

Appendix A. Geometric Interpretation of Gap Quantities

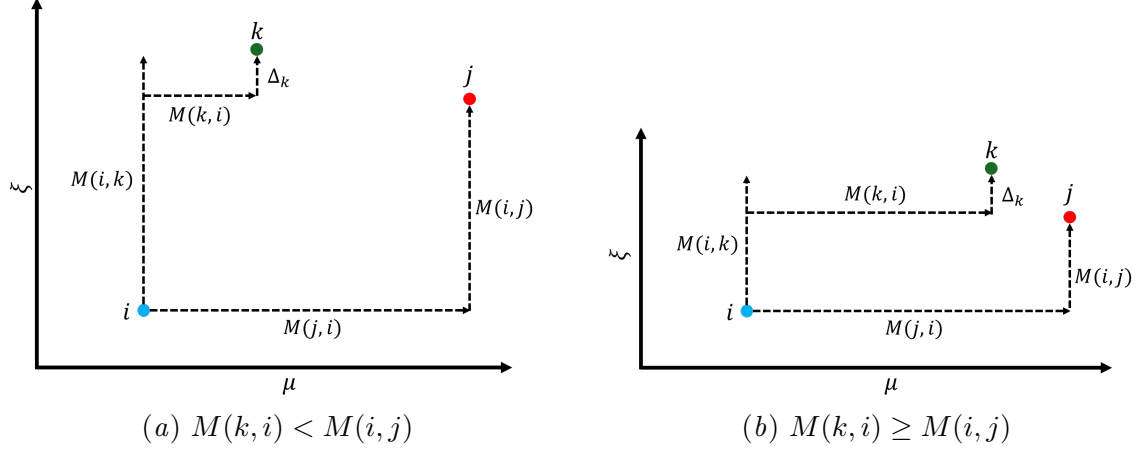


Figure 5: **Geometric illustration of gap quantities in the mean-risk space.** Each point represents an arm, plotted by its expected reward on the horizontal axis (μ) and scaled mean-variance risk on the vertical axis ($\xi := \alpha(\sigma^2 - \rho\mu)$). Here suppose that arms i and j are Pareto optimal while arm k is non-Pareto, suboptimal. In both panels, arm i has lower risk but a smaller mean than that of arm k ; the opposite case (i.e., arm i has a higher mean but a greater risk than those of arm k) can be treated similarly. If arm i moves upward by more than $M(i, k)$ (or if k moves downward), arm i would become dominated by arm k . In these illustrations, the identity $M(i, k) = M(i, j) + (\xi_k - \xi_j)$ holds. Let us suppose that the gap of arm i satisfies $\Delta_i = \min(\min(M(i, j), M(j, i)), M(k, i)^+ + \Delta_k)$ under the existence of other possible Pareto and non-Pareto arms. Panel (a): When $M(k, i) < M(i, j)$, we obtain $M(i, k) = M(i, j) + (\xi_k - \xi_j) > M(k, i) + (\xi_k - \xi_j) = M^+(k, i) + \Delta_k$. The smaller $M(i, k)$ is, the more the samplings from arms i and k are required to discriminate them in practice. The choice of $M^+(k, i) + \Delta_k$ —the lower bound of $M(i, k)$ —as the gap Δ_i corresponds to a “conservative” estimate reflecting not only the suboptimal arm k but also other Pareto arms j via Eq. (1). Panel (b): When $M(k, i) \geq M(i, j)$, the term $\min(M(i, j), M(j, i))$ dominates the expression of Δ_i , indicating that the difficulty in distinguishing i from another Pareto-optimal arm j outweighs that from suboptimal arm k . Thus, the contribution of k to Δ_i becomes negligible in this case.

To clarify the role of the gap quantities introduced in Section 2, we provide a geometric illustration in the mean-risk space. Specifically, we consider the case in which a suboptimal arm $k \notin D^+$ is compared with multiple Pareto-optimal arms $i, j \in D^+$ (see Figure 5). The associated gap quantities characterize distinct sources of uncertainty that affect the accurate identification of ϵ -Pareto optimal arms under sampling processes.

Robustness Against Elimination by Suboptimal Arms: The quantity $M(k, i)^+ + \Delta_k$ represents a “conservative” margin ensuring that a Pareto-optimal arm $i \in D^+$ is not mistakenly eliminated due to statistical fluctuations in empirical estimates. Here, $M(k, i)^+$ measures how close the suboptimal arm k is to dominating the optimal arm i , and Δ_k reflects the dif-

difficulty of confirming the suboptimality of k . Smaller value of this term indicates higher risk of erroneous elimination of i , by chance, along the process of some instance of samplings.

Distinguishability among Pareto-Optimal Arms: The term $\Delta_{ij} := \min\{M(i, j), M(j, i)\}$ captures the closeness between two distinct Pareto-optimal arms $i, j \in D^+$. The smaller Δ_{ij} , the closer the two arms in both objectives. This makes it more difficult to distinguish between them and judge correctly that neither of them dominates another with higher confidence.

Suboptimality Measure for Arm k : For a suboptimal arm $k \notin D^+$, the gap Δ_k represents the minimum shift required in all objectives for k to enter the Pareto set. Smaller Δ_k implies that arm k is closer to the Pareto frontier and, thus, more prone to be misclassified as Pareto-optimal, by chance.

In summary, the gap quantity Δ_i for each arm $i \in [K]$ encodes the difficulty of correctly identifying the Pareto-optimal set under noisy and finite observations. It captures three key aspects: the resilience of optimal arms against being dominated by suboptimal ones, the distinguishability among optimal arms, and the closeness of suboptimal arms to the Pareto frontier. These interpretations provide a concrete understanding of the role and design of the gap-based arm selection strategy in our RAMGapE framework.

Appendix B. Theoretical Analysis

In this section, we present upper bounds on the performance of RAMGapEb and RAMGapEc, as introduced in Section 3. Our analysis follows a similar proof structure as the UGapE algorithm [Gabillon et al. \(2012\)](#), which establishes a unified gap-based analysis framework for fixed-budget and fixed-confidence best arm identification. This similarity allows us to extend the classical regret arguments to the risk-averse multi-objective setting. A key feature of RAMGapE is its unified arm selection strategy, which operates across both fixed-budget and fixed-confidence settings. This shared structure allows for a largely unified theoretical analysis. Appendix B.1 outlines the common components of the proof, while Appendix B.4.2 details the derivation of confidence intervals specific to the fixed-confidence setting. Before presenting the main theoretical results, we introduce the concept of an event \mathcal{E} that will be essential for the following analysis.

$$\mathcal{E} := \left\{ \forall i \in [K], \forall t \in \{2K+1, \dots, T\}, |\hat{\mu}_i(t) - \mu_i| < \beta_i(t) \wedge \left| \hat{\mu}_i^{(2)}(t) - \mu_i^{(2)} \right| < \beta_i(t) \right\}, \quad (10)$$

where the values of T and $\beta_i(t)$ are defined separately for each setting. In particular, for any arm $i \in [K]$ and at any round $t \geq 2K+1$, both $\underline{\mu}_i(t) \leq \mu_i \leq \bar{\mu}_i(t)$ and $\underline{\xi}_i(t) \leq \xi_i \leq \bar{\xi}_i(t)$ *surely* hold when event \mathcal{E} holds (see Appendix B.4.2).

B.1. Analysis of the Arm Selection Strategy

First, we present lower(Lemma 4) and upper(Lemma 6) for $V(t)$ in event \mathcal{E} , which indicates their connection with regret. We prove that for set $\hat{D}_t^+ (\neq D^+)$ at any round $t \in \{2K+1, \dots, T\}$, the quantity $V_{\hat{D}_t^+ \Delta D^+}(t)$ serves as an upper bound on the simple regret of this set $r_{\hat{D}_t^+}$ under the condition that event \mathcal{E} occurs.

Lemma 4 *On event \mathcal{E} , for any round $t \in \{2K+1, \dots, T\}$, we have $V(t) \geq r_{\hat{D}_t^+}$.*

Proof On event \mathcal{E} , for any arm $i \in \widehat{D}_t^+ \triangle D^+ (= D^+ \setminus \widehat{D}_t^+ \cup \widehat{D}_t^+ \setminus D^+)$ and each round $t \in \{2K+1, \dots, T\}$, we prove the lemma for the two cases individually in parallel to the definition of gaps dependent on whether the arm in question belongs to the Pareto set D^+ :

Case 1. $i \in D^+ \setminus \widehat{D}_t^+$:

Case 1.1. $|D^+| = 1$: In this case, for any arm $k(\neq i)$, $i \succ k$ and $M(k, i)^+ = 0$. We can write

$$\begin{aligned}
 V_i(t) &= \min_{j \in \widehat{D}_t^+ \text{ s.t. } j \succ_t i} \max \left(\bar{\mu}_i(t) - \underline{\mu}_j(t), \bar{\xi}_j(t) - \underline{\xi}_i(t) \right) \\
 &\geq \min_{j \neq i} \max \left(\bar{\mu}_i(t) - \underline{\mu}_j(t), \bar{\xi}_j(t) - \underline{\xi}_i(t) \right) \\
 &\stackrel{(A)}{\geq} \min_{j \neq i} \max(\mu_i - \mu_j, \xi_j - \xi_i) \\
 &\geq \min_{j \neq i} \min(\mu_i - \mu_j, \xi_j - \xi_i) \\
 &= \min_{j \notin D^+} m(j, i) \\
 &\stackrel{(B)}{=} \min_{j \notin D^+} \Delta_j \\
 &\stackrel{(C)}{=} \Delta_i = r_i(\widehat{D}_t^+) \tag{11}
 \end{aligned}$$

The inequality (A) holds because of event \mathcal{E} , (B) holds from the definition of gap Δ_j (Eq. 1) for $j \notin D^+$ where only a single Pareto solution exists, and (C) holds from that for $i \in D^+$.

