Probably Anytime-Safe Stochastic Combinatorial Semi-Bandits

Yunlong Hou [1] **Vincent Y. F. Tan** [1 2 3] **Zixin Zhong** [4]

Abstract

Motivated by concerns about making online decisions that incur undue amount of risk at each time step, in this paper, we formulate the probably anytime-safe stochastic combinatorial semi-bandits problem. In this problem, the agent is given the option to select a subset of size at most K from a set of L ground items. Each item is associated to a certain mean reward as well as a variance that represents its risk. To mitigate the risk that the agent incurs, we require that with probability at least $1 - \delta$, over the entire horizon of time T, each of the choices that the agent makes should contain items whose sum of variances does not exceed a certain variance budget. We call this probably anytime-safe constraint. Under this constraint, we design and analyze an algorithm PASCombUCB that minimizes the regret over the horizon of time T. By developing accompanying information-theoretic lower bounds, we show that under both the problem-dependent and problem-independent paradigms, PASCombUCB is almost asymptotically optimal. Experiments are conducted to corroborate our theoretical findings. Our problem setup, the proposed PASCombUCB algorithm, and novel analyses are applicable to domains such as recommendation systems and transportation in which an agent is allowed to choose multiple items at a single time step and wishes to control the risk over the whole time horizon.

[1]Department of Mathematics, National University of Singapore, Singapore [2]Department of Electrical and Computer Engineering, National University of Singapore, Singapore [3]Institute of Operations Research and Analytics, National University of Singapore, Singapore [4]Department of Computing Science, University of Alberta, Canada. Correspondence to: Zixin Zhong <zzhong10@ualberta.ca>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

1. Introduction

Audrey, a burgeoning social media influencer, makes profits by posting advertisements (ads) under her account. The advertiser pays her only if an ad is clicked. Having taken a class in online optimization, Audrey aims to leverage the theory of bandit algorithms to design an exploration-exploitation strategy to ensure that the expected number of clicks of the ads she has posted is maximized. Since the platform is space-limited, Audrey can only post no more than K out of L available ads everyday. Some of these ads, however, include an innocuous-looking lottery or voucher that asks the viewer of the social media platform to provide personal information that may lead to fraud or information leakage. If a user clicks it and becomes a victim of fraud, this may damage Audrey's reputation. Audrey thus has to be circumspect in which and how many ads she posts.

On the one hand, Audrey wants to post as many ads that she believes to have high click-through rates as possible; the *expected reward* she obtains is then the sum of expected rewards of the individual ads. On the other hand, she should balance this with the *total risk* of the ads that are posted over a period of time; similarly, the risk of a set of ads posted is modeled as the sum of the risks of the individual ads. How should Audrey plan the posts of her ads over a period of time to learn their individual expected rewards and risks to ensure that her total expected reward is maximized and, at the same time, with high probability, the risk incurred *at any point in time* in her exploration-exploitation strategy is bounded by some fixed permissible threshold?

In addition to influencers like Audrey, online platforms that make profits by advertising such as YouTube and TikTok also encounter similar problems. We are therefore motivated to formulate the *probably anytime-safe stochastic combinatorial semi-bandits* (PASSCSB) problem which is a *regret minimization* problem with an anytime safety constraint. More precisely, we aim to design and analyze the performance of an algorithm that, with high probability, ensures that the risk (as measured by the variance) *at any time* step is below a given threshold and whose regret is minimized.

Literature review. There is a large body of works that take risk into account while conducting the exploration and/or exploitation of the unknown reward distributions in the stochastic multi-armed bandits (MABs) literature.

Under the risk-constrained pure exploration framework, Hou et al. (2023) and David et al. (2018) attempted to identify the optimal arm within those low-risk (based on their variances or α-quantiles) arms with probability at least $1-\delta$.

Under the *risk-aware* regret minimization setup, Sani et al. (2012), Vakili & Zhao (2016) and Zhu & Tan (2020) consider the mean-variance as the measure to be minimized over a fixed time horizon. Cassel et al. (2018) provided a general and systematic instruction to analyzing risk-aware MABs, i.e., the risk was incorporated in the *Empirical Distribution Performance Measure* and the U-UCB algorithm is adopted to perform "proxy regret minimization". While these risk-aware algorithms reduce the overall risk during the exploration and exploitation process, the risk is not strictly enforced to be below a prescribed threshold; rather the risk measure is penalized within the objective function, similarly to a Lagrangian. Another setup similar to the risk-aware setup is the *constrained* bandits regret minimization. Mahdavi et al. (2012) required that the number of times the constraint can only be violated is at most sublinear in the horizon T. Kagrecha et al. (2023) proposed a CVaR constraint and performed exploration on the feasible arm, followed by exploration among the feasible arm set. Unlike our formulation, these algorithm are permitted to sample risky arms during exploration.

A more stringent constraint can be found in the literature on *conservative bandits* (Wu et al., 2016), which requires the cumulative return at any time step to be above a constant fraction of the return resulting from repeatedly sampling the base arm. Kazerouni et al. (2017) extended this setup to conservative contextual linear bandits and this was further improved by Garcelon et al. (2020). A similar problem is *bandits with knapsacks* (Badanidiyuru et al., 2018), which imposes a budget on the cumulative consumed resources and the algorithm stops when the budget is depleted.

The most stringent constraint can be found in the *safe bandits* problem. Khezeli & Bitar (2020) and Moradipari et al. (2020) presented the SEGE, SCLUCB, and SCLTS algorithms to tackle this problem. This problem demands that the expected reward of the pulled arm at each time step be greater than a prescribed threshold with high probability, i.e., the *"stagewise safety constraint"*. The authors assumed the convexity (continuity) of the arm set and performed exploration around the explored safe arms, starting from a baseline safe arm. This continuity of the (super) arm set does not hold under the combinatorial semi-bandits setup. More comparisons are presented in App. E.

For the (unconstrained) combinatorial semi-bandits (CSB) setup, Chen et al. (2013) presented a UCB-type algorithm ComUCB1 to balance the trade-off between exploration and exploitation. Kveton et al. (2015) improved the analysis of ComUCB1 and achieved a tight upper bound (within a specific set of instances). Kveton et al. (2014) introduced matroid structure to CSB and leveraged the matroid structure to design and analyze a greedy algorithm OMM. The risk-aware CSB problem is less studied by the community. Ayyagari & Dukkipati (2021) utilized CVaR as the risk-aware measure within the CSB problem, where the risk constraint was not explicitly specified.

We observe that the existing literature mentioned above are not directly applicable to Audrey, while our setting (described formally below) dovetails neatly with her problem. Audrey can utilize our algorithm to sequentially and adaptively select different sets of ads everyday and almost always (i.e., with high probability) avoids sets of ads with unacceptably high risks. Beyond any specific applications, we believe that this problem is of fundamental theoretical importance in the broad context of regret minimization in combinatorial multi-armed bandits.

Main Contributions. Our first contribution lies at the formulation of a novel PASSCSB problem. In the PASSCSB problem, there are L items with different reward distributions. At each time step, a random reward is generated from each item's distribution. Based on the previous observations, the learning agent selects a *solution* at each time step. A solution consists of at most K items. The expected return (variance) of a solution is the summation of the reward (variance) of its constituents. Given $T \in \mathbb{N}$, the agent aims to maximize the cumulative return over T time steps and ensure that with probability $1-\delta$ the variance of all selected solutions are below a given threshold.

The key challenge of regret minimization under the PASSCSB lies in handling two distinct tasks—we seek optimality in the mean and safeness in the variance of each chosen solution. Our second contribution is the design and analysis of the Probably Anytime-Safe Combinatorial UCB (or PASComBUCB) algorithm.

Thirdly, we also derive a problem-dependent upper bound on the regret of PASComBUCB, which involves a *hardness* parameter $H(\Delta(\Lambda))$. We see that $H(\Delta(\Lambda))$ characterizes the effectiveness of ascertaining the safety of potential solutions in the regret. To assess the optimality of PASComBUCB, we prove an accompanying problem-dependent lower bound on the regret of any variance-constrained consistent algorithm. The upper and lower problem-dependent bounds match in almost all the parameters (except in K). Additionally, we show that if δ_T decays exponentially fast in T, the problem-dependent regret cannot be logarithmic in T. We further present a problem-independent upper bound on the regret of PASComBUCB and a lower bound for any algorithm. Just as the problem-dependent bounds, these bounds also match in almost all the parameters.

Lastly, experiments are conducted to illustrate the empirical

performance and corroborate our theoretical findings.

In summary, this paper is the first to explore the regret minimization problem in the combinatorial bandits with an *anytime* constraint on the variance. When $\delta \to 1$ and $\bar{\sigma}^2$ is large (so that the optimal safe solution is the one with the highest mean regardless of safety considerations), our problem reduces to the standard combinatorial semi-bandits (Kveton et al., 2015), and the regret incurred by the safety constraint vanishes, resulting in the same upper bound as the unconstrained case. Furthermore, the framework and analysis of PASCOMBUCB can be extended to other risk measures as long as there are appropriate concentration bounds, e.g., Bhat & Prashanth (2019) or Chang & Tan (2022) enables us to use CVaR or certain continuous functions as risk measures within the generic PASCOMBUCB framework.

2. Problem Setup

For $m \in \mathbb{N}$, let $[m] := \{1, 2, \ldots, m\}$. An instance of a *variance-constrained stochastic combinatorial semi-bandit* is a tuple $\Lambda = (E, \mathcal{A}_K, \nu, \bar{\sigma}^2)$. We describe the four elements of Λ in the following. Firstly, the finite set $E = [L]$ is known as the *ground set* in which each $i \in E$ is known as an *item*. Secondly, the family $\mathcal{A}_K \subset \{S \in 2^E : |S| \leq K\}$ is a collection of subsets of E with cardinality at most K. Each element $S \in \mathcal{A}_K$ is known as a *solution* and $\mathcal{A}_K$ satisfies the condition that all subsets of $S \in \mathcal{A}_K$ remain solutions, i.e., $\mathcal{A}_K$ is downward-closed. Thirdly, the vector of probability distributions $\nu = (\nu_1, \nu_2, \ldots, \nu_L)$ contains σ^2-sub-Gaussian distributions $\{\nu_i\}_{i \in E}$ with means $\{\mu_i\}_{i \in E}$ and variances $\{\sigma_i^2\}_{i \in E}$. The final element of an instance $\bar{\sigma}^2 > 0$ denotes the permissible upper bound on the variance. To avoid trivialities, we assume that $\bar{\sigma}^2 > \sigma^2$ and $K \geq 2$.

The *return* of item $i \in E$ is the random variable W_i with distribution ν_i. The *(stochastic) return* of a solution $S \in \mathcal{A}_K$ is $\sum_{i \in S} W_i$ where $W_i \sim \nu_i$. The *expected return* and *variance* of $S \in \mathcal{A}_K$ are

$$\mu_S := \sum_{i \in S} \mu_i \quad \text{and} \quad \sigma_S^2 := \sum_{i \in S} \sigma_i^2$$

respectively. We further assume that every instance Λ satisfies $\sigma_S^2 \neq \bar{\sigma}^2$ for all $S \in \mathcal{A}_K$ and each distribution ν_i is supported in the interval $[0, 1]$.

