A Proof of Theorem 4.1

Consider a BAI instance $\tilde{\nu}$ with distributions and corresponding mean values $\{\mathsf{P}_i : i \in [K]\}$ and $\{\lambda_i : i \in [K]\}$, respectively, for the arms of this BAI instance. Given the adversarial model, we have the following two properties:

i. There exist some adversarial distributions $\{\mathbb{Q}_i : i \in [K]\}$, such that we have

$$\mathsf{P}_i \;=\; (1-\varepsilon)\mathbb{P}_i + \varepsilon\mathbb{Q}_i \quad \forall\, i \in [K]\,. \tag{22}$$

ii. The suboptimality gap of each arm in the BAI instance is equal to the suboptimality gap corresponding to the CBAI instance, i.e.,

$$\lambda_{a^\star} - \lambda_i \;=\; (\mu_{a^\star} - U_{a^\star}) - (\mu_i + U_i)\,, \quad \forall\, i \in [K]\,. \tag{23}$$

Such a choice of suboptimality gap is obtained by choosing $\lambda_{a^\star} = \mu_{a^\star} - U_{a^\star}$ and $\lambda_i = \mu_i + U_i$ for every $i \in [K] \setminus a^\star$.

We follow the same line of analysis as in [4, Theorem 18], which shows that any algorithm that is δ-PAC for a PIBAI instance is also δ-PAC for the counterpart BAI instance. This holds because (i) the samples for the BAI instance are drawn according to the same law as that of the PIBAI instance, and (ii) the suboptimality gaps in the PIBAI instance are the same as those of the BAI instance. Thus, any algorithm operating on the BAI instance requires at least as many samples before stopping as that required by the PIBAI instance. Thus, it is sufficient to find a lower bound on the BAI instance. In order to do so, we invoke [26, Lemma 2], which proves that

$$\lim_{\delta\to 0}\; \frac{\mathbb{E}_{\tilde{\nu}}[\tau]}{\log(1/\delta)} \;\geq\; T^\star(\lambda)\,, \tag{24}$$

where

$$[T^\star(\lambda)]^{-1} \;\triangleq\; \sup_{w\in\mathcal{Q}^K}\; \inf_{\zeta\in\mathsf{Alt}(\lambda)} \sum_{i\in[K]} w_a D_{\mathsf{KL}}(\lambda_i, \zeta_i)\,, \tag{25}$$

where $\mathcal{Q}^K$ denotes the K-dimensional probability simplex, and $\mathsf{Alt}(\lambda)$ denotes the set of bandit instances for which $a^\star$ is not the best arm, i.e., $\mathsf{Alt}(\lambda) \triangleq \{\nu \in \mathsf{SG}_K(\sigma) : a^\star(\nu) \cap a^\star(\lambda) = \emptyset\}$, where $a^\star(\nu)$ denotes the best arm of the bandit instance ν. Furthermore, restricting the class of bandits to Gaussian bandits, it is shown in [26] that we can obtain the following lower bound on $T^\star(\lambda)$

$$T^\star(\lambda) \;\geq\; \sum_{a\in[K]} \left(\frac{\sqrt{2}\sigma}{\max\{(\lambda_{a^\star} - \lambda_{b^\star}), (\lambda_{a^\star} - \lambda_a)\}}\right)^2\,, \tag{26}$$

which proves the desired result by noting that based on property (ii), for every $i \in [K]$ we have

$$\lambda_{a^\star} - \lambda_i \;=\; \Delta_i\,. \tag{27}$$

B Proof of Lemma 4.1

First, let us denote the set of uncontaminated rewards (drawn from the true measure $\mathbb{F}$) obtained up to time t by $\mathcal{G}_t$. Accordingly, denote the set of contaminated rewards drawn from the adversarial models up to time t by $\mathcal{C}_t$. Thus, we have that

$$\mathbf{R}^t \;=\; \mathcal{G}_t \cup \mathcal{C}_t\,, \tag{28}$$

where $\mathbf{R}^t$ is the set of all rewards obtained from the source. Furthermore, let us denote the set of removed rewards for estimation by $\mathcal{R}_t$, and the set of remaining rewards for estimation by $\mathcal{A}_t$. Now, we construct an interval around the true mean μ, denoted by

$$E \;\triangleq\; \left[\mu - \sigma\sqrt{2\log\frac{2}{\alpha}},\; \mu + \sigma\sqrt{2\log\frac{2}{\alpha}}\right]\,. \tag{29}$$

For any random variable X generated according to the true distribution $\mathbb{F}$, we have

$$\mathbb{P}(X \notin E) \;=\; \mathbb{P}\Bigg\{|X-\mu| > \sigma\sqrt{2\log\frac{2}{\alpha}}\Bigg\} \tag{30}$$

$$\leq\; \frac{\alpha}{2}\,, \tag{31}$$

where the inequality follows from X being σ-sub-Gaussian. Furthermore, define the sequence of rewards by $\{X_t : t \in \mathbb{N}\}$, and corresponding to the reward at time t, i.e. X_t, define the random variable $Y_i \;\triangleq\; \mathbb{1}_{\{X_t \notin E, X_t \sim \mathbb{F}\}}$. Then, we have

$$\mathbb{P}\Bigg\{\sum_{s=1}^{t} Y_s > \alpha t\Bigg\} = \;\mathbb{P}\Bigg\{\sum_{s=1}^{t} Y_s - t\mathbb{E}[Y_s] > \alpha t - t\mathbb{E}[Y_s]\Bigg\} \tag{32}$$

$$\leq\; \mathbb{P}\Bigg\{\sum_{s=1}^{t} Y_s - t\mathbb{E}[Y_s] > \frac{\alpha t}{2}\Bigg\} \tag{33}$$

$$\leq \exp\left(-\frac{t\alpha^2}{2}\right), \tag{34}$$

where the first inequality follows from (31), and the second inequality is a result of applying the Hoeffding's inequality. Furthermore, note that from resilience ρ[1] for σ-sub-Gaussian distributions [27], we have that

$$\Big|\mathbb{E}[X|\mathcal{G}_t \cap E] - \mu\Big| \;\leq\; \rho\,, \quad \text{where} \quad \rho \;=\; \mathcal{O}\left(\sigma\alpha\sqrt{\log\frac{1}{\alpha}}\right). \tag{35}$$