Case 1.2. $|D^+| \geq 2$: In this case, we can write

$$\begin{aligned}
 V_i(t) &= \min_{j \in \widehat{D}_t^+ \text{ s.t. } j \succ_t i} \max \left(\bar{\mu}_i(t) - \underline{\mu}_j(t), \bar{\xi}_j(t) - \underline{\xi}_i(t) \right) \\
 &\geq \min_{j \neq i} \max \left(\bar{\mu}_i(t) - \underline{\mu}_j(t), \bar{\xi}_j(t) - \underline{\xi}_i(t) \right) \\
 &\stackrel{(A)}{\geq} \min_{j \neq i} \max(\mu_i - \mu_j, \xi_j - \xi_i) \\
 &= \min \left\{ \min_{j \in D^+ \setminus \{i\}} \max(\mu_i - \mu_j, \xi_j - \xi_i), \min_{j \notin D^+} \max(\mu_i - \mu_j, \xi_j - \xi_i) \right\} \\
 &= \min \left\{ \min_{j \in D^+ \setminus \{i\}} M(i, j), \min_{j \notin D^+} M(i, j) \right\} \\
 &\geq \min \left\{ \min_{j \in D^+ \setminus \{i\}} \min\{M(i, j), M(j, i)\}, \min_{j \notin D^+} M(i, j) \right\} \\
 &= \min \left\{ \min_{j \in D^+ \setminus \{i\}} \min\{M(i, j), M(j, i)\}, M(i, k) \right\}
 \end{aligned}$$

where in the last equality we introduced $k = \operatorname{argmin}_{j \notin D^+} M(i, j)$ for simplicity. The inequality (A) holds because of event \mathcal{E} .

The proof proceeds by considering two separate cases: $k \prec i$ (Case 1.2.1.) and $k \not\prec i$ (Case 1.2.2.).

Case 1.2.1. $k \prec i$:

Suppose that arm h satisfies $\Delta_k = \max_{j \in D^+ \text{ s.t. } j \succ k} m(k, j) = m(k, h)$ in defining the gap of the non-Pareto arm k ($\notin D^+$). Note that $M(i, k) \geq m(k, h)$ holds because otherwise it contradicts the condition of the arm i being a Pareto solution: that is, suppose that $M(i, k) = \max(\mu_i - \mu_k, \xi_k - \xi_i) < \min(\mu_h - \mu_k, \xi_k - \xi_h) = m(k, h)$ holds. Then, when $M(i, k) = \mu_i - \mu_k$ is satisfied, $\mu_i - \mu_k < \min(\mu_h - \mu_k, \xi_k - \xi_h) \leq \mu_h - \mu_k$, that is, $\mu_i < \mu_h$, apparently contradicts $i \in D^+$. This is the same for case $M(i, k) = \xi_k - \xi_i$. Then, we can write

$$\begin{aligned}
 & \min \left\{ \min_{j \in D^+ \setminus \{i\}} \min\{M(i, j), M(j, i)\}, M(i, k) \right\} \\
 & \geq \min \left\{ \min_{j \in D^+ \setminus \{i\}} \min\{M(i, j), M(j, i)\}, m(k, h) \right\} \\
 & = \min \left\{ \min_{j \in D^+ \setminus \{i\}} \min\{M(i, j), M(j, i)\}, \Delta_k \right\} \\
 & = \min \left\{ \min_{j \in D^+ \setminus \{i\}} \min\{M(i, j), M(j, i)\}, M(k, i)^+ + \Delta_k \right\} \quad (M(k, i)^+ = 0 \text{ for } \because k \prec i) \\
 & \geq \min \left\{ \min_{j \in D^+ \setminus \{i\}} \min\{M(i, j), M(j, i)\}, \min_{j \notin D^+} (M(j, i)^+ + \Delta_j) \right\} \\
 & = \Delta_i = r_i(\widehat{D}_t^+) \tag{12}
 \end{aligned}$$

Case 1.2.2. $k \not\prec i$: Because arm k belongs not to the Pareto set, there should exist an arm $h \in D^+ \setminus \{i\}$ such that $h \succ k$, that is, $\mu_h > \mu_k$ and $\xi_h < \xi_k$. Hence,

$$\begin{aligned}
 M(i, k) &= \max\{\mu_i - \mu_k, \xi_k - \xi_i\} \\
 &\geq \max\{\mu_i - \mu_h, \xi_h - \xi_i\} \\
 &= M(i, h) \\
 &\geq \min_{j \in D^+ \setminus \{i\}} \min\{M(i, j), M(j, i)\}.
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 & \min \left\{ \min_{j \in D^+ \setminus \{i\}} \min\{M(i, j), M(j, i)\}, M(i, k) \right\} \\
 & = \min_{j \in D^+ \setminus \{i\}} \min\{M(i, j), M(j, i)\} \\
 & \geq \min \left\{ \min_{j \in D^+ \setminus \{i\}} \min\{M(i, j), M(j, i)\}, \min_{j \notin D^+} (M(j, i)^+ + \Delta_j) \right\} \\
 & = r_i(\widehat{D}_t^+) \tag{13}
 \end{aligned}$$

Case 2. $i \in \widehat{D}_t^+ \setminus D^+$:

$$\begin{aligned}
 V_i(t) &= \max_{j \neq i} \min \left(\bar{\mu}_j(t) - \underline{\mu}_i(t), \bar{\xi}_i(t) - \underline{\xi}_j(t) \right) \\
 &\stackrel{(A)}{=} \max_{j \neq i} \min(\mu_j - \mu_i, \xi_i - \xi_j) \\
 &\geq \max_{j \in D^+ \text{ s.t. } j \succ i} \min(\mu_j - \mu_i, \xi_i - \xi_j) \\
 &= \max_{j \in D^+ \text{ s.t. } j \succ i} m(i, j) = \Delta_i = r_i(\widehat{D}_t^+) \tag{14}
 \end{aligned}$$

The inequality (A) holds because of event \mathcal{E} .

Using Eq. 11, 12, 13 and 14, we have

$$V(t) \geq V_{\widehat{D}_t^+ \triangle D^+}(t) = \max_{i \in \widehat{D}_t^+ \triangle D^+} V_i(t) \geq \max_{i \in \widehat{D}_t^+ \triangle D^+} r_i(\widehat{D}_t^+) \stackrel{(A)}{=} \max_{i \in [K]} r_i(\widehat{D}_t^+) = r_{\widehat{D}_t^+},$$

where the equality (A) follows from that $r_i(\widehat{D}_t^+) = 0$ for any $i \notin \widehat{D}_t^+ \triangle D^+$. \blacksquare

Lemma 5 *On event \mathcal{E} , for any round $t \in \{2K+1, \dots, T\}$, if arm $i \in \{m_t, p_t\}$ is pulled, we have $V(t) \leq 2\beta_i(t)$.*

Proof The proof proceeds by case analysis, depending on whether $m_t \in \widehat{D}_t^+$ or not.

Case 1. $m_t \in \widehat{D}_t^+$: In this case, we can write

$$\begin{aligned}
 V(t) &= \max_{j \neq m_t} \min \left(\bar{\mu}_j(t) - \underline{\mu}_{m_t}(t), \bar{\xi}_{m_t}(t) - \underline{\xi}_j(t) \right) \\
 &= \min \left(\bar{\mu}_{p_t}(t) - \underline{\mu}_{m_t}(t), \bar{\xi}_{m_t}(t) - \underline{\xi}_{p_t}(t) \right) \\
 &= \min \left(\hat{\mu}_{p_t}(t) - \hat{\mu}_{m_t}(t), \hat{\xi}_{m_t}(t) - \hat{\xi}_{p_t}(t) \right) + \beta_{m_t}(t) + \beta_{p_t}(t) \\
 &\leq 2\beta_i(t)
 \end{aligned}$$

Case 2. $m_t \notin \widehat{D}_t^+$: In this case, we can write

$$\begin{aligned}
 V(t) &= \min_{j \in \widehat{D}_t^+ \text{ s.t. } j \succ m_t} \max \left(\bar{\mu}_{m_t}(t) - \underline{\mu}_j(t), \bar{\xi}_j(t) - \underline{\xi}_{m_t}(t) \right) \\
 &= \max \left(\bar{\mu}_{m_t}(t) - \underline{\mu}_{p_t}(t), \bar{\xi}_{p_t}(t) - \underline{\xi}_{m_t}(t) \right) \\
 &= \max \left(\hat{\mu}_{m_t}(t) - \hat{\mu}_{p_t}(t), \hat{\xi}_{p_t}(t) - \hat{\xi}_{m_t}(t) \right) + \beta_{m_t}(t) + \beta_{p_t}(t) \\
 &\leq 2\beta_i(t)
 \end{aligned}$$

The proof of Lemma 5 is completed through the analysis of the two cases. \blacksquare

Lemma 6 *On event \mathcal{E} , if arm $i \in \{m_t, p_t\}$ is pulled at time $t \in \{2K+1, \dots, T\}$, we have*

$$V(t) \leq \min(0, -r_i(\widehat{D}_t^+) + 2\beta_i(t)) + 2\beta_i(t). \quad (15)$$

Proof On event \mathcal{E} , for any round $t \in \{2K+1, \dots, T\}$, if arm i is pulled, from Lemma 4 and Lemma 5, we have the inequalities:

$$r_i(\widehat{D}_t^+) \leq V(t) \leq 2\beta_i(t).$$

Rearranging the left inequality yields:

$$0 \leq -r_i(\widehat{D}_t^+) + V(t) \leq -r_i(\widehat{D}_t^+) + 2\beta_i(t) \Rightarrow 0 \leq -r_i(\widehat{D}_t^+) + 2\beta_i(t).$$

Together with the right inequality,

$$V(t) \leq 2\beta_i(t),$$

we combine these two inequalities to obtain

$$V(t) \leq \min(0, -r_i(\widehat{D}_t^+) + 2\beta_i(t)) + 2\beta_i(t).$$

This concludes the proof. ■

B.2. Regret Bound for the Fixed-Budget Setting

Here we prove an upper-bound on the simple regret of RAMGapEb. Since the setting considered by the algorithm is fixed-budget, we may say $T = n$. From the definition of the confidence interval $\beta_i(t)$ in Eq. 9 and a union bound, we have that $\mathbb{P}[\mathcal{E}] \geq 1 - 4Kn \exp(-2a)$. We now have all the tools needed to prove the performance of RAMGapE for the ϵ -Pareto set identification problem.