Define $\mathcal{S} := \{S \in \mathcal{A}_K : \sigma_S^2 < \bar{\sigma}^2\}$ to be the *safe set* which contains all the *safe* solutions. Let the complement of $\mathcal{S}$ be the *unsafe set* $\mathcal{S}^c$. Denote the *optimal safe solution* as $S^\star := \arg\max\{\mu_S : S \in \mathcal{S}\}$ with return $\mu^\star$. For simplicity, we assume that $S^\star$ is unique. Denote the *suboptimal set* $\mathcal{B} := \{S \in \mathcal{A}_K : \mu_S < \mu^\star\}$ and the *risky set* $\mathcal{R} := \{S \in \mathcal{A}_K : \mu_S \geq \mu^\star, S \neq S^\star\}$. For a solution S, let the mean gap $\Delta_S := \mu^\star - \mu_S$ and the variance gap $\Delta_S^{\mathrm{v}} := |\sigma_S^2 - \bar{\sigma}^2|$.

An instance Λ, time horizon $T \in \mathbb{N}$ and confidence parameter $\delta \in (0, 1)$ are specified. An agent, who knows $E, \mathcal{A}_K$ and $\bar{\sigma}^2$ but not the vector of probability distributions ν, interacts adaptively with the instance over T time steps as follows. At time step $t \in [T]$, the agent uses a stochastic function π_t that selects a solution $S_t \in \mathcal{A}_K$ based on the observation history $\mathcal{H}_{t-1} := ((S_s, \{W_i(s)\}_{i \in S_s}))_{s \in [t-1]}$. In other words, $S_t = \pi_t(\mathcal{H}_{t-1})$ is a stochastic function of the history $\mathcal{H}_{t-1}$. The agent receives the random return $\sum_{i \in S_t} W_i(t)$, where $\{W(s) = \{W_i(s)\}_{i \in E}\}_{s \in [T]}$ are i.i.d. according to ν across time. The weights of the selected items $\{W_i(t) : i \in S_t\}$ are observed by the agent at each time $t \in [T]$. The collection of stochastic functions $\pi = \{\pi_t\}_{t \in [T]}$ is known as the agent's *policy*.

The goal of the agent is to minimize *the expected cumulative regret* (or simply *regret*) $\mathrm{Reg}(T)$ over the horizon T, subject to a certain risk constraint. More precisely, the *regret* suffered by a policy π employed by the agent is defined as

$$\mathrm{Reg}^\pi(T) := \mathbb{E}_\pi \left[\sum_{t=1}^{T} \left(\sum_{i \in S^\star} W_i(t) - \sum_{i \in S_t} W_i(t) \right) \right]$$

The policy π should satisfy the condition that all the solutions chosen $\{S_t^\pi\}_{t \in [T]} \subset \mathcal{A}_K$ are safe with probability at least $1 - \delta$, i.e.,

$$\mathbb{P}_\pi \left[\forall\, t \in [T], S_t^\pi \in \mathcal{S} \right] \geq 1 - \delta. \tag{1}$$

This is referred to as the *probably anytime-safe* constraint.

In the problem-dependent lower bounds, we will refer to a certain class of "good" policies that operate as the time horizon $T \to \infty$ and the probability of being safe in the sense of (1) tends to 1. This is formalized in the following.

Definition 2.1. *Fix an instance ν and a vanishing sequence $\{\delta_T\}_{T=1}^\infty \subset (0, 1)$. A policy $\pi = \{\pi_t\}_{t=1}^\infty$ is said to be a $\{\delta_T\}_{T=1}^\infty$-*variance-constrained consistent algorithm *if*

- $\mathrm{Reg}^\pi(T) = o(T^a)$ *for all* $a > 0$ *and*
- $\mathbb{P}_\pi \left[\forall\, t \in [T], S_t^\pi \in \mathcal{S} \right] \geq 1 - \delta_T$ *for all* $T \in \mathbb{N}$.

We often omit the superscripts π in Reg^π, S_t^π (or A_t^π and $A_{t,r}^\pi$ in PASCOMBUCB) and the subscripts π in the probabilities and expectations if there is no risk of confusion.

3. Our Algorithm: PASCOMBUCB

Our algorithm Probably Anytime-Safe Combinatorial UCB (or PASCOMBUCB), presented in Algorithm 1, is designed to satisfy the probably anytime-safe constraint. In particular, we apply (and analyze) the GREEDY-SPLIT subroutine in Line 11; this subroutine has not been involved in an algorithm designed for standard combinatorial semi-bandits such as COMBUCB1 (Chen et al., 2013).

Statistics. Since each item $i \in E$ is σ^2-sub-Gaussian, any solution that contains at most $q := \lfloor \frac{\bar{\sigma}^2}{\sigma^2} \rfloor$ items is safe with probability (w.p.) 1. We call such a solution *absolutely*

Algorithm 1 PASCOMBUCB

1: **Input**: An instance Λ (with unknown ν), the horizon T and the confidence parameter $\delta \in (0,1)$.
2: Set phase counter $p = 1$ and time step counter $t = 1$.
3: **while** $\exists\, i \in E$ such that $T_i(p-1) < 2$ **do**
4: Pull $A_p = \arg\max_{S:|S|\le q} |\{i \in S : T_i(p-1) < 2\}|$.
5: $p \leftarrow p+1$, $t \leftarrow t+1$.
6: **end while**
7: Update the sample mean, sample variance and confidence bounds according to (4).
8: Update the empirically safe set $\mathcal{S}_p$ and possibly safe set $\bar{\mathcal{S}}_p$ according to (5) and (6) respectively.
9: **while** $t < T$ **do**
10: Identify a solution $A_p = \arg\max_{A \in \bar{\mathcal{S}}_{p-1}} U_A^{\mu}(p-1)$.
11: Invoke GREEDY-SPLIT to split the solution A_p into n_p sub-solutions $\{A_{p,1}, \ldots, A_{p,n_p}\} \subset \mathcal{S}_{p-1}$.
12: Set $n_p \leftarrow \min\{n_p, T-t\}$.
13: Choose solution $\{A_{p,1}, \ldots, A_{p,n_p}\}$.
14: Update the statistics of all solutions based on (4).
15: Update the empirical sets based on (5) and (6).
16: Set $t = t + n_p$ and $p = p+1$,
17: **end while**

safe. Algorithm 1 (PASCOMBUCB) is conducted in *phases*, where each phase consists of multiple time steps and each item can be pulled at most once during each phase. Thus we adopt a different notation "A" to denote the solution in our algorithm. Define $T_i(p) := \sum_{s=1}^{p} \mathbb{1}\{i \in A_p\}$ as the number of times item i is pulled up to and including phase p. Denote the sample mean and sample variance of item i at phase p respectively as

$$\hat{\mu}_i(p) := \frac{1}{T_i(p)} \sum_{s=1}^{p} W_i(s) \cdot \mathbb{1}\{i \in A_s\}, \quad \text{and}$$
$$\hat{\sigma}_i^2(p) := \frac{1}{T_i(p)} \sum_{s=1}^{p} \left(W_i(s) - \hat{\mu}_i(p)\right)^2 \cdot \mathbb{1}\{i \in A_s\}.$$

The bound based on the Law of Iterated Logarithms (LIL) is used to construct the confidence radii. For a fixed $\epsilon \in (0,1)$, define $\mathrm{lil}(t,\rho) := (1+\sqrt{\epsilon})\left(\frac{1+\epsilon}{2t} \ln\left(\frac{\ln((1+\epsilon)t)}{\rho}\right)\right)^{1/2}$ and denote the confidence radius for the mean as

$$\alpha(t) := \mathrm{lil}(t, \omega_\mu), \tag{2}$$

where ω_μ is a parameter to be chosen. The confidence radii for the variance are asymmetric about the empirical variance and are parameterized by ω_{v} and ω'_{v} that may not necessarily be the same. They are defined as

$$\beta_{\mathrm{u}}(t) := 3 \cdot \mathrm{lil}(t, \omega_{\mathrm{v}}) \quad \text{and} \quad \beta_{\mathrm{l}}(t) := 3 \cdot \mathrm{lil}(t, \omega'_{\mathrm{v}}). \tag{3}$$

We denote the *upper* and *lower confidence bounds* (UCB and LCB) for the mean of item i as

$$U_i^{\mu}(p) := \hat{\mu}_i(p) + \alpha(T_i(p)) \quad \text{and}$$
$$L_i^{\mu}(p) := \hat{\mu}_i(p) - \alpha(T_i(p))$$

respectively. The UCB and LCB for the variance of item i are defined as

$$U_i^{\mathrm{v}}(p) := \min\{\hat{\sigma}_i^2(p) + \beta_{\mathrm{u}}(T_i(p)), \sigma^2\} \quad \text{and}$$
$$L_i^{\mathrm{v}}(p) := \max\{\hat{\sigma}_i^2(p) - \beta_{\mathrm{l}}(T_i(p)), 0\}$$

respectively. With the sample mean, sample variance, and confidence bounds for the items, we define the following statistics for all solution $S \in \mathcal{A}_K$:

$$\begin{aligned} \hat{\mu}_S(p) &= \sum_{i \in S} \hat{\mu}_i(p), & \hat{\sigma}_S^2(p) &= \sum_{i \in S} \hat{\sigma}_i^2(p), \\ U_S^{\mu}(p) &= \sum_{i \in S} U_i^{\mu}(p), & L_S^{\mu}(p) &= \sum_{i \in S} L_i^{\mu}(p), \\ U_S^{\mathrm{v}}(p) &= \sum_{i \in S} U_i^{\mathrm{v}}(p), & L_S^{\mathrm{v}}(p) &= \sum_{i \in s} L_i^{\mathrm{v}}(p). \end{aligned} \tag{4}$$

Denote the *empirically safe set* as

$$\mathcal{S}_p := \{S \in \mathcal{A}_K : U_S^{\mathrm{v}}(p) < \bar{\sigma}^2\} \tag{5}$$

and the *possibly safe set* as

$$\bar{\mathcal{S}}_p := \{S \in \mathcal{A}_K : L_S^{\mathrm{v}}(p) < \bar{\sigma}^2\}. \tag{6}$$

The solutions in $\mathcal{S}_t$ and $\bar{\mathcal{S}}_t$ are called *empirically safe* and *possibly safe* solutions respectively.

Dynamics. In the *initialization stage* (lines 3 to 6), PASCOMBUCB greedily pulls the absolutely safe solutions. When each item has been pulled at least twice, this stage is terminated. After initialization, during phase p, PASCOMBUCB **firstly** identifies a solution $A_p = \arg\max_{A \in \bar{\mathcal{S}}_p} U_A^{\mu}(p-1)$ via an *optimization oracle* (Line 10). It **then** calls a subroutine GREEDY-SPLIT to greedily partition the solution A_p into empirically safe sub-solutions (Line 11, see Figure 1 for illustration). **Subsequently**, these solutions are chosen and the stochastic rewards from the corresponding items are observed (Line 13). **Lastly**, the empirical estimates, the confidence bounds, and the empirical sets are updated (Lines 14 and 15).

Illustration. Figures 2 and 3 illustrate the regret accumulated during phase p and over the whole T horizon respectively. As shown in Figure 2, the regret accumulated during phase p can be decomposed into two parts

$$\sum_{r=1}^{n_p} (\mu^\star - \mu_{A_{p,r}}) = \Delta_{A_p} + \mu^\star(n_p - 1)$$

where Δ_{A_p} is *the (phase-wise) instantaneous regret due to suboptimality* and $\mu^\star(n_p-1)$ is *the instantaneous regret due to safeness-checking*; the latter term results from

Algorithm 2 GREEDY-SPLIT

Input: A solution A_p and the upper confidence bound on the variance $U^{\mathrm{v}}(p-1)$ at phase $p-1$.
Set $n_p = 1, s = 1$ and $A_{p,1} = \emptyset$.
Index the items in A_p by $i_1, \ldots, i_{|A_p|}$.
while $s \leq |A_p|$ **do**
  if $U^{\mathrm{v}}_{A_{p,n_p}}(p-1) + U^{\mathrm{v}}_{i_s}(p-1) \leq \bar{\sigma}^2$ **then**
    Set $A_{p,n_p} \leftarrow A_{p,n_p} \cup \{i_s\}$.
  else
    $n_p \leftarrow n_p + 1$ and $A_{p,n_p} = \{i_s\}$.
  end if
  $s \leftarrow s + 1$.
end while
return $\{A_{p,1}, \ldots, A_{p,n_p}\}$.