Additionally, we know that due to the σ-sub-Gaussian assumption, for any set $\mathcal{A}$ of samples from distribution $\mathbb{F}$,

$$\mathbb{P}\left\{\left|\frac{1}{|\mathcal{A}|}\sum_{x\in\mathcal{A}} x - \mu\right| > \sigma\sqrt{\frac{2}{|\mathcal{A}|}\log\frac{1}{\delta}}\right\} \;\leq\; \delta\,. \tag{36}$$

Thus, combining (35) and (36), we obtain

$$\mathbb{P}\left\{\left|\frac{1}{|\mathcal{G}_t \cap E|}\sum_{x\in\mathcal{G}_t\cap E} x - \mu\right| > \mathcal{O}\left(\sigma\alpha\sqrt{\log\frac{1}{\alpha}}\right) + \sigma\sqrt{\frac{2}{|\mathcal{G}_t\cap E|}\log\frac{1}{\delta}}\right\} \;\leq\; \delta\,. \tag{37}$$

Furthermore, since $|E \cap \mathcal{G}_t| \leq t$, from (37) we obtain

$$\mathbb{P}\left[\left|\sum_{x\in\mathcal{G}_t\cap E}(x-\mu)\right| > t\left\{\mathcal{O}\left(\sigma\alpha\sqrt{\log\frac{1}{\alpha}}\right) + \sigma\sqrt{\frac{2}{t}\log\frac{1}{\delta}}\right\}\right] \;\leq\; \delta\,. \tag{38}$$

Now, let us define the event

$$\mathcal{S}_t \;\triangleq\; \left\{\sum_{s=1}^{t} Y_s > \alpha t\right\}. \tag{39}$$

It can be readily verified that, conditioned on $\bar{\mathcal{S}}_t$, defined as the complement of $\mathcal{S}_t$, all the elements of $\mathcal{A}_t$ fall in the interval E. This is because of the fact that under the event $\bar{\mathcal{S}}_t$, the number of points drawn from the true distribution $\mathbb{F}$ until time t that belong to the interval E is at least $t(1-\alpha)$, whereas for constructing $\mathcal{A}_t$, we trim $2\alpha t$ points from the sequence of samples obtained up to t. Furthermore, for every $t > T(\alpha,\delta)$, we have $\mathbb{P}(\mathcal{S}_t) < \delta$, where we have defined

$$T(\alpha,\delta) \;\triangleq\; \frac{2}{\alpha^2}\log\frac{1}{\delta}\,. \tag{40}$$

[1]A distribution $\mathbb{P}$ over $\mathbb{R}^d$ is called (ρ,α)-resilient (with respect to some norm $\|\cdot\|$) if $\|\mathbb{E}_{X\sim\mathbb{P}}[X \mid E] - \mathbb{E}_{X\sim\mathbb{P}}[X]\| \leq \rho$ for all events E with $\mathbb{P}(E) \geq 1-\alpha$.

Thus, for all $t > T(\alpha, \delta)$, we have

$$\mathbb{P}\left\{\left|\sum_{x\in\mathcal{A}_t}(x-\mu)\right| > \left|\sum_{x\in\mathcal{A}_t\cap E\cap\mathcal{G}_t}(x-\mu)\right| + \left|\sum_{x\in\mathcal{A}_t\cap E\cap\mathcal{C}_t}(x-\mu)\right|\right\}$$

$$= \mathbb{P}\left\{\left|\sum_{x\in\mathcal{A}_t}(x-\mu)\right| > \left|\sum_{x\in\mathcal{A}_t\cap E\cap\mathcal{G}_t}(x-\mu)\right| + \left|\sum_{x\in\mathcal{A}_t\cap E\cap\mathcal{C}_t}(x-\mu)\right| \,\middle|\, \mathcal{S}_t\right\} \times \mathbb{P}(\mathcal{S}_t)$$

$$+ \mathbb{P}\left\{\left|\sum_{x\in\mathcal{A}_t}(x-\mu)\right| > \left|\sum_{x\in\mathcal{A}_t\cap E\cap\mathcal{G}_t}(x-\mu)\right| + \left|\sum_{x\in\mathcal{A}_t\cap E\cap\mathcal{C}_t}(x-\mu)\right| \,\middle|\, \bar{\mathcal{S}}_t\right\} \times \mathbb{P}(\bar{\mathcal{S}}_t) \quad (41)$$

$$< \delta\,, \quad (42)$$

where the last inequality is a consequence of the fact that $\mathbb{P}(\mathcal{S}_t) < \delta$. Note that

$$\left|\sum_{x\in\mathcal{A}_t\cap E\cap\mathcal{G}_t}(x-\mu)\right| \le \left|\sum_{x\in E\cap\mathcal{G}_t}(x-\mu)\right| + \left|\sum_{x\in\mathcal{R}_t\cap E\cap\mathcal{G}_t}(x-\mu)\right|. \quad (43)$$

Using (38), in conjunction with the first term on the right hand side of (43) we have

$$\mathbb{P}\left[\left|\sum_{x\in E\cap\mathcal{G}_t}(x-\mu)\right| \le t\left\{\mathcal{O}\left(\sigma\alpha\sqrt{\log\frac{1}{\alpha}}\right) + \sigma\sqrt{\frac{2}{t}\log\frac{1}{\delta}}\right\}\right] \ge 1-\delta\,. \quad (44)$$