Theorem 2 *If we run RAMGapEb with parameter $0 < a \leq \frac{n-2K}{16K}\epsilon^2$, its simple regret $r_{\widehat{D}_n^+}$ satisfies*

$$\widetilde{\delta} = \mathbb{P}\left[r_{\widehat{D}_n^+} \geq \epsilon\right] \leq 4Kn \exp(-2a),$$

and in particular this probability is minimized for $a = \frac{n-2K}{16K}\epsilon^2$.

Proof This proof is by contradiction. We assume that $r_{\widehat{D}_n^+} > \epsilon$ on event \mathcal{E} and consider the following two steps:

Step 1: Here we indicate that on the event \mathcal{E} , we have the following upper-bound on the number of pulls of any arm $i \in [K]$:

$$T_i(n) < \frac{4a}{\max\left(\frac{r_i(\widehat{D}_n^+) + \epsilon}{2}, \epsilon\right)^2} + 2. \quad (16)$$

Let t_i be the last round that arm i is pulled. If arm i has been pulled only during the initialization phase, $T_i(n) = 2$ and Eq. 16 trivially holds. If PullArm has selected i , then we have

$$\min(0, -r_i(\widehat{D}_{t_i}^+) + 2\beta_i(t_i)) + 2\beta_i(t_i) \stackrel{(A)}{\geq} V(t_i) \stackrel{(B)}{\geq} r_{\widehat{D}_{t_i}^+} > \epsilon. \quad (17)$$

(A) and (B) hold because of Lemmas 4 and 6.

We derive the following transformation by applying $\beta_i(t_i)$ and Eq. 17.

$$\begin{aligned} 2\beta_i(t_i) &\geq \max\left(\frac{r_i(\widehat{D}_{t_i}^+) + \epsilon}{2}, \epsilon\right) \Rightarrow 4\beta_i^2(t_i) = \frac{4a}{T_i(t_i)} \geq \max\left(\frac{r_i(\widehat{D}_{t_i}^+) + \epsilon}{2}, \epsilon\right)^2 \\ \Leftrightarrow T_i(t_i) &\leq \frac{4a}{\max\left(\frac{r_i(\widehat{D}_{t_i}^+) + \epsilon}{2}, \epsilon\right)^2} < \frac{4a}{\max\left(\frac{r_i(\widehat{D}_{t_i}^+) + \epsilon}{2}, \epsilon\right)^2} + 2 \end{aligned}$$

As a result of the final transformation, we obtain Eq. 16.

Step 2: Using Eq. 16, we have $n = \sum_{i=1}^K T_i(n) < \sum_{i=1}^K \frac{4a}{\max\left(\frac{r_i(\widehat{D}_n^+) + \epsilon}{2}, \epsilon\right)^2} + 2K$ on event \mathcal{E} .

It is easy to see that by selecting $a \leq \frac{n-2K}{16K}\epsilon^2$, the right-hand-side of this inequality will be smaller than or equal to n , which is a contradiction. Thus, we conclude that $r_{\widehat{D}_n^+} \leq \epsilon$ on event \mathcal{E} . The final result follows from the probability of event \mathcal{E} defined at the beginning of this section. \blacksquare

B.3. Regret Bound for Fixed-Confidence Setting

We establish an upper bound on the simple regret of RAMGapEc. As the algorithm is analyzed in the fixed-confidence framework, we set $T = +\infty$ without loss of generality. By applying a union bound over all possible values of $T_i(t) \in \{2, \dots, t\}$ for $t = 2K + 1, \dots, \infty$, and utilizing the confidence intervals $\beta_i(t)$ defined in Eq. 9, it follows that the event \mathcal{E} occurs with probability at least $1 - \delta$, i.e., $\mathbb{P}(\mathcal{E}) \geq 1 - \delta$ (see Theorem 8).

Theorem 3 *The RAMGapEc algorithm stops after \tilde{n} rounds and returns an ϵ -Pareto set, $\widehat{D}_{\tilde{n}}^+$, that satisfies*

$$\mathbb{P}\left[r_{\widehat{D}_{\tilde{n}}^+} \leq \epsilon \wedge \tilde{n} \leq N\right] \geq 1 - \delta,$$

where $N = 2K + \mathcal{O}\left(\frac{K}{\epsilon^2} \log\left(\frac{K \log_2^2(1/\epsilon)}{\delta}\right)\right)$.

Proof We first prove an upper bound on the simple regret of RAMGapEc. Using Lemma 4, we have that on the event \mathcal{E} , the simple regret of RAMGapEc upon stopping satisfies $V(t) \geq r_{\widehat{D}_n^+}$. Since the algorithm stops when $V(t) < \epsilon$, this implies that $r_{\widehat{D}_{\tilde{n}}^+} < \epsilon$ on \mathcal{E} , and hence

$$\mathbb{P}\left[r_{\widehat{D}_{\tilde{n}}^+} \leq \epsilon\right] \geq \mathbb{P}(\mathcal{E}) \geq 1 - \delta.$$

Next, we derive an upper bound on the number of times each arm is pulled. Let t_i be the last round at which arm i is selected. If arm i is pulled only during the initialization phase, then $T_i(\tilde{n}) = 2$ and the following bound holds trivially. We now consider the case where arm i is selected at some round $t_i > 2K$ by the PullArm procedure. On the event \mathcal{E} , by Lemma 5, we have $V(t_i) \leq 2\beta_i(t_i)$. Combining this with the stopping condition $V(t_i) < \epsilon$, we obtain:

$$\beta_i(t_i) < \frac{\epsilon}{2}.$$

Recall that the confidence interval is defined as

$$\beta_i(t) = \sqrt{\frac{4 \log \left(\frac{8K(\log_2 T_i(t))^2}{\delta} \right)}{T_i(t)}}.$$

Since RAMGapEc must hold for any arm i , $T_i(t_i) \geq \Omega(1/\epsilon^2)$ because of stopping criteria. Thus, it is natural to bound $\log_2 T_i(t_i) \leq \log_2(1/\epsilon^2)$, and substitute accordingly.

$$\log_2 T_i(t_i) \leq \log_2 \left(\frac{1}{\epsilon^2} \right) = 2 \log_2(1/\epsilon),$$

and thus

$$(\log_2 T_i(t_i))^2 \leq 4 \log_2^2(1/\epsilon).$$

Substituting this into the expression for $\beta_i(t_i)$ gives

$$\beta_i(t_i) \leq \sqrt{\frac{4 \log \left(\frac{32K \log_2^2(1/\epsilon)}{\delta} \right)}{T_i(t_i)}}.$$

To ensure $\beta_i(t_i) < \epsilon/2$, it suffices to require

$$T_i(t_i) > \frac{16}{\epsilon^2} \log \left(\frac{32K \log_2^2(1/\epsilon)}{\delta} \right).$$

Hence, the number of pulls for any arm i is upper-bounded as

$$T_i(\tilde{n}) \leq \frac{16}{\epsilon^2} \log \left(\frac{32K \log_2^2(1/\epsilon)}{\delta} \right) + 2.$$

Finally, summing over all arms $i \in [K]$, the total number of rounds before stopping satisfies

$$\tilde{n} = \sum_{i=1}^K T_i(\tilde{n}) \leq \sum_{i=1}^K \left(\frac{16}{\epsilon^2} \log \left(\frac{32K \log_2^2(1/\epsilon)}{\delta} \right) + 2 \right) = \mathcal{O} \left(\frac{K}{\epsilon^2} \log \left(\frac{K \log_2^2(1/\epsilon)}{\delta} \right) \right).$$

This completes the proof. ■

B.4. Other Theorems on RAMGapE

B.4.1. OUTPUT ACCURACY

Theorem 7 *Given allowance $\epsilon > 0$, on event \mathcal{E} , at any round $t \geq 2K + 1$, if $V(t) < \epsilon$, \widehat{D}_t^+ is ϵ -Pareto set.*

Proof The proof is completed by the following two conditions:

- (1) if $i \in \widehat{D}_t^+$ and $V(t) < \epsilon$, then $\forall j \in [K], \neg(\mu_i \leq \mu_j - \epsilon \wedge \xi_i \geq \xi_j + \epsilon)$
- (2) if $i \notin \widehat{D}_t^+$ and $V(t) < \epsilon$, then $\exists j \in [K] \setminus \{i\}, \mu_i \leq \mu_j + \epsilon \wedge \xi_i \geq \xi_j - \epsilon$

Note that $V(t) \left(= \max_{k \in [K]} V_k(t) \right) < \epsilon$ implies $V_k(t) < \epsilon$ for any arm k .

(1) Remind $V_i(t) = \max_{j \neq i} \min \left(\bar{\mu}_j(t) - \underline{\mu}_i(t), \bar{\xi}_i(t) - \underline{\xi}_j(t) \right)$ for arm $i \in \widehat{D}_t^+$ (Eq. 4). For arm $i \in \widehat{D}_t^+$, if $V_i(t) < \epsilon$, then $\bar{\mu}_k(t) - \underline{\mu}_i(t) < \epsilon \vee \bar{\xi}_i(t) - \underline{\xi}_k(t) < \epsilon$ holds for any arm $k (\neq i)$. On event \mathcal{E} where the true mean and the true risk are surely within their confidence interval, $\mu_k - \mu_i < \epsilon \vee \xi_i - \xi_k < \epsilon$ holds. This implies (1).

(2) $V_i(t) = \min_{j \in \widehat{D}_t^+ \text{ s.t. } j \succ_t i} \max \left(\bar{\mu}_i(t) - \underline{\mu}_j(t), \bar{\xi}_j(t) - \underline{\xi}_i(t) \right)$ for arm $i \notin \widehat{D}_t^+$ (Eq. 4). For arm $i \notin \widehat{D}_t^+$, if $V_i(t) < \epsilon$, there exists an arm $k \succ_t i$ such that $\bar{\mu}_i(t) - \underline{\mu}_k(t) < \epsilon \wedge \bar{\xi}_k(t) - \underline{\xi}_i(t) < \epsilon$. On event \mathcal{E} , $\mu_i - \mu_k < \epsilon \wedge \xi_k - \xi_i < \epsilon$ holds. This implies (2).

This completes the proof. ■

B.4.2. DERIVATION OF HIGH PROBABILITY CONFIDENCE INTERVAL UNDER FIXED CONFIDENCE SETTING

Here, we provide a proof of the theorem regarding the high-probability confidence intervals used in RAMGapEc.

