat{\mu}_a(n)) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta) \leq \log n\right), \tag{D.2}$$

$\sum_n \mathbb{P}(F_n(a))$ is equal to

$$\mathbb{E}\Big(\sum_{n=KN_{\min}}^{T} \sum_{t=N_{\min}}^{n} \mathbb{1}\Big(A_n = a,\ \mathrm{Med}(\hat{\mu}_{1,t}) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2},$$
$$I_*(n) \leq t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) + \log t \leq tM + \log(t)\Big)\Big). \tag{D.3}$$

and the third term, i.e. $\sum_n \mathbb{E}(G_n(a))$, is at most

$$\sum_{n=1}^{T} \bigcup_{t=1}^{n} \left\{ \mathbb{P}\left(\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \geq M, N_1(n) = t\right\}\right). \tag{D.4}$$

Here, Equation (D.2) corresponds to the deviation of sub-optimal arm a, which will contribute to main term in the total regret, while Equation (D.3) corresponds to the deviation of the optimal arm, whose total contribution to the regret will at most be a constant and Equation (D.4) corresponds to large deviations of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ on the optimal arm.

First, we bound the probability of event $E_n(a)$ occurring with the following Lemma, this will give us the main term in the regret.

Lemma 12 *For any* $\delta < \min\left(1, \Delta_a + \Delta_{\min}, \frac{1}{4(\Delta_a+\Delta_{\min})}\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m(\mu_a), m(\mu_1)\right)\right)$*, we have that*

$$\sum_n \mathbb{P}(E_n(a)) \leq \frac{\log(T)}{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m(\mu_a), m(\mu_1)\right) - 2\delta\left(\Delta_a + \delta + \Delta_{\min}\right)} + \frac{4}{1-\exp\left(-\delta^2/s_{\varepsilon}^2\right)},$$

Then, we bound the probability of both event $F_n(a)$ and $G_n(a)$ in the two following lemmas. These two events will have a negligible probability compared to the probability of event $E_n(a)$ occuring.

Lemma 13 *Suppose* $\delta < 1$ *and set* $M = \delta^2/(2s_{\varepsilon}^2)$*. Then, we have*

$$\sum_{n=1}^{T} \mathbb{P}(F_n(a)) \leq \frac{e^{-\frac{\delta^2}{s_{\varepsilon}^2}}}{\left(1-\exp\left(-\frac{\delta^2}{s_{\varepsilon}^2}\right)\right)^2} + \frac{2}{\left(1-\exp\left(-\frac{\delta^2}{2s_{\varepsilon}^2}\right)\right)^2} \leq \frac{4}{\left(1-\exp\left(-\frac{\delta^2}{2s_{\varepsilon}^2}\right)\right)^2}.$$

Lemma 14 *Let* $N_{\min}$ *be given by*

$$N_{\min} = \left\lceil \frac{2\log(T)s_{\varepsilon}^2}{\log(1+\log(T)^{0.99})\delta^2} \right\rceil,$$

Then, for any value of $M > 0$*, we have*

$$\sum_{n=1}^{T} \mathbb{P}(G_n(a)) \leq 1 + \log(T)^{0.99}.$$

Using the Lemmas 12, 13, 14 and injecting into Equation (D.1), we get

$$\begin{aligned}\mathbb{E}[N_a(T)] \leq & \frac{\log(T)}{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m(\mu_a), m(\mu_1)\right) - 2\delta\left(\Delta_a + \delta + \Delta_{\min}\right)} \\ & + \left\lceil \frac{2\log(T)s_{\varepsilon}^2}{\log(1+\log(T)^{0.99})\delta^2} \right\rceil + \sqrt{\log(T)} + \frac{4}{\left(1-\exp\left(-\frac{\delta^2}{2s_{\varepsilon}^2}\right)\right)^2} + \frac{4}{1-\exp\left(-\delta^2/s_{\varepsilon}^2\right)}.\end{aligned}$$

Then, choose $\delta^2 = \log(1+\log(1+\log(T)))^{-1}$, which satisfies the constraints for Lemma 13 for T sufficiently large. Remark that δ satisfies the condition $\delta \leq 1$ needed for Lemma 13, and it is such that $\delta \xrightarrow[T\to\infty]{} 0$, and

$$\left\lceil \frac{2\log(T)s_{\varepsilon}^2}{\log(\sqrt{\log(T)})\delta^2} \right\rceil + \sqrt{\log(T)} + \frac{4}{\left(1-\exp\left(-\frac{\delta^2}{2s_{\varepsilon}^2}\right)\right)^2} + \frac{4}{1-\exp\left(-\delta^2/s_{\varepsilon}^2\right)} = o(\log(T)).$$

Hence, we have shown that

$$\lim_{T\to\infty} \frac{\mathbb{E}[N_a(T)]}{\log(T)} \leq \frac{1}{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m(\mu_a), m(\mu_1)\right)},$$

which concludes the proof of Theorem 2.