Furthermore, we have

$$\left|\sum_{x\in\mathcal{R}_t\cap E\cap\mathcal{G}_t}(x-\mu)\right| \le \sigma|E\cap\mathcal{G}_t\cap\mathcal{R}_t|\sqrt{2\log\frac{2}{\alpha}} \le t\sigma\alpha\sqrt{2\log\frac{2}{\alpha}}\,, \quad (45)$$

based on the definition of interval E, and

$$\left|\sum_{x\in\mathcal{A}_t\cap E\cap\mathcal{C}_t}(x-\mu)\right| \le 2|\mathcal{C}_t|\sigma\sqrt{\log\frac{2}{\alpha}} \le 2\sigma\varepsilon t\sqrt{\log\frac{2}{\varepsilon}}\,. \quad (46)$$

Now, let us define the event $\mathcal{L}_t$ as

$$\mathcal{L}_t \triangleq \left\{\left|\sum_{x\in\mathcal{G}_t\cap E}(x-\mu)\right| \le t\left[\mathcal{O}\left(\sigma\alpha\sqrt{\log\frac{1}{\alpha}}\right) + \sigma\sqrt{\frac{2}{t}\log\frac{1}{\delta}}\right]\right\}. \quad (47)$$

From (42), we have

$$\mathbb{P}\left\{\left|\sum_{x\in\mathcal{A}_t}(x-\mu)\right| \le \left|\sum_{x\in\mathcal{A}_t\cap E\cap\mathcal{G}_t}(x-\mu)\right| + \left|\sum_{x\in\mathcal{A}_t\cap E\cap C_t}(x-\mu)\right|\right\} \ge 1-\delta\,. \quad (48)$$

Thus,

$$1-\delta \le \mathbb{P}\left\{\left|\sum_{x\in\mathcal{A}_t}(x-\mu)\right| \le \left|\sum_{x\in\mathcal{A}_t\cap E\cap\mathcal{G}_t}(x-\mu)\right| + \left|\sum_{x\in\mathcal{A}_t\cap E\cap C_t}(x-\mu)\right| \,\middle|\, \mathcal{L}_t\right\} \times \mathbb{P}(\mathcal{L}_t)$$

$$+ \quad \mathbb{P}\left\{\left|\sum_{x\in\mathcal{A}_t}(x-\mu)\right| \le \left|\sum_{x\in\mathcal{A}_t\cap E\cap\mathcal{G}_t}(x-\mu)\right| + \left|\sum_{x\in\mathcal{A}_t\cap E\cap C_t}(x-\mu)\right| \,\middle|\, \bar{\mathcal{L}}_t\right\} \times \mathbb{P}(\bar{\mathcal{L}}_t) \quad (49)$$

$$\le \mathbb{P}\left\{\left|\sum_{x\in\mathcal{A}_t}(x-\mu)\right| \le tC_1\sigma\alpha\sqrt{\log\frac{1}{\alpha}} + t\sigma\sqrt{\frac{2}{t}\log\frac{1}{\delta}} + 2\sigma\varepsilon t\sqrt{\log\frac{2}{\varepsilon}}\right.$$

$$\left. + t\sigma\alpha\sqrt{2\log\frac{2}{\alpha}}\right\} + \delta\,, \quad (50)$$

where (50) is a result of (44), (45), and (46). Further, let us set $\alpha = \varepsilon/2$. Thus, using this in conjunction with (50), there exists a constant $C_1 \in \mathbb{R}_+$, such that for all $t > T(\alpha, \delta)$, we have

$$\mathbb{P}\left\{\left|\sum_{x\in\mathcal{A}_t}(x-\mu)\right| \le tC_1\frac{\varepsilon}{2}\sigma\sqrt{\log\frac{2}{\varepsilon}} + 2\sigma\varepsilon t\sqrt{\log\frac{2}{\varepsilon}} + t\sigma\frac{\varepsilon}{2}\sqrt{2\log\frac{4}{\varepsilon}} + \sigma t\sqrt{\frac{2}{t}\log\frac{1}{\delta}}\right\} \ge 1-2\delta\,. \tag{51}$$

Furthermore, dividing the term inside (51) throughout by $t(1-\varepsilon)$, we obtain that

$$\begin{aligned}
1-2\delta \;\le\; & \mathbb{P}\left\{\left|\hat{\mu}_t-\mu\right| \le \frac{C_1\sigma\varepsilon}{2(1-\varepsilon)}\sqrt{\log\frac{2}{\varepsilon}} + \frac{\varepsilon\sigma}{2(1-\varepsilon)}\sqrt{2\log\frac{4}{\varepsilon}} + \frac{2\sigma\varepsilon}{1-\varepsilon}\sqrt{\log\frac{2}{\varepsilon}} \right.\\
& \qquad \left. + \frac{\sigma}{1-\varepsilon}\sqrt{\frac{2}{t}\log\frac{1}{\delta}}\right\} && (52)\\
\le\; & \mathbb{P}\left\{\left|\hat{\mu}_t-\mu\right| \le C_1\sigma\varepsilon\sqrt{\log\frac{2}{\varepsilon}} + \varepsilon\sigma\sqrt{2\log\frac{4}{\varepsilon}} + 4\sigma\varepsilon\sqrt{\log\frac{2}{\varepsilon}} + \frac{\sigma}{1-\varepsilon}\sqrt{\frac{2}{t}\log\frac{1}{\delta}}\right\} && (53)\\
\le\; & \mathbb{P}\left\{\left|\hat{\mu}_t-\mu\right| \le \mathcal{O}\left(\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right) + \frac{\sigma}{1-\varepsilon}\sqrt{\frac{2}{t}\log\frac{1}{\delta}}\right\}, && (54)
\end{aligned}$$

where (53) is a result of our assumption that $\varepsilon < 1/2$. Finally, replacing δ with $\delta/2$ in (54), we obtain the desired result.

