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

Shubhada Agrawal SAGRAWAL362@GATECH.EDU
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, USA.

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

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

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

Editors: Claire Vernade and Daniel Hsu

Abstract

We investigate the regret-minimisation problem in a multi-armed bandit setting with arbitrary corruptions. Similar to the classical setup, the agent receives rewards generated independently from the distribution of the arm chosen at each time. However, these rewards are not directly observed. Instead, with a fixed $\varepsilon \in (0, \frac{1}{2})$, the agent observes a sample from the chosen arm's distribution with probability $1 - \varepsilon$, or from an arbitrary corruption distribution with probability ε. Importantly, we impose no assumptions on these corruption distributions, which can be unbounded. In this setting, accommodating potentially unbounded corruptions, we establish a problem-dependent lower bound on regret for a given family of arm distributions. We introduce `CRIMED`, an asymptotically-optimal algorithm that achieves the exact lower bound on regret for bandits with Gaussian distributions with known variance. Additionally, we provide a finite-sample analysis of `CRIMED`'s regret performance. Notably, `CRIMED` can effectively handle corruptions with ε values as high as $\frac{1}{2}$. Furthermore, we develop a tight concentration result for medians in the presence of arbitrary corruptions, even with ε values up to $\frac{1}{2}$, which may be of independent interest. We also discuss an extension of the algorithm for handling misspecification in Gaussian model.

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

1. Introduction

Multi-armed bandits are a widely-used statistical model in which an agent (or algorithm) interacts with the environment by selecting actions based on past observations and receives a *reward* for each chosen action. A classical objective is to minimise the *regret*, defined as the difference between the rewards accumulated by the algorithm and those obtained by choosing the best action in hindsight at each step. In this paper, we delve into the problem of sequential decision-making under partial information, where the observations resulting from actions are susceptible to arbitrary yet stochastic corruption. Specifically, we explore a variant of the stochastic multi-armed bandit problem with *unbounded stochastic corruption*.

The algorithm is presented with K arms, or K unknown probability distribution from a given family $\mathcal{L}$, denoted by $\mu := (\mu_1, \dots, \mu_K)$, where $\forall a \in [K], \mu_a \in \mathcal{L}$. When it selects an arm A_n at time n, an independent sample, Y_n, is drawn from the corresponding distribution μ_{A_n}. This

corresponds to the reward of the algorithm for pulling the chosen arm. However, unlike in the classical setup, Y_n is not directly observed by the algorithm. Instead, the observations are subject to corruption with a known probability $\varepsilon \in (0, 0.5)$: 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 an arbitrary distribution $H_{A_n,n}$. Here, $\mathbf{H}_n := (H_{1,n}, \ldots, H_{K,n})$ denotes the set of corruption distributions at time n, which is assumed oblivious in that it can depend on the previous reward observations but not on the past algorithmic actions. Further, we place absolutely no assumptions on the distributions $H_{\cdot,n}$. Following the classical regret-minimisation setting, the algorithm's goal in this partial-information setting is to sequentially sample the arms to maximise the expected cumulative reward when the observations are corrupted.

The bandit problem forms the theoretical cornerstone of modern Reinforcement Learning (RL) and serves as the algorithmic foundation for recommender systems. As both bandit problems and RL are increasingly finding practical applications, the question of robustness against corrupted or externally perturbed observations has gained considerable significance. This is particularly relevant because, in real-life applications such as finance, medicine, advertising, or recommender systems, algorithms often need to contend with corrupted data, as observations collected from multiple sources are susceptible to measurement or recording errors and inaccuracies.

In finance, the observed payoff data frequently contains outliers due to data contamination (Adams et al., 2019). In clinical research trials for new drugs and medical devices, outlier data can lead to false positive interventions and conclusions (Thabane et al., 2013). In online platforms and recommender systems, the ranking of products can get skewed by the appearance of small number of fake users (Golrezaei et al., 2021). The classical corruption-oblivious bandit algorithms, however, cannot effectively decide the arms to pull, when a possibly small fraction of the data may be subject to measurement errors or corruption. Specifically, a recent line of works (Jun et al., 2018; Liu and Shroff, 2019; Xu et al., 2021; Azize and Basu, 2023) show that one can make a classical bandit algorithms to incur linear regret by contaminating only a small amount of observations (logarithmic on horizon or even lower). These findings motivate us to study the bandit setup in the presence of *arbitrary* corruption, and design algorithms robust to it.

Researchers have broadly studied three types of settings: *adversarial bandits* (Auer et al., 1995, 2002b), *stochastic bandits with bounded adversarial corruptions*, in which an adversary shifts the rewards under constraint on the total shift budget (Lykouris et al., 2018; Gupta et al., 2019; Zimmert and Seldin, 2019), and more recently, *unbounded stochastic corruption* (Altschuler et al., 2019; Mukherjee et al., 2021; Basu et al., 2022). To the best of our knowledge, there is *no established generic lower bound on regret in the context of unbounded stochastic corruptions*, unlike in the first two settings. Furthermore, there is *no known algorithm capable of yielding an appropriate upper bound* on regret while also maintaining robustness. This paper aims to fill these two gaps by investigating bandits with unbounded stochastic corruptions.

Regret. For $\mu \in \mathcal{L}^K$, let $m^*(\mu)$ denote the mean of the optimal arm in μ (arm with the maximum mean), and let $m(\mu_a)$ denote the mean of arm a. For an arm a, let $\Delta_a := m^*(\mu) - m(\mu_a)$ denote the instantaneous mean regret incurred by pulling it. Recall that Y_n denotes the independent (uncorrupted) sample drawn from the distribution associated with arm A_n. Let $Y_{a,j}$ denote the j^{th} independent sample drawn from arm a.

Since in our setup, the observations are corrupted (while the rewards are uncorrupted), we define the *expected regret under corruption* till time T as $\mathbb{E}[R_T] := \mathbb{E}[\sum_{n=1}^{T}(m^*(\mu) - Y_n)]$. Here, the expectation is with respect to all the randomness present in the system, including the impact of

corruption on action selection. See also (Kapoor et al., 2019; Basu et al., 2022) for a similar notion of regret. We further observe that $\mathbb{E}[R_T] = \sum_{a=1}^{K} \mathbb{E}[N_a(T)]\,\Delta_a$, 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) $\Delta_a = 0$, minimising the expected regret reduces to minimising the expected number of pulls of the suboptimal arms $\mathbb{E}[N_a(T)]$.

Notation. Let $\mathbb{R}$ and $\mathbb{R}^+$ denote the set of real numbers and non-negative real numbers, respectively, and let $\mathcal{P}(\mathbb{R})$ denote the set of all probability distributions on $\mathbb{R}$. For any set S, we denote by 2^S the set of subsets of S. For μ, $n \in \mathbb{N}$, and $\mathbf{H}_n \in \mathcal{P}(\mathbb{R})^K$, let $\mu \odot_\varepsilon \mathbf{H}_n := (1-\varepsilon)\mu + \varepsilon \mathbf{H}_n$ denote the vector of distributions in μ corrupted by corruption distributions in $\mathbf{H}_n$ with corruption proportion ε. We use a similar notation for each component $\mu_a, H_{a,n} \in \mathcal{P}(\mathbb{R})$, i.e. $\mu_a \odot_\varepsilon H_{a,n} := (1-\varepsilon)\mu_a + \varepsilon H_{a,n}$ for $a \in [K]$ and we call $\mu_a \odot_\varepsilon H_{a,n}$ a corrupted distribution. Additionally, by $\mathbf{H_T}$, we denote the $T \times K$ matrix of corruption distributions with $\{\mathbf{H}_n : n \in [T]\}$ as rows. Finally, we denote by $\mathcal{G}$ the set of all Gaussian distributions with variance 1, by φ the Gaussian pdf, and by Φ the Gaussian CDF.

1.1. Contributions

In this paper, we investigate two question:

- *Can we derive a problem-dependent lower bound on regret for a given set of reward distributions and the corresponding worst-case corruption distributions?*
- *Can we leverage this lower bound to design an asymptotically-optimal algorithm that is robust to unbounded stochastic corruptions?*

In this section, we briefly describe the main contributions of this work.

1. *A Generic Lower Bound on Regret:* To the best of our knowledge, we establish the first instance-dependent lower bound on regret that is applicable to any given family of reward distributions and arbitrary corruption distributions (Section 2.1). Specifically, in Theorem 1, we demonstrate that any algorithm performing well across all bandit instances within a given class must, in expectation, pull each suboptimal arm at least $\Omega(\log T)$ times over the course of T trials. This result aligns with the known $\Omega(\log T)$ problem-dependent lower bound in the classical setting (Lai and Robbins, 1985). Moreover, when $\epsilon = 0$, our proposed lower bound reduces to that of the uncorrupted setting (Lai and Robbins, 1985; Burnetas and Katehakis, 1996).

2. *An Impossibility Result:* We demonstrate in Appendix B that constructing confidence intervals for the mean of the true distribution in the presence of corruption, a classical problem in statistics, is not feasible without prior knowledge of a bound on corruption probability ε. This resolves the open problem discussed in Wang and Ramdas (2023, Remark 3) and also justifies the assumption about the knowledge of ε in the current work (Remark 8).

3. *An Analytical Quantifier of Hardness:* The lower bound in Theorem 1 is in terms of an optimisation problem that takes the given bandit instance μ as an input. In order to explicitly bring out the structure of this problem and the hard corruption distributions for the given bandit instance, we undertake an in-depth study for the specific setting of Gaussian reward distributions with known variance, while still allowing for unbounded and arbitrary corruptions (see also Chen et al. (2018)). For this setting, we characterise the hardest corruption distributions associated with pairs of arms in μ that lead to the maximum regret in any algorithm. In addition, we show that for each suboptimal arm a, Δ_a should be at least $2\Phi^{-1}\left(\frac{1}{2}(1-\varepsilon)\right)$ for any algorithm to

achieve a sub-logarithmic regret in presence of corruption, and also observe a non-convexity in the lower bound (Section 2.3 and in Appendix D). These observations stand in stark contrast to the classical bandit setup, necessitating a careful treatment in our analysis.

4. *Algorithm Design:* In Section 3, we leverage the formulation and properties of the lower bound to propose an index-based algorithm, namely `CRIMED` (Corruption Robust IMED, Algorithm 1), for unbounded corruptions and Gaussian reward distributions with known variance (we discuss extension to *misspecified* Gaussian distributions in Appendix F). This is an extension of the IMED Algorithm proposed by Honda and Takemura (2015), with two main changes in the index design. First, it replaces the classical information-theoretic quantities that appear in the IMED index with their pessimistic versions in order to account for the presence of corruptions. Second, it uses median as a robust estimate for mean in the presence of corruption. In Section 3.2, we give a finite-sample analysis of the regret of `CRIMED` (Theorem 2). Notably, `CRIMED` is asymptotically (as $T \to \infty$) optimal for any corruption level $\varepsilon < \frac{1}{2}$, which is a significant improvement over the previous works of Kapoor et al. (2019) and Basu et al. (2022) allowing only much smaller ϵ.

5. *Median as the Robust Estimator and its Impact:* Bandits involving arbitrary corruptions present significantly greater challenges compared to their classical counterparts. In the presence of arbitrary corruptions, it is well-known that no consistent estimators for the mean of distributions can exist (Chen et al., 2018). To address this challenge, we draw from the robust estimation literature and opt for the median as a robust estimate of the mean. This choice is motivated by the fact that for *symmetric* distributions, median is optimal because it incurs the smallest bias among all robust estimators of the location parameter (see Section C). In Theorem 3, we establish a novel concentration bound for the empirical median of corrupted Gaussian rewards that applies to any value of ε less than $\frac{1}{2}$.