D.2. Proof of Lemma 12: controlling deviations of sub-optimal arm (event $E_n(a)$)

Let us first handle the summation from Equation (D.2), this term will give us the main term in regret. Consider the following inequalities:

$$\begin{aligned}\sum_{n=1}^{T} \mathbb{1}\left(E_n(a)\right) &= \sum_{n=1}^{T} \mathbb{1}\left(A_n = a, N_a(n)\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_a(n)) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq \log n\right) \\ &\leq \sum_{n=1}^{T}\sum_{t=1}^{n} \mathbb{1}\left(A_n = a, t\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq \log T, N_a(n) = t\right) \\ &\leq \sum_{t=1}^{T} \mathbb{1}\left(t\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq \log T\right).\end{aligned}$$

The last line follows from the fact that for a given t, there exists only one n such that the two events $A_n = a$ and $N_a(n) = t$ are true. Thus, to bound $\sum_n \mathbb{P}(E_n(a))$, it suffices to bound

$$\sum_{t=1}^{\infty} \mathbb{P}\left(t\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq \log T\right). \tag{D.5}$$

Each summand in the above expression is bounded by

$$\mathbb{P}\left(t\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1)\right) - t\int_{m(\mu_1)-\delta}^{m(\mu_1)} \frac{d\,\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, z\right)}{dz} dz \leq \log T\right).$$

Using that for $x \in \mathbb{R}$, $\frac{d\,\mathrm{kl}_{\mathcal{G}}^{\varepsilon}(x,x+\Delta)}{d\Delta} \leq \Delta$ (see Lemma 3), and injecting the probability back into Equation (D.5), we get

$$\begin{aligned}&\sum_{t=1}^{\infty} \mathbb{P}\left(t\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq \log T\right) \\ &\leq \sum_{t=1}^{\infty} \mathbb{P}\left(t\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1)\right) - t\int_{m(\mu_1)-\delta}^{m(\mu_1)} \left(z - \mathrm{Med}(\hat{\mu}_{a,t}) + \frac{\Delta_{\min}}{2}\right) dz \leq \log T\right) \\ &\leq \sum_{t=1}^{\infty} \mathbb{P}\left(t\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1)\right) - t\delta\left(m(\mu_1) - \mathrm{Med}(\hat{\mu}_{a,t}) + \frac{\Delta_{\min}}{2}\right) \leq \log T\right).\end{aligned}$$

Using Theorem 3 with $y = \delta \leq 1$ to bound the probability that the median have deviations larger than δ, we get

$$\begin{aligned}&\sum_{t=1}^{\infty} \mathbb{P}\left(t\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq \log T\right) \\ &\leq \sum_{t=1}^{\infty} \mathbb{P}\left(t\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1)\right) - t\delta\left(\Delta_a + \delta + \Delta_{\min}\right) \leq \log T\right) \\ &\quad + \sum_{t=1}^{\infty} 2\exp\left(-\frac{t\delta^2}{s_{\varepsilon}^2}\right).\end{aligned}$$

At this point, let us introduce

$$t_0 = \left\lceil \frac{\log(T)}{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m(\mu_a), m(\mu_1)\right) - 2\delta\left(\Delta_a + \delta + \Delta_{\min}\right)} \right\rceil.$$

where, because of the inequality $\delta \leq \min(\Delta_a + \Delta_{\min}, \frac{1}{4(\Delta_a+\Delta_{\min})}\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m(\mu_a), m(\mu_1)\right))$, we can conclude that

$$\begin{aligned}\delta\left(\Delta_a + \delta + \Delta_{\min}\right) &\leq 2\delta\left(\Delta_a + \Delta_{\min}\right) \\ &\leq \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m(\mu_a), m(\mu_1)\right).\end{aligned}$$

Hence the denominator in t_0 is positive.