Theorem 8 *Under the fixed confidence setting, if let $\beta_i(t) = \sqrt{4 \log(8K(\log_2 T_i(t))^2 / \delta) / T_i(t)}$, event \mathcal{E} holds with probability at least $1 - \delta$.*

Proof We define the events \mathcal{E}_1 and \mathcal{E}_2 as

$$\begin{aligned} \mathcal{E}_1 &= \left\{ \exists i \in [K], \exists t \geq 2K + 1, |\hat{\mu}_i(t) - \mu_i| \geq \beta_i(t) \right\}, \\ \mathcal{E}_2 &= \left\{ \exists i \in [K], \exists t \geq 2K + 1, |\hat{\mu}_i^{(2)}(t) - \mu_i^{(2)}| \geq \beta_i(t) \right\}. \end{aligned}$$

From the definition of \mathcal{E} , the probability that \mathcal{E} occurs is bounded from below by $\mathbb{P}[\mathcal{E}] = 1 - \mathbb{P}[\mathcal{E}^C] \geq 1 - (\mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_2])$.

We derive upper bounds of each event \mathcal{E}_1 and \mathcal{E}_2 .

First, we provide an upper bound of the event \mathcal{E}_1 . From Hoeffding-Azuma inequality [Azuma \(1967\)](#); [Tropp \(2012\)](#), for any integer γ and a positive function $x(\gamma)$, the following inequality holds.

$$\mathbb{P} \left[\exists i, \exists s \in \{1, \dots, 2^\gamma\}, \left| \sum_{l=1}^s (X_i(l) - \mu_i) \right| > x(\gamma) \right] \leq 2 \exp \left(-\frac{x(\gamma)^2}{2^\gamma} \right). \quad (18)$$

Since the $\beta_i(t)$ depends only on $T_i(t)$ under given K and δ we rewrite $\beta_i(t)$ as β_s when $s = T_i(t)$, for convenience. Using Eq. 18, $\mathbb{P}[\mathcal{E}_1]$ is bounded as follows:

$$\begin{aligned}
 \mathbb{P}[\mathcal{E}_1] &\leq \sum_{i=1}^K \mathbb{P}[\exists t \geq 2K+1, |\hat{\mu}_i(t) - \mu_i| \geq \beta_i(t)] \\
 &\leq \sum_{i=1}^K \mathbb{P}\left[\exists s \geq 2, \left|\sum_{l=1}^s (X_i(l) - \mu_i)\right| \geq s\beta_s\right] \\
 &\leq \sum_{i=1}^K \sum_{\gamma=1}^{\infty} \mathbb{P}\left[\exists s \in \{2^{\gamma-1}, \dots, 2^\gamma\}, \left|\sum_{l=1}^s (X_i(l) - \mu_i)\right| \geq 2^{\gamma-1}\beta_{2^\gamma}\right] \\
 &\leq \sum_{i=1}^K \sum_{\gamma=1}^{\infty} \mathbb{P}\left[\exists s \in \{1, \dots, 2^\gamma\}, \left|\sum_{l=1}^s (X_i(l) - \mu_i)\right| \geq 2^{\gamma-1}\beta_{2^\gamma}\right] \\
 &\leq \sum_{i=1}^K \sum_{\gamma=1}^{\infty} 2 \exp\left(-\frac{(2^{\gamma-1}\beta_{2^\gamma})^2}{2^\gamma}\right) \\
 &= \sum_{i=1}^K \sum_{\gamma=1}^{\infty} 2 \exp(-2^{\gamma-2}\beta_{2^\gamma}^2) \\
 &= \sum_{i=1}^K \sum_{\gamma=1}^{\infty} 2 \exp\left(-\log\left(\frac{8K(\log_2 2^\gamma)^2}{\delta}\right)\right) \\
 &= 2K \sum_{\gamma=1}^{\infty} \frac{\delta}{8K\gamma^2} = \frac{\delta}{4} \frac{\pi^2}{6} < \frac{\delta}{2}
 \end{aligned}$$

In this study, we consider that arm i ' reward distribution ν_i is bounded in $[0, 1]$, so we can say for each $i \in [K]$, $X_i^2 \in [0, 1]$. So we can say $\mathbb{P}[\mathcal{E}_2] < 2/\delta$ and conclude $\mathbb{P}[\mathcal{E}] \geq 1 - \delta$. ■

Corollary 9 *As a consequence of Theorem 8, $\mathbb{P}\{\forall i, \forall t, |\widehat{\text{MV}}_i(t) - \text{MV}_i| < (3 + \rho)\beta_i(t)\} \geq 1 - \delta$ holds.*

Proof On event \mathcal{E} , we derive the following sequence of transformations:

$$\begin{aligned}
 (3 + \rho)\beta_i(t) &= \beta_i(t) + (2 + \rho)\beta_i(t) \\
 &> |\hat{\mu}_i^{(2)}(t) - \mu_i^{(2)}| + (2 + \rho)|\hat{\mu}_i(t) - \mu_i| \\
 &= |\hat{\mu}_i^{(2)}(t) - \mu_i^{(2)}| + 2|\hat{\mu}_i(t) - \mu_i| + \rho|\hat{\mu}_i(t) - \mu_i| \\
 &\geq |\hat{\mu}_i^{(2)}(t) - \mu_i^{(2)}| + |\hat{\mu}_i(t) + \mu_i||\hat{\mu}_i(t) - \mu_i| + \rho|\hat{\mu}_i(t) - \mu_i| \\
 &= |\hat{\mu}_i^{(2)}(t) - \mu_i^{(2)}| + |\hat{\mu}_i^2(t) - \mu_i^2| + \rho|\hat{\mu}_i(t) - \mu_i| \\
 &\geq |\hat{\mu}_i^{(2)}(t) - \mu_i^{(2)} - \hat{\mu}_i^2(t) + \mu_i^2 - \rho\hat{\mu}_i(t) + \rho\mu_i| \\
 &= |\hat{\mu}_i^{(2)}(t) - \hat{\mu}_i^2(t) - \rho\hat{\mu}_i(t) - \mu_i^{(2)} + \mu_i^2 + \rho\mu_i| \\
 &= |\hat{\sigma}_i^2(t) - \rho\hat{\mu}_i(t) - \sigma_i^2 + \rho\mu_i| = |\widehat{\text{MV}}_i(t) - \text{MV}_i|
 \end{aligned}$$

From the final result of the above derivation, it follows that under the event \mathcal{E} , the inequality $\forall i, \forall t, |\widehat{\text{MV}}_i(t) - \text{MV}_i| < (3 + \rho)\beta_i(t)$ holds with probability at least $1 - \delta$. ■

ANALYSIS FOR FIXED CONFIDENCE SETTING

We now present the main theoretical guarantee of RAMGapEc in the fixed-confidence setting, showing that the returned solution is an ε -Pareto set with high probability.

We first prove that RAMGapEc terminates in finite time.

Theorem 10 (termination of RAMGapEc) *For the stopping time of RAMGapEc $\tau := \inf \{t > 0 | V(t) < \epsilon\}$,*

$$\mathbb{P}[\tau < \infty] = 1.$$

Proof In this proof, as in the proof of Theorem 8, we use the notation $\beta_n = \sqrt{4 \log(8K(\log_2 T_i(t))^2/\delta)/T_i(t)}$. At time round t , let n be the smaller one of $T_{m_t}(t)$ and $T_{p_t}(t)$. Then, if $m_t \in \hat{D}_t^+$, we have:

$$\begin{aligned} \min \left\{ \bar{\mu}_{p_t}(t) - \underline{\mu}_{m_t}(t), \bar{\xi}_{m_t}(t) - \underline{\xi}_{p_t}(t) \right\} &\leq \min \{ \hat{\mu}_{p_t} - \hat{\mu}_{m_t}, \hat{\xi}_{m_t} - \hat{\xi}_{p_t} \} + \beta_{T_{m_t}(t)} + \beta_{T_{p_t}(t)} \\ &\leq 2\beta_n. \end{aligned}$$

On the other hand, if $m_t \notin \hat{D}_t^+$, then:

$$\begin{aligned} \max \left\{ \bar{\mu}_{m_t}(t) - \underline{\mu}_{p_t}(t), \bar{\xi}_{p_t}(t) - \underline{\xi}_{m_t}(t) \right\} &\leq \max \{ \hat{\mu}_{p_t} - \hat{\mu}_{m_t}, \hat{\xi}_{m_t} - \hat{\xi}_{p_t} \} + \beta_{T_{m_t}(t)} + \beta_{T_{p_t}(t)} \\ &\leq 2\beta_n. \end{aligned}$$

Therefore, we have $V(t) = V_{m_t}(t) \leq 2\beta_n$.

Since $\beta_n \rightarrow 0$ as $n \rightarrow \infty$, there exists some N_0 such that for all $n \geq N_0$, $V(t) < \epsilon$.

Assume that RAMGapEc has not yet stopped at time t , that is, $V_{m_t}(t) > \epsilon$. Then, at least one of $T_{m_t}(t)$ or $T_{p_t}(t)$ is less than N_0 . Since RAMGapEc selects the arm among m_t and p_t that has been pulled fewer times, the selected arm always has been pulled fewer than N_0 times. Therefore, RAMGapEc must stop by time round $t = KN_0$ at the latest. ■

Theorem 11 (correctness of RAMGapEc) *Let $\varepsilon > 0$ and $\delta \in (0, 1)$ be user-specified accuracy and confidence parameters, respectively. Then, the output \hat{D}_n^+ of RAMGapEc satisfies:*

$$\mathbb{P} \left[\hat{D}_n^+ \text{ is an } \varepsilon\text{-Pareto set} \right] \geq 1 - \delta.$$

Proof Let \mathcal{E} denote the high-probability event under which the empirical mean and second moment estimates are within the confidence intervals, as defined in Eq. 10. Theorem 8 ensures that this event holds with probability at least $1 - \delta$, i.e.,

$$\mathbb{P}[\mathcal{E}] \geq 1 - \delta.$$

On event \mathcal{E} , the true values μ_i and ξ_i of each arm $i \in [K]$ lie within the respective confidence bounds constructed by RAMGapEc at each round t . In particular, for all $t \geq 2K + 1$ and $i \in [K]$, we have:

$$\mu_i \in [\hat{\mu}_i(t) \pm \beta_i(t)], \quad \xi_i \in [\hat{\xi}_i(t) \pm \beta_i(t)].$$

From Theorem 7, we know that if $V(t) < \varepsilon$, then the current set \widehat{D}_t^+ is guaranteed to be an ε -Pareto set under event \mathcal{E} . RAMGapEc terminates at the first time \tilde{n} such that $V(\tilde{n}) < \varepsilon$, and thus returns $\widehat{D}_{\tilde{n}}^+$.