1.2. Related work

Our work connects and relates to several research areas, which we now briefly summarise.

Multi-armed bandits. The problem of bandits was first introduced in the context of designing adaptive clinical trials by Thompson (1933), and later popularised 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-minimisation framework introduced earlier, asymptotic instance-dependent lower bounds on regret are well known (Lai and Robbins, 1985; Burnetas and Katehakis, 1996).

Index-based (UCB) algorithms for this setting were popularised by the work of Auer et al. (2002a). Cappé et al. (2013); Agrawal et al. (2021) proposed asymptotically-optimal UCB 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) developed a different style of (IMED) algorithms that have a lower computational cost and are also statistically optimal. Alternative optimal algorithms relying on Bayesian posteriors to sample arms (Thompson sampling) 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 owing to its optimality and computational simplicity.

Bandits with bounded corruption. In the adversarial bandits setting, the rewards are assumed to be generated by an adaptive adversary from a bounded interval, e.g. $[0, 1]$. See, for example, Auer

et al. (1995, 2002b); Abernethy and Rakhlin (2009); Audibert et al. (2009); Neu (2015). 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 discussed in the previous paragraph, and are of parallel interest (Bubeck and Slivkins, 2012; Seldin and Slivkins, 2014; Seldin and Lugosi, 2017; Abbasi-Yadkori et al., 2018; Pogodin and Lattimore, 2020).

In the stochastic setting with bounded adversarial corruptions, whenever an arm is pulled at time n, a reward r_n is stochastically generated from the corresponding distribution. But an adversary switches the reward to r'_n such that over the horizon T, $\sum_{n=1}^{T} |r'_n - r_n| \leq C$, for a non-negative constant C. This setting and its variants have also been extensively studied in literature (Lykouris et al., 2018; Gajane et al., 2018; Gupta et al., 2019; Zimmert and Seldin, 2019; Kapoor et al., 2019). Here, the bound C plays a critical role, and the existing regret bounds are linearly dependent on it. These existing regret bounds and algorithms are unfit to handle large amounts of corruptions. This propels the study of bandits that are robust to unbounded corruptions.

Robust estimation under unbounded stochastic corruption. A robust estimator is an estimator that perform well even in the presence of anomalous data. The corruption model considered in this work has a long history in robust statistics. Given a data generating distribution P and a corruption budget ε, a corruption neighbourhood of P is the collection of all distributions of the form $(1-\varepsilon)P + \varepsilon H$, for $H \in \mathcal{P}(\mathbb{R})$. Huber (1964) developed an asymptotic theory of minimax optimality of estimators for distributions in corruption neighbourhood of P. Since then, several methods have been devised to assess the asymptotic robustness of estimators (see, Huber and Ronchetti (2009); Hampel et al. (1986)), in particular, in terms of the stability of the limit of an estimator when the samples come from a corrupted distribution. Lately, a non-asymptotic notion of robustness has gained interest. Here, the goal is to obtain estimators that concentrate fast, either when the data-generating distribution P is heavy-tailed (Catoni, 2012; Devroye et al., 2016; Lugosi and Mendelson, 2019; Agrawal, 2023), or corrupted (Wang and Ramdas, 2023; Chen et al., 2018).

These two concepts (asymptotic and non-asymptotic robustness) are closely linked, and the estimators that perform well in the asymptotic sense have also been shown to perform well in the non-asymptotic setting. Huber's contamination model has also been widely-studied in computer science (Diakonikolas et al., 2018; Charikar et al., 2017). In this work, we use concentration of median to control the regret of `CRIMED`, that receives samples from a corruption neighbourhood of the arm distributions (or from a misspecified model). 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-minimisation 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 Altschuler et al. (2019), Mukherjee et al. (2021), and Basu et al. (2022). Altschuler et al. (2019) and Mukherjee et al. (2021) study the best-arm identification problem with a goal to find the arm with the largest median and mean, respectively, in presence of corruptions. While adhering to the same corruption model, Basu et al. (2022) consider the regret minimisation problem, and devise a UCB-type algorithm that incurs $O(\log(T))$ instance-dependent regret that is within a constant of the lower bound. Significantly improving on their work, we devise an algorithm whose regret exactly matches the lower bound asymptotically, as $T \to \infty$. We also demonstrate the superiority of the proposed algorithm experimentally.

2. Lower bound and KL-divergence in corrupted neighbourhoods

Given a class $\mathcal{L}$ of probability distributions, we want algorithms that perform uniformly well on all the K-armed bandit instances with arms from $\mathcal{L}$, when the observations are corrupted with probability $\varepsilon \in [0, 1/2)$. To meet this requirement, the algorithm needs to generate sufficient samples from each arm. In this section, we present a lower bound on the number of samples that the algorithm needs to generate from each arm.

2.1. Problem-dependent lower bound

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

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

Here, $\mathbb{E}_{\mu \odot_\varepsilon \mathbf{H_T}}[\cdot]$ denotes the expectation with respect to both the corrupted bandit process $\mu \odot_\varepsilon \mathbf{H}_n$, for each $n \in [T]$, and the possible randomness of the algorithm (omitted from notation). Definition 1 is similar to the notion of consistent algorithms considered in the classical setup (Lattimore and Szepesvári, 2020, Definition 16.1). Observe that unlike in that setting, for every instance, the algorithm should perform well with respect to *every sequence* of K corruption distributions.

The lower bound on the expected number of times a uniformly-good algorithm pulls a suboptimal arm involves an optimisation problem, which we present first.

Corrupted KL-inf. The corrupted KL-inf is a function $\mathrm{KL}^\varepsilon_{\inf} : \mathcal{P}(\mathbb{R}) \times \mathbb{R} \times 2^{\mathcal{P}(\mathbb{R})} \to \mathbb{R}^+$, that for $\eta \in \mathcal{P}(\mathbb{R})$, $x \in \mathbb{R}$, and $\mathcal{L} \subset \mathcal{P}(\mathbb{R})$, equals

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

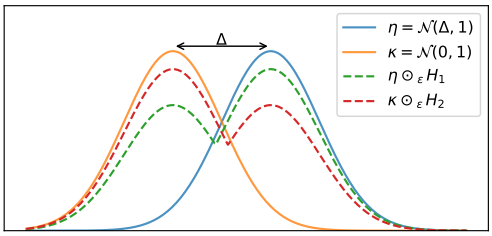
For $\varepsilon = 0$, this is equivalent to the optimisation problem that appears in the lower bound of the uncorrupted setting, leading to the traditional $\mathrm{KL}_{\inf} := \min_\kappa \{\mathrm{KL}(\eta, \kappa) : \ \kappa \in \mathcal{L}, \ m(\kappa) \geq x\}$ (c.f. Burnetas and Katehakis (1996), Lattimore and Szepesvári (2020, Chapter 16)). The additional optimisation over the corruption distributions H and H' makes $\mathrm{KL}^\varepsilon_{\inf}$ smaller than $\mathrm{KL}_{\inf}$. Moreover, we observe (Figure 1(*b*)) that for $\varepsilon > 0$, $\mathrm{KL}^\varepsilon_{\inf}$ can be non-convex in the second argument, unlike for $\varepsilon = 0$ (Agrawal et al., 2021, Lemma 10). As we will see later, these imply that the problem in presence of corruption is inherently harder than the classical setting.

In the reminder of this paper, ε denotes a known and fixed constant in $(0, 0.5)$. See Appendix B for a negative result, and a justification for the need to know ε (Remark 8).

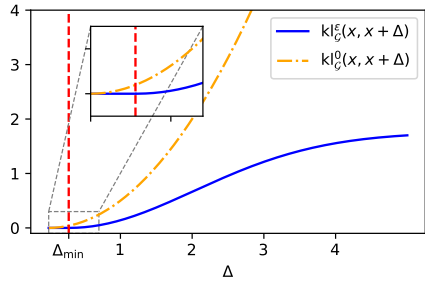
Theorem 1 (Lower bound) *For $\varepsilon > 0$, $\mathcal{L} \subset \mathcal{P}(\mathbb{R})$, and a bandit instance $\mu \in \mathcal{L}^K$, for any suboptimal arm a in μ, a uniformly-good algorithm satisfies*

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

A few remarks are in order. First, since $\mathrm{KL}^\varepsilon_{\inf} \leq \mathrm{KL}_{\inf}$ for $\varepsilon \geq 0$, the lower bound above is higher than that in the classical setting. Second, we show in discussion around Remark 8 that for $\varepsilon = \frac{1}{2}$, $\mathrm{KL}^\varepsilon_{\inf} = 0$, implying that logarithmic regret cannot be achieved if a bound on the corruption



(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 $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$. $\mathrm{Supp}(H)$ denotes the support of distribution H. $\Delta_{\min}$ is defined in Definition 3.

probability is unknown. Next, we show in Lemma 2 that a separation between $m(\mu_a)$ and $m^*(\mu)$ is required, without which $\mathrm{KL}^{\varepsilon}_{\inf} = 0$. Next, recall that $\mathbf{H}_n$ is allowed to possibly depend on the previous reward observations but not on the past algorithmic actions. Since the setting with corruption $\mathbf{H}$ fixed across time is simpler than one allowing for different $\mathbf{H}_n$ at each time n, the lower bound in Theorem 1 holds for the general setting considered in this paper (Remark 6).

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, which is of independent interest. We refer the reader to Section A.1 for a complete proof of Theorem 1.

2.2. Huber's pair and corrupted KL-inf

In Section 3, the proposed algorithm computes $\mathrm{KL}^{\varepsilon}_{\inf}$ using samples. To facilitate computation, in this section, we characterise the optimisers for $\mathrm{KL}^{\varepsilon}_{\inf}(\eta, x; \mathcal{L})$, specifically, the optimal 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 $\mathrm{KL}^{\varepsilon}_{\inf}$. Define

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

and

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

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 $x \in \mathrm{Sp}\,(\eta) \cap \mathrm{Sp}\,(\kappa)^c$, $\frac{\mathrm{d}\eta}{\mathrm{d}\kappa}(x) \coloneqq \infty$, and for $x \in \mathrm{Sp}\,(\eta)^c \cap \mathrm{Sp}\,(\kappa)$, $\frac{\mathrm{d}\eta}{\mathrm{d}\kappa}(x) \coloneqq 0$. 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 corruption distributions H_1 and H_2 (Remark 7).

Lemma 1 (Optimal corruption pair) *For $\eta \in \mathcal{P}(\mathbb{R})$, $\kappa \in \mathcal{P}(\mathbb{R})$, and $\varepsilon \in (0, \frac{1}{2})$, H_1 and H_2 defined by Equations (2.2) and (2.3), respectively, are the optimal corruption pair in Equation (2.1), i.e., $(H_1, H_2) \in \operatorname{argmin} \{\mathrm{KL}(\eta \odot_{\varepsilon} H, \kappa \odot_{\varepsilon} H') : H \in \mathcal{P}(\mathbb{R}),\ H' \in \mathcal{P}(\mathbb{R})\}$.*

To prove the above result, we show that the directional derivative (for an 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. A similar pair of corruption distributions were considered by Huber (1965) in a hypothesis testing setup.