The required sum-of-probabilities (D.5) can further be bounded by:

$$\begin{aligned}\sum_{t=t_0}^{\infty} \mathbb{P}\left(t_0\left(\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1)\right) - \delta\left(\Delta_a + \delta + \Delta_{\min}\right)\right) \leq \log T\right) \\ + 2\sum_{t=t_0}^{\infty} \exp\left(-\frac{t\delta^2}{s_{\varepsilon}^2}\right) + t_0 - 1,\end{aligned}$$

which is further less than

$$\begin{aligned}\sum_{t=t_0}^{\infty} \mathbb{P}\left(\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1)\right) \leq \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m(\mu_a), m(\mu_1)\right) - \delta\left(\Delta_a + \delta + \Delta_{\min}\right)\right) \\ + 2\sum_{t=t_0}^{\infty} \exp\left(-\frac{t\delta^2}{s_{\varepsilon}^2}\right) + t_0 - 1.\end{aligned}$$

Now, using Lemma 4 under the condition $\delta \leq 1$, we bound the probability in the summation above as below:

$$\sum_{t=1}^{\infty} \mathbb{P}\left(t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq \log T\right) \leq 4\sum_{t=t_0}^{\infty} \exp\left(-\frac{t\delta^2}{2s_{\varepsilon}^2}\right) + t_0 - 1,$$

which is at most

$$\frac{4}{1 - \exp\left(-\delta^2/s_{\varepsilon}^2\right)} + t_0 - 1.$$

Hence for $\delta \leq 1$, we have

$$\sum_n \mathbb{P}(E_n(a)) \leq t_0 - 1 + \frac{4}{1 - \exp\left(-\delta^2/s_{\varepsilon}^2\right)}, \tag{D.6}$$

where it can be checked, by definition, that

$$t_0 - 1 \leq \frac{\log(T)}{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m(\mu_a), m(\mu_1)\right) - 2\delta\left(\Delta_a + \delta + \Delta_{\min}\right)}.$$

D.3. Proof of Lemma 13: controlling deviation of the optimal arm (event $F_n(a)$)

Since each arm is pulled at least $N_{\min}$ times till time $n \geq KN_{\min}$, we have

$$\sum_{n=KN_{\min}}^{T} \mathbb{P}(F_n(a)) \;=\; \mathbb{E}\Big(\sum_{n=KN_{\min}}^{T} \sum_{t=N_{\min}}^{n} \mathbb{1}\Big(A_n = a,\; \mathrm{Med}(\hat{\mu}_{1,t}) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2},\\ I_*(n) \leq t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) + \log t \leq tM + \log(t)\Big)\Big).$$

By changing the order of summation in the above expression, it can be shown to equal

$$\sum_{t=1}^{T} \mathbb{E}\Big(\sum_{n=t}^{T} \mathbb{1}\Big(A_n = a,\; \mathrm{Med}(\hat{\mu}_{1,t}) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2},\\ I_*(n) \leq t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta) + \log t \leq tM + \log(t)\Big)\Big),$$

which is at most

$$\sum_{t=1}^{T} \mathbb{E}\left(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq M\right)\right.\\ \left. \times \sum_{n=t}^{T} \mathbb{1}\left(A_n = a,\; I_*(n) \leq t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) + \log t\right)\right).$$

Recall that for time n such that $A_n = a$, $I_*(n) = N_a(n)\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(\mathrm{Med}(\hat{\mu}_a(n)) - \Delta_{\min}, \mathrm{Med}_*(n)) + \log N_a(n)$, which is at least $\log N_a(n)$. Using this, the above summation is bounded by

$$\sum_{t=1}^{T} \mathbb{E}\left(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq M\right)\right.\\ \left. \times \sum_{n=t}^{T} \mathbb{1}\left(A_n = a,\; \log N_a(n) \leq t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) + \log t\right)\right),$$

which is at most (also see Honda and Takemura (2015, Lemma 13))

$$\sum_{t=1}^{T} \mathbb{E}\Big(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq M\right)\\ \times e^{t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) + \log t}\Big). \quad \text{(D.7)}$$

Then, we show the following Lemma Using the bound on $\sum_n \mathbb{P}(F_n(a))$ from Equation (D.7) and observing that the expectation in the bound is for a non-negative random variable, we get the following bound on $\sum_{n=1}^{T} \mathbb{P}(F_n(a))$:

$$\sum_{t=1}^{T} t \int_0^{\infty} \mathbb{P}\Big(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \leq m_1(\mu) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq M\right)\\ \times e^{t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right)} \geq x\Big)\mathrm{d}x. \quad \text{(D.8)}$$

Let us control the integral above separately on $[0, 1]$ and $[1, \infty)$.