C Proof of Theorem 4.2

Let us begin by defining the event

$$\begin{aligned}
\mathcal{E} \;\triangleq\; \Bigg\{\forall t > T(\alpha,\delta),\ \forall i \in [K]\setminus a^\star \;:\; & \left|\hat{\mu}_i(t)-\mu_i\right| \le U_i + \beta_i(t,\delta) \quad \text{and} \\
& \left|\hat{\mu}_{a^\star}(t)-\mu_{a^\star}\right| \le U_{a^\star} + \beta_{a^\star}(t,\delta)\Bigg\}, \qquad (55)
\end{aligned}$$

where we have defined

$$\beta_i(t,\delta) \;\triangleq\; \frac{\sigma}{1-\varepsilon}\sqrt{\frac{2}{N_i(t)}\log\frac{(K-1)Ct^\beta}{\delta}}\,. \tag{56}$$

Now, noting that we have defined the maximum overlap in confidence intervals between the best arm $a^\star$ and the most ambiguous arm j_t as B_t in (7), for all $\tau > T(\alpha,\delta)$, we have that

$$\begin{aligned}
& \mathbb{P}\left\{\mu_{\hat{a}_\tau} + U_{\hat{a}_\tau} < \mu_{a^\star} - U_{a^\star}\right\}\\
& \le\; \mathbb{P}\left\{(\mu_{a^\star}-U_{a^\star}) - (\mu_{\hat{a}_\tau}+U_{\hat{a}_\tau}) > B_\tau\right\} && (57)\\
& \le\; \mathbb{P}\left\{(\mu_{a^\star}-U_{a^\star}) - (\mu_{\hat{a}_\tau}+U_{\hat{a}_\tau}) > \hat{\mu}_{a^\star_\tau}(\tau) + \beta_{a^\star}(\tau,\delta) - \left(\hat{\mu}_{\hat{a}_\tau}(\tau) - \beta_{\hat{a}_\tau}(\tau,\delta)\right)\right\} && (58)\\
& =\; \mathbb{P}\left\{(\mu_{a^\star}-\hat{\mu}_{a^\star_\tau}(\tau)) - (\mu_{\hat{a}_\tau}-\hat{\mu}_{\hat{a}_\tau}(\tau)) > \beta_{a^\star}(\tau,\delta) + \beta_{\hat{a}_\tau}(\tau,\delta) + U_{\hat{a}_\tau} + U_{a^\star} \;\middle|\; \mathcal{E}\right\}\mathbb{P}(\mathcal{E})\\
& \quad +\; \mathbb{P}\left\{(\mu_{a^\star}-\hat{\mu}_{a^\star_\tau}(\tau)) - (\mu_{\hat{a}_\tau}-\hat{\mu}_{\hat{a}_\tau}(\tau)) > \beta_{a^\star}(\tau,\delta) + \beta_{\hat{a}_\tau}(\tau,\delta) + U_{\hat{a}_\tau} + U_{a^\star} \;\middle|\; \bar{\mathcal{E}}\right\}\mathbb{P}(\bar{\mathcal{E}}) && (59)\\
& \le\; \mathbb{P}(\bar{\mathcal{E}})\,, && (60)
\end{aligned}$$

where the first inequality is a result of the stopping criterion and the second inequality follows from the definition of the overlap B_t in (7). Hence, we have

$$\mathbb{P}(\bar{\mathcal{E}}) \;=\; \mathbb{P}\Bigg\{\exists t > T(\alpha,\delta),\; \exists a \in [K] \setminus a^\star,\; \Big|\hat{\mu}_a(t) - \mu_a\Big| > U_a + \beta_a(t,\delta) \quad \text{or} \\ \Big|\hat{\mu}_{a^\star}(t) - \mu_{a^\star}\Big| > U_{a^\star} + \beta_{a^\star}(t,\delta)\Bigg\} \tag{61}$$

$$\leq \sum_{a\in[K]\setminus a^\star} \sum_{t=1}^{\infty} \mathbb{P}\Bigg\{\Big|\hat{\mu}_a(t) - \mu_a\Big| > U_a + \beta_a(t,\delta)\Bigg\} + \mathbb{P}\Bigg\{\Big|\hat{\mu}_{a^\star}(t) - \mu_{a^\star}\Big| > U_{a^\star} + \beta_{a^\star}(t,\delta)\Bigg\} \tag{62}$$

$$\leq \sum_{a\in[K]\setminus a^\star} \sum_{t=1}^{\infty} \frac{\delta}{(K-1)Ct^\beta} \tag{63}$$

where (63) is a result of Lemma 4.1. If we choose C such that

$$C \;\geq\; \sum_{t=1}^{\infty} \frac{1}{t^\beta}\,, \tag{64}$$

then we obtain that

$$\mathbb{P}(\bar{\mathcal{E}}) \;\leq\; \delta\,. \tag{65}$$

Note that a choice of C always exists for any $\beta > 1$, since we have that

$$\sum_{t=1}^{\infty} \frac{1}{t^\beta} \;\leq 1 + \int_1^{\infty} \frac{1}{t^\beta}\,\mathrm{d}t \;=\; 1 + (\beta-1)^{-1}\,. \tag{66}$$

Furthermore, note that by the design of our stopping rule, τ is always greater than $T(\alpha,\delta)$. Thus, choosing $C = (1 + (\beta-1)^{-1})$ ensures that (65) holds. This completes the proof.