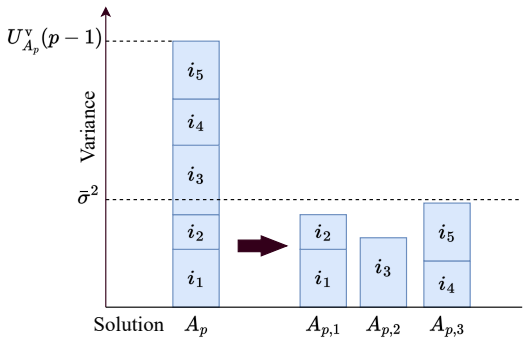


Figure 1. A diagram of a split to a solution A_p containing 5 items.

the safeness constraint. At the beginning, since the upper confidence bounds of the variances of all solutions are large, each solution will be split into up to $2Q$ sub-solutions, where $Q := \lceil \frac{K}{q} \rceil$, and hence the regret due to safeness checking can be large. As the algorithm progresses, we obtain more observations of items and get more confident about their variances ($U^{\mathrm{v}}_i(p)$ decreases). Hence, during some later phase, it suffices to split some solutions into fewer sub-solutions and the regret due to safeness-checking reduces. Furthermore, when most items are sampled sufficiently many times, the unsafe solutions are excluded from the possibly safe set $\bar{\mathcal{S}}_p$, and the only contribution to the regret is via the suboptimality of the solution A_p.

Remark 3.1. *The two parameters ω_{v} and ω'_{v} determine the confidence radii of variances and do not necessarily have to be the same. The confidence parameter ω'_{v} is solely a parameter of* PASCOMBUCB*; its choice does not rely on the confidence parameter δ and only affects $L^{\mathrm{v}}_S(p)$, the lower confidence bound of the variance, which determines when we ascertain a solution to be unsafe. The choice of ω_{v} depends on δ and it influences $U^{\mathrm{v}}_S(p)$, the upper confidence bound of the variance, which guides* PASCOMBUCB *to split the solution to satisfy the probably anytime-safe constraint.*

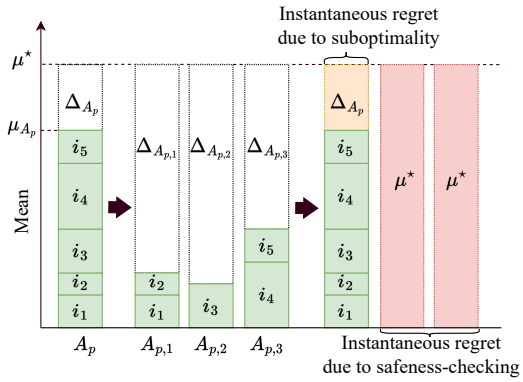


Figure 2. Solution A_p is split into $n_p = 3$ sub-solutions, the instantaneous regret at phase p can be divided into the instantaneous regret due to suboptimality and the instantaneous regret due to safeness-checking.

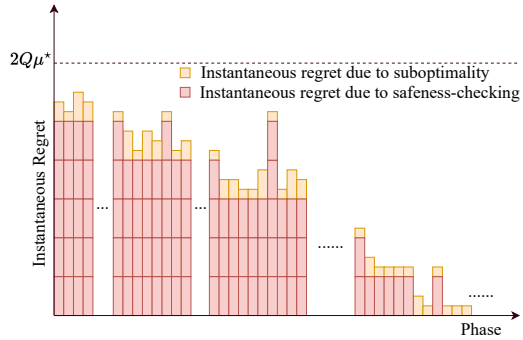


Figure 3. An illustration of the instantaneous regret yielded by PASCOMBUCB. As the variances of the items are more determined, less regret due to safeness-checking is generated.

4. Problem-dependent Bounds

For simplicity, when a time horizon T and a confidence parameter $\delta = \delta_T$ are given, we set the confidence parameters $\omega_\mu = \omega'_{\mathrm{v}} = \frac{1}{T^2}$ and $\omega_{\mathrm{v}} = \frac{\delta_T}{T^2}$.

We introduce various suboptimality gaps that contribute to the regret due to the suboptimality.

- for $i \in E \setminus S^\star$, let the *minimum safe-suboptimal gap* be
$$\Delta_{i,\mathcal{S}\cap\mathcal{B},\min} := \min_{S \ni i, S \in \mathcal{S}\cap\mathcal{B}} \Delta_S;$$
- for $i \in E$, let the *minimum unsafe-suboptimal gap* be
$$\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min} := \min_{S \ni i,\ S \in \mathcal{S}^c\cap\mathcal{B}} \Delta_S;$$
and let the *tension parameter between the mean gap Δ_S and variance gap Δ^{v}_S* be
$$c_i := \max_{S \ni i,\ S \in \mathcal{S}^c\cap\mathcal{B}} \left(\frac{\Delta_S}{\max\{\Delta_S, \Delta^{\mathrm{v}}_S/3\}} \right)^2.$$

We also define following safeness gaps that induce the conservative sampling strategy to guarantee the probably anytime-safe constraint. For $i \in E$, and

- for the risky set $\mathcal{R}$, define *the minimum unsafeness gap* $\Delta^{\mathrm{v}}_{i,\mathcal{R}} := \min_{S \ni i, S \in \mathcal{R}} \Delta^{\mathrm{v}}_S$.
- for the safe and suboptimal set $\mathcal{S} \cap \mathcal{B}$, let

$$\Psi_{i,\mathcal{S}\cap\mathcal{B}} := \max_{S \ni i,\ S \in \mathcal{S}\cap\mathcal{B}} \min\left\{\frac{\ln T}{\Delta_S^2}, \frac{9\ln(T/\delta_T)}{(\Delta^{\mathrm{v}}_S)^2}\right\}$$

  which characterizes the order of the number of times that item i needs to be sampled in order to identify the suboptimality of all safe and suboptimal solutions $A \ni i$ while satisfying the safeness constraint. We further define a variant of $\Psi_{i,\mathcal{S}\cap\mathcal{B}}$ as

$$\Psi'_{i,\mathcal{S}\cap\mathcal{B}} := \max_{S \ni i, S \in \mathcal{S}\cap\mathcal{B}} \min\left\{\frac{\ln T}{\Delta_S^2}, \frac{9\ln(1/\delta_T)}{(\Delta^{\mathrm{v}}_S)^2}\right\}$$

  which will be used to characterize the lower bound.
- for the unsafe and suboptimal set $\mathcal{S}^c \cap \mathcal{B}$, let

$$\Phi_{i,\mathcal{S}^c\cap\mathcal{B}} := \max_{S \ni i,\ S \in \mathcal{S}^c\cap\mathcal{B}} \min\left\{\frac{\ln T}{\Delta_S^2}, \frac{9\ln T}{(\Delta^{\mathrm{v}}_S)^2}\right\}$$

  which characterizes the hardness of identifying the unsafeness of suboptimality of all unsafe and suboptimal solutions that contain item i.

Define $\xi(\omega) := \frac{2+\epsilon}{\epsilon}\left(\frac{\omega}{\ln(1+\epsilon)}\right)^{1+\epsilon}$, where $\epsilon \in (0,1)$ is fixed.

4.1. Problem-dependent Upper Bound

Theorem 4.1 (Problem-dependent upper bound)**.** *Let $\Lambda = (E, \mathcal{A}_K, \nu, \bar{\sigma}^2)$ be an instance and let $\{\delta_T\}_{T=1}^{\infty} \in o(1)$ be a sequence that satisfies $\ln(1/\delta_T) = o(T^b)$ for all $b > 0$ (i.e., $\{\delta_T\}$ is not exponentially decaying). Then,* PASCOMBUCB *is a $\{\delta_T\}_{T=1}^{\infty}$-variance-constrained consistent algorithm. More precisely, given a time budget T, the probably anytime-safe constraint is satisfied and the regret of* PASCOMBUCB $\mathrm{Reg}(T)$ *is upper bounded by*

$$\min\{T\mu^\star, \mathrm{Reg}_1(T) + \mathrm{Reg}_2(T)\} + \mathrm{Reg}_3(T),$$

where

$$\mathrm{Reg}_1(T) = O\Bigg(\sum_{i\in E\setminus S^\star} \frac{K\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}} + \sum_{i\in E} \frac{c_i K \ln T}{\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min}}\Bigg)$$

$$\mathrm{Reg}_2(T) = 2\mu^\star H\left(\Delta(\Lambda)\right), \quad \mathrm{Reg}_3(T) = 2\mu^\star(L+1)$$

where $\Delta(\Lambda) = \{\Delta^{\mathrm{v}}_{S^\star}\} \cup \{\Delta^{\mathrm{v}}_{i,\mathcal{R}}, \Psi_{i,\mathcal{S}\cap\mathcal{B}}, \Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\}_{i\in E}$ and $H\left(\Delta(\Lambda)\right) := H(1,\Lambda)$ is defined in (26) in App. B.4.

Remark 4.2. *If the gaps in $\Delta(\Lambda)$ are sufficiently small and $\delta_T = T^{-\lambda}$ for a fixed $\lambda > 0$,*

$$H\left(\Delta(\Lambda)\right) = O\Bigg(\frac{(\lambda+1)K^2\ln T}{(\Delta^{\mathrm{v}}_{S^\star})^2} + K\sum_{i\in E}\Bigg(\frac{\ln T}{(\Delta^{\mathrm{v}}_{i,\mathcal{R}})^2} + \max_{\substack{S\ni i,\\ S\in\mathcal{S}\cap\mathcal{B}}} \min\left\{\frac{\ln T}{\Delta_S^2}, \frac{(\lambda+1)\ln T}{(\Delta^{\mathrm{v}}_S)^2}\right\} + \Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\Bigg)\Bigg).$$

See (26) for more details of this calculation.

The first term $T\mu^\star$ in the regret bound provides a naïve upper bound for the expected regret conditional on the variance constraint holds. The order of the regret ($o(T^a)$ for all $a > 0$) implies the regret will be asymptotically bounded by the second term when the time budget T is sufficiently large. The second term is comprised of two parts—*the regret due to suboptimality* $\mathrm{Reg}_1(T)$ and *the regret due to safeness-checking* $\mathrm{Reg}_2(T)$. The intuition for the regret due to suboptimality $\mathrm{Reg}_1(T)$ is that

- Each item in any safe and suboptimal solution will be sampled $O(\frac{K\ln T}{\Delta^2_{i,\mathcal{S}\cap\mathcal{B},\min}})$ times to ascertain the suboptimality of all safe and suboptimal solutions to which this item belongs to.
- Each item in an unsafe and suboptimal solution S will be sampled $O\left(\frac{K\ln T}{\max\{\Delta_S, \Delta^{\mathrm{v}}_S/3\}^2}\right)$ times to ascertain either the suboptimality or the unsafeness of S. As this should be done for all the unsafe and suboptimal solutions, we need to take the maximum of the above time complexity. More precisely, when $c_i = 1$, suboptimality identification of the unsafe and suboptimal solutions to which item i belongs dominates the regret; and when $c_i < 1$, the ascertaining of the unsafeness dominates the regret.