Integral on $[0,1]$ On $[0,1]$ we only control the deviations of the empirical median and we do not care about the deviations of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$:

$$\int_0^1 \mathbb{P}\Big(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \leq m_1(\mu) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta \right) \leq M \right) \times e^{t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right)} \geq x \Big) \mathrm{d}x$$

$$\leq \int_0^1 \mathbb{P}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \leq m_1(\mu) - \delta - \frac{\Delta_{\min}}{2} \right) \mathrm{d}x = \mathbb{P}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \leq m_1(\mu) - \delta - \frac{\Delta_{\min}}{2} \right) \leq 2e^{-t\frac{\delta^2}{s^2_{\varepsilon}}}.$$

Using Theorem 3 for the last line, for $\delta < 1$. Then, we get

$$\sum_{t=N_{\min}}^{T} \int_0^1 t\mathbb{P}\Big(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \leq m_1(\mu) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta \right) \leq M \right)$$

$$\leq 2\sum_{t=1}^{\infty} te^{-t\frac{\delta^2}{s^2_{\varepsilon}}} = 2\frac{e^{-\frac{\delta^2}{s^2_{\varepsilon}}}}{\left(1 - e^{-\frac{\delta^2}{s^2_{\varepsilon}}}\right)^2}. \tag{D.9}$$

Next, we bound the integral on $[1,\infty)$.

Integral on $[1,\infty)$ We use that the deviations of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ are bounded by M in the indicator function to bound simplify the probability as follows.

$$\sum_{t=1}^{T} t \int_1^{\infty} \mathbb{P}\Big(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \leq m_1(\mu) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta \right) \leq M \right) \times e^{t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right)} \geq x \Big) \mathrm{d}x$$

$$\leq \sum_{t=1}^{T} t \int_1^{\infty} \mathbb{P}\left(e^{tM} \geq e^{t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right)} \geq x \right) \mathrm{d}x$$

$$= \sum_{t=1}^{T} t \int_1^{\exp(tM)} \mathbb{P}\left(t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta \right) \geq \log x \right) \mathrm{d}x$$

Then, we use a change of variable $x \leftarrow e^y$ to show that the above is smaller than

$$\sum_{t=1}^{T} t \int_0^{tM} \mathbb{P}\left(t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta) \geq y \right) e^y \mathrm{d}y.$$

Next, we use the first case of Lemma 4 with $y = \delta$ and bound the probability that $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta)$ is strictly positive. We have,

$$\mathbb{P}\left(\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) > 0\right) \leq 2\exp\left(-\frac{t\delta^2}{s_\varepsilon^2}\right)$$

Using this bound, we get the following control

$$\sum_{t=1}^{T} t \int_1^{\infty} \mathbb{P}\Big(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \leq m_1(\mu) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq M\right)$$
$$\times\, e^{t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right)} \geq x\Big)\mathrm{d}x$$
$$\leq 2\sum_{t=N_{\min}}^{T} \int_0^{tM} t\exp\left(-\frac{t\delta^2}{s_\varepsilon^2}\right) e^y \mathrm{d}y \leq 2\sum_{t=N_{\min}}^{T} t\exp\left(-\frac{t\delta^2}{s_\varepsilon^2}\right) e^{Mt}.$$

Now, take $M = \frac{\delta^2}{2s_\varepsilon^2}$, to keep the exponent of the exponential negative, we get

$$\sum_{t=1}^{T} t \int_1^{\infty} \mathbb{P}\Big(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \leq m_1(\mu) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq M\right)$$
$$\times\, e^{t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right)} \geq x\Big)\mathrm{d}x$$
$$\leq 2\sum_{t=1}^{T} t\exp\left(-\frac{t\delta^2}{2s_\varepsilon^2}\right) \leq \frac{2\exp\left(-\frac{\delta^2}{2s_\varepsilon^2}\right)}{\left(1-\exp\left(-\frac{\delta^2}{2s_\varepsilon^2}\right)\right)^2} \leq \frac{2}{\left(1-\exp\left(-\frac{\delta^2}{2s_\varepsilon^2}\right)\right)^2}.$$

Wrap-up: bounding $\sum_{n=1}^{T}\mathbb{P}(F_n(a))$ Combining Equation (D.8), Equation (D.9) and Equation (D.8), and choosing $M = \delta^2/(2s_\varepsilon^2)$, we finally obtain

$$\sum_{n=1}^{T}\mathbb{P}(F_n(a)) \leq \frac{e^{-\frac{\delta^2}{s_\varepsilon^2}}}{\left(1-\exp\left(-\frac{\delta^2}{s_\varepsilon^2}\right)\right)^2} + \frac{2}{\left(1-\exp\left(-\frac{\delta^2}{2s_\varepsilon^2}\right)\right)^2} \leq \frac{4}{\left(1-\exp\left(-\frac{\delta^2}{2s_\varepsilon^2}\right)\right)^2}. \quad \text{(D.10)}$$