It follows from Lemma 1 that the optimal corruption pair depends on the two input distributions. In particular, the corruption always stays within the support of the input pair of distributions. For illustration, we present these 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 for Gaussian model with corruption are formally proven in Lemma 3 and Appendix D.2, and will be used later in our analysis.

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 $\mathcal{L} = \mathcal{G}$, the collection of all Gaussian distributions with a unit variance, 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, which also holds under time-varying corruption (Remark 6).

Recall that $m(\mu_a)$ denotes the mean reward of arm a with distribution μ_a.

Proposition 2 (Lower bound for Gaussian bandits) *For $\mu \in \mathcal{G}^K$, any uniformly-good algorithm satisfies for any suboptimal 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 state the necessary and sufficient conditions to have a finite lower bound on regret in Gaussian bandits (Proposition 2) for a known and fixed $\varepsilon \in \left(0, \frac{1}{2}\right)$. Thus, Lemma 2 states the (necessary and sufficient) conditions to achieve logarithmic regret for Gaussian bandits.

Lemma 2 (Disjoint corruption neighbourhoods) *For $\eta \in \mathcal{G}$, $\kappa \in \mathcal{G}$, the following are equivalent:*

$$(\mathbf{1})\ \forall (H_1, H_2) \in \mathcal{P}(\mathbb{R})^2,\ \kappa \odot_\varepsilon H_1 \neq \eta \odot_\varepsilon H_2, \quad (\mathbf{2})\ |m(\kappa) - m(\eta)| > 2\Phi^{-1}\left(\frac{1}{2(1-\varepsilon)}\right).$$

The condition (1) above states that for any corruption distribution H_1, there doesn't exist a distribution H_2 such that the corrupted distributions $\kappa \odot_\varepsilon H_1$ and $\eta \odot_\varepsilon H_2$ are the same, rendering $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(m(\kappa), m(\eta)) = 0$, i.e., the corruption neighbourhoods (Definition 9) of κ and η are disjoint. The lemma above shows that this condition is equivalent to separation in the means of the two Gaussian distributions η and κ. This is also related to the fact that in the presence of corruption, the mean of the true distributions can only be estimated up to an unavoidable error (Appendix C). We postpone the proof of Lemma 2 to Section C.1.

Lemma 2, when combined with Proposition 2, implies that a suboptimal condition to ensure logarithmic regret is a separation between the means of the optimal arm and suboptimal arms. This justifies formally introducing the required minimum gap.

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

For Gaussian distributions, 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).*

(a) The normalisation constants c_1 and c_2 are equal, i.e., $c \coloneqq c_1 = c_2$, and uniquely solve

$$1/(1-\varepsilon) = c\Phi(\Delta_-/2) + \Phi(\Delta_+/2), \tag{2.5}$$

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

(b) $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ has an almost closed-form expression (Equation (D.1)) (up to c defined as in Part (a)). For $0 \le \Delta \le \Delta_{\min}$, $x \in \mathbb{R}$, $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x+\Delta) = 0$. Moreover, it is invariant, i.e.,

$$\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x+\Delta, y) = \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, y-\Delta), \quad \text{for } y \ge x+\Delta.$$

(c) For $x \in \mathbb{R}$ and $\Delta \ge 0$, the function $\Delta \mapsto \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x+\Delta)$ is continuously differentiable. For $\varepsilon > 0$ and $\Delta \le \Delta_{\min}$, $\partial\, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x+\Delta)/\partial\Delta = 0$. For $\Delta > \Delta_{\min}$,

$$\frac{\partial\, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x+\Delta)}{\partial \Delta} = (1-\varepsilon)\,\Delta\,(\Phi(\Delta_+/2) - \Phi(\Delta_-/2)) > 0.$$

They constitute the key properties of the corrupted divergence used in our regret analysis. We refer the reader to Appendix D for a complete proof of Lemma 3, plus other interesting properties of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$.

Consequences of Lemma 3. Part (b) shows that there is a flat region below $\Delta = \Delta_{\min}$, where $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x+\Delta)$ equals 0 (Fig. 1(*b*)). We use this property in regret analysis of the algorithm for proving fast convergence of the empirical $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ to 0, as well as to avoid certain computations at each step. Part (c) shows that $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ is strictly increasing in the second argument for values larger than the first argument. This was used in Proposition 2 to conclude that $\mathrm{KL}^{\varepsilon}_{\inf}(\eta, x; \mathcal{G}) = \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(m(\eta), x)$.

Computational remarks. The normalising constant c from Part (a) implicitly depends on ε and Δ, and so do Δ_- and Δ_+. Here, Δ_- and Δ_+ are related to support sets of the optimal corruption pair H_1 and H_2 (Lemma 7 in Appendix D). 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 Equation (2.5). In this limit, from Lemma 3(c), it follows that the derivative of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ converges to 0.

We also note that unlike the classical Gaussian bandit, from Fig. 1(*b*) we see that $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x, x+\Delta)$ is non-convex in Δ, implying that after a point, increasing Δ does not substantially decrease the number of pulls of suboptimal arms, and hence the regret, in presence of corruption.

3. `CRIMED`: Algorithm and analysis

In this section, we leverage the lower bound in Proposition 2 to propose an algorithm robust to corruption, namely `CRIMED`. We then give a finite-sample upper bound on the regret of `CRIMED`, showing its asymptotic optimality for Gaussian bandits with unbounded stochastic corruption. Finally, we explicate the technical novelty of our regret analysis. We discuss an extension of `CRIMED` for handling model-misspecifications in Appendix F.

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^*,

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

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

end

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

First, we present our algorithm design. For $n \in \mathbb{N}$ and $a \in [K]$, let $\hat{\mu}_a(n)$ denote the empirical distribution constructed using $N_a(n)$ samples from arm a. We use median of the corrupted observations as an estimator for the mean of underlying reward distributions. This choice is natural in the case of Gaussian distributions, since it is known 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 C). The fact that we use the median is also closely linked to the symmetry of the Gaussian distribution for which median is the same as mean.

Let $\mathrm{Med}(\cdot)$ denote the median of the input distribution, and define the maximum estimated median at time n as $\mathrm{Med}_*(n) := \max_a \mathrm{Med}(\hat{\mu}_a(n))$. We present CRIMED in Algorithm 1. Note that in Algorithm 1 we introduce a forced exploration for $N_{\min}$ steps, where for $T > 0$,

$$N_{\min} := \left\lceil \frac{2\log(T)\log(1+\log(1+\log(T)))^2 s_\varepsilon^2}{\log(1+\log(T)^{0.99})} \right\rceil, \; s_\varepsilon := \frac{\left(\frac{\varepsilon/2}{\log\frac{1}{1-2\varepsilon}}\right)^{\frac{1}{2}} + \left(\frac{1-2\varepsilon}{4\log\left(\frac{1-\varepsilon}{\varepsilon}\right)}\right)^{\frac{1}{2}}}{(1-\varepsilon)\,\varphi\left(\frac{\Delta_{\min}}{2}+1\right)}. \quad (3.1)$$

Here, s_ε is a proxy of the variance of the empirical median from Theorem 3. It converges to a constant, $\frac{1}{2\varphi(1)}$, as $\varepsilon \to 0$. The amount of forced-exploration $N_{\min}$ is $o(s_\varepsilon^2 \log T)$ as $T \to \infty$. 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 use the explicit formulation for $\mathrm{kl}_{\mathcal{G}}^{\varepsilon}$ (Lemma 3(b)), while we compute the constant c using a root-finding algorithm on Equation (2.5).

3.2. Theoretical results: Regret upper bound and concentration results

We now present the theoretical guarantees for the proposed algorithm, as well as the refined concentration inequality for median that play a key role in our analysis, and are of independent interest.

Theorem 2 (Finite-sample regret upper bound) *For $\varepsilon \in (0, \frac{1}{2})$ and $\mu \in \mathcal{G}^K$ such that for each suboptimal arm a, $\Delta_a > \Delta_{\min}$, CRIMED satisfies*

$$\mathbb{E}[N_a(T)] \le N_{\min} + \frac{\log(T)}{\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(m(\mu_a), m^*(\mu)\right) - 2\delta\left(\Delta_a + \delta + \Delta_{\min}\right)} + O((\log T)^{0.99}),$$

where $m^*(\mu) = \max_a m(\mu_a)$, and $\delta^2 := (\log(1+\log(1+\log T)))^{-1}$.

The exact $O((\log T)^{0.99})$ term can be found in Equation (E.5) in the Appendix E. 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. Indeed, in Section 4, we numerically compare `CRIMED` and an aggressive version with $N_{\min} = 1$, called `CRIMED*`, and observe a smaller regret for `CRIMED*`.

Corollary 4 (Asymptotic optimality) *For $\varepsilon \in (0, \frac{1}{2})$ and $\mu \in \mathcal{G}^K$ such that for each suboptimal arm a, $\Delta_a > \Delta_{\min}$, `CRIMED` is asymptotically optimal, i.e.,*

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

As a consequence, we also have

$$\lim_{T\to\infty} \frac{\mathbb{E}[R_T]}{\log T} = \sum_{a=1}^{K} \frac{\Delta_a}{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(m(\mu_a), m^*(\mu))}$$

The proof of Theorem 2, which we detail in Appendix E, proceeds by controlling the probability of selecting a suboptimal arm a at each step n. `CRIMED` pulls arm a at time n if its index $I_a(n)$ is the smallest. Thus, the probability of pulling an arm can be bounded by controlling the deviations of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ 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 $X_1, \ldots, X_n$, $\mathrm{Med}(X_1^n)$ denotes the empirical median. This can be alternatively seen as $\mathrm{Med}(\hat{P}_n)$, where $\hat{P}_n$ denotes their empirical distribution of $X_1, \ldots, X_n$.

Theorem 3 (Concentration of median for corrupted Gaussians) *Let $X_1, \ldots, X_n$ be random samples sampled with $X_i \sim \mathcal{N}(m, 1) \odot_{\varepsilon} H_i$, for $H_i \in \mathcal{F}(X_1, \ldots, X_{i-1})$, i.e. the smallest filtration spanned by $X_1, \ldots, X_{i-1}$, and let $\varepsilon < \frac{1}{2}$. For $y \in [0, 1]$,*

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

Note that $H_i \in \mathcal{F}(X_1, \ldots, X_{i-1})$, which means the outliers can depend on the past observations. We remark that Theorem 3 does not require i.i.d. samples. In particular, the corruption distribution is allowed to be time-varying. Further, observe from Equation (3.1) that when ε goes to 0, s_{ε} goes to $\frac{1}{2\varphi(1)}$. On the other hand, when ε goes to $\frac{1}{2}$, $\varphi(\frac{\Delta_{\min}}{2} + 1)$ goes to 0, hence s_{ε} goes to ∞, and we get a trivial bound in the theorem. In particular, our bound adapts to ε. We refer the reader to Appendix E.5 for a proof of the theorem.

Remark 5 (A refined concentration result) Theorem 3 is an improvement over the concentration in Altschuler et al. (2019, Lemma 7) in which the variance term does not depend on ε. Furthermore, Theorem 3 allows for an ε that is arbitrarily close to $\frac{1}{2}$, which is an improvement over the existing bounds for robust mean estimators featuring an upper limit on ε away from $\frac{1}{2}$. For example, $\varepsilon \leq \frac{1}{7}$ in (Wang and Ramdas, 2023, Theorem 2), and $\varepsilon \leq \frac{1}{15}$ in (Altschuler et al., 2019, Theorem 18).