Therefore, on event \mathcal{E} , the returned set $\widehat{D}_{\tilde{n}}^+$ satisfies the ε -Pareto optimality condition. Combining this with the probability bound on \mathcal{E} , we conclude:

$$\mathbb{P}\left[\widehat{D}_{\tilde{n}}^+ \text{ is an } \varepsilon\text{-Pareto set}\right] \geq \mathbb{P}[\mathcal{E}] \geq 1 - \delta.$$

■

Appendix C. Arm Setting for Experiments

In this section, we describe the parameters of the arms used in the experiments.

Table 1: (a, b) values of each pattern for Experiment 3

Index	(a, b)	Index	(a, b)
1	(0.6462, 0.4308)	2	(0.9146, 0.6684)
3	(0.3139, 0.2511)	4	(3.7333, 3.2667)
5	(0.2028, 0.1940)	6	(0.4050, 0.4234)
7	(1.1172, 1.2768)	8	(1.6569, 2.0712)
9	(9.8779, 13.5171)	10	(0.0800, 0.1200)

Table 2: (a, b) values of each pattern for Experiment 4

Index	(a, b)	Index	(a, b)	Index	(a, b)
1	(3.1125, 2.0750)	2	(0.3721, 0.2502)	3	(0.7343, 0.4978)
4	(2.7665, 1.8914)	5	(1.5858, 1.0933)	6	(0.1750, 0.1217)
7	(0.4844, 0.3396)	8	(7.4555, 5.2703)	9	(0.5855, 0.4173)
10	(0.2011, 0.1445)	11	(1.0850, 0.7864)	12	(0.1700, 0.1242)
13	(1.0451, 0.7700)	14	(0.8003, 0.5946)	15	(0.4112, 0.3080)
16	(0.9397, 0.7098)	17	(0.1496, 0.1139)	18	(0.7701, 0.5913)
19	(0.1644, 0.1273)	20	(0.1433, 0.1118)	21	(0.7420, 0.5839)
22	(4.1542, 3.2963)	23	(0.2368, 0.1894)	24	(8.1278, 6.5557)
25	(0.9541, 0.7758)	26	(0.8345, 0.6842)	27	(1.3518, 1.1174)
28	(0.5351, 0.4459)	29	(1.3968, 1.1735)	30	(0.2526, 0.2140)
31	(0.4995, 0.4266)	32	(0.1937, 0.1668)	33	(0.2171, 0.1885)
34	(10.6034, 9.2780)	35	(0.2873, 0.2535)	36	(0.5752, 0.5115)
37	(2.0854, 1.8697)	38	(0.5376, 0.4859)	39	(0.4057, 0.3697)
40	(0.4173, 0.3834)	41	(0.7743, 0.7170)	42	(1.1156, 1.0416)
43	(0.1817, 0.1710)	44	(0.3196, 0.3033)	45	(0.6128, 0.5861)
46	(0.3385, 0.3265)	47	(0.6274, 0.6099)	48	(2.5964, 2.5444)
49	(4.8632, 4.8046)	50	(1.2070, 1.2021)	51	(2.3015, 2.3108)
52	(1.0011, 1.0134)	53	(0.1484, 0.1514)	54	(0.5733, 0.5897)
55	(4.3481, 4.5092)	56	(5.7451, 6.0063)	57	(2.1309, 2.2458)
58	(3.1815, 3.3804)	59	(0.2459, 0.2633)	60	(5.0978, 5.5048)
61	(11.4694, 12.4856)	62	(1.3892, 1.5246)	63	(0.5330, 0.5897)
64	(2.8987, 3.2332)	65	(1.7749, 1.9959)	66	(1.4193, 1.6090)
67	(0.4206, 0.4807)	68	(0.7348, 0.8466)	69	(1.2393, 1.4396)
70	(0.2078, 0.2433)	71	(0.2217, 0.2617)	72	(0.8207, 0.9769)
73	(0.7387, 0.8865)	74	(0.1879, 0.2274)	75	(0.3771, 0.4600)
76	(0.5618, 0.6909)	77	(0.2282, 0.2830)	78	(0.1043, 0.1303)
79	(0.2410, 0.3037)	80	(1.8028, 2.2908)	81	(3.5191, 4.5084)
82	(0.4082, 0.5273)	83	(0.1622, 0.2112)	84	(0.2656, 0.3488)
85	(0.1022, 0.1354)	86	(1.5086, 2.0139)	87	(0.3365, 0.4530)
88	(1.3058, 1.7721)	89	(0.2386, 0.3266)	90	(0.2890, 0.3988)
91	(0.9339, 1.2993)	92	(0.8941, 1.2543)	93	(0.2227, 0.3151)
94	(0.8204, 1.1703)	95	(1.3010, 1.8714)	96	(0.2659, 0.3856)
97	(2.2502, 3.2913)	98	(2.7231, 4.0166)	99	(1.3219, 1.9663)
100	(6.5372, 9.8058)				

pattern	Index (a, b)									
	6	(0.5683, 0.5941)	7	(0.2695, 0.3080)	8	(3.0829, 3.8536)	9	(0.0928, 0.1270)	10	(0.1366, 0.2050)
24	1	(0.6462, 0.4308)	2	(0.1270, 0.0928)	3	(1.3150, 1.0520)	4	(0.2087, 0.1826)	5	(0.8412, 0.8046)
	6	(3.4377, 3.5940)	7	(0.3832, 0.4379)	8	(10.5295, 13.1619)	9	(0.2306, 0.3156)	10	(1.4383, 2.1574)
25	1	(2.1574, 1.4383)	2	(0.9146, 0.6684)	3	(0.3139, 0.2511)	4	(12.7407, 11.1481)	5	(3.5940, 3.4377)
	6	(0.4050, 0.4234)	7	(0.1141, 0.1304)	8	(0.5052, 0.6315)	9	(0.1536, 0.2101)	10	(0.9091, 1.3636)
26	1	(0.4537, 0.3024)	2	(0.9146, 0.6684)	3	(13.1619, 10.5295)	4	(0.2087, 0.1826)	5	(1.9345, 1.8504)
	6	(0.2854, 0.2983)	7	(1.1172, 1.2768)	8	(0.5052, 0.6315)	9	(0.0928, 0.1270)	10	(2.6857, 4.0286)
27	1	(0.1200, 0.0800)	2	(2.1213, 1.5501)	3	(3.8536, 3.0829)	4	(0.6154, 0.5385)	5	(12.2604, 11.7273)
	6	(0.1940, 0.2028)	7	(0.7631, 0.8722)	8	(0.3585, 0.4482)	9	(0.9823, 1.3443)	10	(0.2085, 0.3127)
28	1	(13.8000, 9.2000)	2	(3.9527, 2.8885)	3	(0.3139, 0.2511)	4	(2.0085, 1.7574)	5	(0.2028, 0.1940)
	6	(0.8046, 0.8412)	7	(0.3832, 0.4379)	8	(0.1043, 0.1303)	9	(0.9823, 1.3443)	10	(0.4308, 0.6462)
29	1	(0.6462, 0.4308)	2	(0.3156, 0.2306)	3	(1.3150, 1.0520)	4	(0.1304, 0.1141)	5	(1.9345, 1.8504)
	6	(11.7273, 12.2604)	7	(0.1826, 0.2087)	8	(0.3585, 0.4482)	9	(0.6684, 0.9146)	10	(2.6857, 4.0286)
30	1	(0.9247, 0.6165)	2	(0.2101, 0.1536)	3	(0.3139, 0.2511)	4	(0.4379, 0.3832)	5	(0.5941, 0.5683)
	6	(1.1770, 1.2305)	7	(3.2667, 3.7333)	8	(1.6569, 2.0712)	9	(0.0928, 0.1270)	10	(9.2000, 13.8000)
31	1	(0.9247, 0.6165)	2	(0.2101, 0.1536)	3	(0.3139, 0.2511)	4	(0.4379, 0.3832)	5	(3.5940, 3.4377)
	6	(0.5683, 0.5941)	7	(1.7574, 2.0085)	8	(0.1043, 0.1303)	9	(9.8779, 13.5171)	10	(0.9091, 1.3636)
32	1	(4.0286, 2.6857)	2	(0.2101, 0.1536)	3	(0.4482, 0.3585)	4	(12.7407, 11.1481)	5	(0.5941, 0.5683)
	6	(1.1770, 1.2305)	7	(1.7574, 2.0085)	8	(0.7175, 0.8969)	9	(0.0928, 0.1270)	10	(0.2085, 0.3127)
33	1	(0.3127, 0.2085)	2	(0.2101, 0.1536)	3	(0.4482, 0.3585)	4	(0.1304, 0.1141)	5	(12.2604, 11.7273)
	6	(3.4377, 3.5940)	7	(0.5385, 0.6154)	8	(1.6569, 2.0712)	9	(0.9823, 1.3443)	10	(0.6165, 0.9247)
34	1	(1.3636, 0.9091)	2	(0.1270, 0.0928)	3	(3.8536, 3.0829)	4	(0.8722, 0.7631)	5	(0.2028, 0.1940)
	6	(0.5683, 0.5941)	7	(0.3832, 0.4379)	8	(0.2511, 0.3139)	9	(1.5501, 2.1213)	10	(9.2000, 13.8000)
Go to next page										

pattern	Index (a, b)									
	6	(1.8504, 1.9345)	7	(0.5385, 0.6154)	8	(0.1690, 0.2113)	9	(0.3314, 0.4536)	10	(0.9091, 1.3636)
47	1	(0.2050, 0.1366)	2	(0.1270, 0.0928)	3	(13.1619, 10.5295)	4	(1.2768, 1.1172)	5	(0.5941, 0.5683)
	6	(0.8046, 0.8412)	7	(3.2667, 3.7333)	8	(0.2511, 0.3139)	9	(0.3314, 0.4536)	10	(1.4383, 2.1574)
48	1	(13.8000, 9.2000)	2	(0.1270, 0.0928)	3	(0.6315, 0.5052)	4	(0.8722, 0.7631)	5	(1.2305, 1.1770)
	6	(0.4050, 0.4234)	7	(0.1826, 0.2087)	8	(1.6569, 2.0712)	9	(2.8885, 3.9527)	10	(0.2085, 0.3127)
49	1	(2.1574, 1.4383)	2	(0.4536, 0.3314)	3	(0.6315, 0.5052)	4	(0.2087, 0.1826)	5	(0.2983, 0.2854)
	6	(3.4377, 3.5940)	7	(1.1172, 1.2768)	8	(0.1043, 0.1303)	9	(0.6684, 0.9146)	10	(9.2000, 13.8000)
50	1	(13.8000, 9.2000)	2	(2.1213, 1.5501)	3	(0.2113, 0.1690)	4	(0.4379, 0.3832)	5	(0.1275, 0.1219)
	6	(0.5683, 0.5941)	7	(0.2695, 0.3080)	8	(1.0520, 1.3150)	9	(2.8885, 3.9527)	10	(0.6165, 0.9247)

Appendix D. Comparison Methods

To evaluate the effectiveness of the proposed method, we selected several representative comparison methods. In this chapter, we provide an overview of these methods.