D.4. Proof of Lemma 14: controlling large deviations of the kl (event $G_n(a)$)

Let us now control $\mathbb{P}(G_n(a))$. We have for any $M > 0$,

$$\begin{aligned}
&\sum_{n=KN_{\min}}^{T} \mathbb{P}(G_n(a)) \\
&= \sum_{n=N_{\min}}^{T} \mathbb{P}\left(\bigcup_{t=N_{\min}}^{n} \left\{ A_n = a, \mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta \right) \geq M, N_1(n) = t \right\} \right) \\
&\leq T^2 \mathbb{P}\left(\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{1,N_{\min}}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta \right) \geq \frac{\delta^2}{2s_{\varepsilon}^2} \right) \\
&\leq T^2 \mathbb{P}\left(\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(\hat{\mu}_{1,N_{\min}}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta \right) \geq 0 \right) = T^2 e^{-N_{\min}\frac{\delta^2}{s_{\varepsilon}^2}}.
\end{aligned}$$

This leads us to choose

$$N_{\min} = \left\lceil \frac{2\log(T) s_{\varepsilon}^2}{\log(1+\log(T)^{0.99})\delta^2} \right\rceil,$$

which ensures that

$$\sum_{n=1}^{T} \mathbb{P}(G_n(a)) \leq 1 + \log(T)^{0.99}. \tag{D.11}$$

D.5. Proof of Theorem 3: concentration of empirical median

Without loss of generality, by doing the change of variable $X \leftarrow X - m$, we assumes in the proof that $m = 0$. For any $\lambda > 0$, we have

$$\mathbb{P}\left(\mathrm{Med}(X_1^n) > \lambda\right) \leq \mathbb{P}\left(\#\{i : X_i \geq \lambda\} \geq \frac{n}{2} \right) = \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \geq \lambda\} \geq \frac{1}{2} \right).$$

Let $W_1, \ldots, W_n$ i.i.d $\mathrm{Ber}(\varepsilon)$, $Y_1, \ldots, Y_n$ i.i.d $\sim P$ and $O_1, \ldots, O_n$ be i.i.d from H, with the W's, the Y's and the O's all independents. By caracterization of mixture of distributions, we have that X_i is equal in distribution to $(1 - W_i)Y_i + W_i O_i$. Hence,

$$\begin{aligned}
\mathbb{P}\left(\mathrm{Med}(X_1^n) \geq \lambda\right) &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} (\mathbb{1}\{(1-W_i)Y_i + W_i O_i \geq \lambda\}) \geq \frac{1}{2} \right) \\
&= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} ((1-W_i)\mathbb{1}\{Y_i \geq \lambda\} + W_i \mathbb{1}\{O_i \geq \lambda\}) \geq \frac{1}{2} \right) \\
&\leq \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} (1-W_i)\mathbb{1}\{Y_i \geq \lambda\} + \frac{1}{n}\sum_{i=1}^{n} W_i \geq \frac{1}{2} \right).
\end{aligned} \tag{D.12}$$

The quantities appearing in the right-hand-side of Equation (D.12) are all with values in $\{0, 1\}$.

Concentration of Bernoulli random variables
$W_1, \dots, W_n$ are i.i.d Bernoulli random variables with mean ε. From Bourel et al. (2020, Lemma 6), for any $\gamma \in (0,1)$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} W_i \geq \varepsilon + \sqrt{\frac{(1-2\varepsilon)\log(1/\gamma)}{4n\log((1-\varepsilon)/\varepsilon)}}\right) \leq \gamma. \tag{D.13}$$

Similarly, for $1 \leq i \leq n$, $(1-W_i)\mathbb{1}\{Y_i > \lambda\}$ are also Bernoulli random variables with mean $\mathbb{E}[(1-W_i)\mathbb{1}\{Y_i > \lambda\}] = (1-\varepsilon)(1-\Phi(\lambda)) \leq (1-\Phi(\lambda)) \leq 1/2$. Again using the sub-gaussian concentration from Bourel et al. (2020, Lemma 6), we have with probability larger than $1-\gamma$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(1-W_i)\mathbb{1}\{Y_i \geq \lambda\} &\leq (1-\varepsilon)(1-\Phi(\lambda)) + \sqrt{\frac{(1-2(1-\varepsilon)(1-\Phi(\lambda)))\log(1/\gamma)}{4n\log\left(\frac{1-(1-\varepsilon)(1-\Phi(\lambda))}{(1-\varepsilon)(1-\Phi(\lambda))}\right)}} \\ &\leq (1-\varepsilon)(1-\Phi(\lambda)) + \sqrt{\frac{(1-2(1-\varepsilon)(1-\Phi(\lambda)))\log(1/\gamma)}{4n\log\left(\frac{\Phi(\lambda)}{1-\Phi(\lambda)}\right)}},\end{aligned} \tag{D.14}$$