The intuition for the regret due to safeness checking $\mathrm{Reg}_2(T)$ is that $H\left(\Delta(\Lambda)\right)$ provides an upper bound for the number of time steps needed for guaranteeing the safeness of all solutions. PASCOMBUCB achieves this in a judicious manner since it does not check the safeness of all the solutions at the start, followed by exploration and exploitation of the possibly high-return safe solutions. Instead, it takes advantage of the fact that when a (safe or unsafe) suboptimal solution is ascertained to be suboptimal, its safeness can be disregarded, as reflected in the terms $\Psi_{i,\mathcal{S}\cap\mathcal{B}}$ and $\Psi_{i,\mathcal{S}^c\cap\mathcal{B}}$. In addition, it will not sample an unsafe solution if it is identified as unsafe w.p. at least $1 - 2\xi(\omega'_{\mathrm{v}})$. The last term $\mathrm{Reg}_3(T)$ corresponds to the regret due to failure of the "good" event and at the initialization stage. A proof sketch is presented in Section 6.

4.2. Problem-dependent Lower Bound

Theorem 4.3 (Problem-dependent lower bound)**.** *Let $\{\delta_T\}_{T=1}^{\infty} \in o(1)$ be a sequence that satisfies $\ln(1/\delta_T) = o(T^b)$ for all $b > 0$. There exists an instance Λ such that for any $\{\delta_T\}_{T\in\mathbb{N}}$-variance-constrained consistent algorithm π, the regret is lower bounded by*

$$\Omega\Bigg(\sum_{i\in E}\frac{\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}}\Bigg) + \frac{\mu^\star}{K}\cdot\Omega\Bigg(\frac{K\ \ln(1/\delta_T)}{(\Delta^{\mathrm{v}}_{S^\star})^2} + \sum_{i\in E}\left(\Psi'_{i,\mathcal{S}\cap\mathcal{B}} + \frac{\ln T}{(\Delta^{\mathrm{v}}_{i,\mathcal{R}})^2} + \Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\right)\Bigg).$$

The proof is presented at App. B.5. With Theorem 4.3, the

problem-dependent upper bound is tight for polynomially decaying $\{\delta_T\}_{T\in\mathbb{N}}$.

Corollary 4.4 (Tightness of problem-dependent bounds)**.** *Let $\delta_T = T^{-\lambda}$ with a fixed $\lambda > 0$, the regret*

$$\mathrm{Reg}(T) \in \Omega\left(\sum_{i\in E} \frac{\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}} + \frac{\mu^\star}{K^2} H\left(\Delta(\Lambda)\right)\right) \cap O\left(\sum_{i\in E} \frac{K \ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}} + \mu^\star H\left(\Delta(\Lambda)\right)\right)$$

where $H\left(\Delta(\Lambda)\right)$ is defined in Remark 4.2. The upper bound above is achieved by PASCOMBUCB.

Under different rates of decay of $\{\delta_T\}_{T\in\mathbb{N}}$ (see App. D for the cases where $\ln(1/\delta_T) = \omega(\ln T)$ and $o(\ln T)$), the upper bound of the regret due to suboptimality $\mathrm{Reg}_1(T)$ (the first term in the total regret) and the upper bound of the regret due to safeness-checking $\mathrm{Reg}_2(T)$ (the latter term) match their corresponding lower bounds up to factors of K and K^2 respectively; this gap is acceptable as K (e.g., number of ads displayed) is usually small relative to L (total number of ads). More discussions are postponed to App. E. One may naturally wonder whether we can tolerate a much more stringent probably anytime-safe constraint. The following theorem (with $b = 1$) indicates no algorithm is $\{\delta_T\}_{T\in\mathbb{N}}$-variance-constrained consistent if δ_T decays *exponentially fast* in T. Detailed proofs are postponed to App. C.

Theorem 4.5 (Impossibility result)**.** *Let $\{\delta_T\}_{T=1}^\infty \in o(1)$ be a sequence that satisfies that there exists $b \in (0, 1]$ such that $\ln(1/\delta_T) = \Omega(T^b)$. For any instance Λ, the regret of any algorithm is lower bounded by $\Omega(T^b)$.*

5. Problem-independent Bounds

We can derive a problem-independent upper bound on the regret of PASCOMBUCB from the problem-dependent one in Theorem 4.1 with some delicate calculations (see App. B.5).

Theorem 5.1 (Problem-independent upper bound)**.** *Let $\{\delta_T\}_{T=1}^\infty \in o(1)$ be a sequence that satisfies $\ln(1/\delta_T) = o(T^b)$ for all $b > 0$. If $T > L$, for any instance Λ with variance gaps lower bounded by $\Delta^{\mathrm{v}} \le \min_{S\in\mathcal{A}_K} \Delta^{\mathrm{v}}_S$, the regret of* PASCOMBUCB *is upper bounded by*

$$O\left(\sqrt{KLT\ln T} + \frac{LK^2}{(\Delta^{\mathrm{v}})^2} \ln\left(\frac{1}{\delta_T}\right)\right).$$

Theorem 5.2 (Problem-independent lower bound)**.** *Let the minimum variance gap be $\Delta^{\mathrm{v}} := \min_{S\in\mathcal{A}_K} \Delta^{\mathrm{v}}_S$. When $K^3 \ge L^2$, we have*

$$\mathrm{Reg}(T) = \Omega\left(\sqrt{KLT} + \min\left\{\frac{L}{(\Delta^{\mathrm{v}})^2} \ln\left(\frac{1}{\delta_T}\right), T\right\}\right).$$

Remark 5.3. *The assumption that the variance gaps of all solutions are lower bounded by Δ^{v} is needed to achieve a non-vacuous problem-independent bound. Given any algorithm and time budget T, the variance gap of $S^\star$ can be arbitrarily small if Δ^{v} is not bounded away from zero, so the* min *in Theorem 5.2 will be dominated by the linear term T, and hence, no algorithm can attain sublinear regret.*

Corollary 5.4 (Tightness of problem-independent bounds)**.** *Let $K^3 \le L^2$, and $\{\delta_T\}_{T=1}^\infty \in o(1)$ satisfies $\ln(1/\delta_T) = o(T^b)$ for all $b > 0$. We have*

$$\mathrm{Reg}(T) \in \Omega\left(\sqrt{KLT} + \frac{L}{(\Delta^{\mathrm{v}})^2} \ln\left(\frac{1}{\delta_T}\right)\right) \cap O\left(\sqrt{KLT\ln T} + \frac{LK^2}{(\Delta^{\mathrm{v}})^2} \ln\left(\frac{1}{\delta_T}\right)\right).$$

The upper bound is achieved by PASCOMBUCB.

We observe that the gap between the upper and lower bounds is manifested on $\sqrt{\ln T}$ and K^2. The presence of $\sqrt{\ln T}$ is not unexpected as it is also involved in the gap between the bounds on the regret for the (unconstrained) combinatorial bandits (Kveton et al., 2015). Besides, the term K^2 is induced by the design of PASCOMBUCB. Additional discussions are provided in App. E.

6. Proof Sketch of the Problem-Dependent Upper Bound (Theorem 4.1)

Assume that PASCOMBUCB has processed T' phases with T time steps, we have $\mathbb{P}[T' \le T] = 1$ since each phase is composed by multiple time steps. Denote the expected regret of PASCOMBUCB with p phases as $\mathbb{E}[\mathrm{R}(p)]$. The expected regret of PASCOMBUCB after T time steps is

$$\mathbb{E}[\mathrm{R}(T')] := \mathbb{E}\left[\sum_{p=1}^{T'} \sum_{r=1}^{n_p} (\mu^\star - \mu_{A_{p,r}})\right].$$

In the proof of Theorem 4.1, we first show a regret decomposition lemma (Lemma 6.1) that separates the total regret into *the regret due to suboptimality* $\mathbb{E}[\mathrm{R}_1(T')]$, *the regret due to safeness-checking* $\mathbb{E}[\mathrm{R}_2(T')]$ and the regret due to the failure of the "good" event and the initialization. Then we upper bound $\mathrm{R}_1(T')$ and $\mathrm{R}_2(T')$ separately. To elucidate the dependence of the regret on the confidence parameters $\omega_\mu, \omega_{\mathrm{v}}$ and ω'_{v}, we retain these notations henceforth. Detailed proofs are presented in App. B.

For $p \in [T], i \in E$, define the "good" events that the sample mean and the sample variance are near their ground truths: $\mathcal{E}^\mu_{i,T_i(p)} := \{\hat{\mu}_i(p) - \alpha(T_i(p)) \le \mu_i \le \hat{\mu}_i(p) + \alpha(T_i(p))\}$ and $\mathcal{E}^{\mathrm{v}}_{i,T_i(p)}(\rho) := \{\hat{\sigma}^2_i(p) - 3\cdot\mathrm{lil}(T_i(p), \rho) \le \sigma^2_i \le \hat{\sigma}^2_i(p) + 3\cdot\mathrm{lil}(T_i(p), \rho)\}$ and

$$\mathcal{E}_{i,T_i(p)} := \mathcal{E}^\mu_{i,T_i(p)} \cap \mathcal{E}^{\mathrm{v}}_{i,T_i(p)}(\omega_{\mathrm{v}}) \cap \mathcal{E}^{\mathrm{v}}_{i,T_i(p)}(\omega'_{\mathrm{v}})$$

$$\mathcal{E} := \bigcap_{i\in E} \bigcap_{p\in[T']} \mathcal{E}_{i,T_i(p-1)}.$$

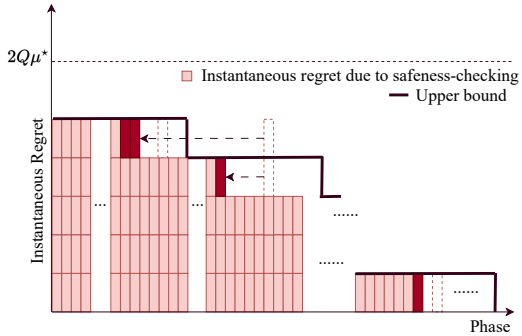


Figure 4. We assume the algorithm will sample those solutions with large $U_A^{\mathrm{v}}(p)$, i.e., those phases in which more sub-solutions are sampled are moved forward (the dark red ones). Based on this, an upper bound can be derived (the thick black lines).

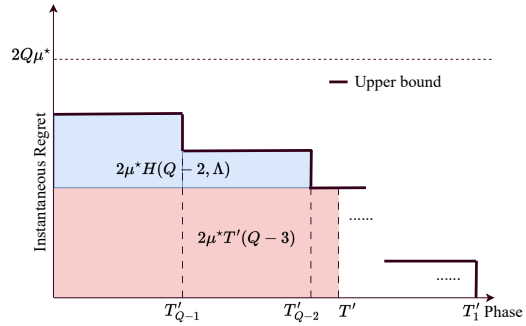


Figure 5. An illustration of the upper bound of $\mathrm{R}_2(T')$ for phase $T'_{Q-2} \le T' < T'_{Q-3}$. When $r' = 1$, $2\mu^\star H(1,\Lambda)$ is the area below the thick line, i.e., the upper bound for $\mathrm{R}_2(T')$ for any T'.

For $r \in [Q-1]$, define $\mathcal{U}_p(r) := \{U_{A_p}^{\mathrm{v}}(p-1) > r\bar{\sigma}^2\}$. When event $\mathcal{U}_p(r)$ occurs at phase p, it indicates at least $r+1$ sub-solutions are needed in order to sample the items in A_p and guarantee the safeness constraint.