D Proof of Theorem 4.3

Define T as the first time such that $\hat{a}_t = a^\star$ for every $t \geq T$. We have

$$\mathbb{P}(T > t) \;=\; \sum_{s=t}^{\infty} \mathbb{P}(\hat{a}_s \neq a^\star, \hat{a}_u = a^\star,\; \forall u > s) \tag{67}$$

$$\leq \sum_{s=t}^{\infty} \mathbb{P}(\hat{a}_s \neq a^\star) \tag{68}$$

$$= \sum_{s=t}^{\infty} \mathbb{P}(\hat{\mu}_{\hat{a}_s}(s) > \hat{\mu}_{a^\star}(s)) \tag{69}$$

$$\leq \sum_{s=t}^{\infty} \sum_{i\in[K]\setminus a^\star} \mathbb{P}(\hat{\mu}_i(s) > \hat{\mu}_{a^\star}(s))\,. \tag{70}$$

Furthermore, by the concentration of the α-trimmed mean estimator in Lemma 4.1, for every $i \in [K]$ and for all $t > T(\alpha,\delta)$, we have

$$\mathbb{P}\Bigg\{|\hat{\mu}_i(t) - \mu_i| > U_i + \Delta_i/2\Bigg\} \;\leq\; \exp\Bigg\{-\frac{\Delta_i^2(1-\varepsilon)^2 N_i(t)}{8\sigma^2}\Bigg\}\,, \tag{71}$$

where we have defined

$$\Delta_i \;\triangleq\; (\mu_{a^\star} - U_{a^\star}) - (\mu_i + U_i)\,. \tag{72}$$

Let us define the event

$$\mathcal{H}(t) \;\triangleq\; \Bigg\{\forall i \in [K] :\; |\hat{\mu}_i(t) - \mu_i| > U_i + \frac{\Delta_i}{2}\Bigg\}\,. \tag{73}$$

Thus, for all $t > T(\alpha, \delta)$, we have

$$\begin{aligned}\mathbb{P}(\hat{\mu}_i(t) > \hat{\mu}_{a^\star}(t)) &= \mathbb{P}\{\hat{\mu}_i(t) > \hat{\mu}_{a^\star}(t),\ \mathcal{H}(t)\} \\ &\quad + \mathbb{P}\{\hat{\mu}_i(t) > \hat{\mu}_{a^\star}(t),\ \bar{\mathcal{H}}(t)\} \qquad (74)\\ &\leq \sum_{i\in[K]} \exp\left\{ - \frac{\Delta_i^2 (1-\varepsilon)^2 N_i(t)}{8\sigma^2} \right\}, \qquad (75)\end{aligned}$$

where the inequality is a result of (71) followed by a union bound. Furthermore, by the sampling strategy of our algorithm, we have $N_i(t) > \sqrt{t}$ for every $i \in [K]$. Thus, combining (70) and (75), for all $t > T(\alpha, \delta)$, we have

$$\begin{aligned}\mathbb{P}(T > t) &\leq \sum_{s=t}^{\infty} \sum_{i\in[K]\setminus a^\star} \sum_{i\in[K]} \exp\left\{ - \frac{\Delta_i^2 (1-\varepsilon)^2}{8\sigma^2} \sqrt{t} \right\} \qquad (76)\\ &\leq K^2 \int_{t-1}^{\infty} \exp\Big(- M\sqrt{s} \Big)\, \mathrm{d}s \qquad (77)\\ &= \frac{2K^2}{M^2} \Big(M\sqrt{t-1} + 1 \Big) \exp\Big(- M\sqrt{t-1} \Big), \qquad (78)\end{aligned}$$

where we have set

$$M \triangleq \frac{\Delta_{b^\star}^2 (1-\varepsilon)^2}{8\sigma^2}. \qquad (79)$$

Now, under the event that $\{T \leq t\}$ and the event $\mathcal{E}$ defined in (55), for all $t > T(\alpha, \delta)$ with probability at least $1 - \delta$, we have

$$\begin{aligned}B_t &= \hat{\mu}_{j_t}(t) + \beta_{j_t}(t,\delta) - (\hat{\mu}_{\hat{a}_t}(t) - \beta_{\hat{a}_t}(t,\delta)) \qquad (80)\\ &\leq (\mu_{j_t} + U_{j_t}) - (\mu_{\hat{a}_t} - U_{\hat{a}_t}) + 2(\beta_{j_t}(t,\delta) + \beta_{\hat{a}_t}(t,\delta)) \qquad (81)\\ &= -\Delta_{j_t} - ((\mu_{\hat{a}_t} - U_{\hat{a}_t}) - (\mu_{a^\star} - U_{a^\star})) + 2(\beta_{j_t}(t,\delta) + \beta_{\hat{a}_t}(t,\delta)) \qquad (82)\\ &\leq -\max(\Delta_{A_t}, \Delta_{b^\star}) + 4\beta_{A_t}(t,\delta), \qquad (83)\end{aligned}$$

where the first inequality is obtained due to the event $\mathcal{E}$ and the last inequality is a result of the fact that $T \leq t$ combined with the arm selection strategy. Furthermore, note that our sampling strategy and stopping rule ensure that $N_i(t) > T(\alpha, \delta)$ for every $i \in [K]$. Let $t_i \in \mathbb{N}$ denote the last time that arm $i \in [K]$ is pulled before stopping. Then, as a consequence of the stopping criterion, (83), and along with the choice of the confidence intervals

$$\beta_i(t) \triangleq \frac{\sigma}{(1-\varepsilon)} \sqrt{\frac{2}{N_i(t)} \log \frac{(K-1)Ct^\beta}{\delta}}, \qquad (84)$$

we obtain

$$\mathbb{P}\left\{ N_i(t_i) \leq \log \frac{(K-1)Ct_i^\beta}{\delta} \cdot \frac{32\sigma^2}{(1-\varepsilon)^2 \max\{\Delta_{b^\star}, \Delta_i\}^2} \right\} > 1 - \delta, \qquad (85)$$

which yields

$$\mathbb{P}\left\{ N_i(\tau) \leq \log \frac{(K-1)C\tau^\beta}{\delta} \cdot \frac{32\sigma^2}{(1-\varepsilon)^2 \max\{\Delta_{b^\star}, \Delta_i\}^2} + 1 \right\} > 1 - \delta. \qquad (86)$$