where in the last line, we used that $p \mapsto (1-p)/p$ is decreasing on $(0,1)$. Then, from Equation(D.14) and Equation(D.13), we get with probability larger than $1-2\gamma$,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}(1-W_i)\mathbb{1}\{Y_i \geq \lambda\} + \frac{1}{n}\sum_{i=1}^{n} W_i \\ &\leq (1-\varepsilon)(1-\Phi(\lambda)) + \sqrt{\frac{(1-2(1-\varepsilon)(1-\Phi(\lambda)))\log(1/\gamma)}{4n\log\left(\frac{\Phi(\lambda)}{1-\Phi(\lambda)}\right)}} + \varepsilon + \sqrt{\frac{(1-2\varepsilon)\log(1/\gamma)}{4n\log((1-\varepsilon)/\varepsilon)}}.\end{aligned}$$

In this equation, there are two free parameters: λ and γ. Next, we choose λ so that $\frac{1}{n}\sum_{i=1}^{n}(1-W_i)\mathbb{1}\{Y_i \geq \lambda\} + \frac{1}{n}\sum_{i=1}^{n} W_i$ is smaller than $1/2$ with high probability. This choice of λ will then allow us to control the probability in Equation (D.12).

Choice of λ
First, we state some basic inequalities for $\Phi(\lambda)$. We have for $\lambda = \frac{\Delta_{\min}}{2} + L$, using Taylor's inequality,

$$\Phi(\lambda) - \frac{1}{2(1-\varepsilon)} \geq L\varphi\left(\frac{\Delta_{\min}}{2} + L\right)$$

and from monotonicity of $x \mapsto x/(1-x)$ on $[0,1)$,

$$\frac{\Phi(\lambda)}{1-\Phi(\lambda)} \geq \frac{\Phi(\frac{\Delta_{\min}}{2})}{1-\Phi(\frac{\Delta_{\min}}{2})} = \frac{\frac{1}{2(1-\varepsilon)}}{\frac{1-2\varepsilon}{2(1-\varepsilon)}} = \frac{1}{1-2\varepsilon}.$$

Then,

$$
\begin{aligned}
&(1-\varepsilon)(1-\Phi(\lambda)) + \sqrt{\frac{(1-2(1-\varepsilon)(1-\Phi(\lambda)))\log(1/\gamma)}{4n\log\left(\frac{\Phi(\lambda)}{1-\Phi(\lambda)}\right)}} + \varepsilon + \sqrt{\frac{(1-2\varepsilon)\log(1/\gamma)}{4n\log((1-\varepsilon)/\varepsilon)}} - \frac{1}{2} \\
&\leq (1-\varepsilon)\left(\frac{1-2\varepsilon}{2(1-\varepsilon)} - L\varphi\left(\frac{\Delta_{\min}}{2}+L\right)\right) \\
&+ \sqrt{\frac{\left(1-2(1-\varepsilon)\left(\frac{1-2\varepsilon}{2(1-\varepsilon)} + L\varphi(\frac{\Delta_{\min}}{2}+L)\right)\right)\log(1/\gamma)}{4n\log\left(\frac{1}{1-2\varepsilon}\right)}} + \varepsilon + \sqrt{\frac{(1-2\varepsilon)\log(1/\gamma)}{4n\log((1-\varepsilon)/\varepsilon)}} - \frac{1}{2} \\
&\leq \frac{1}{2} - \varepsilon - (1-\varepsilon)L\varphi\left(\frac{\Delta_{\min}}{2}+L\right) + \sqrt{\frac{\varepsilon\log(1/\gamma)}{2n\log\left(\frac{1}{1-2\varepsilon}\right)}} + \varepsilon + \sqrt{\frac{(1-2\varepsilon)\log(1/\gamma)}{4n\log((1-\varepsilon)/\varepsilon)}} - \frac{1}{2} \\
&\leq -(1-\varepsilon)L\varphi\left(\frac{\Delta_{\min}}{2}+L\right) + \sqrt{\frac{\varepsilon\log(1/\gamma)}{2n\log\left(\frac{1}{1-2\varepsilon}\right)}} + \sqrt{\frac{(1-2\varepsilon)\log(1/\gamma)}{4n\log((1-\varepsilon)/\varepsilon)}}.
\end{aligned}
$$