Lemma 6.1. *Assume that* PASCombUCB *has processed T' phases with T time steps, the expected regret of* PASCombUCB *can be decomposed into three parts as follows*

$$\mathbb{E}[\mathrm{R}(T')] \le \mathbb{E}[\mathrm{R}_1(T')|\mathcal{E}] + \mathbb{E}[\mathrm{R}_2(T')|\mathcal{E}] + \mathrm{R}_3(T)$$

$$\textit{where} \quad \mathrm{R}_1(T') := \sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{B}\}\Delta_{A_p}$$

$$\mathrm{R}_2(T') := \mu^\star \sum_{p=1}^{T'} \left[2 \sum_{r=1}^{Q-1} \mathbb{1}\{\mathcal{U}_p(r)\}\right]$$

$$\mathrm{R}_3(T) := 2\mu^\star L\big(1 + T\big(\xi(\omega_\mu) + 2\xi(\omega_v) + 2\xi(\omega'_v)\big)\big)$$

In Lemma 6.1, the first term $\mathrm{R}_1(T')$ is the *(high-probability) regret due to suboptimality*, in the sense that only the mean gaps of the suboptimal solutions contribute to $\mathrm{R}_1(T)$. The second term $\mathrm{R}_2(T')$ is called *the (high-probability) regret due to safeness-checking*, since it depends on the variance gaps and goes to 0 if $\bar{\sigma}^2$ is sufficiently large. The last term $\mathrm{R}_3(T)$ contains the regret from the initialization stage and the regret results from the failure of the "good" event $\mathcal{E}$.

The regret due to suboptimality can be bounded in terms of the minimum safe/unsafe-suboptimal gaps as follows.

Lemma 6.2. *Conditioned on event $\mathcal{E}$, the regret due to suboptimality $\mathrm{R}_1(T')$ can be bounded by*

$$O\Big(\sum_{i\in E\setminus S^\star} \frac{K}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}} \ln\frac{1}{\omega_\mu} + \sum_{i\in E} \frac{c_i K}{\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min}} \ln\frac{1}{\omega'_{\mathrm{v}}}\Big).$$

The regret due to safeness-checking involves more critical parameters of the instance and we encode them in $T'_{r'}$ and $H(r',\Lambda)$ for $r' \in [Q]$ (see Figure 5); these terms are defined formally in (25) and (26) respectively.

Lemma 6.3. *On the event $\mathcal{E}$, if $T' \in [T'_{r'}, T'_{r'-1})$ then*

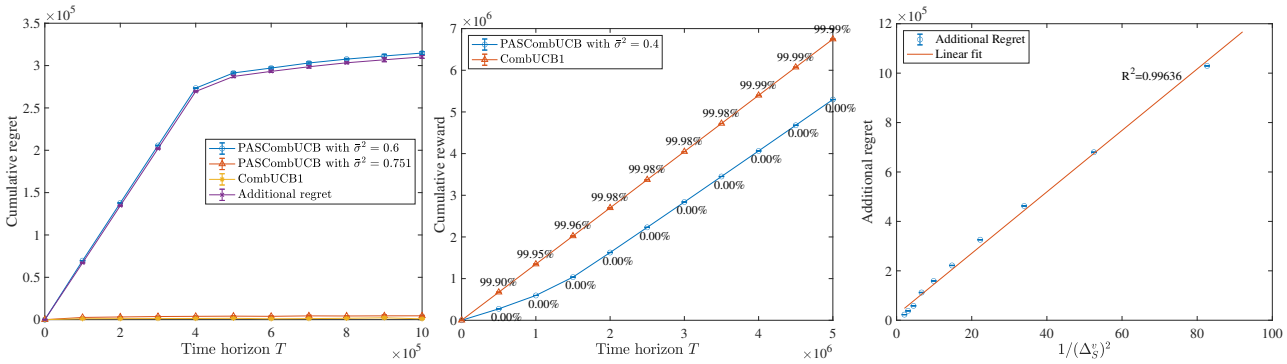
$$\mathrm{R}_2(T') \le 2\mu^\star[T'(r'-1) + H(r',\Lambda)] \le 2\mu^\star H(1,\Lambda)$$

To upper bound $\mathrm{R}_2(T')$, we assume the algorithm samples solutions with large $U_A^{\mathrm{v}}(p)$ in $\bar{\mathcal{S}}_p$, which will then be split into several sub-solutions (see Figure 4). Furthermore, for $r' = Q-1, Q-2, \ldots, 1$, we derive an upper bound for the number of phases in which event $\mathcal{U}_p(r') \cap (\mathcal{U}_p(r'+1))^c$ occurs (at most $2r'+1$ sub-solutions are being pulled in these phases). To be more specific (see Figure 5), for $r' = Q-1$, we compute the maximum number of phases T'_{Q-1} in which at most $2Q-1$ sub-solutions are sampled. Then for $r' = Q-2$, we compute the maximum number of phases $T'_{Q-2} - T'_{Q-1}$ in which at most $2Q-3$ sub-solutions are sampled. We do this until the time budget runs out. As T' increases, r' decreases and $H(r',\Lambda)$ increases. When $r' = 1$, i.e. $T' \ge T'_1$, $H(1,\Lambda)$ is an upper bound for the total number of sub-solutions being pulled (up to a constant) for the safeness-checking or the price of satisfying the probably anytime-safe constraint. The upper bound for the regret due to safeness-checking is the instance-dependent constant $2\mu^\star H(1,\Lambda)$ when $T' \ge T'_1$. More discussions are postponed to Step 3 in the proof in App. B.4.

7. Experiments

In this section, we ran 3 sets of experiments to illustrate the empirical performance of PASCombUCB and to corroborate its theoretical guarantees. As CombUCB1 (Kveton et al., 2015; Khezeli & Bitar, 2020; Amani et al., 2019) has tight regret guarantees, we adopt it as the benchmark in the unconstrained case. Codes are accessible at `https://github.com/Y-Hou/PASSCSB.git`.

Experimental Design: We design two instances where the rewards are Beta distributed with means and variances as in Table 1. There are $L = 10$ base arms and the admissible solution set $\mathcal{A}_K$ contains all subsets of $[L]$ with cardinality no greater than $K = 3$ (so $|\mathcal{A}_K| = 175$). Since the arm distributions are supported on $[0,1]$, the sub-Gaussian parameter $\sigma^2 = 0.25$. The confidence parameter $\delta = 0.05$.



(a) Experiment 1 (b) Experiment 2 (c) Experiment 3

Figure 6. Results of the Experiments: (a) Experiment 1: Cumulative regret v.s. Time horizon; (b) Experiment 2: Cumulative reward v.s. Time horizon with the percentages of violations besides the data points; (c) Experiment 3: Additional regret v.s. $1/(\Delta_S^{\mathrm{v}})^2$.

Item index	1	2	3	4	5 to 10
Means	0.5	0.45	0.4	0.35	0.3
Variances (**Set 1**)	0.24	0.24	0.04	0.01	0.01
Variances (**Set 2**)	0.01	0.01	0.01	0.01	0.01

Table 1. Two sets of items with equal means for each item.

In Experiment 1, we quantify the additional regret due to the safeness checking and evaluate the performance of PASCombUCB with **Set 1** under the unconstrained case. We run (1) PASCombUCB with $\bar{\sigma} = 0.6$ which needs to check the safeness of the solutions; (2) PASCombUCB with $\bar{\sigma} = 0.751$ which can be regarded as a variant of PASCombUCB without the safeness constraint, since our algorithm is aware of the safeness of all solutions; (3) CombUCB1 which is a baseline algorithm.

In Experiment 2, we illustrate the effectiveness of PASCombUCB in satisfying the safety constraint. Furthermore, we show that if an algorithm ignores the safety constraint, it will violate the safety constraint $\Omega(T)$ times if there exists a risky solution. We run PASCombUCB and CombUCB1 with **Set 1** under the constrained case where $\bar{\sigma} = 0.4$, the optimal safe solution is $\{1, 3, 4\}$, and the optimal solution under the unconstrained case $\{1, 2, 3\}$ is unsafe (risky).

In Experiment 3, we empirically verify the dependence of the additional regret on the hardness parameter $H(\Delta(\Lambda))$ in (26) using **Set 2**. We fix the time horizon $T = 2 \times 10^6$ and vary the threshold on the variance from 0.14 to 0.72 (i.e., $\bar{\sigma}^2 = 0.14 \times 1.2^k$ for $k = 0, 1, \ldots, 9$). As any solution that is comprised of 3 items has variance 0.03, we have $\Delta_S^v = \bar{\sigma}^2 - 0.03$. We compare the *additional regret* with respect to $1/(\Delta_S^v)^2$, which is proportional to $H(\Delta(\Lambda))$ under this setup according to (26).

Experimental Results: For **Experiment 1**, we present the results in Figure 6(a). We first observe when $\bar{\sigma}^2 = 0.751$, the regret incurred by PASCombUCB is similar to that by CombUCB1 for all T considered, which suggests that PASCombUCB is comparable to CombUCB1 under the unconstrained case, and hence in the following experiments we refer the difference between the regret of PASCombUCB and the regret of CombUCB1 as the "additional regret". Secondly, when $\bar{\sigma}^2 = 0.6$, the regret of PASCombUCB increases rapidly at the beginning and plateaus when $T > 4 \times 10^5$. This corroborates the design of PASCombUCB : (i) at the beginning, PASCombUCB pulls solutions conservatively to meet the anytime-safe constraint w.h.p.; (ii) after a number of time steps ($T > 4 \times 10^5$), the safeness of the optimal (safe) solution can be ascertained, it then exploits the optimal solution aggressively and eventually matches the performance of CombUCB1.

For **Experiment 2**, we plot the percentage of times each algorithm violates the safeness constraint $\sigma_{A_t}^2 < \bar{\sigma}^2$ as well as the cumulative rewards in Figure 6(b). The reward of PASCombUCB increases slowly at the start and then more rapidly when $T > 1.5 \times 10^6$, when the safeness of the optimal safe solution has been ascertained. However, while the reward of CombUCB1 increases linearly (as it pulls the risky solution $\{1, 2, 3\}$ $\Omega(T)$ times), it violates the safeness constraint $\sigma_{S_t}^2 < \bar{\sigma}^2$ at almost all times. This implies that the safety constraint is almost always violated by CombUCB1 ($\Omega(T)$ times) whereas PASCombUCB can meet the probably anytime-safe requirement.

For **Experiment 3**, the results are in Figure 6(c). As suggested by Theorem 4.1, the regret due to safeness checking is proportional to $H(\Delta(\Lambda))$. Figure 6(c) indicates that empirically, the additional regret scales linearly in $1/(\Delta_S^{\mathrm{v}})^2$, which corroborates our theoretical results.

Additional discussions on the tightness results, the problem formulation and comparisons with other literature, as well as future research directions, are presented in App. E.

Acknowledgements

The authors are supported by Singapore Ministry of Education (MOE) grants (Grant Numbers: A-0009042-01-00, A-8000189-01-00, A-8000980-00-00, A-8000423-00-00) and funding from CIFAR through Amii and NSERC.

References

Amani, S., Alizadeh, M., and Thrampoulidis, C. Linear stochastic bandits under safety constraints. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, volume 32, pp. 9256–9266, 2019.

Ayyagari, S. and Dukkipati, A. Risk-aware algorithms for combinatorial semi-bandits, 2021.

Badanidiyuru, A., Kleinberg, R., and Slivkins, A. Bandits with knapsacks. *Journal of the ACM*, 65(3), 2018.

Bhat, S. P. and Prashanth, L. A. Concentration of risk measures: a wasserstein distance approach. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, volume 32, pp. 11762–11771. Curran Associates, Inc., 2019.

Cassel, A., Mannor, S., and Zeevi, A. A general approach to multi-armed bandits under risk criteria. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 1295–1306. PMLR, 2018.

Chang, J. Q. L. and Tan, V. Y. F. A unifying theory of Thompson sampling for continuous risk-averse bandits. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, 2022.

Chen, W., Wang, Y., and Yuan, Y. Combinatorial multi-armed bandit: General framework and applications. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28, pp. 151–159. PMLR, 2013.