Now, taking the limit of $\delta \to 0$, if $N_i(\tau) \geq T$,

$$\begin{aligned}&\lim_{\delta\to 0} \mathbb{P}\left\{ N_i(\tau) \leq \log \frac{(K-1)C\tau^\beta}{\delta} \cdot \frac{32\sigma^2}{(1-\varepsilon)^2 \max\{\Delta_{b^\star}, \Delta_i\}^2} + 1 \right\} \qquad (87)\\ &= \mathbb{P}\left\{ \lim_{\delta\to 0} N_i(\tau) \leq \log \frac{(K-1)C\tau^\beta}{\delta} \cdot \frac{32\sigma^2}{(1-\varepsilon)^2 \max\{\Delta_{b^\star}, \Delta_i\}^2} + 1 \right\} \qquad (88)\\ &= 1, \qquad (89)\end{aligned}$$

where the transition from (87) to (88) is a result of the monotone convergence theorem. Next, note that for any $i \in [K]$, we have

$$N_i(\tau) \;=\; N_i(\tau)\mathbb{1}_{\{N_i(\tau)<T\}} + N_i(\tau)\mathbb{1}_{\{N_i(\tau)\geq T\}}\,. \tag{90}$$

Thus, in the limit of $\delta \to 0$, from (90), we almost surely have

$$N_i(\tau) \;\leq\; T + \log\frac{(K-1)C\tau^\beta}{\delta}\cdot\frac{32\sigma^2}{(1-\varepsilon)^2\max\{\Delta_{b^\star},\Delta_i\}^2} + 1\,. \tag{91}$$

Furthermore, from the fact that $\tau = \sum_{i\in[K]} N_i(\tau)$, in the limit of $\delta \to 0$ we almost surely have

$$\tau \;\leq\; KT + \frac{16H}{(1-\varepsilon)^2}\log\frac{(K-1)C\tau^\beta}{\delta} + K\,. \tag{92}$$

Since $f(x) = x - \frac{1}{C_1}\log C_2 x^\alpha$ is a monotonically increasing function in x, there exists $x_{\max}$ such that for all $x \geq x_{\max}$, we have $f(x) \geq 0$. Next, we will find a choice $\bar{x}$ such that $f(\bar{x}) \geq 0$. This implies that $\bar{x} \geq x_{\max}$. To this end, we use [[26], Lemma 18], which states that

Lemma D.1 ([26], Lemma 18)**.** *For every $\alpha \in [1, e/2]$ and any two constants $C_1, C_2 > 0$ the identity*

$$x \;=\; \frac{\alpha}{C_1}\left[\log\left(\frac{C_2 e}{C_1^\alpha}\right) + \log\log\left(\frac{C_2}{C_1^\alpha}\right)\right] \tag{93}$$

indicates that $C_1 x \geq \log(C_2 x^\alpha)$.

In the above lemma, by choosing $C_1 = \frac{(1-\varepsilon)^2}{16H}$ and $C_2 = \frac{(K-1)C\exp(K(T+1)(1-\varepsilon)^2/16H)}{\delta}$, in the limit of $\delta \to 0$, we almost surely have

$$\begin{aligned}\tau \;\leq \frac{16\beta H}{(1-\varepsilon)^2}\Bigg[&\log\frac{(K-1)Ce(16H/(1-\varepsilon)^2)^\beta}{\delta} + \log\log\frac{(K-1)C(16H/(1-\varepsilon)^2)^\beta}{\delta}\\ &+ (K+TK) + \log(K+TK)\Bigg]\,.\end{aligned} \tag{94}$$

Thus, taking expectation on both sides of the above inequality, we have

$$\begin{aligned}\lim_{\delta\to 0}\frac{\mathbb{E}[\tau]}{\log(1/\delta)} \;\leq \lim_{\delta\to 0}\;&\frac{16\beta H\log\frac{(K-1)Ce(16H/(1-\varepsilon)^2)^\beta}{\delta}}{(1-\varepsilon)^2\log(1/\delta)} + \frac{16\beta H\log\log\frac{(K-1)C(16H/(1-\varepsilon)^2)^\beta}{\delta}}{(1-\varepsilon)^2\log(1/\delta)}\\ &+ \lim_{\delta\to 0}\;\frac{K+K\mathbb{E}[T]}{\log(1/\delta)} + \frac{\mathbb{E}[\log(K+TK)]}{\log(1/\delta)}\,.\end{aligned} \tag{95}$$

Next, by recalling the definition of M in (79), note that

$$\begin{aligned}\mathbb{E}[T] \;&=\; \sum_{t=1}^{\infty}\mathbb{P}(T\geq t) && (96)\\ &\leq\; \frac{2K^2}{M^2}\left\{1 + \lim_{x\to\infty}\int_0^x (M\sqrt{t}+1)\exp(-M\sqrt{t})\,\mathrm{d}t\right\} && (97)\\ &\leq\; \frac{2K^2}{M^2}\left(1+\frac{6}{M^2}\right) && (98)\\ &<\; +\infty\,, && (99)\end{aligned}$$

where the first inequality follows from (76). Thus, combing (95) and (99), we obtain

$$\lim_{\delta\to 0}\frac{\mathbb{E}[\tau]}{\log(1/\delta)} \;\leq \frac{16\beta H}{(1-\varepsilon)^2}\,. \tag{100}$$

Finally, using the fact that $\varepsilon < 1/2$, we obtain the desired result. Furthermore, it can be readily verified that the first part of the maximum operation in Theorem 4.3 is a direct consequence of the stopping rule by choosing $\alpha = \varepsilon/2 < 1/4$. This completes the proof.