Now, suppose $L \leq 1$ and choose

$$
\begin{aligned}
L &= \frac{1}{(1-\varepsilon)\varphi\left(\frac{\Delta_{\min}}{2}+1\right)}\left(\sqrt{\frac{\varepsilon}{2\log\left(\frac{1}{1-2\varepsilon}\right)}} + \sqrt{\frac{(1-2\varepsilon)}{4\log((1-\varepsilon)/\varepsilon)}}\right)\sqrt{\frac{\log(1/\gamma)}{n}} \\
&= s_\varepsilon\sqrt{\frac{\log(1/\gamma)}{n}},
\end{aligned} \tag{D.15}
$$

with $s_\varepsilon\sqrt{\frac{\log(1/\gamma)}{n}} \leq 1$. After this choice of λ, there is only one free parameter remaining: γ.

Injection of chosen λ in Equation (D.12)
From the choice of $\lambda = \frac{\Delta_{\min}}{2} + L$ from Equation (D.15), we have

$$
\mathbb{P}\left(\mathrm{Med}(X_1^n) \geq \frac{\Delta_{\min}}{2} + s_\varepsilon\sqrt{\frac{\log(1/\gamma)}{n}}\right) \leq 2\gamma.
$$

Under the condition that $\gamma \geq \exp\left(-n/s_\varepsilon^2\right)$. Let us now reformulate this result by solving the following equation for γ:

$$
y = s_\varepsilon\sqrt{\frac{\log(1/\gamma)}{n}},
$$

we get for any $0 \leq y \leq 1$

$$
\mathbb{P}\left(\mathrm{Med}(X_1^n) \geq \frac{\Delta_{\min}}{2} + y\right) \leq 2\exp\left(-ny^2/s_\varepsilon^2\right).
$$

To get the other direction, remark that X is equal in distribution to $-X$ and inject in the above concentration.

D.6. Proof of Lemma 4: concentration of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$

The two proofs are very similar except that we don't concentrate around the same quantity.

Case $m_b = m_a - \delta$
We write that from Lemma 10,

$$\begin{aligned}
&\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(X_1^n) - \frac{\Delta_{\min}}{2}, m_a - \delta\right) \\
&\quad = \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(X_1^n) - \frac{\Delta_{\min}}{2}, m_a - \delta\right) - \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m_a - \Delta_{\min} - \delta, m_a - \delta\right) \\
&\quad \leq \left(m_a - \mathrm{Med}(X_1^n) - \frac{\Delta_{\min}}{2} - \delta\right)_+ \max\left(m_a - \mathrm{Med}(X_1^n) - \delta + \frac{\Delta_{\min}}{2}, \Delta_{\min}\right).
\end{aligned}$$

Then, from Theorem 3, with probability larger than $1 - 2\exp\left(\frac{-ny^2}{s_\varepsilon^2}\right)$, we have for any $y \leq 1$,

$$\begin{aligned}
\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(X_1^n) - \frac{\Delta_{\min}}{2}, m_a - \delta\right) &\leq (y-\delta)_+ \max\left(y - \delta + \Delta_{\min}, \Delta_{\min}\right) \\
&= (y-\delta)_+ \left(|y-\delta| + \frac{\Delta_{\min}}{2}\right), \qquad \text{(D.16)}
\end{aligned}$$

where the last line comes from the fact that when $y \leq \delta$, the bound is 0 anyway.

Case $m_b > m_a + \Delta_{\min}$
From Lemma 10,

$$\begin{aligned}
&\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m_a, m_b\right) - \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(X_1^n) - \frac{\Delta_{\min}}{2}, m_b\right) \\
&\quad \leq \left(\mathrm{Med}(X_1^n) - m_a - \frac{\Delta_{\min}}{2}\right)_+ \max\left(m_b - \mathrm{Med}(X_1^n) + \frac{\Delta_{\min}}{2}, m_b - m_a\right).
\end{aligned}$$