David, Y., Szörényi, B., Ghavamzadeh, M., Mannor, S., and Shimkin, N. PAC bandits with risk constraints. In *ISAIM*, 2018.

Garcelon, E., Ghavamzadeh, M., Lazaric, A., and Pirotta, M. Improved algorithms for conservative exploration in bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 3962–3969, 2020.

Hou, Y., Tan, V. Y. F., and Zhong, Z. Almost optimal variance-constrained best arm identification. *IEEE Transactions on Information Theory*, 2023.

Jamieson, K., Malloy, M., Nowak, R., and Bubeck, S. lil'UCB: An optimal exploration algorithm for multi-armed bandits. In *Proceedings of the 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pp. 423–439, Barcelona, Spain, 2014. PMLR.

Kagrecha, A., Nair, J., and Jagannathan, K. Constrained regret minimization for multi-criterion multi-armed bandits. *Machine Learning*, 2023.

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

Kazerouni, A., Ghavamzadeh, M., Abbasi-Yadkori, Y., and Van Roy, B. Conservative contextual linear bandits. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 3913–3922. Curran Associates Inc., 2017.

Khezeli, K. and Bitar, E. Safe linear stochastic bandits. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(06):10202–10209, 2020.

Kveton, B., Wen, Z., Ashkan, A., Eydgahi, H., and Eriksson, B. Matroid bandits: Fast combinatorial optimization with learning. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, pp. 420–429, 2014.

Kveton, B., Wen, Z., Ashkan, A., and Szepesvari, C. Tight Regret Bounds for Stochastic Combinatorial Semi-Bandits. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, volume 38, pp. 535–543. PMLR, 2015.

Mahdavi, M., Jin, R., and Yang, T. Trading regret for efficiency: Online convex optimization with long term constraints. *Journal of Machine Learning Research*, 13 (1):2503–2528, 2012.

Moradipari, A., Thrampoulidis, C., and Alizadeh, M. Stage-wise conservative linear bandits. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, volume 33, pp. 11191–11201. Curran Associates, Inc., 2020.

Sani, A., Lazaric, A., and Munos, R. Risk-aversion in multi-armed bandits. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pp. 3275–3283. Curran Associates Inc., 2012.

Vakili, S. and Zhao, Q. Risk-averse multi-armed bandit problems under mean-variance measure. *IEEE Journal of Selected Topics in Signal Processing*, 10(6):1093–1111, 2016.

Wu, Y., Shariff, R., Lattimore, T., and Szepesvari, C. Conservative bandits. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, pp. 1254–1262. PMLR, 2016.

Zhong, Z., Cheung, W. C., and Tan, V. Y. F. Thompson sampling algorithms for cascading bandits. *Journal of Machine Learning Research*, 22(218):1–66, 2021.

Zhu, Q. and Tan, V. Y. F. Thompson sampling algorithms for mean-variance bandits. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 11599–11608. PMLR, 2020.

Appendices

The contents of the appendices are organized as follows:

- In App. A, we list 3 useful lemmas concerning the LIL concentration bound.
- In App. B, we present detailed proofs of the upper bounds.
  - App. B.1: preliminary results for the proof of the upper bound;
  - App. B.2: the proof of the decomposition lemma Lemma 6.1;
  - App. B.3: the proof of Lemma 6.2 (the regret due to suboptimality);
  - App. B.4: the proof of Lemma 6.3 (the regret due to safeness-checking);
  - App. B.5: the proofs of Theorem 4.1 (problem-dependent upper bound) and Theorem 5.1 (problem-independent upper bound).
- In App. C, we present detailed proofs of the lower bounds.
  - App. C.1: preliminary results for the proof of the lower bound and the proof of the impossibility result Theorem 4.5.
  - App. C.2: the proof of Theorem 4.3 (problem-dependent lower bound);
  - App. C.3: the proof of Theorem 5.2 (problem-independent lower bound);
- In App. D, we present a corollary characterizing the tightness of the upper bound in Theorem 4.1.
- In App. E, we provide additional discussions on the tightness results, the problem formulation and comparisons with other literature, as well as future research directions.

A. Auxiliary results

Lemma A.1 (Lemma 3 in (Jamieson et al., 2014))**.** *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. centered sub-Gaussian random variables with scale parameter σ. Fix any $\epsilon \in (0,1)$ and $\delta \in (0, \ln(1+\epsilon)/e)$. Then one has*

$$\mathbb{P}\left[\forall t \in \mathbb{N} : \sum_{s=1}^{t} X_s \leq (1+\sqrt{\epsilon})\sqrt{2\sigma^2\,(1+\epsilon)\,t \ln\left(\frac{\ln\left((1+\epsilon)t\right)}{\delta}\right)}\right] \geq 1 - \xi(\delta),$$

where $\xi(\delta) := \frac{2+\epsilon}{\epsilon}\left(\frac{\delta}{\ln(1+\epsilon)}\right)^{1+\epsilon}$.

Lemma A.2. *For $t \geq 1, \epsilon \in (0,1), \omega \in (0,1]$ and $u > 0$, let $\gamma := \frac{(1+\epsilon)(1+\sqrt{\epsilon})^2}{2}$, $c := \frac{(u\cdot s)^2}{\gamma} = \frac{2(u\cdot s)^2}{(1+\epsilon)(1+\sqrt{\epsilon})^2}$ and $m := \frac{\gamma}{u^2\cdot s^2}\left(2\ln\frac{1}{\omega} + \ln\ln_+\frac{1}{s^2} + \ln\frac{2\gamma(1+\epsilon)}{u^2}\right)$. If $t > m$, it holds that*

$$s > \mathrm{lil}(t,\omega) = (1+\sqrt{\epsilon})\sqrt{\frac{1+\epsilon}{2t}\ln\left(\frac{\ln\left((1+\epsilon)t\right)}{\omega}\right)}.$$

Proof of Lemma A.2. Note that fact that

$$u\cdot s \leq (1+\sqrt{\epsilon})\sqrt{\frac{1+\epsilon}{2t}\ln\left(\frac{\ln\left((1+\epsilon)t\right)}{\omega}\right)} \iff c = \frac{2(u\cdot s)^2}{(1+\epsilon)(1+\sqrt{\epsilon})^2} \leq \frac{1}{t}\ln\left(\frac{\ln\left((1+\epsilon)t\right)}{\omega}\right)$$

According to the computations in Jamieson et al. (2014) equation (1), i.e.,

$$\frac{1}{t}\ln\left(\frac{\ln((1+\epsilon)t)}{\omega}\right) \geq c' \Rightarrow t \leq \frac{1}{c'}\ln\left(\frac{2\ln((1+\epsilon)/(c'\omega))}{\omega}\right)$$

for $t \geq 1, \epsilon \in (0,1), c' > 0, \omega \in (0,1]$. We take $c' = c$, thus

$$\begin{aligned}
t &\leq \frac{1}{c}\ln\left(\frac{2\ln((1+\epsilon)/(c\omega))}{\omega}\right) \\
&= \frac{1}{c}\left(\ln\frac{2}{\omega} + \ln\left(\ln\frac{\gamma(1+\epsilon)}{u^2\cdot\omega} + \ln\frac{1}{s^2}\right)\right) \\
&\overset{(a)}{\leq} \frac{1}{c}\left(\ln\frac{2}{\omega} + \ln\frac{\gamma(1+\epsilon)}{u^2\cdot\omega} + \ln\ln_+\frac{1}{s^2}\right) \\
&= \frac{\gamma}{u^2\cdot s^2}\left(2\ln\frac{1}{\omega} + \ln\ln_+\frac{1}{s^2} + \ln\frac{2\gamma(1+\epsilon)}{u^2}\right) = m
\end{aligned}$$

where we adopt $\ln(x+y) \le x + \ln\ln_+ y, \forall x, y \in \mathbb{R}_+$ in (a). Therefore, if $t > m$, we must have

$$u \cdot s > (1+\sqrt{\epsilon})\sqrt{\frac{1+\epsilon}{2t}\ln\left(\frac{\ln\left((1+\epsilon)t\right)}{\omega}\right)}.$$

□

Lemma A.3. *With the choice of the confidence radii in (2) and (3), for all $i \in E$, we have*

$$\mathbb{P}\left[\forall\, p{\in}\mathbb{N} : |\hat{\mu}_i(p) - \mu_i| \le \alpha(T_i(p))\right] \ge 1 - 2\xi(\omega_\mu) \tag{7}$$

$$\mathbb{P}\left[\forall\, p{\in}\mathbb{N} : |\hat{\sigma}_i^2(p) - \sigma_i^2| \le \beta_{\mathrm{u}}(T_i(p))\right] \ge 1 - 4\xi(\omega_{\mathrm{v}}) \tag{8}$$

$$\mathbb{P}\left[\forall\, p{\in}\mathbb{N} : |\hat{\sigma}_i^2(p) - \sigma_i^2| \le \beta_{\mathrm{l}}(T_i(p))\right] \ge 1 - 4\xi(\omega'_{\mathrm{v}}) \tag{9}$$

Proof. Note the fact that any distribution supported on $[0,1]$ is $1/4$-sub-Gaussian. By a direct application of Lemma A.1 to the sample mean $\hat{\mu}_i(p)$ and the sample second moment $\hat{M}_{2,i}(p) := \frac{1}{T_i(p)}\sum_{s=1}^{p} W_i(s)^2 \mathbb{1}\{i \in A_s\}$ of arm $i \in [L]$, (7) can be derived and

$$\mathbb{P}\left[\forall\, p{\in}\mathbb{N} : |\mu_i - \hat{\mu}_i(p)| \le \mathrm{lil}(T_i(p), \omega'_{\mathrm{v}})\right] \ge 1 - 2\xi(\omega'_{\mathrm{v}}), \quad \text{and}$$
$$\mathbb{P}\left[\forall\, p{\in}\mathbb{N} : |\hat{M}_{2,i}(p) - (\mu_i^2 + \sigma_i^2)| \le \mathrm{lil}(T_i(p), \omega'_{\mathrm{v}})\right] \ge 1 - 2\xi(\omega'_{\mathrm{v}})$$

Since the rewards are in $[0,1]$, $|\mu_i^2 - \hat{\mu}_i^2(p)| = |\mu_i + \hat{\mu}_i(p)| \cdot |\mu_i - \hat{\mu}_i(p)| \le 2 \cdot \mathrm{lil}(T_i(p), \omega'_{\mathrm{v}})$. Using this and the triangle inequality, we obtain for every $p \ge 1$,

$$\begin{aligned}|\hat{\sigma}_i^2(p) - \sigma_i^2| &= |\mu_i^2 - \hat{\mu}_i^2(p)| + |(\mu_i^2 + \sigma_i^2) - \hat{M}_{2,i}(p)| \\ &\le 2 \cdot \mathrm{lil}(T_i(p), \omega_{\mathrm{v}}) + \mathrm{lil}(T_i(p), \omega_{\mathrm{v}}) = \beta_{\mathrm{u}}(T_i(p)).\end{aligned}$$

Therefore, (8) is proved. (9) can be similarly obtained. □

B. Proof of the Upper Bound

B.1. Proof scheme of the problem-dependent upper bound

In this subsection, we provide technical lemmas that can upper bound the components in $\mathrm{R}_1(T')$ and $\mathrm{R}_2(T')$.