E Proof of Theorem 4.4

Based on the estimator concentration in Lemma 4.1, for every $t > T(\alpha, \delta)$ and for any arm $i \in [K]$, we have

$$\mathbb{P}\left\{ \left| \hat{\mu}_i(t) - \mu \right| > U_i + \frac{\sigma}{(1-\varepsilon)} \sqrt{\frac{2}{t} \log \frac{Kt^2\pi^2}{12\delta}} \right\} \;\leq\; \frac{6\delta}{Kt^2\pi^2} \,. \tag{101}$$

Earlier, we had defined $\mathcal{M}_t$ as the set of active arms, which were not yet been eliminated by the SE-CBAI algorithms at time t (Algorithm 2, line 4). Based on that, let us define the event $\mathcal{E}$ such that

$$\mathcal{F} \;\triangleq\; \left\{ \left| \hat{\mu}_i(t) - \mu_i \right| \leq U_i + \gamma_t,\; \forall t \geq 1,\; \forall i \in \mathcal{M}_t \right\} . \tag{102}$$

We obtain

$$\begin{aligned}
\mathbb{P}(\bar{\mathcal{F}}) \;&\leq\; \sum_{i \in [K]} \sum_{t=1}^{\infty} \mathbb{P}\Big\{ \left| \hat{\mu}_i(t) - \mu_i \right| \leq U_i + \gamma_t \Big\} && (103)\\
&\leq\; \sum_{i \in [K]} \sum_{t=1}^{\infty} \frac{6\delta}{Kt^2\pi^2} && (104)\\
&\leq\; \delta \,, && (105)
\end{aligned}$$

where the second inequality is a consequence of (101) and the last inequality holds due to the Basel identity. Event $\mathcal{F}$ implies that with probability at least $1-\delta$, for all t and for every $j \in \mathcal{M}_t$, we have

$$\begin{aligned}
\hat{\mu}_{a^\star}(t) \;&\geq\; \mu_{a^\star} - U_{a^\star} - \gamma_t && (106)\\
&=\; \Delta_j + (\mu_j + U_j) - \gamma_t && (107)\\
&\geq\; \mu_j + U_j - \gamma_t && (108)\\
&\geq\; \hat{\mu}_j(t) - 2\gamma_t \,. && (109)
\end{aligned}$$

This proves that the best arm $a^\star$ is contained in $\mathcal{M}_t$ with probability at least $1-\delta$ for every $t > T(\alpha, \delta)$. Finally, by the choice of our stopping rule τ, we have $\tau > T(\alpha, \delta)$. This completes the proof.

F Proof of Theorem 4.5

First, note that due to the successive elimination strategy, we have

$$\tau \;\leq\; 2 \sum_{i \in [K] \setminus a^\star} N_i(\tau) \,. \tag{110}$$

Furthermore, by the choice of the active set $\mathcal{M}_t$ defined in Algorithm 2 line 4, any arm $i \in [K] \setminus a^\star$ is eliminated no later than the time t such that

$$\hat{\mu}_i(t) \;<\; \hat{\mu}_{a^\star}(t) - 2\gamma_t \,. \tag{111}$$

Combining (110) with the event $\mathcal{F}$ defined in (102), with probability at least $1-\delta$ for all $t > T(\alpha, \delta)$, we have

$$\hat{\mu}_i(t) \;<\; \mu_{a^\star} - U_{a^\star} - 3\gamma_t \,, \quad \forall\, i \in [K] \setminus a^\star \,. \tag{112}$$

This indicates that for $t > T(\alpha, \delta)$ with probability at least $1-\delta$, we have

$$\mu_i + U_i + \gamma_t \;<\; \mu_{a^\star} - U_{a^\star} - 3\gamma_t \,, \quad \forall\, i \in [K] \setminus a^\star \,, \tag{113}$$

which, in turn, indicates that

$$\mathbb{P}\Big(\Delta_i \;>\; 4\gamma_t \Big) \geq 1 - \delta \,, \quad \forall\, i \in [K] \setminus a^\star \,. \tag{114}$$

The above inequality holds with equality by setting γ_t as

$$\gamma_t \;\triangleq\; \frac{\sigma}{(1-\varepsilon)} \sqrt{\frac{2}{t} \log \frac{Kt^2\pi^2}{12\delta}} \,. \tag{115}$$

Hence, for some universal constant $L > 0$, we have

$$N_i(\tau) \;\leq\; \frac{L\sigma^2}{\Delta_i^2(1-\varepsilon)^2} \log \frac{K}{\delta \Delta_i} \;, \quad \forall\, i \in [K] \setminus a^\star \;. \tag{116}$$

Finally, combining (110) and (116), in conjunction with the fact that from the sampling rule we know $N_i(\tau) > T(\alpha, \delta)$, we find that with probability at least $1 - \delta$, we have

$$\tau \;\leq\; \max \left\{ 32K \log \frac{1}{\delta},\; \mathcal{O}\left(\sum_{i \in [K] \setminus a^\star} \frac{1}{\Delta_i^2} \log \frac{K}{\delta \Delta_i} \right) \right\} , \tag{117}$$

where we have used $\varepsilon < 1/2$. This completes the proof.

G Experimental Details

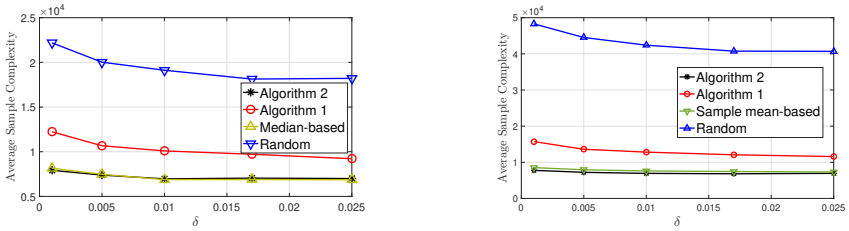
Experiments with real data. In this section, we provide the details of the experiments with real data. Specifically, we use two real-world datasets, one of which considers the application of content recommendation, and the other considers the applications of drug discovery. For content recommendation, we use the New Yorker Caption Contest dataset, and for drug discovery, we use the PKIS2 datset. Each experiment is averaged over 1000 Monte Carlo trials. For each experiment, the adversarial distribution is assumed to have a uniform distribution with a randomly generated mean such that the index of the best arm does not change as a result of corruption.