Then, from Theorem 3, with probability larger than $1 - 2\exp\left(\frac{-ny^2}{s_\varepsilon^2}\right)$, we have for any $y \leq 1$,

$$\begin{aligned}
&\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m_a, m_b\right) - \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(X_1^n) - \frac{\Delta_{\min}}{2}, m_b\right) \\
&\quad \leq y \max\left(m_b - m_a + y + \Delta_{\min}, m_b - m_a\right) \\
&\quad = y(m_b - m_a + y + \Delta_{\min}).
\end{aligned}$$

D.7. Proof of Lemma 11

The event $\{A_n = a\}$ can be written as a disjoint union of

$$\left\{A_n = a,\ \mathrm{Med}_*(n) > m(\mu_1) - \delta - \frac{\Delta_{\min}}{2}\right\} \qquad \text{(D.17)}$$

and

$$\left\{A_n = a,\ \mathrm{Med}_*(n) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2}\right\}. \qquad \text{(D.18)}$$

Of these, intuitively, the second event in Equation (D.17) should not be rare. However, once sufficient samples have been allocated to arm a, the event $\{A_n = a\}$ becomes rare when $\mathrm{Med}_*(n)$ is close to $m_1(\mu)$. This is because after sufficient samples, $\hat{\mu}_a(n) \approx \mu_a$, which implies that $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(\mathrm{Med}(\hat{\mu}_a(n)) - \Delta_{\min}, \mathrm{Med}_*(n))$ should be large. For the event in Equation (D.18), for large n, the event $\{\mathrm{Med}_*(n) \leq m_1(\mu) - \delta - \Delta_{\min}/2\}$ should be rare. We will show that the probability of Equation (D.17) occurring, summed across time, contributes to the main term in regret.

Define $I_*(n) := \min_a I_a(n)$ to be the minimum index. Recall that $a^*(n)$ denotes the arm with the maximum estimated mean, i.e.,

$$a^*(n) \in \arg\max_{b \in [K]} \mathrm{Med}(\hat{\mu}_b(n)).$$

Since $A_n = a$ implies that $I_a(n) = I_*(n)$. Then,

$$\begin{aligned} I_a(n) &= I_*(n) \\ &\leq I_{a^*(n)}(n) \\ &= \log N_{a^*(n)}(n) \\ &\leq \log n. \end{aligned}$$

Thus, $\{A_n = a\}$ implies that $I_a(n) \leq \log n$ and Equation (D.17) is contained in

$$\left\{ A_n = a,\ N_a(n)\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_a(n)) - \Delta_{\min}, \mathrm{Med}_*(n)\right) \leq \log n,\ \mathrm{Med}_*(n) > m(\mu_1) - \delta - \frac{\Delta_{\min}}{2} \right\}.$$

Next, using the monotonicity of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ in the second argument and its translation invariance (Lemma 3) in the above containment, we have that $\left\{ A_n = a,\ \mathrm{Med}_*(n) > m_1(\mu) - \delta - \frac{\Delta_{\min}}{2} \right\}$ is contained in

$$\left\{ A_n = a, N_a(n)\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_a(n)) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq \log n \right\}. \tag{D.19}$$

Next, observe that the event in Equation (D.18) satisfies

$$\begin{aligned} &\left\{ A_n = a,\ \mathrm{Med}_*(n) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2} \right\} \\ &\subset \left\{ A_n = a,\ \mathrm{Med}_*(n) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_1(n)) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq M \right\} \\ &\qquad \cup \left\{ \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_1(n)) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \geq M \right\} \end{aligned}$$

which is included in

$$\begin{aligned} \bigcup_{t=1}^{n} \Bigg(& A_n = a,\ \mathrm{Med}_*(n) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2}, \\ & \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_1(n)) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq M, N_1(n) = t \Bigg) \\ & \cup \left\{ \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_1(n)) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \geq M, N_1(n) = t \right\} \end{aligned}$$

Let $\hat{\mu}_{1,t}$ denote the empirical distribution for arm 1 with t samples. Now, since $A_n = a$ implies that

$$I_a(n) = I_*(n) \leq I_1(n) = N_1(n)\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_1(n)) - \Delta_{\min}, \mathrm{Med}_*(n)\right) + \log N_1(n),$$

the above union-of-events is further contained in

$$\bigcup_{t=1}^{n} \Big\{ A_n = a,\ \mathrm{Med}(\hat{\mu}_{1,t}) \leq \mathrm{Med}_*(n) \leq m(\mu_1) - \delta - \frac{\Delta_{\min}}{2},$$
$$I_*(n) \leq t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) + \log t \leq tM + \log(t)\Big\}$$
$$\cup \left\{ \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \geq M, N_1(n) = t \right\},$$

which is the union of $F_n(a)$ and $G_n(a)$.