Note that at phase p, the identified solution A_p belongs to one of the 4 disjoint sets: (1) $A_p = S^\star$; (2) $\mathcal{S} \cap \mathcal{B}$; (3) $\mathcal{R}$ and (4) $\mathcal{S}^c \cap \mathcal{B}$, i.e.

$$1 = \mathbb{1}\{A_p = S^\star\} + \mathbb{1}\{A_p \in \mathcal{S} \cap \mathcal{B}\} + \mathbb{1}\{A_p \in \mathcal{R}\} + \mathbb{1}\{A_p \in \mathcal{S}^c \cap \mathcal{B}\}$$

and $\mathbb{1}\{A_p \in \mathcal{B}\} = \mathbb{1}\{A_p \in \mathcal{S} \cap \mathcal{B}\} + \mathbb{1}\{A_p \in \mathcal{S}^c \cap \mathcal{B}\}$. Define two events (the $\mathcal{F}$ events) that connect the instance and the confidence radii

$$\mathcal{F}_p^\mu := \left\{\Delta_{A_p} \le 2 \sum_{i \in A_p \setminus S^\star} \alpha(T_i(p-1))\right\}$$

$$\mathcal{F}_p(x, \rho) := \left\{x \le 2 \sum_{i \in A_p} \mathrm{lil}(T_i(p-1), \rho)\right\}$$

where x is a constant and ω is a confidence parameter. When $A_p \in \mathcal{B}$, it indicates solution A_p has not been sampled sufficiently many times and its suboptimality has not been ascertained. When $A_p \in \mathcal{S}^c$, it implies the unsafeness of A_p has not been recognized. We formalize this in the following lemma.

Lemma B.1. *Conditional on the event $\mathcal{E}$, given any $p \in [T]$, we have*

- $S^\star \in \bar{\mathcal{S}}_{p-1}$;

- *If* $A_p \in \mathcal{S} \cap \mathcal{B}$,

$$\mathbb{1}\left\{A_p \in \mathcal{S} \cap \mathcal{B}\right\} \leq \mathbb{1}\left\{\mathcal{F}_p^{\mu}\right\};$$

- *If* $A_p \in \mathcal{R}$,

$$\mathbb{1}\left\{A_p \in \mathcal{R}\right\} \leq \mathbb{1}\left\{\mathcal{F}_p\left(\frac{\Delta_{A_p}^{\mathrm{v}}}{3}, \omega_{\mathrm{v}}'\right)\right\};$$

- *If* $A_p \in \mathcal{S}^c \cap \mathcal{B}$,

$$\mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \leq \mathbb{1}\left\{\mathcal{F}_p^{\mu}, \mathcal{F}_p\left(\frac{\Delta_{A_p}^{\mathrm{v}}}{3}, \omega_{\mathrm{v}}'\right)\right\}.$$

Proof of Lemma B.1. By the design of PASCOMBUCB , $A_p \in \bar{\mathcal{S}}_{p-1}$.

(1) We firstly prove that $S \in \bar{\mathcal{S}}_{p-1}, \forall S \in \mathcal{S}$. On the event $\mathcal{E}$, we have

$$\begin{aligned} L_S^{\mathrm{v}}(p-1) &= \sum_{i \in A} \max\{\hat{\sigma}_i^2(p-1) - \beta_{\mathrm{l}}(T_i(p-1)), 0\} \\ &\leq \sum_{i \in A} \max\{\sigma_i^2, 0\} \\ &= \sigma_S^2 < \bar{\sigma}^2 \end{aligned}$$

Thus, $S \in \bar{\mathcal{S}}_{p-1}$, and in particular, $S^\star \in \bar{\mathcal{S}}_{p-1}$.

(2) If $A_p \in \mathcal{B}$, according to the sampling strategy in Line 10 of PASCOMBUCB and $S^\star \in \bar{\mathcal{S}}_{p-1}$, we have $U_{S^\star}^{\mu}(p-1) \leq U_{A_p}^{\mu}(p-1)$ which indicates $\sum_{i \in S^\star \setminus A_p} U_i^{\mu}(p-1) \leq \sum_{i \in A_p \setminus S^\star} U_i^{\mu}(p-1)$. Thus,

$$\begin{aligned} \sum_{i \in S^\star \setminus A_p} \mu_i &\leq \sum_{i \in S^\star \setminus A_p} U_i^{\mu}(p-1) \\ &\leq \sum_{i \in A_p \setminus S^\star} U_i^{\mu}(p-1) \\ &\leq \sum_{i \in A_p \setminus S^\star} \mu_i + 2\alpha_i(T_i(p-1)) \\ \implies \quad \Delta_{A_p} &\leq 2 \sum_{i \in A_p \setminus S^\star} \alpha(T_i(p-1)) \end{aligned} \tag{10}$$

(3) If $A_p \in \mathcal{S}^c$, according to the sampling strategy, we have $A_p \in \bar{\mathcal{S}}_{p-1}$ which indicates $L_{A_p}^{\mathrm{v}}(p-1) = \sum_{i \in A_p} L_i^{\mathrm{v}}(p-1) < \bar{\sigma}^2$. Thus,

$$\begin{aligned} \bar{\sigma}^2 &> \hat{\sigma}_{A_p}^2(p-1) - \sum_{i \in A_p} \beta_{\mathrm{l}}(T_i(p-1)) \\ &\geq \sigma_{A_p}^2 - 2 \sum_{i \in A_p} \beta_{\mathrm{l}}(T_i(p-1)) \\ \implies \quad \Delta_{A_p}^{\mathrm{v}} &\leq 2 \sum_{i \in A_p \setminus S^\star} \beta_{\mathrm{l}}(T_i(p-1)) \end{aligned} \tag{11}$$

Note that if $A_p \in \mathcal{S} \cap \mathcal{B}$, according to (10),

$$\mathbb{1}\left\{A_p \in \mathcal{S} \cap \mathcal{B}\right\} \leq \mathbb{1}\left\{\mathcal{F}_p^{\mu}\right\}.$$

If $A_p \in \mathcal{R} \subset \mathcal{S}^c$, by (11)

$$\mathbb{1}\left\{A_p \in \mathcal{R}\right\} \leq \mathbb{1}\left\{\mathcal{F}_p\left(\frac{\Delta_{A_p}^{\mathrm{v}}}{3}, \omega_{\mathrm{v}}'\right)\right\}.$$

If $A_p \in \mathcal{S}^c \cap \mathcal{B}$, by (11) and (10)

$$\mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \leq \mathbb{1}\left\{\mathcal{F}_p^{\mu}, \mathcal{F}_p\left(\frac{\Delta_{A_p}^{\mathrm{v}}}{3}, \omega_{\mathrm{v}}'\right)\right\}$$

□

At phase p, we define two sequences of mutually-exclusive events $\{\mathcal{G}_{j,p}^{\mu}\}_{j\in\mathbb{N}}$ and $\{\mathcal{G}_{j,p}(x,\omega\}_{j\in\mathbb{N}}$ (the $\mathcal{G}$ events) which can further bound the number of times $\mathcal{F}_p^{\mu}$ and $\mathcal{F}_p^{\mathrm{v}}(x,\omega)$ occur respectively. These events are indexed by two strictly-decreasing sequences of constants:

$$a_1 > a_2 > \ldots > a_k > \ldots$$
$$1 > b_1 > b_2 > \ldots > b_k > \ldots$$

where $\lim_{j\to\infty} a_j = \lim_{j\to\infty} b_j = 0$. For simplicity, we set $a_j = \frac{4}{9^{j-2}}$, $b_j = \frac{1}{4^j}, \forall j \in \mathbb{N}$ and denote the constant $C = \sum_{j\in\mathbb{N}} \frac{a_j}{b_j} = 259.2$. For $x \in \mathbb{R}_+$ and $\omega \in (0, \ln(1+\epsilon)/e)$, define

$$m_j(x,\omega) := \frac{a_j \cdot \gamma K^2}{x^2}\left(2\ln\frac{1}{\omega} + \ln\ln_+\frac{1}{x^2} + D\right)$$

and $m_j(x,\omega) := \infty$ otherwise, where (1) $\gamma = \frac{(1+\epsilon)(1+\sqrt{\epsilon})^2}{2}$ and ϵ is the constant in the confidence bounds (3), (2) $\ln\ln_+(x) = \ln\ln x$ if $x \geq e$ and it equals to 0 otherwise, (3) $D = \ln\left(324K^2(1+\epsilon)^2(1+\sqrt{\epsilon})^2\right)$. Denote

$$G_{j,p}^{\mu} := \left\{i \in A_p \setminus S^\star : T_i(p-1) \leq m_j(\Delta_{A_p}, \omega_\mu)\right\} \quad \text{and}$$
$$G_{j,p}(x,\omega) := \left\{i \in A_p : T_i(p-1) \leq m_j(x,\omega)\right\}$$

as the sets of items that were not chosen sufficiently often. For $j \in \mathbb{N}$, the events at phase p are sequentially defined as

$$\begin{aligned}\mathcal{G}_{j,p}^{\mu} := \Big\{&\text{at least } b_jK \text{ items in } A_p \setminus S^\star \text{ were chosen}\\ &\text{at most } m_j(\Delta_{A_p}, \omega_\mu) \text{ times}\Big\} \bigcap \left(\bigcup_{k\in[j-1]} \mathcal{G}_{k,p}^{\mu}\right)^c\\ &= \left\{\left|G_{j,p}^{\mu}\right| \geq b_jK\right\} \bigcap \left(\bigcup_{k\in[j-1]} \mathcal{G}_{k,p}^{\mu}\right)^c \qquad \text{and}\end{aligned}$$

$$\begin{aligned}\mathcal{G}_{j,p}(x,\omega) := \Big\{&\text{at least } b_jK \text{ items in } A_p \text{ were chosen}\\ &\text{at most } m_j(x,\omega) \text{ times}\Big\} \bigcap \left(\bigcup_{k\in[j-1]} \mathcal{G}_{k,p}(x,\omega)\right)^c\\ &= \left\{\left|G_{j,p}(x,\omega)\right| \geq b_jK\right\} \bigcap \left(\bigcup_{k\in[j-1]} \mathcal{G}_{k,p}(x,\omega)\right)^c\end{aligned}$$

Lemma B.2. *With our choice of $\{a_j\}_{j\in\mathbb{N}}$ and $\{b_j\}_{j\in\mathbb{N}}$,*

- *if $\mathcal{F}_p^{\mu}$ occurs, $\mathcal{G}_{j,p}^{\mu}$ occurs for some j, i.e.*

$$\mathbb{1}\left\{\mathcal{F}_p^{\mu}\right\} \leq \mathbb{1}\left\{\bigcup_{j\in\mathbb{N}} \mathcal{G}_{j,p}^{\mu}\right\}.$$

- *if $\mathcal{F}_p(x,\omega)$ occurs, $\mathcal{G}_{j,p}(x,\omega)$ occurs for some j, i.e.*

$$\mathbb{1}\left\{\mathcal{F}_p(x,\omega)\right\} \leq \mathbb{1}\left\{\bigcup_{j\in\mathbb{N}} \mathcal{G}_{j,p}(x,\omega)\right\}. \tag{12}$$

Proof of Lemma B.2. We prove (12) in the following and the other statement can be proved by the same procedures.