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

Lemma 4 (Concentration of $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$.) *Let $\delta > 0$, $x \in \mathbb{R}$. Let $X_1, \dots, X_n$ be n random samples such that $X_i \sim \mathcal{N}(m_a, 1) \odot_\varepsilon H_i$, for $H_i \in \mathcal{F}(X_1, \dots, X_{i-1})$.*

(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 $1 - 2\exp\left(-ny^2/s_\varepsilon^2\right)$,

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

A useful thresholding. 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, from Lemma 4(a). This contrasts with the uncorrupted-Gaussian setting, for which this is a zero probability event, except when $y = 0$. We extensively use this key property 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 $m_a - y - \mathrm{Med}(X_1^n) \leq \frac{\Delta_{\min}}{2}$. We refer the reader to Appendix E.6 for a proof of the lemma.

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 arises 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 $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ is large.

We decompose the event $A_n = a$ as a union of three disjoint events. (i) $E_n(a)$: when the suboptimal arm is not well estimated. (ii) $F_n(a)$: when the optimal arm is not well estimated and $G_n(a)$ when $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ has large deviations (ref. Lemma 10 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 illustration

In this section, we numerically illustrate the efficiency of our algorithm. The computation of `CRIMED` indexes depends on the threshold c, that we evaluate using Equation (2.5) and the default *scipy* (Virtanen et al., 2020) root-finding algorithm. We consider arm distributions as $\mathcal{N}(m_a, \sigma_a^2)$, i.e., Gaussian inliers. The corruption distributions (outliers) are of form $\mathcal{N}(m_o, \sigma_o^2)$ in Setting 1 and 2, and standard Cauchy in Setting 3. Table below details the parameters used. Arm 3 is optimal.

Parameters	Horizon	σ_a	means arms m_a	ε	medians outliers m_b	σ_o outliers
Setting 1	10,000	0.5	[0.8, 0.9, 1]	0.01	[1, 1, 0.8]	1
Setting 2	10,000	0.5	[0.8, 0.9, 1]	0.01	[10, 10, -20]	1
Setting 3	10,000	0.5	[0.8, 0.9, 1]	0.01	[10, 10, -20]	∞

Setting 1 corresponds to a mild corruption, in which the corrupted distribution of arm 3 still has the largest mean of 0.98. In Setting 2, the corruption causes a change in the order of the arms (arm 2

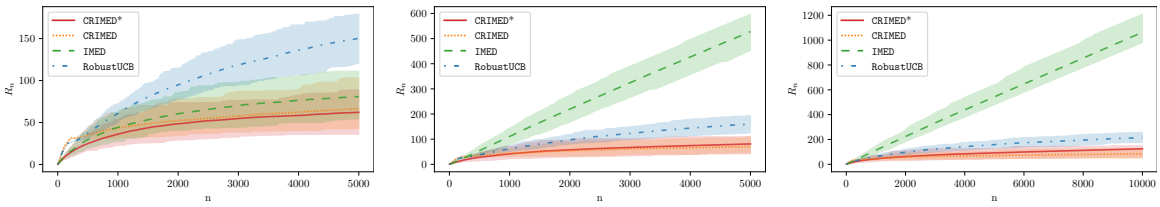


Figure 2: Cumulative regret for 100 repetitions on Settings 1 (left), 2 (middle), and 3 (right). Solid lines represent the means and shaded area are 90% percentile intervals.

is optimal according to the corrupted distributions). Hence, robustness is needed to identify correctly the optimal arm. In Setting 3, the outliers are heavy-tailed. We compare 4 algorithms: `CRIMED` (Algorithm 1) with $N_{\min}$ set to Equation (3.1), `CRIMED`*, an 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 means are estimated using the empirical mean, and finally `RobustUCB` (Basu et al., 2022).

Figure 2 illustrates that all the algorithms, except `IMED`, feature a logarithmic regret. `IMED` incurs a linear regret in large-corruption settings (Settings 2, 3). On the other hand, `CRIMED` and `CRIMED`* perform comparably, and they are both performing better than `RobustUCB`. We also observe no significant difference in performance when the corruptions are heavy-tailed vs Gaussian.

5. Discussion and open questions

We studied a variant of the stochastic multi-armed bandits with unbounded stochastic corruption and analysed the behaviour of the KL divergence within a corruption neighbourhood in details. This served as the foundation for the proposed algorithm `CRIMED`, that asymptotically achieves this lower bound for Gaussian bandits. We developed a new concentration result for median allowing `CRIMED` to tackle corruption proportion up to $\frac{1}{2}$, which was not possible with the existing algorithms. We discussed a modification of `CRIMED` to handle misspecifications in Gaussian model.

We believe that the Gaussian reward assumption is not essential, and we believe it is possible to generalise these results to at least symmetric, unimodal distributions. `CRIMED` can be appropriately adjusted to handle more general reward distributions with the following modifications: 1) replacing median with any robust estimator for mean that concentrates fast, for example, median-of-means, and 2) adjusting $I_a(n)$, the index for arm a, to replace $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$ by the corresponding KL-inf. These changes would necessitate establishing concentration results similar to Theorem 3 and Lemma 4.

Note that the Gaussian assumption enabled us to obtain an (almost) closed-form expression of the $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}$. Further, its symmetry property allowed us to use the known exact optimality (including constants) of the median for robust mean estimation. Generalisation to non-symmetric and possibly non-parametric inliers is likely to be much more challenging. This is because in the non-symmetric case, robust estimators approximate location parameters that may be far from the mean and we have to trade off between the error due to corruption and the error due to asymmetry. This requires the study of a non-trivial trade-off between robustness and asymptotic distance to the mean, using robust mean estimators such as median-of-means or M-estimators.

Similarly, we believe that the proof techniques we introduce for corruption robust bandits are complementary to those for some structured bandit problems (such as unimodal or Lipschitz structure) and, hence, could yield extensions to such settings, provided that an efficient and optimal multivariate robust mean estimator is used.

Acknowledgments

This work commenced when S. Agrawal was a PhD student at TIFR, Mumbai, India. S. Agrawal acknowledges support from the Department of Atomic Energy, Government of India, under project no. RTI4001. This work has also been supported by the French Ministry of Higher Education and Research, the Hauts-de-France region, Inria, the MEL, the I-Site ULNE regarding project R-PILOTE-19-004-APPRENF, and the Inria A.Ex. SR4SG project. D. Basu, O.-A. Maillard, and T. Mathieu acknowledge the Inria-Kyoto University Associate Team "RELIANT" for supporting the project. D. Basu also acknowledges the ANR JCJC grant for the REPUBLIC project (ANR-22-CE23-0003-01). Additionally, S. Agrawal acknowledges support from the Sarojini Damodaran Fellowship and the Google PhD Fellowship in Machine Learning.

References

Yasin Abbasi-Yadkori, Peter Bartlett, Victor Gabillon, Alan Malek, and Michal Valko. Best of both worlds: Stochastic & adversarial best-arm identification. In *Conference on Learning Theory*, pages 918–949. PMLR, 2018.

Jacob Abernethy and Alexander Rakhlin. Beating the adaptive bandit with high probability. In *2009 Information Theory and Applications Workshop*, pages 280–289. IEEE, 2009.

John Adams, Darren Hayunga, Sattar Mansi, David Reeb, and Vincenzo Verardi. Identifying and treating outliers in finance. *Financial Management*, 48(2):345–384, 2019.

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, 2023.

Shubhada Agrawal, Sandeep Juneja, and Peter Glynn. Optimal δ-correct best-arm selection for heavy-tailed distributions. In *Algorithmic Learning Theory*, pages 61–110. PMLR, 2020.

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.

Jean-Yves Audibert, Sébastien Bubeck, et al. Minimax policies for adversarial and stochastic bandits. In *COLT*, volume 7, pages 1–122, 2009.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of IEEE 36th annual foundations of computer science*, pages 322–331. IEEE, 1995.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002a.

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, 2002b.

Achraf Azize and Debabrota Basu. Interactive and concentrated differential privacy for bandits. *arXiv:2309.00557*, 2023.

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

Sébastien Bubeck and Aleksandrs Slivkins. The best of both worlds: Stochastic and adversarial bandits. In *Conference on Learning Theory*, pages 42–1. JMLR Workshop and Conference Proceedings, 2012.

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.

Moses Charikar, Jacob Steinhardt, and Gregory Valiant. Learning from untrusted data. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 47–60, 2017.

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.

Ilias Diakonikolas, Gautam Kamath, Daniel M Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robustly learning a gaussian: Getting optimal error, efficiently. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2683–2702. SIAM, 2018.

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.

Dylan J Foster, Claudio Gentile, Mehryar Mohri, and Julian Zimmert. Adapting to misspecification in contextual bandits. *Advances in Neural Information Processing Systems*, 33:11478–11489, 2020.

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.

Avishek Ghosh, Sayak Ray Chowdhury, and Aditya Gopalan. Misspecified linear bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

Negin Golrezaei, Vahideh Manshadi, Jon Schneider, and Shreyas Sekar. Learning product rankings robust to fake users. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 560–561, 2021.

Anupam Gupta, Tomer Koren, and Kunal Talwar. Better algorithms for stochastic bandits with adversarial corruptions. In *Conference on Learning Theory*, pages 1562–1578. PMLR, 2019.

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: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.

Peter J. Huber. Robust estimation of a location parameter. *Ann. Math. Statist.*, 35(1):73–101, 03 1964. ISSN 0003-4851.

Peter J. Huber. A Robust Version of the Probability Ratio Test. *The Annals of Mathematical Statistics*, 36(6):1753 – 1758, 1965. 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.

Kwang-Sung Jun, Lihong Li, Yuzhe Ma, and Jerry Zhu. Adversarial attacks on stochastic bandits. *Advances in Neural Information Processing Systems*, 31, 2018.

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.

Fang Liu and Ness Shroff. Data poisoning attacks on stochastic bandits. In *International Conference on Machine Learning*, pages 4042–4050. PMLR, 2019.

Gábor Lugosi and Shahar Mendelson. Mean estimation and regression under heavy-tailed distributions: A survey. *Foundations of Computational Mathematics*, 19(5):1145–1190, 2019.

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.

Arpan Mukherjee, Ali Tajer, Pin-Yu Chen, and Payel Das. Mean-based best arm identification in stochastic bandits under reward contamination. *Advances in Neural Information Processing Systems*, 34:9651–9662, 2021.

Gergely Neu. Explore no more: Improved high-probability regret bounds for non-stochastic bandits. *Advances in Neural Information Processing Systems*, 28, 2015.

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 Gábor Lugosi. An improved parametrization and analysis of the exp3++ algorithm for stochastic and adversarial bandits. In *Conference on Learning Theory*, pages 1743–1759. PMLR, 2017.

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.

Lehana Thabane, Lawrence Mbuagbaw, Shiyuan Zhang, Zainab Samaan, Maura Marcucci, Chenglin Ye, Marroon Thabane, Lora Giangregorio, Brittany Dennis, Daisy Kosa, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC medical research methodology*, 13(1):1–12, 2013.

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.

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

Yinglun Xu, Bhuvesh Kumar, and Jacob D Abernethy. Observation-free attacks on stochastic bandits. *Advances in Neural Information Processing Systems*, 34:22550–22561, 2021.

Julian Zimmert and Yevgeny Seldin. An optimal algorithm for stochastic and adversarial bandits. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 467–475. PMLR, 2019.