- **Round-Robin:** A simple uniform sampling strategy that cycles through all arms regardless of observed outcomes. It serves as a fundamental baseline for fair but non-adaptive allocation, and is typically sample-inefficient.
- **Dominated Elimination Round-Robin (DE Round-Robin):** An enhanced version of Round-Robin that progressively eliminates arms empirically dominated in both mean and risk. It aims to reduce unnecessary sampling of clearly suboptimal arms.
- **Least-Important Elimination Round-Robin (LIE Round-Robin):** A fixed-budget strategy that eliminates arms contributing least to the current Pareto frontier. Initially allocates samples uniformly, then gradually focuses on promising arms.
- **Risk-Averse LUCB (RA-LUCB):** An extension of the classical LUCB algorithm to the risk-aware, multi-objective setting. It selects two arms each round—a potentially optimal arm and a challenger—and pulls both to reduce uncertainty near the Pareto frontier.
- **ξ Lower Confidence Bound (ξ -LCB):** A Lower Confidence Bound (LCB)-based approach that targets arms with the lowest risk-adjusted performance, computed via the mean-variance trade-off. It encourages conservative exploration to identify risk-averse Pareto-optimal arms.
- **Hypervolume Improvement-based Pareto set Exploration (HVI-Pareto):** A method based on hypervolume improvement, selecting arms that most contribute to expanding the estimated Pareto front. This promotes diverse and well-distributed sampling across objectives. We set the reference point for hypervolume computation to $(R_\mu, R_\xi) = (0, \frac{0.25}{3+\rho})$, which corresponds to the worst-case scenario under our problem setting: since rewards are scaled to $[0, 1]$, the minimum possible mean is 0 and the maximum possible variance is 0.25. Thus, the maximum value of the risk measure $\xi = \alpha(\sigma^2 - \rho\mu)$ is $\frac{0.25}{3+\rho}$ when $\mu = 0$ and $\sigma^2 = 0.25$.
- **Empirical Gap-based Pareto Set Exploration (EGP):** EGP is a fixed-budget algorithm that aims to efficiently identify the Pareto-optimal arms by leveraging empirical dominance gaps. At each round, for each arm i , an empirical gap value $\widehat{V}_i(t)$ is computed as

$$\widehat{V}_i(t) = \begin{cases} \max_{j \neq i} \min \left(\widehat{\mu}_i(t) - \widehat{\mu}_j(t), \widehat{\xi}_j(t) - \widehat{\xi}_i(t) \right) & \text{if } i \in \widehat{D}_t^+, \\ \min_{j \neq i} \max \left(\widehat{\mu}_j(t) - \widehat{\mu}_i(t), \widehat{\xi}_i(t) - \widehat{\xi}_j(t) \right) & \text{if } i \notin \widehat{D}_t^+. \end{cases}$$

The arm to be pulled is selected as $I(t) := \arg \max_i \left\{ -\widehat{V}_i(t) + \beta_i(t) \right\}$, where $\beta_i(t)$ denotes the confidence width for arm i . This sampling strategy focuses on arms close to the empirical Pareto frontier, effectively reducing uncertainty in the decision boundary and improving the quality of the final selection.

Algorithm 3: Round-Robin

Input: K, a, n, ϵ, ρ Pull each arm $i \in [K]$ twice and update $\hat{\mu}_i(t), \hat{\xi}_i(t), \underline{\mu}_i(t), \bar{\mu}_i(t), \underline{\xi}_i(t), \bar{\xi}_i(t)$ Set $T_i(K) = 2$ for all i , and $t \leftarrow 2K + 1$ **while** $t \leq n$ **do** Identify $i_t := t \bmod K$ Draw $X_{i_t}(T_{i_t}(t) + 1) \sim \nu_{i_t}$ $t \leftarrow t + 1$ Update $\hat{\mu}_{i_t}(t), \hat{\xi}_{i_t}(t), \beta_{i_t}(t)$, and $T_{i_t}(t)$

; // (Fixed-Budget Setting)

if $t > n$ **then** | **break** **end**

; // (Fixed-Confidence Setting)

if $t > 2K$ **and** $V(t) < \epsilon$ **then** | **break** **end****end****return** \hat{D}_n^+

Algorithm 4: Dominated Elimination Round-Robin (DE Round-Robin)

Input: K, ϵ, ρ Pull each arm $i \in [K]$ twice and update $\hat{\mu}_i(t), \hat{\xi}_i(t), \underline{\mu}_i(t), \bar{\mu}_i(t), \underline{\xi}_i(t), \bar{\xi}_i(t)$ Set $T_i(K) = 2$ for all i , $t \leftarrow 2K + 1$ and $I \leftarrow 1$ **while** $V(t) > \epsilon$ **do** **for** $i = I \bmod K$ **do** **if** $i \in \hat{D}_t^+ \cup \left\{ k \notin \hat{D}_t^+ \mid V_k(t) > \epsilon \right\}$ **then** | Draw $X_i(T_i(t) + 1) \sim \nu_i$ | Update $\hat{\mu}_i(t), \hat{\xi}_i(t), \underline{\mu}_i(t), \bar{\mu}_i(t), \underline{\xi}_i(t), \bar{\xi}_i(t)$ | $t \leftarrow t + 1$ | $I \leftarrow I + 1$ **end** **else** | $I \leftarrow I + 1$ **end** **end****end****return** \hat{D}_n^+

Algorithm 5: Least-important elimination Round-Robin (LIE Round-Robin)

Input: K, a, n, ϵ, ρ
 Pull each arm $i \in [K]$ twice and update $\hat{\mu}_i(t), \hat{\xi}_i(t), \underline{\mu}_i(t), \bar{\mu}_i(t), \underline{\xi}_i(t), \bar{\xi}_i(t)$
 Set $T_i(K) = 2$ for all i , $t \leftarrow 2K + 1$ and $I \leftarrow 1$
while $t \leq n$ **do**
 Identify $i_I := I \bmod K$
 if $i_I \neq \arg \min_{i \in [K]} V_i(t)$ **then**
 Draw $X_{i_I}(T_{i_I}(t) + 1) \sim \nu_{i_I}$
 $t \leftarrow t + 1$
 $I \leftarrow I + 1$
 Update $\hat{\mu}_{i_I}(t), \hat{\xi}_{i_I}(t), \beta_{i_I}(t)$, and $T_{i_I}(t)$
 end
 else
 $I \leftarrow I + 1$
 end
 if $t > n$ **then**
 break
 end
end
return \hat{D}_n^+

Algorithm 6: Risk-Averse Lower and Upper Confidence Bounds (RA-LUCB)

Input: K, a, n, ϵ, ρ
 Pull each arm $i \in [K]$ twice and update $\hat{\mu}_i(t), \hat{\xi}_i(t), \underline{\mu}_i(t), \bar{\mu}_i(t), \underline{\xi}_i(t), \bar{\xi}_i(t)$
 Set $T_i(K) = 2$ for all i and $t \leftarrow 2K + 1$
while $t \leq n$ **do**
 Identify m_t and p_t by Eq. 6 and Eq. 7
 Draw $X_{m_t}(T_{m_t}(t) + 1) \sim \nu_{m_t}$
 Draw $X_{p_t}(T_{p_t}(t) + 1) \sim \nu_{p_t}$
 $t \leftarrow t + 2$
 Update $\hat{\mu}_{m_t}(t), \hat{\mu}_{p_t}(t), \hat{\xi}_{m_t}(t), \hat{\xi}_{p_t}(t), \beta_{m_t}(t), \beta_{p_t}(t), T_{m_t}(t)$, and $T_{p_t}(t)$
 ; // (Fixed-Budget Setting)
 if $t > n$ **then**
 break
 end
 ; // (Fixed-Confidence Setting)
 if $t > 2K \wedge V(t) < \epsilon$ **then**
 break
 end
end
return \hat{D}_n^+

Algorithm 7: ξ Lower Confidence Bound (ξ -LCB)

Input: K, a, n, ϵ, ρ Pull each arm $i \in [K]$ twice and update $\hat{\mu}_i(t), \hat{\xi}_i(t), \underline{\mu}_i(t), \bar{\mu}_i(t), \underline{\xi}_i(t), \bar{\xi}_i(t)$ Set $T_i(K) = 2$ for all i and $t \leftarrow 2K + 1$ **while** $t \leq n$ **do** Identify $i_t := \arg \min_{i \in [K]} \underline{\xi}_i(t)$ Draw $X_{i_t}(T_{i_t}(t) + 1) \sim \nu_{i_t}$ $t \leftarrow t + 1$ Update $\hat{\mu}_{i_t}(t), \hat{\xi}_{i_t}(t), \beta_{i_t}(t)$, and $T_{i_t}(t)$ **if** $t > n$ **then** | **break** **end****end****return** \hat{D}_n^+