New Yorker Caption Contest: This repository contains data gathered from the cartoon caption contest, in which users are asked to write captions for a given cartoon. The dataset is constructed using several cartoons (along with the captions) and the user ratings corresponding to each of them, where the users were asked to rate each caption as "funny" (3), "somewhat funny" (2) and "unfunny" (1). We choose contest 651 for our simulation, while several other contests are available in the repository, which can be found here. For simplicity, we select $K = 4$ captions from the contest with the aim of finding the caption which is the most highly rated. For this, we compute the empirical mean score for each caption, and then rewards are generated according to a Gaussian distribution with the corresponding empirical means.

Protein Kinase Inhibitors for Cancer Drug Discovery: For this experiment, we use the PKIS2 dataset, which is available in [28], and it is an extended version of the PKIS1 dataset published by Glaxo-SmithKline in 2013. The dataset contains a collection of protein kinase and a list of small molecule compounds (kinase inhibitors), and it enumerates how strongly each inhibitor reacts with each kinase. This is an important problem in cancer drug discovery, where researchers are interested in finding targeted kinase inhibitors for treating cancer cells. The dataset can be downloaded from this link. For our experiment, we select one specific kinase ACVRL1, which is present in the dataset. PKIS2 tests 641 inhibitors against different kinase, out of which a total of 189 are tested against ACVRL1. For simplicity, we select $K = 4$ of these 189 inhibitors. The dataset provides a "percentage inhibition" for each compound, which is averaged over several trials. For each of these entries, we normalize it to be between 0 and 1, and then find out the percentage control by subtracting each of the normalized entries from 1. The percentage control forms an interesting measure for understanding how effective the compound is against the targeted kinase. Furthermore, following the setup in [29], we take the logarithm of each percentage control, which has been seen to have a Gaussian distribution whose variance is bounded by 1. Finally, our goal is to find the compound that exhibits the highest percentage control against ACVRL1.

Experiments with synthetic data. Next, we present two more experiments with synthetic data, which illustrate the looseness of the theoretical confidence interval for the proposed gap-based algorithm (Algorithm 1). The adversarial model is the same as that of the real-world experiments. Specifically, for the first experiment (Figure 4a), we use the confidence interval prescribed by theory (16), which is observed to be loose empirically. For this experiment, we use the same set-up of a 4-armed Gaussian bandit, with the mean vector $\boldsymbol{\mu} = [2.5,\; 2.3,\; 2,\; 0.6]$, where the probability of attack is set to $\varepsilon = 0.1$. Clearly, as discussed, in this setting, the median-based successive elimination procedure prescribed in [4] outperforms all other methods due to the better uncertainty $U_i = \mathcal{O}\Big(\frac{\varepsilon}{1-\varepsilon}\Big)$. However, we observe that our proposed successive elimination algorithm based on the α-trimmed mean estimator very closely follows the performance of the median-based algorithm.

Furthermore, in the case of exponentially distributed bandit instances, the median-based procedure no longer works, since the exponential distribution is not unimodal. The second experiment (Figure 4b) is based on this set-up, where we use an 8-armed exponential bandit instance whose mean vector is given by $\boldsymbol{\mu} = [2.5,\ 2.3,\ 2,\ 1.4,\ 1,\ 0.6,\ 0.2,\ 0.05]$. In Figure 4b, the "sample mean-based strategy" refers to the successive elimination algorithm, where the estimator is replaced by the sample mean. All the other parameters remain the same as in the previous experiment, and we have averaged both the experiments over 1000 Monte Carlo trials. In this case, we observe that the theoretical confidence interval for the gap-based procedure described in (16) is loose, and the proposed successive-elimination based algorithm outperforms the gap-based procedures in identifying the best arm.

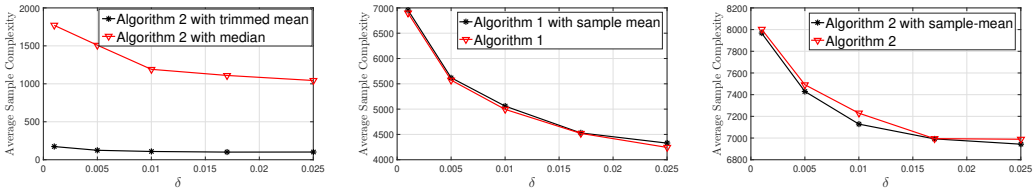


(a) Gaussian bandit (b) Exponential bandit

Figure 4: More experiments with synthetic data

Comparison with the sample median estimator. To clarify our rationale of choosing the trimmed mean estimator over the sample median, we perform further experiments. Specifically, we have the following setup. We consider a simple bandit instance with $K = 2$ arms, where the arms generate rewards drawn from a log-normal distribution (which is a heavy-tailed distribution). The parameters used for the two arms are $\mu = [1, 1.05]$ and $\sigma = [1, 1.2]$. The goal of the learner is to identify the arm with the highest mean, where the mean of any arm $i \in [K]$ is given by $\theta_i = \exp\left(\mu + \frac{\sigma^2}{2}\right)$. The superior performance of the trimmed mean estimator can be found in Figure 5a.

Comparison with the sample mean estimator. We also perform more experiments to show the robustness of the trimmed mean estimator over the sample-mean used in algorithms for the corruption-free setting. For this purpose, we use a corruption level of $\varepsilon = 0.1$ to compare the performance of the algorithms 1 and 2 against the attack-free counterparts. For this experiment, we have used a Gaussian bandit with $K = 8$ arms, where the mean vector is given by $\mu = [2.5, 2.3, 2, 1.4, 1, 0.6, 0.2, 0.05]$. The corresponding results for algorithms 1 and 2 can be found in Figures 5b and 5c, which clearly show the robustness of the trimmed mean estimator compared to the sample-mean.



(a) Log-normal bandits (b) Gaussian bandit (c) Gaussian bandit

Figure 5: Experiments for showing the efficacy of the trimmed mean estimator