Appendix A. Proofs for results in Section 2.1 and Section 2.2: Lower bound and hardest corruption pair

Remark 6 Observe that

$$\sup_{\mathbf{H_T}\in\mathcal{P}(\mathbb{R})^{T\times K}} \mathbb{E}_{\mu\odot_\varepsilon\mathbf{H_T}}\left[N_a(T)\right] \geq \sup_{\mathbf{H}\in\mathcal{P}(\mathbb{R})^{K}} \mathbb{E}_{\mu\odot_\varepsilon\mathbf{H}}\left[N_a(T)\right],$$

where $\mathbf{H_T}$ is a $T\times K$ matrix of corruption distributions with row n being $\mathbf{H}_n$, the vector of corruption distributions at time n. Note that the inequality follows since, in the RHS above, the corruption distributions associated with each arm are the same across time. Below, we establish the lower bound for this simpler setting, where the corruption doesn't change with time. Since the lower bound for fixed $\mathbf{H}$ across time will only be smaller, this proves that the same lower bound holds even for the more general setting considered in this work that allows for time-varying corruption.

A.1. Proof of Theorem 1

Recall that the given bandit instance is denoted by $\mu\in\mathcal{L}^K$. For $a\in[K]$ and $i\in\{1,\ldots,N_a(T)\}$, let $\{X_{a,i}\}_{a,i}$ denote the $N_a(T)$ corrupted observations from arm a till time T. Under μ with corruption distributions $\mathbf{H}=\{H_1,\ldots,H_K\}$, the likelihood of observing 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. We refer the reader to Lattimore and Szepesvári (2020, Chapter 4) for a formal

introduction to stochastic multi-armed bandits. An application of the data-processing inequality (see, 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 means x and y. Since the RHS above is true for all events $\mathcal{E}_T$ that are $\mathcal{F}_T$ measurable, optimising 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

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

Since for every $\mathbf{H}$, the corresponding $\mathbf{H}'$ with $H'_b = H_b$ for all $b \in [K]$ and $b \neq 2$ is feasible, 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-minimisation 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

$$\mathbb{P}_{\nu \odot_\varepsilon \mathbf{H}'}\left(\mathcal{E}_T^c\right) = \mathbb{P}_{\nu \odot_\varepsilon \mathbf{H}'}\left(N_1(T) \geq T - \sqrt{T}\right) \leq \frac{\sum_{a \neq 2} \mathbb{E}_{\nu \odot_\varepsilon \mathbf{H}'}\left[N_a(T)\right]}{T - \sqrt{T}} =: 1 - Q_T^{\mathbf{H}'} = (Q_T^{\mathbf{H}'})^c.$$

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

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

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

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

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 optimise 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 Equations (2.2) and (2.3) are optimal for $\mathrm{KL}^{\varepsilon}_{\inf}$, 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

$$H_{i,t} = (1-t)H_i + tH_i', \quad \text{for } i \in \{1,2\},$$

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 minimised at $t=0$. To see this,

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

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

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

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 $t=0$. To this end, we now define the sets

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

Then,

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

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.

Remark 7 Observe that Equations (2.2) and (2.3) implicitly define the probability measures H_1 and H_2. To see this, we first argue that H_1 is non-negative. Recall that

$$\mathrm{d}(\eta \odot_\varepsilon H_1) = (1-\varepsilon)\mathrm{d}\eta + \varepsilon \mathrm{d}H_1.$$

From Equation (2.2) it is clear that $\mathrm{d}H_1(x) = 0$ for $\frac{\mathrm{d}\eta}{\mathrm{d}\kappa} \geq c_1$. Thus, H_1 is supported only on the complement set, where it is non-negative by choice of c_1. Next, since $\mathrm{d}(\eta \odot_\varepsilon H_1)$ integrates to 1 (by choice of c_1) and so does η, H_1 too integrates to 1. One can similarly argue that H_2 defined by Equation (2.3) is a probability measure.

Appendix B. Discussion on the knowledge of ε and an impossibility result

We note that in the current work, we do not require the knowledge of the precise value of corruption probability ε. Instead, knowing an upper bound on ε suffices. In this section, we will show that without such knowledge, no uniformly good algorithm can achieve logarithmic regret (Remark 8).

Let us remind the reader that this is also a standard assumption in the robust statistics literature. On a related note, Wang and Ramdas (2023, Remark 3) mention as an open problem whether it is possible to construct confidence intervals using corrupted samples without the knowledge of ε (or a bound on it). In this section, we provide a negative answer to this question.

Our approach is to relate the problem of constructing anytime-valid confidence intervals in the presence of corruption to a specific sequential hypothesis testing problem using corrupted data, which can, in turn, be formulated as the best-arm identification (BAI) problem with corruptions in the multi-armed bandit framework. We then use the machinery from the proof of Theorem 1 to arrive at a lower bound for this problem in terms of $\mathrm{KL}^{\varepsilon}_{\inf}$, defined in Equation (2.1). We show that if a bound on ε is not known, then $\mathrm{KL}^{\varepsilon}_{\inf} = 0$, rendering the BAI problem un-learnable, hence, the impossibility for the existence of a sequential test, and hence, the impossibility for the existence of non-trivial anytime confidence intervals. The approach for proving the impossibility result for BAI is reminiscent of a similar negative result for classical BAI in bandits with heavy-tailed distributions (uncorrupted setting), proven in (Agrawal et al., 2020, Theorem 3).

For $\delta > 0$, and a collection $\mathcal{L}$ of probability measures, we only prove the negative result for constructing an anytime valid upper bound that holds with probability at least $1-\delta$, using samples generated from a corruption neighbourhood (Definition 9) of a distribution $\mu_1 \in \mathcal{L}$. Symmetric arguments (including a symmetric sequential test) give a corresponding negative result for the lower bound that holds with probability at least $1-\delta$. We now introduce the specific sequential test for this.

Sequential setting (hypothesis testing) of interest. Consider the problem of testing whether the mean of a distribution $\mu_1 \in \mathcal{L}$ is below a given threshold ζ in a δ-correct framework. To be more specific, let $m(\mu_1)$ denote its mean, and let $m(\mu_1) < \zeta$ (unknown to the algorithm). The algorithm can generate samples from μ_1. However, on doing so, it observes the true sample with probability $1-\varepsilon$, and receives a sample from an arbitrary corruption distribution with the remaining ε probability, i.e., it observes samples from an ε corruption neighbourhood of μ_1. In presence of ε corruption, the goal of the algorithm is to generate finite samples (possibly random number of samples, depending on observations made), and declare that $m(\mu_1) < \zeta$ with probability at least $1-\delta$. While ensuring this δ-correctness property, the algorithm's goal is also to minimise its expected stopping time. Let us denote the δ-correct algorithm for this problem by $\mathcal{A}(\delta, \zeta)$.

Equivalence of δ-correct anytime-valid upper bound and sequential test described above. Observe that given a δ-correct algorithm for the above described problem, $\mathcal{A}(\delta, \zeta)$, one can construct an anytime-valid upper bound on the true mean for μ_1 that holds with probability at least $1-\delta$ using ε-corrupted samples as follows: at any time n, define the set

$$U_n := \{\zeta : \mathcal{A}(\delta, \zeta) \text{ has not stopped in } n \text{ samples}\}.$$

Then, U_n is an anytime-valid upper bound on $m(\mu_1)$ that holds with probability at least $1-\delta$. The reverse implication also holds, i.e., given any sequence of δ-correct anytime-valid upper bound on $m(\mu_1)$ constructed using ε-corrupted samples, say $\bar{U}_n$, one can design a δ-correct algorithm for the above described problem, as described next. Consider the algorithm that stops and declares $m(\mu_1) < \zeta$ at time n if $\bar{U}_n \leq \zeta$. Else, it generates a sample, computes U_{n+1}, and proceeds.

With the above equivalence at hand, it suffices to show that any δ-correct algorithm for identifying $m(\mu_1) \leq \zeta$ would require an unbounded number of samples if ε is not known. For this, following arguments similar to those in the proof of Theorem 1, one can show the following lower bound on the expected number of samples $\mathbb{E}[N]$ any δ-correct algorithm would require to generate, when the corruption proportion ε is known:

$$\mathbb{E}[N] \geq \frac{\log \frac{1}{\delta}}{\mathrm{KL}^{\varepsilon}_{\inf}(\mu_1, \zeta; \mathcal{L})}, \tag{B.1}$$

where, recall that

$$\mathrm{KL}^{\varepsilon}_{\inf}(\mu_1, \zeta; \mathcal{L}) := \min_{H, H', \nu_1} \left\{ \mathrm{KL}\left(\mu_1 \odot_{\varepsilon} H, \nu_1 \odot_{\varepsilon} H'\right) : \nu_1 \in \mathcal{L}, H, H' \in \mathcal{P}(\mathbb{R}), m(\nu_1) \geq \zeta \right\}. \tag{B.2}$$

When ε is not known to the algorithm, a further optimisation over $\varepsilon < \frac{1}{2}$ would feature in the lower bound. Otherwise, for every fixed $\varepsilon < \frac{1}{2}$ for which the algorithm is δ-correct, there exists an $\varepsilon' > 0$ such that $\tilde{\varepsilon} := \varepsilon + \varepsilon' < \frac{1}{2}$, and the algorithm wouldn't be δ-correct for some distribution in $\mathcal{L}$ with corruption proportions being $\tilde{\varepsilon}$. This follows from the corresponding lower bound in Equation (B.1) for $\tilde{\varepsilon}$ instead of ε, and monotonicity of $\mathrm{KL}^{\varepsilon}_{\inf}$ in ε (larger ε implies a smaller $\mathrm{KL}^{\varepsilon}_{\inf}$, and a higher lower bound). Thus, when ε is unknown, $\mathrm{KL}^{\varepsilon}_{\inf}$ with $\varepsilon = \frac{1}{2}$ would feature in the lower bound in Equation (B.1).

Unknown ε implies $\mathrm{KL}^{\varepsilon}_{\inf} = 0$. As discussed above, if ε is unknown, $\mathrm{KL}^{\varepsilon}_{\inf}$ features with $\varepsilon = \frac{1}{2}$ in the lower bound. Now, consider any distribution $\kappa \in \mathcal{L}$ such that $m(\kappa) \geq \zeta$. In the definition of $\mathrm{KL}^{\varepsilon}_{\inf}(\mu_1, \zeta; \mathcal{L})$ in Equation (B.2), $\nu_1 = \kappa$, $H = \kappa$, $H' = \eta$ are feasible solutions for $\varepsilon = \frac{1}{2}$, and satisfy $\mu_1 \odot_{\varepsilon} \kappa = \kappa \odot_{\varepsilon} \mu_1$, implying that $\mathrm{KL}^{\varepsilon}_{\inf} = 0$. Hence, the lower bound in Equation (B.1) is unbounded, implying non-existence of a δ-correct algorithm, hence δ-correct anytime-valid upper bound.

Remark 8 Observe that from the above discussion, we have that for $\varepsilon = \frac{1}{2}$, $\mathrm{KL}^{\varepsilon}_{\inf} = 0$. This also implies that the lower bound in Theorem 1 in unbounded for our regret-minimisation setting in presence of corruption. Thus, we also conclude that a logarithmic regret is not possible without a knowledge of a bound on ε that is strictly smaller than $\frac{1}{2}$.

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

Let $\mathcal{T}$ be the set of all functions A of probability distributions, $A : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$, such that A is translation equivariant, meaning that for any $\tilde{\delta} > 0$, if $X \sim \kappa$, and $X+\tilde{\delta} \sim \eta$, then $A(\eta) = A(\kappa)+\tilde{\delta}$. The class $\mathcal{T}$ of translation equivariant functions is interesting because common functions like mean, median, and quantiles belong to this class.