Algorithm 8: Hypervolume Improvement-based Pareto set Exploration (HVI-Pareto)

Input: K, a, n, ϵ, ρ , reference point (R_μ, R_ξ) Pull each arm $i \in [K]$ twice and update $\hat{\mu}_i(t), \hat{\xi}_i(t), \underline{\mu}_i(t), \bar{\mu}_i(t), \underline{\xi}_i(t), \bar{\xi}_i(t)$ Set $T_i(K) = 2$ for all i and $t \leftarrow 2K + 1$ **while** $t \leq n$ **do** Compute $HVI_i(t) := (\hat{\mu}_i(t) - R_\mu)(R_\xi - \hat{\xi}_i(t))$ for each i Identify $d_t := \arg \min_{i \in \hat{D}_t^+} HVI_i(t)$ and $d'_t := \arg \max_{i \notin \hat{D}_t^+} HVI_i(t)$ Draw $X_{d_t}(T_{d_t}(t) + 1) \sim \nu_{d_t}$ Draw $X_{d'_t}(T_{d'_t}(t) + 1) \sim \nu_{d'_t}$ $t \leftarrow t + 2$ Update $\hat{\mu}_{d_t}(t), \hat{\mu}_{d'_t}(t), \hat{\xi}_{d_t}(t), \hat{\xi}_{d'_t}(t), \beta_{d_t}(t), \beta_{d'_t}(t), T_{d_t}(t)$, and $T_{d'_t}(t)$ **if** $t > n$ **then** | **break** **end****end****return** \hat{D}_n^+

Algorithm 9: Empirical Gap-based Pareto Set Exploration (EGP)

Input: K, a, n, ϵ, ρ

Pull each arm $i \in [K]$ twice and update $\hat{\mu}_i(t), \hat{\xi}_i(t), \underline{\mu}_i(t), \bar{\mu}_i(t), \underline{\xi}_i(t), \bar{\xi}_i(t)$

Set $T_i(K) = 2$ for all i and $t \leftarrow 2K + 1$

while $t \leq n$ **do**

Compute $\hat{V}_i(t) = \begin{cases} \max_{j \neq i} \min \left(\hat{\mu}_i(t) - \hat{\mu}_j(t), \hat{\xi}_j(t) - \hat{\xi}_i(t) \right) & \text{if } i \in \hat{D}_t^+ \\ \min_{j \neq i} \max \left(\hat{\mu}_j(t) - \hat{\mu}_i(t), \hat{\xi}_i(t) - \hat{\xi}_j(t) \right) & \text{if } i \notin \hat{D}_t^+ \end{cases}$, for $i \in [K]$

Identify $I(t) := \operatorname{argmax}_{i \in [K]} \left(-\hat{V}_i(t) + \beta_i(t) \right)$

Draw $X_{I(t)}(T_{I(t)}(t) + 1) \sim \nu_{I(t)}$

$t \leftarrow t + 1$

Update $\hat{\mu}_{I(t)}(t), \hat{\xi}_{I(t)}(t), \beta_{I(t)}(t)$, and $T_{I(t)}(t)$

if $t > n$ **then**

| **break**

end

end

return \hat{D}_n^+

Appendix E. Additional Experimental Results

E.1. Performance under Unbounded Gaussian Rewards

In this section, we provide additional experimental results to further demonstrate performance and robustness when rewards are not bounded in $[0, 1]$. Here, we replicate Experiment 1 (Stopping Time Comparison) using Gaussian rewards.

For each arm $i \in [K]$, the reward distribution is given by

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2),$$

where the mean μ_i is in $[0.4, 0.6]$ and the variance σ_i^2 is in $[0.01, 0.2]$, consistent with the setting of the original Beta distribution experiments.

CONFIDENCE INTERVAL FOR MEAN, VARIANCE, AND MV

For unbounded Gaussian rewards, we derive high-probability confidence intervals for mean and variance. In the following, we employ a Hoeffding-type union bound over time that is less tight than what we presented for bounded rewards in the main text (Note that tighter bounds can further improve the RAMGapE performance). Here, we applied the concentration inequalities introduced in [Wainwright \(2019\)](#), in which the mean is treated using the standard sub-Gaussian concentration inequality and for the variance we used the centered squared deviations $Y_i := (X_i - \mu_i)^2 - \sigma_i^2$, and (ν^2, B) -sub-exponential property. Following [Honorio and Jaakkola \(2014\)](#), we adopt the sub-exponential parameters $(\nu, B) = (4\sqrt{2}\sigma_i^2, 4\sigma_i^2)$ when applying the Bernstein-type inequality to derive the confidence interval for the variance.

As results, for each arm i , with probability at least $1 - \delta$, the following confidence intervals are obtained for mean μ_i and variance σ_i^2 :

$$\begin{aligned} |\hat{\mu}_i(t) - \mu_i| &\leq \sigma_{\max} \sqrt{\frac{2}{T_i(t)} \ln \frac{4KT_i^2(t)}{\delta}}, \\ |\hat{\sigma}_i^2(t) - \sigma_i^2| &\leq 8\sigma_{\max}^2 \max \left(\sqrt{\frac{1}{T_i(t)} \ln \frac{4KT_i^2(t)}{\delta}}, \frac{1}{T_i(t)} \ln \frac{4KT_i^2(t)}{\delta} \right) + \frac{2\sigma_{\max}^2}{T_i(t)} \ln \frac{4KT_i^2(t)}{\delta}, \end{aligned}$$

where $\hat{\mu}_i(t)$ and $\hat{\sigma}_i^2(t)$ are the empirical mean and variance over $T_i(t)$ samples. In the derivation, the formal solutions are acquired in terms of the true σ_i at the location of σ_{\max} in the right-hand sides. Because the true σ_i is unknown a priori to the algorithm, we replace it with a possible maximum value of σ_i , keeping the property of upper bounds. In the current experimental setup, we can simply use $\sigma_{\max} = \sqrt{0.2}$.

Finally, we can derive the corresponding confidence interval for MV for Gaussian rewards. For each arm i and for each round t , we denote the upper and lower bounds of MV_i as $\overline{MV}_i(t) := \overline{\sigma}_i(t) - \rho \underline{\mu}_i(t)$ and $\underline{MV}_i(t) := \underline{\sigma}_i^2(t) - \rho \overline{\mu}_i(t)$ (recall $\rho > 0$), where the underlines and overlines denote the lower and upper confidence bounds of the respective quantities.

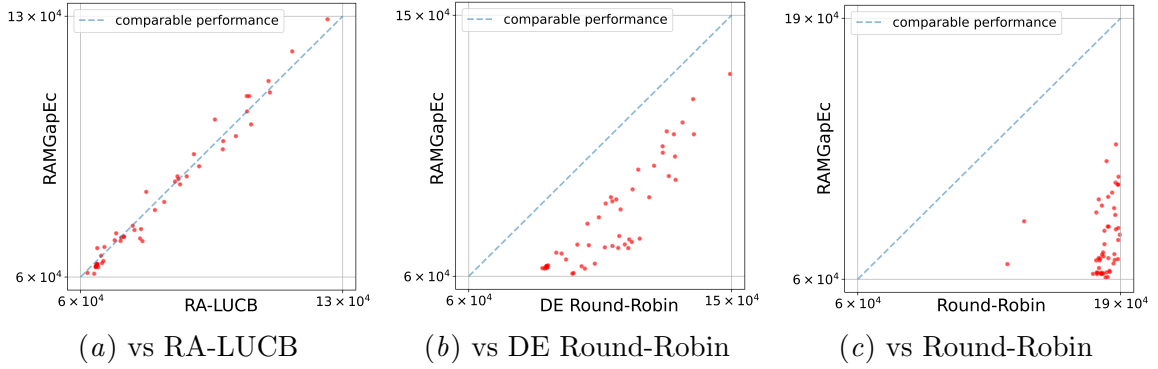


Figure 6: Stopping Time Comparison of Experiment 1 with $(\delta, \epsilon, \rho) = (0.05, 0.1, 0.01)$ for Gaussian rewards. The blue dashed line corresponds to the identity line, i.e., the set of points where both methods terminate at the same time, indicating comparable performance. Points located below this line signify that the proposed method stops earlier than the baseline.

EXPERIMENT PART

We replicate Experiment 1 (Stopping Time Comparison) under Gaussian rewards. The setup is identical to that of Experiment 1 with Beta distributions, except for the reward generation process. Specifically, we conduct simulations across 50 problem instances, each comprising $K = 10$ arms. For each instance, the mean and variance of arms are drawn from the patterns detailed in Table 4. The algorithmic parameters are fixed at $(\delta, \epsilon, \rho) = (0.05, 0.1, 0.01)$.

The results are shown in Fig. 6. The plot demonstrates a trend consistent with our findings for Beta rewards: RAMGapEc terminates significantly faster than the Round-Robin-based baseline algorithms, and as fast as RA-LUCB. This result supports that RAMGapE is a robust algorithm whose effectiveness is not necessarily confined to bounded reward settings.

pattern		Index (μ, σ^2)									
47	6	(0.4889, 0.0522)	7	(0.4667, 0.1156)	8	(0.4444, 0.1789)	9	(0.4222, 0.1367)	10	(0.4000, 0.0733)	
	1	(0.6000, 0.1789)	2	(0.5778, 0.2000)	3	(0.5556, 0.0100)	4	(0.5333, 0.0733)	5	(0.5111, 0.1156)	
	6	(0.4889, 0.0944)	7	(0.4667, 0.0311)	8	(0.4444, 0.1578)	9	(0.4222, 0.1367)	10	(0.4000, 0.0522)	
48	1	(0.6000, 0.0100)	2	(0.5778, 0.2000)	3	(0.5556, 0.1156)	4	(0.5333, 0.0944)	5	(0.5111, 0.0733)	
	6	(0.4889, 0.1367)	7	(0.4667, 0.1789)	8	(0.4444, 0.0522)	9	(0.4222, 0.0311)	10	(0.4000, 0.1578)	
49	1	(0.6000, 0.0522)	2	(0.5778, 0.1367)	3	(0.5556, 0.1156)	4	(0.5333, 0.1789)	5	(0.5111, 0.1578)	
	6	(0.4889, 0.0311)	7	(0.4667, 0.0733)	8	(0.4444, 0.2000)	9	(0.4222, 0.0944)	10	(0.4000, 0.0100)	
50	1	(0.6000, 0.0100)	2	(0.5778, 0.0522)	3	(0.5556, 0.1789)	4	(0.5333, 0.1367)	5	(0.5111, 0.2000)	
	6	(0.4889, 0.1156)	7	(0.4667, 0.1578)	8	(0.4444, 0.0733)	9	(0.4222, 0.0311)	10	(0.4000, 0.0944)	