It is well known that in the presence of corruption with ε probability, consistent estimation of any translation equivariant function, including mean, is not possible even with infinite samples (Chen et al., 2018). The presence of corruption introduces a bias that is unavoidable, that we refer to as *bias due to corruption*. We recall this in what follows.

Definition 9 (Corruption neighbourhood.) For a fixed corruption proportion ε and a fixed distribution κ, its corruption neighbourhood is defined as the collection of all distributions in the following set (denoted as κ_ε):

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

Observe that the mean of distributions in the corruption neighbourhood of any distribution κ can be arbitrarily large or small.

The bias due to using the family of functions $\mathcal{T}$ defined above, for distribution κ in presence of corruption with probability ε, is defined as

$$b_\kappa(\varepsilon) := \inf_{A \in \mathcal{T}} \sup_{\kappa' \in \kappa_\varepsilon} |A(\kappa') - A(\kappa)|. \tag{C.1}$$

The definition above quantifies the mini-max bias suffered by any translation-equivariant function of κ, in presence of ε corruption. The lemma below, taken from Huber and Ronchetti (2009, Section 4.2), shows that when κ is symmetric and unimodal, then the function that suffers the minimum bias is median. We refer the reader to Huber and Ronchetti (2009) for a proof of the lemma.

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.

C.1. Proof of Lemma 2

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

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

We first prove that if for all $(H_1, H_2) \in \mathcal{P}(\mathbb{R})^2$ such that $\kappa \odot_\varepsilon H_1 \neq \eta \odot_\varepsilon H_2$, then

$$|m(\kappa) - m(\eta)| > 2b_0(\epsilon).$$

For this, we show the contrapositive of the above. To this end, let us assume that $|m(\kappa) - m(\eta)| \leq 2b_0(\varepsilon)$. Then

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

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

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

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

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

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

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

Hence $(p' - (1-\varepsilon')\kappa)(x) \geq 0$.

Similarly, if $x - m(\kappa) \geq 2b_0(\varepsilon')$, then

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

giving $(p' - (1-\varepsilon')\kappa)(x) \geq 0$.

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.

Since

$$|m(\kappa) - m(\eta)| = 2b_0(\varepsilon') = m(\eta) - m(\kappa),$$

we have

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

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

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

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

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

implying that $p'(x) - (1-\varepsilon')\eta(x) \ge 0$.

On the other hand, if $x - m(\eta) \ge 2b_0(\varepsilon')$, then

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

and then $p'(x) - (1-\varepsilon')\eta(x) \ge 0$.

Additionally, $p' - (1-\varepsilon')\eta$ sums to ε', as shown below:

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

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 $\kappa' \in \kappa_\varepsilon$,

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

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 D. 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 characterise its solutions in Lemma 1, and we 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 characterises the optimal H and H' for this problem, and are defined by Equation (2.2) and Equation (2.3). Later, in Lemma 6, we identify the support sets for the optimal corruption pair (H_1, H_2) in the specific setting 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. In particular, we show Equation (D.1) below which gives a closed-form expression for $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,y)$, for $x < y$, once we know the optimal c from Lemma 3(a), which we compute using a root-finding algorithm. For any $x < y$, let $\Delta = y - x$, c, Δ_+ and Δ_- be as in Lemma 3(a). Then,

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

D.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 normalisation constants for the optimal corruption distributions in $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,y)$ (Lemma 1).

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 (presented later), we have

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

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 the uniqueness of this common value c. From the discussion in the previous paragraph, c solves the following equation:

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

Observe that c is uniquely defined by Equation (D.2), 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 $x < y$,

$$\frac{\mathrm{kl}^\varepsilon_{\mathcal{G}}(x,y)}{1-\varepsilon} = \int_{A_c} c\varphi(t-y)\log(c) + \int_{D_c}\varphi(t-x)\log\frac{1}{c} + \int_{\mathbb{R}\setminus A_c\cup D_c}\varphi(t-x)\log\left(\frac{\varphi(t-x)}{\varphi(t-y)}\right)\mathrm{d}t. \tag{D.3}$$

On simplifying, it then equals $1-\varepsilon$ times

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

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

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

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

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

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 (D.3) we have that $\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,y)$ equals $(1-\varepsilon)$ times

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

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 c from Equation (D.2). 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 8,

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

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 (D.4), $\Delta\mapsto \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,x+\Delta)$ is also differentiable with continuous derivative on $(0,\infty)$.

Differentiating Equation (D.4) with respect to Δ (after setting $y=x+\Delta$), and substituting for c' from Lemma 8, we have that

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

Next, using

$$c\varphi(\Delta_-/2) = \varphi(\Delta_+/2),$$

and collecting the coefficients of like-terms, the required derivative scaled by $1-\varepsilon$ equals

$$\Delta\left(\Phi\left(\frac{\Delta_+}{2}\right)-\Phi\left(\frac{\Delta_-}{2}\right)\right)+\frac{\varphi^2(\Delta_+/2)}{\Phi(\Delta_-/2)}\left(\frac{1}{c\Delta}\log\frac{1}{c}-\frac{1}{c^2\Delta}\log\frac{1}{c}+\frac{\Delta}{2c}+\frac{\Delta}{2c^2}-\frac{\Delta_-}{2c^2}-\frac{\Delta_+}{2c}\right)$$
$$+\varphi\left(\frac{\Delta_+}{2}\right)\left(\log\frac{1}{c}-\frac{\Delta_+}{2\Delta}\log\frac{1}{c}+\frac{\Delta_+}{2\Delta c}\log\frac{1}{c}+\frac{\Delta\Delta_-}{4}-\frac{\Delta\Delta_+}{4c}+\frac{\Delta_+\Delta_-}{4c}-\frac{\Delta_+\Delta_-}{4}\right).$$

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. □

D.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

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

Similarly, we have the rewriting

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

■

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 normalisation constant. $\mathrm{Sp}\,(H_1) = A_c$ *and* $\mathrm{Sp}\,(H_2) = D_c$*, where*

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

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

Lemma 8 *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 (D.2),

$$\begin{aligned}0 =& c'(\Delta)\Phi\left(\frac{\Delta_-}{2}\right) + c(\Delta)\left(\frac{\Delta'_-}{2}\right)\varphi\left(\frac{\Delta_-}{2}\right) + \left(\frac{\Delta'_+}{2}\right)\varphi\left(\frac{\Delta_+}{2}\right)\\ =& c'(\Delta)\left(\Phi\left(\frac{\Delta_-}{2}\right) + \frac{1}{\Delta}\varphi\left(\frac{\Delta_-}{2}\right) - \frac{1}{c(\Delta)\Delta}\varphi\left(\frac{\Delta_+}{2}\right)\right)\\ &+ c(\Delta)\frac{\Delta_+}{2\Delta}\varphi\left(\frac{\Delta_-}{2}\right) + \frac{\Delta_-}{2\Delta}\varphi\left(\frac{\Delta_+}{2}\right).\end{aligned} \tag{D.5}$$

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

$$\varphi\left(\frac{\Delta_+}{2}\right) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\Delta_+^2}{8}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{\Delta_-^2}{8}+\log(c)} = c(\Delta)\varphi\left(\frac{\Delta_-}{2}\right). \tag{D.6}$$

Plugging this in Equation (D.5), we have

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

Hence, we deduce that

$$c'(\Delta) = -\frac{c(\Delta)\varphi\left(\frac{\Delta_-}{2}\right)}{\Phi\left(\frac{\Delta_-}{2}\right)}. \tag{D.7}$$

■

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

Lemma 9 (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,

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

On the other hand, if $m_a \geq m_b$, then $\mathrm{KL}^{\varepsilon}_{\inf}(\nu_a, m_*) - \mathrm{KL}^{\varepsilon}_{\inf}(\nu_b, m_*) \leq 0$. ■

Observe that Lemma 9 gives a bound very similar to that in the Gaussian setting without corruptions. Indeed, in the latter case,

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

Lemma 9 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 E. Proofs of results from Section 3

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

For $a \in [K]$ and $t \in \mathbb{N}$, let $\hat{\mu}_{a,t}$ denote the empirical distribution obtained using t samples observed (corrupted samples) from arm a. 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 10 below.

Lemma 10 (Decomposition of bad event) *For any $M > 0$,*

$$\{A_n = a\} \subset E_n(a) \cup F_n(a) \cup G_n(a),$$

where $E_n(a)$, $F_n(a)$ and $G_n(a)$ are disjoint events defined by

$$\begin{aligned}
E_n(a) &= \left\{ A_n = a, N_a(n)\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_a(n)) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \le \log n \right\}, \\
F_n(a) &= \bigcup_{t=N_{\min}}^{n} \Big\{ A_n = a,\ \mathrm{Med}(\hat{\mu}_{1,t}) \le m(\mu_1) - \delta - \frac{\Delta_{\min}}{2}, \\
&\qquad I_*(n) \le t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) + \log t \le tM + \log t \Big\}, \\
G_n(a) &= \bigcup_{t=N_{\min}}^{n} \left\{ A_n = a, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \ge M, N_1(n) = t \right\}.
\end{aligned}$$

Using Lemma 10, observe that for $T \ge KN_{\min}$,

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

Thus, to bound the average number of pulls of suboptimal arm a, it suffices to bound the summation of the probabilities of the above indicator functions since

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

In the above inequality,

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

the second term is equal to

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

and the third term satisfies

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

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

First, we bound the probability of the event $E_n(a)$ occurring with the following lemma. This gives us the main term in our regret upper bound.

Lemma 11 *For $\delta>0$ satisfying,*

$$\delta<\min\left(1,\Delta_a+\Delta_{\min},\frac{\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(m(\mu_a),m(\mu_1)\right)}{4(\Delta_a+\Delta_{\min})}\right),$$

we have

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

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

Lemma 12 *For $\delta<1$ and $M=\frac{\delta^2}{2s_{\varepsilon}^2}$,*

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

Lemma 13 *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

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

Substituting the bounds from Lemmas 11, 12, 13 in Equation (E.1), we get

$$\mathbb{E}[N_a(T)] \leq \frac{\log(T)}{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m(\mu_a), m(\mu_1)\right) - 2\delta\left(\Delta_a + \delta + \Delta_{\min}\right)} + \left\lceil \frac{2\log(T)s_{\varepsilon}^2}{\log(1+\log(T)^{0.99})\delta^2} \right\rceil + (\log T)^{0.99} + \frac{4}{\left(1-\exp\left(-\frac{\delta^2}{2s_{\varepsilon}^2}\right)\right)^2} + \frac{4}{1-\exp\left(-\delta^2/s_{\varepsilon}^2\right)}. \tag{E.5}$$

Next, choose

$$\delta^2 = \frac{1}{\log(1+\log(1+\log(T)))},$$

which also satisfies the constraints for Lemma 12 for T sufficiently large, and is such that $\delta \xrightarrow[T\to\infty]{} 0$. It satisfies,

$$\left\lceil \frac{2\log(T)s_{\varepsilon}^2}{\log(1+(\log T)^{0.99})\delta^2} \right\rceil + (\log T)^{0.99} + \frac{4}{\left(1-\exp\left(-\frac{\delta^2}{2s_{\varepsilon}^2}\right)\right)^2} + \frac{4}{1-\exp\left(-\delta^2/s_{\varepsilon}^2\right)} = o(\log(T)).$$

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.

E.2. Proof of Lemma 11: controlling deviations of suboptimal arm (event $E_n(a)$)

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

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

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

$$\sum_{t=1}^{\infty} \mathbb{P}\left(t\mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(\mathrm{Med}(\hat{\mu}_{a,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \leq \log T\right). \tag{E.6}$$

Each summand in the above expression is bounded by

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

Using that for $x \in \mathbb{R}$, $\frac{d\, \mathrm{kl}^{\varepsilon}_{\mathcal{G}}(x,x+\Delta)}{d\Delta} \leq \Delta$ (see Lemma 3), and injecting the probability back into Equation (E.6), we get

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

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

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

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

$$\begin{aligned}
2\delta\left(\Delta_a + \delta + \Delta_{\min}\right) &\leq 4\delta\left(\Delta_a + \Delta_{\min}\right) \\
&\leq \mathrm{kl}^{\varepsilon}_{\mathcal{G}}\left(m(\mu_a), m(\mu_1)\right).
\end{aligned}$$

Hence the denominator in t_0 is positive.

The required sum-of-probabilities (E.6) can further be bounded by:

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

which is further less than

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

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

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

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

$$\sum_{n}\mathbb{P}(E_n(a))\leq t_0-1+\frac{4}{1-\exp\left(-\delta^2/s_{\varepsilon}^2\right)}, \tag{E.7}$$

where it can be checked, by definition, that

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

E.3. Proof of Lemma 12: 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

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

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

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

which is smaller than

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

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

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

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

$$\sum_{t=1}^{T}\mathbb{E}\Big(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \le m(\mu_1) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^\varepsilon_\mathcal{G}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \le M\right)$$
$$\times\, e^{t\mathrm{kl}^\varepsilon_\mathcal{G}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) + \log t}\Big). \quad \text{(E.8)}$$

Using the bound on $\sum_n \mathbb{P}(F_n(a))$ from Equation (E.8) and observing that the expectation in the bound is for a non-negative random variable, we get the following bound on $\sum_{n=1}^{T}\mathbb{P}(F_n(a))$:

$$\sum_{t=1}^{T} t\int_0^\infty \mathbb{P}\Big(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \le m_1(\mu) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^\varepsilon_\mathcal{G}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \le M\right)$$
$$\times\, e^{t\mathrm{kl}^\varepsilon_\mathcal{G}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right)} \ge x\Big)\mathrm{d}x. \quad \text{(E.9)}$$

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 $\mathrm{kl}^\varepsilon_\mathcal{G}$:

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

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

$$\sum_{t=N_{\min}}^{T}\int_0^1 t\mathbb{P}\Big(\mathbb{1}\left(\mathrm{Med}(\hat{\mu}_{1,t}) \le m_1(\mu) - \delta - \frac{\Delta_{\min}}{2}, \mathrm{kl}^\varepsilon_\mathcal{G}\left(\mathrm{Med}(\hat{\mu}_{1,t}) - \frac{\Delta_{\min}}{2}, m(\mu_1) - \delta\right) \le M\right)$$
$$\le 2\sum_{t=1}^{\infty} te^{-t\frac{\delta^2}{s_\varepsilon^2}} = 2\frac{e^{-\frac{\delta^2}{s_\varepsilon^2}}}{\left(1 - e^{-\frac{\delta^2}{s_\varepsilon^2}}\right)^2}. \quad \text{(E.10)}$$

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

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

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

$$\leq 2\sum_{t=1}^{T} t \exp\left(-\frac{t\delta^2}{2s_\varepsilon^2}\right) \leq \frac{2\exp\left(-\frac{\delta^2}{2s_\varepsilon^2}\right)}{\left(1-\exp\left(-\frac{\delta^2}{2s_\varepsilon^2}\right)\right)^2} \leq \frac{2}{\left(1-\exp\left(-\frac{\delta^2}{2s_\varepsilon^2}\right)\right)^2}.$$

Wrap-up: bounding $\sum_n \mathbb{P}(F_n(a))$: Combining Equations (E.9), (E.10), and (E.9), and choosing

$$M = \frac{\delta^2}{2s_\varepsilon^2},$$

we finally obtain

$$\sum_{n=1}^{T} \mathbb{P}(F_n(a)) \leq \frac{e^{-\frac{\delta^2}{s_\varepsilon^2}}}{\left(1-\exp\left(-\frac{\delta^2}{s_\varepsilon^2}\right)\right)^2} + \frac{2}{\left(1-\exp\left(-\frac{\delta^2}{2s_\varepsilon^2}\right)\right)^2} \leq \frac{4}{\left(1-\exp\left(-\frac{\delta^2}{2s_\varepsilon^2}\right)\right)^2}. \quad \text{(E.11)}$$

E.4. Proof of Lemma 13: controlling large deviations of the kl (event $G_n(a)$)

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

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

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

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

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

Without loss of generality, by doing the change of variable $X \leftarrow X - m$, we assume 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 $\sim \mathcal{N}(m, 1)$ 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 characterization of a mixture of distributions, we have that X_i is equal in distribution to $(1 - W_i)Y_i + W_i O_i$. Hence,

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

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

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

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

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 $1 - \gamma$,

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

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

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

In this equation, there are two free parameters: λ and γ. Next, we choose λ so that

$$\frac{1}{n}\sum_{i=1}^{n}(1-W_i)\mathbb{1}\{Y_i \geq \lambda\} + \frac{1}{n}\sum_{i=1}^{n} W_i \leq \frac{1}{2}$$

with high probability. This choice of λ will then allow us to control the probability in Equation (E.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, we have

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

Now, suppose $L \leq 1$ and choose

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

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 (E.12)
From the choice of $\lambda = \frac{\Delta_{\min}}{2} + L$ from Equation (E.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.

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

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 9,

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

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

$$\mathrm{kl}_{\mathcal{G}}^{\varepsilon}\left(\mathrm{Med}(X_1^n) - \frac{\Delta_{\min}}{2}, m_a - \delta\right) \leq (y-\delta)_+ \max\left(y - \delta + \Delta_{\min}, \Delta_{\min}\right)$$
$$= (y-\delta)_+\left(|y-\delta| + \frac{\Delta_{\min}}{2}\right), \tag{E.16}$$

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 9,

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

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

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

E.7. Proof of Lemma 10

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

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

and

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

Of these, intuitively, the second event in Equation (E.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 (E.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 (E.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 (E.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

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

Next, observe that the event in Equation (E.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

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

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

E.8. Proof of Corollary 4

Taking the limsup in Theorem 2 yield the upper bound $\limsup_{T\to\infty} \frac{\mathbb{E}[N_a(T)]}{\log T} \leq \frac{1}{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(m(\mu_a), m^*(\mu))}$. Then, from this upper bound and the definition of regret, we have $\limsup_{T\to\infty} \frac{\mathbb{E}[R_T]}{\log T} \leq \sum_{a=1}^{K} \frac{\Delta_a}{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(m(\mu_a), m^*(\mu))}$. Similarly, applying the same reasoning with Proposition 2 we get $\liminf_{T\to\infty} \frac{\mathbb{E}[R_T]}{\log T} \leq \sum_{a=1}^{K} \frac{\Delta_a}{\mathrm{kl}^{\varepsilon}_{\mathcal{G}}(m(\mu_a), m^*(\mu))}$, which proves the conclusion of the corollary.

Appendix F. CRIMED for misspecified Gaussian model

In the main text, we assumed that the arm distributions followed a Gaussian distribution with unit variance and observations were corrupted. In this section, we explore a modification of CRIMED that demonstrates a strong numerical performance even in the presence of model misspecification (with corruption). Impacts of model misspecification in bandits has attracted increasing attention, where the existing studies focus mostly on the linear bandits and linear contextual bandits (Ghosh et al., 2017; Foster et al., 2020). In contrast, we study multi-armed stochastic bandits with misspecification about reward distributions being Gaussian.

We commence by detailing the misspecified model and the adaptation of CRIMED, providing a brief outline of its theoretical rationale. The section concludes with numerical analyses of the proposed algorithm.

F.1. Misspecified Gaussian model: Regret lower bound and algorithm design

Fix $\varepsilon > 0$, the fraction of observations that are corrupted. Recall that we place absolutely no assumptions on corruption distributions. In the current section, we consider misspecification of the Gaussian models. Fix $\varepsilon^{(m)} > 0$, —this denotes the fraction of samples that are misspecified. Fix $\delta > 0$, —this represents a bound on the difference of the mean of the Gaussian distribution and the misspecification distribution. Unlike for the corruption distributions, which are allowed to perturb the mean arbitrarily, we need this bound for the misspecification model. To be more specific, we consider the following perturbation of the Gaussian models for the arm distributions, with fixed and known $\varepsilon^{(m)}$, and δ.

$$\mathcal{I}^{\delta}_{\varepsilon^{(m)}} := \left\{(1-\varepsilon^{(m)})\mathcal{N}(x,1) + \varepsilon^{(m)}\eta^{(m)} : \; x \in \mathbb{R}, \; |x - m(\eta^{(m)})| \leq \delta, \; \eta^{(m)} \in \mathcal{P}(\mathbb{R})\right\},$$

where, recall that $\mathcal{P}(\mathbb{R})$ denotes the collection of all probability measures on $\mathbb{R}$. Here, $\eta^{(m)}$ is the misspecification distribution, which is restricted to have a mean close to that of the true Gaussian distribution, but otherwise can be arbitrary.

Bandit model. Each arm a in the bandit instance is associated with a distribution $\mu_a \in \mathcal{I}^{\delta}_{\varepsilon^{(m)}}$, which is a $(1-\varepsilon^{(m)}, \varepsilon^{(m)})$ mixture of $\mathcal{N}(m_a, 1)$ and $\eta^{(m)}_a$, i.e.,

$$\mu_a = (1-\varepsilon^{(m)})\mathcal{N}(m_a, 1) + \varepsilon^{(m)}\eta^{(m)}_a.$$

On pulling an arm a, the algorithm receives a reward which is an independent sample drawn from μ_a. However, as in the main text, with probability $1-\varepsilon$, it observes this independent sample, but with the remaining ε probability, it observes a sample drawn from an arbitrary corruption distribution. However, unlike in the main text, here the uncorrupted sample is not from a Gaussian distribution, but a mixture of a Gaussian and a misspecification distribution.