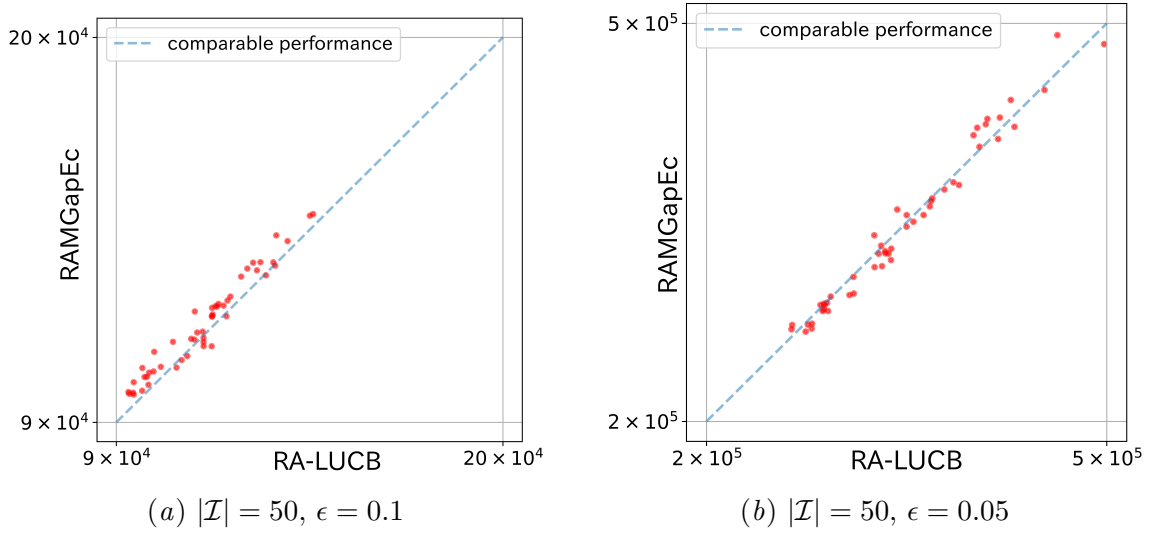


Figure 7: **Stopping time between RAMGapE and RA-LUCB for different numbers of problem instances ($|\mathcal{I}|$) and tolerance levels (ϵ).**

E.2. Stopping time between RAMGapE and RA-LUCB based on ϵ

To further investigate the comparative performance of RAMGapE against RA-LUCB in the fixed-confidence setting, we conducted additional experiments based on Experiment 1 by varying the tolerance level, ϵ .

Fig. 7 presents scatter plots of stopping times for three settings:

- (a) $|\mathcal{I}| = 50, \epsilon = 0.1$: This is identical to the setting in Experiment 1.
- (b) $|\mathcal{I}| = 50, \epsilon = 0.05$: We used the original 50 instances but with a smaller tolerance ϵ to create a more challenging identification task.

As shown in the plots, the points remain tightly clustered around the identity line ($y = x$) in all settings. This indicates that there are no statistically significant differences in sample efficiency between RAMGapE and RA-LUCB, even when the problem is made harder with smaller ϵ . This result confirms our initial interpretation that both algorithms perform almost equally in the fixed-confidence setting we tested.

E.3. Comparison of Pulling Ratios of Pareto and non-Pareto Arms

To support the analysis in the main paper regarding RAMGapE’s exploration strategy, this section provides a detailed visualization of the pulling ratios. Figs. 8 and 9 show the proportion of samples allocated to true Pareto-optimal arms (in red) versus non-Pareto arms (in blue) for each trial in Experiments 3 and 4, respectively.

These figures illustrate the outcome of RAMGapE’s adaptive exploration. While other algorithms may continue to explore suboptimal arms or overly exploit a subset of arms, RAMGapE efficiently prunes non-Pareto arms. As a result, it ultimately allocates a significantly higher proportion of its budget to the Pareto set. The longer red bars for RAMGapE

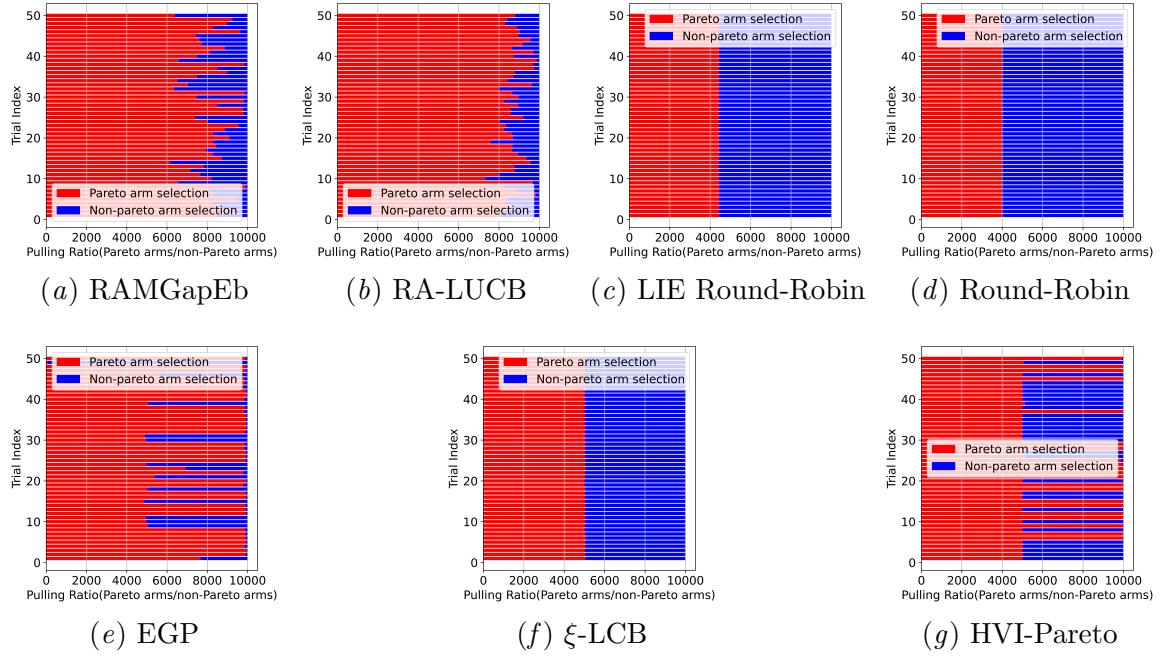


Figure 8: **Comparison of pulling ratios of Pareto and non-Pareto arms in Experiment 3.** The vertical and horizontal axes correspond to the trial index and the pulling ratio about Pareto and non-Pareto arms per $T = 10,000$, respectively. For each trial, the total number of pulling Pareto-optimal (in red), and non-Pareto arms (in blue) within $T = 10,000$ is shown as bars for each method. The longer the red bars, the more frequently Pareto-optimal arms were pulled, which indicates that the algorithm focuses more effectively on the exploration and exploitation of Pareto optimal solutions.

confirm that its strategy effectively focuses the sampling effort on the most promising regions of the decision space, which is a key factor for its strong performance.

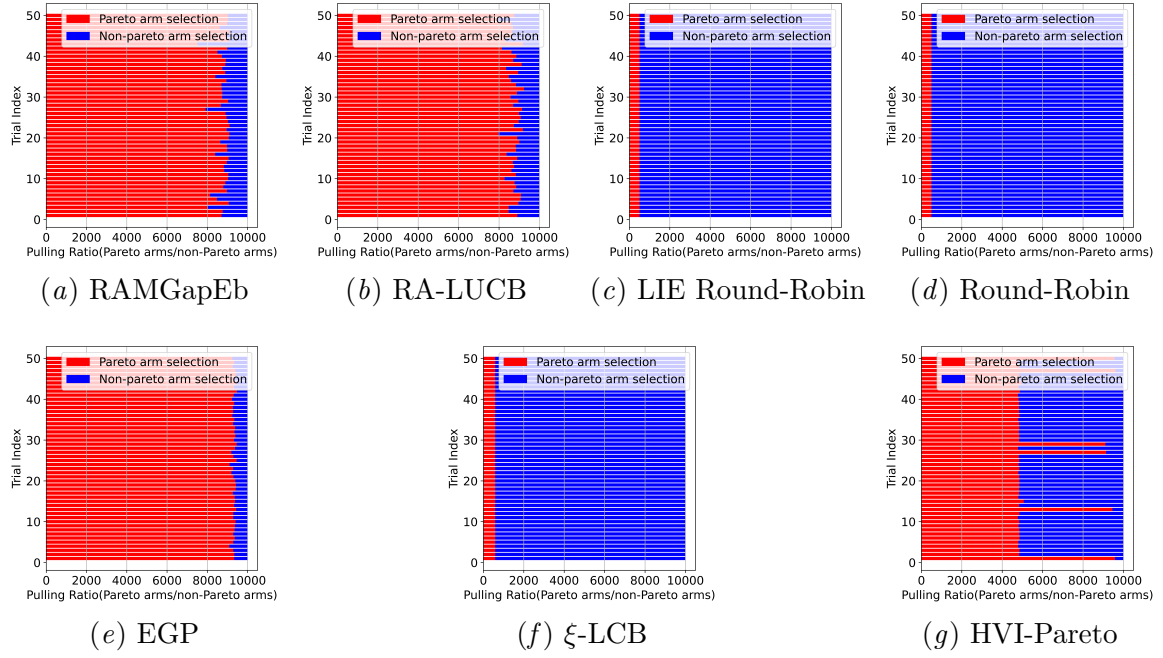


Figure 9: **Comparison of pulling ratios for Pareto and non-Pareto arms in Experiment 4.** The meanings of the plot is the same as Fig. 8 except $K = 100$ arms.