Next, consider the ε corruption neighbourhood of distributions in $\mathcal{I}^{\delta}_{\varepsilon^{(m)}}$, given by $\mathcal{C}^{\delta}_{\varepsilon,\varepsilon^{(m)}}$ below.

$$\mathcal{C}^{\delta}_{\varepsilon,\varepsilon^{(m)}} = \left\{\kappa \odot_{\varepsilon} H : \; \kappa \in \mathcal{I}^{\delta}_{\varepsilon^{(m)}}, \; H^{(m)} \in \mathcal{P}(\mathbb{R})\right\}.$$

For an arm a with distribution $\mu_a \in \mathcal{I}^{\delta}_{\varepsilon^{(m)}}$ and corruption distribution H, the observations from these arms are distributed as $\mu_a \odot_{\varepsilon} H \in \mathcal{C}^{\delta}_{\varepsilon,\varepsilon^{(m)}}$. This can be re-expressed as

$$(1-\varepsilon^{(m)})(1-\varepsilon)\mathcal{N}(m_a, 1) + \varepsilon^{(m)}(1-\varepsilon)\eta^{(m)}_a + \varepsilon H.$$

Regret. For this setting, $\bar{\Delta}_a := m^*(\mu) - m(\mu_a)$, where recall that $m^*(\mu)$ denotes the maximum mean of distributions in μ, and define $\Delta_a := \max_b(m_b - m_a)$, where m_a is the mean of the Gaussian distribution associated with arm a. Then, $\bar{\Delta}_a$ equals

$$\max_b \left\{(1-\varepsilon^{(m)})m_b + \varepsilon^{(m)} m(\eta_b^{(m)}) - (1-\varepsilon^{(m)})m_a - \varepsilon^{(m)} m(\eta_a^{(m)})\right\}.$$

Since η_a is assumed to satisfy $\left|m(\eta_a^{(m)}) - m_a\right| \leq \delta$, we have

$$\bar{\Delta}_a \leq \max_b \left\{m_b - m_a + 2\varepsilon^{(m)}\delta\right\} = \Delta_a + 2\varepsilon^{(m)}\delta.$$

The expected regret incurred by the algorithm in T trials can then be shown to satisfy

$$\mathbb{E}[R_T] = \sum_{a=1}^K \mathbb{E}[N_a(T)]\bar{\Delta}_a \leq \sum_{a=1}^K \mathbb{E}[N_a(T)]\left(\Delta_a + 2\varepsilon^{(m)}\delta\right).$$

Here, the expectation is with respect to the randomness in the algorithm, arm distributions, as well as the corruption distributions. As in Theorem 1, one can then obtain the following lower bound on regret for an appropriate definition of uniformly-good algorithms (that know both ε and $\varepsilon^{(m)}$):

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

where $\mathrm{KL}^\varepsilon_{\inf}$ is defined as earlier, and is given below for completeness. For $\eta \in \mathcal{P}(\mathbb{R})$, $x \in \mathbb{R}$,

$$\mathrm{KL}^\varepsilon_{\inf}\left(\mu_a, m^*(\mu); \mathcal{I}^\delta_{\varepsilon^{(m)}}\right) = \min\left\{\mathrm{KL}(\mu_a \odot_\varepsilon H, \kappa \odot_\varepsilon H') : \kappa \in \mathcal{I}^\delta_{\varepsilon^{(m)}}, m(\kappa) \geq x, H, H' \in \mathcal{P}(\mathbb{R})\right\}. \tag{F.1}$$

For $x \in \mathbb{R}$ and $y \geq x$, recall the definition of Gaussian $\mathrm{KL}^\varepsilon_{\inf}$, $\mathrm{kl}^\varepsilon_{\mathcal{G}}(x,y)$, from Equation (2.4). We now show that the $\mathrm{KL}^\varepsilon_{\inf}$ with respect to the misspecified model $\mathcal{I}^\delta_{\varepsilon^{(m)}}$ defined above, is lower bounded by that for a Gaussian class with a unit variance, with a blown-up corruption proportion.

Lemma 14 *Let* $\mu_a = (1-\varepsilon^{(m)})\mathcal{N}(m_a, 1) + \varepsilon^{(m)}\eta_a^{(m)} \in \mathcal{I}^\delta_{\varepsilon^{(m)}}$, *and* $\tilde{\varepsilon} := \varepsilon + \varepsilon^{(m)} - \varepsilon\varepsilon^{(m)}$. *Then*

$$\mathrm{KL}^\varepsilon_{\inf}\left(\mu_a, x; \mathcal{I}^\delta_{\varepsilon^{(m)}}\right) \geq \mathrm{kl}^{\tilde{\varepsilon}}_{\mathcal{G}}\left(m_a, x - \varepsilon^{(m)}\delta\right).$$

Proof Observe from Equation (F.1) that $\mathrm{KL}^\varepsilon_{\inf}(\mu_a, x; \mathcal{I}^\delta_{\varepsilon^{(m)}})$ equals

$$\min\Big\{\mathrm{KL}(((1-\varepsilon^{(m)})\mathcal{N}(m_a,1) + \varepsilon^{(m)}\eta_a^{(m)}) \odot_\varepsilon H, ((1-\varepsilon^{(m)})\mathcal{N}(y,1) + \varepsilon^{(m)}\kappa^{(m)}) \odot_\varepsilon H') : \\ (1-\varepsilon^{(m)})y + \varepsilon^{(m)} m(\kappa^{(m)}) \geq x, H, H', \kappa^{(m)} \in \mathcal{P}(\mathbb{R}), \left|y - m(\kappa^{(m)})\right| \leq \delta\Big\},$$

where the minimisation is over y, $\kappa^{(m)}$, H, and H'. The inequalities on $m(\kappa^{(m)})$ in the constraints above imply

$$y \geq x - \varepsilon^{(m)}\delta.$$

Using this in the constraints instead, and further optimising over $\eta_a^{(m)}$, $\mathrm{KL}_{\inf}^{\varepsilon}$ is lower bounded as below:

$$\min \Big\{ \mathrm{KL}(((1-\varepsilon^{(m)})\mathcal{N}(m_a,1) + \varepsilon^{(m)}\eta_a^{(m)}) \odot_\varepsilon H, ((1-\varepsilon^{(m)})\mathcal{N}(y,1) + \varepsilon^{(m)}\kappa^{(m)}) \odot_\varepsilon H') : \\ y \geq x - \varepsilon^{(m)}\delta, H, H', \kappa^{(m)}, \eta_a^{(m)} \in \mathcal{P}(\mathbb{R}) \Big\},$$

where the minimisation is over y, $\kappa^{(m)}$, $\eta_a^{(m)}$, H, and H'. Let $\tilde{\varepsilon} := \varepsilon + \varepsilon^{(m)} - \varepsilon\varepsilon^{(m)}$. Then, the lower bound obtained above equals,

$$\mathrm{KL}_{\inf}^{\tilde{\varepsilon}}(\mathcal{N}(m_a,1), x - \varepsilon^{(m)}\delta; \mathcal{G}),$$

which equals

$$\mathrm{kl}_{\mathcal{G}}^{\tilde{\varepsilon}}(m_a, x - \varepsilon^{(m)}\delta),$$

proving the desired bound. ■

We now present the modification of CRIMED for this setting. We do not use the knowledge of δ in algorithm design.

Algorithm. We increase the value of the parameter ε in CRIMED to

$$\tilde{\varepsilon} = \varepsilon + \tilde{\varepsilon}^{(m)} - \varepsilon\tilde{\varepsilon}^{(m)}$$

to encompass both corruption and the misspecification distributions as outliers (corruptions), i.e., we modify CRIMED to use $\tilde{\varepsilon}$ in place of ε everywhere (index as well as $N_{\min}$). We call CRIMED$^{(m)}$ the resulting algorithm (similarly CRIMED $*^{(m)}$).

Regret bound. Following the proof of Theorem 2, we get the following upper bound on regret of the modified algorithm. For $\mu \in \mathcal{I}^{\delta}_{\varepsilon^{(m)}}$ such that for each sub-optimal arm a, $\Delta_a - 2\varepsilon^{(m)}\delta \geq \Delta_{\min}$, where $\Delta_{\min}$ is the minimum gap (Definition 3) corresponding to $\tilde{\varepsilon}$, CRIMED$^{(m)}$ satisfies

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

where the arguments of $\mathrm{kl}_{\mathcal{G}}^{\tilde{\varepsilon}}$ are means of the Gaussian parts of the arm distributions. Further, recall that the condition on the misspecification distributions, gives the following:

$$\left| m_a - m(\eta_a^{(m)}) \right| \leq \delta \implies m(\mu_a) \geq m_a - \varepsilon^{(m)}\delta.$$

Since $\mathrm{kl}^{\tilde{\varepsilon}^{(m)}}$ is non-decreasing in its second argument (Lemma 3), we get the following upper bound on regret:

$$\lim_{T\to\infty} \frac{\mathbb{E}[N_a(T)]}{\log(T)} \leq \frac{1}{\mathrm{kl}_{\mathcal{G}}^{\tilde{\varepsilon}}\left(m_a, \max_b m_b\right)} \leq \frac{1}{\mathrm{kl}_{\mathcal{G}}^{\tilde{\varepsilon}}\left(m_a, m^*(\mu) - \varepsilon^{(m)}\delta\right)},$$

where $m^*(\mu)$ denotes the maximum mean of the arms in μ. This establishes a logarithmic regret for the misspecified setting. In the next section, we present some numerical results to justify the logarithmic bound.

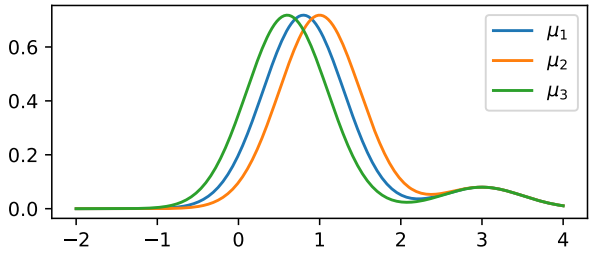


Figure 3: Reward distributions for arms in Settings 4 and 5.

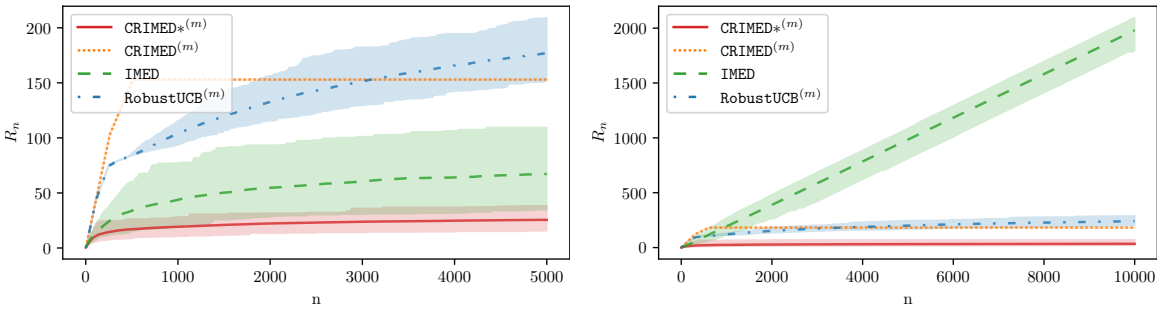


Figure 4: Cumulative regret for 100 repetitions on Settings 4 (left) and 5 (right). Solid lines represent the means and shaded area are 90% percentile intervals.

F.2. Experimental illustration

In this section, we present experiments for the misspecified setting. We consider two settings of bandit with 3 arms: Setting 4 and Setting 5. In Setting 4, there are no outliers, the law is misspecified Gaussian with means m_a having values $[0.6, 0.8, 1]$, and standard deviation 0.5; the misspecification distribution for each arm is Gaussian with means $[3, 3, 3]$ and standard deviation 0.5; the misspecification weight $\varepsilon^{(m)} = 0.1$. Plots of the three distributions can be found on Figure 3.

In Setting 5, in addition to model misspecification, we also have corruption. The arm distributions are the same as in Setting 4. In addition, the corruption proportion is set to $\varepsilon = 0.01$, with corruption distributions for each arm being Gaussian with means $[10, 10, -20]$, and standard deviation 1. The results are plotted in Figure 4.

In Figure 4 we see that `CRIMED` $*^{(m)}$ performs well in a misspecified setting. In particular, it is better than `IMED` which (mistakenly) considers a Gaussian model. This shows that using corruption to tackle model misspecification is worthwhile. As in experiments from Section 4, `CRIMED` $*^{(m)}$ is also better than `RobustUCB`, mainly due to the non-optimality of `RobustUCB`.