To ease the notations, we omit the parameters x, ω and p in $\mathcal{F}_p(x,\omega), \mathcal{G}_{j,p}(x,\omega)$ and $m_j(x,\omega)$, since they are fixed when given $\mathcal{F}_p(x,\omega)$. The event $\mathcal{G}_j$ can be rewritten as

$$\begin{aligned}
\mathcal{G}_j &= \Big\{|G_j| \geq b_j K\Big\} \bigcap \left(\bigcap_{k\in[j-1]} \mathcal{G}_k^c\right) \\
&= \Big\{|G_j| \geq b_j K\Big\} \bigcap \left(\bigcap_{k\in[j-1]} \Big\{|G_k| < b_k K\Big\}\right)
\end{aligned}$$

The statement is proved by contradiction. We assume that when $\mathcal{F}\left(\mathcal{F}_p(x,\omega)\right)$ occurs, none of event $\mathcal{G}_j$ occurs. Hence,

$$\begin{aligned}
&\left(\bigcup_{j\in\mathbb{N}} \mathcal{G}_j\right)^c = \bigcap_{j\in\mathbb{N}} \mathcal{G}_j^c \\
&= \bigcap_{j\in\mathbb{N}} \left[\Big\{|G_j| < b_j K\Big\} \bigcup \left(\bigcup_{k\in[j-1]} \Big\{|G_k| \geq b_k K\Big\}\right)\right] \\
&= \bigcap_{j\in\mathbb{N}} \Big\{|G_j| < b_j K\Big\}
\end{aligned}$$

Let $\bar{G}_j := A_p \setminus G_j$ and define $G_0 = A_p$. According to the definition of G_j, we have $G_j \subset G_{j-1}$ and $\bar{G}_{j-1} \subset \bar{G}_j, \forall j \in \mathbb{N}$. Because $\lim_{j\to\infty} m_j = 0$, there exists j_0 such that $\bar{G}_j = A_p, \forall j \geq j_0$. Therefore, we can write A_p by the "telescoping" sum, i.e., $A_p = \cup_{j\in\mathbb{N}} \left(\bar{G}_j \setminus \bar{G}_{j-1}\right)$. $\mathcal{F}_p(x,\omega)$ indicates

$$\begin{aligned}
x &\leq 2\sum_{i\in A_p} \mathrm{lil}(T_i(p-1),\omega) \\
&= 2\sum_{j\in\mathbb{N}} \sum_{i\in\bar{G}_j\setminus\bar{G}_{j-1}} \mathrm{lil}(T_i(p-1),\omega)
\end{aligned} \tag{13}$$

Note that for $i \in \bar{G}_j \setminus \bar{G}_{j-1} = G_{j-1} \setminus G_j$, we have $T_i(p-1) \in (m_j, m_{j-1}]$. By Lemma A.2 with parameters $t = T_i(p-1), s = x, u = \sqrt{\frac{1}{a_j K^2}}$ and note $a_1 > a_j$, we have

$$\sqrt{\frac{1}{a_j K^2}} \cdot x > \mathrm{lil}(T_i(p-1),\omega).$$

Note our choice of a_j and b_j satisfy

$$2\sum_{j\in\mathbb{N}} \frac{b_{j-1} - b_j}{\sqrt{a_j}} \leq 1$$

Thus, (13) can be further bounded by

$$\begin{aligned} x &\leq 2\sum_{j\in\mathbb{N}}\sum_{i\in\bar{G}_j\setminus\bar{G}_{j-1}} \mathrm{lil}(T_i(p-1),\omega) \\ &< 2\sum_{j\in\mathbb{N}} |\bar{G}_j\setminus\bar{G}_{j-1}|\sqrt{\frac{1}{a_jK^2}}\cdot x \\ &\leq 2\sum_{j\in\mathbb{N}}\frac{(b_{j-1}-b_j)K}{\sqrt{a_j}K}x \\ &\leq x \end{aligned}$$

which constitutes a contradiction. Thus when $\mathcal{F}$ occurs, there must exists $j\in\mathbb{N}$ such that $\mathcal{G}_j$ occurs. □

When none of the $\mathcal{G}^{\mu}_{j,p}$ (resp. $\mathcal{G}_{j,p}(x,\omega)$) occurs, $\mathcal{F}^{\mu}_p$ (resp. $\mathcal{F}_p(x,\omega)$) must not occur, which indicates all of the items in A_p have been sampled sufficiently many times such that the suboptimality (resp. unsafeness) of A_p is identified, thus A_p will not been sampled in future phases.

As there will be multiple $\mathcal{F}$ events happening, we provide the following useful lemma that merges all $\mathcal{F}$ events.

Lemma B.3. *Given two confidence parameters $\omega_1\geq\omega_2\in(0,\ln(1+\epsilon)/e)$*

- *If $\mathcal{F}^{\mu}_p$ occurs, then $\mathcal{F}_p(\Delta_{A_p},\omega_\mu)$ occurs.*
- *If both events $\mathcal{F}_p(x,\omega_1)$ and $\mathcal{F}_p(y,\omega_2)$ occur, then event $\mathcal{F}_p\left(\max\{x,\sqrt{\frac{\ln\frac{1}{\omega_1}}{\ln\frac{1}{\omega_2}}}y\},\omega_1\right)$ occurs.*

Proof of Lemma B.3. If $\mathcal{F}^{\mu}_p$ occurs, we have

$$\Delta_{A_p}\leq 2\sum_{i\in A_p\setminus S^\star}\alpha(T_i(p-1))\leq 2\sum_{i\in A_p}\alpha(T_i(p-1)) = 2\sum_{i\in A_p}\mathrm{lil}(T_i(p-1),\omega_\mu)$$

Thus, $\mathcal{F}_p(\Delta_{A_p},\omega_\mu)$ occurs.

For the second statement, notice the fact that

$$\begin{aligned} \frac{\mathrm{lil}(T_i(p-1),\omega_1)}{\mathrm{lil}(T_i(p-1),\omega_2)} &= \frac{(1+\sqrt{\epsilon})\left(\frac{1+\epsilon}{2T_i(p-1)}\ln\left(\frac{\ln((1+\epsilon)T_i(p-1))}{\omega_1}\right)\right)^{1/2}}{(1+\sqrt{\epsilon})\left(\frac{1+\epsilon}{2t}\ln\left(\frac{\ln((1+\epsilon)T_i(p-1))}{\omega_2}\right)\right)^{1/2}} \\ &= \sqrt{\frac{\ln\frac{1}{\omega_1}+\ln\ln((1+\epsilon)T_i(p-1))}{\ln\frac{1}{\omega_2}+\ln\ln((1+\epsilon)T_i(p-1))}} \\ &\overset{(a)}{\geq} \sqrt{\frac{\ln\frac{1}{\omega_1}}{\ln\frac{1}{\omega_2}}} =: \rho \end{aligned}$$

where (a) utilizes the trick that $\frac{a+c}{b+c}\geq\frac{a}{b},\forall a,b,c\in\mathbb{R}_+$ and $a\leq b$. When event $\mathcal{F}_p(y,\omega_2)$ occurs,

$$\begin{aligned} & y\leq 2\sum_{i\in A_p}\mathrm{lil}(T_i(p-1),\omega_2)\leq 2\sum_{i\in A_p}\frac{1}{\omega}\mathrm{lil}(T_i(p-1),\omega_1) \\ \implies\quad & \rho\cdot y\leq 2\sum_{i\in A_p}\mathrm{lil}(T_i(p-1),\omega_1) \\ \implies\quad & \mathcal{F}_p(\rho y,\omega_1) \end{aligned}$$

Thus if both events $\mathcal{F}_p(x,\omega_1)$ and $\mathcal{F}_p(y,\omega_2)$ occur, we must have that event $\mathcal{F}_p\left(\max\{x,\rho y\},\omega_1\right)) = \mathcal{F}_p\left(\max\{x,\sqrt{\frac{\ln\frac{1}{\omega_1}}{\ln\frac{1}{\omega_2}}}y\},\omega_1)\right)$ occurs.

□

B.2. Upper bound decomposition

Lemma 6.1. *Assume that* PASCombUCB *has processed T' phases with T time steps, the expected regret of* PASCombUCB *can be decomposed into three parts as follows*

$$\mathbb{E}[\mathrm{R}(T')] \leq \mathbb{E}[\mathrm{R}_1(T')|\mathcal{E}] + \mathbb{E}[\mathrm{R}_2(T')|\mathcal{E}] + \mathrm{R}_3(T)$$

$$\textit{where} \quad \mathrm{R}_1(T') := \sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{B}\}\Delta_{A_p}$$

$$\mathrm{R}_2(T') := \mu^\star \sum_{p=1}^{T'} \left[2 \sum_{r=1}^{Q-1} \mathbb{1}\{\mathcal{U}_p(r)\}\right]$$

$$\mathrm{R}_3(T) := 2\mu^\star L\big(1 + T\big(\xi(\omega_\mu) + 2\xi(\omega_v) + 2\xi(\omega_v')\big)\big)$$

Proof of Lemma 6.1. The expected regret can be decomposed as:

$$\begin{aligned}
\mathbb{E}\left[\mathrm{R}(T')\right] &= \mathbb{E}\left[\sum_{p=1}^{T'}\sum_{r=1}^{n_p}(\mu^\star - \mu_{A_{p,r}})\right] \\
&= \mathbb{E}\left[\sum_{p=1}^{T'}\sum_{r=1}^{n_p}(\mu^\star - \mu_{A_{p,r}})\mathbb{1}\left\{\mathcal{E}\right\}\right] + \mathbb{E}\left[\sum_{p=1}^{T'}\sum_{r=1}^{n_p}(\mu^\star - \mu_{A_{p,r}})\mathbb{1}\left\{\mathcal{E}^c\right\}\right] \\
&= \mathbb{E}\left[\sum_{p=1}^{T'}\sum_{r=1}^{n_p}(\mu^\star - \mu_{A_{p,r}})\middle|\mathcal{E}\right]\mathbb{P}[\mathcal{E}] + \mathbb{E}\left[\sum_{p=1}^{T'}\sum_{r=1}^{n_p}(\mu^\star - \mu_{A_{p,r}})\middle|\mathcal{E}^c\right]\mathbb{P}[\mathcal{E}^c]
\end{aligned}$$

In the initialization stage, it will take at most $2L$ time steps, since each pulled solution A_p contains at least one item i with $T_i(p) < 2$. Thus the regret is at most $2L \cdot \mu^\star$.

The expected regret when the good events fail can be upper bounded by

$$\begin{aligned}
\mathbb{E}\left[\sum_{p=1}^{T'}\sum_{r=1}^{n_p}(\mu^\star - \mu_{A_{p,r}})\middle|\mathcal{E}^c\right]\mathbb{P}\left[\mathcal{E}^c\right] &\leq \mathbb{E}\left[\sum_{p=1}^{T'}\sum_{r=1}^{n_p}\mu^\star\middle|\mathcal{E}^c\right]\mathbb{P}[\mathcal{E}^c] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T}\mu^\star\middle|\mathcal{E}^c\right] L \cdot 2(\xi(\omega_\mu) + 2\xi(\omega_v) + 2\xi(\omega_v')) \\
&\leq 2\mu^\star T L \cdot (\xi(\omega_\mu) + 2\xi(\omega_v) + 2\xi(\omega_v'))
\end{aligned}$$

where $\mathbb{P}[\mathcal{E}^c]$ can be bounded using Lemma A.3.

Conditional on the good event $\mathcal{E}$, the high-probability regret can be upper bounded by

$$\begin{aligned}
\hat{\mathrm{R}}(T') &:= \sum_{p=1}^{T'}\sum_{r=1}^{n_p}(\mu^\star - \mu_{A_{p,r}}) \qquad (14) \\
&= \sum_{p=1}^{T'}\left[\Delta_{A_p} + \mu^\star(n_p - 1)\right] \\
&\overset{(a)}{\leq} \sum_{p=1}^{T'}\left[\Delta_{A_p} + \mu^\star\sum_{r=0}^{Q-1}\left(2 \cdot \mathbb{1}\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\} - 2\right)\right] \\
&= \sum_{p=1}^{T'}\left[\Delta_{A_p} + \mu^\star\sum_{r=1}^{Q-1} 2 \cdot \mathbb{1}\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\}\right]
\end{aligned}$$

where (a) makes use of Lemma B.4 and the fact if $U^{\mathrm{v}}_{A_p}(p-1) \in ((m-1)\bar{\sigma}^2, m\bar{\sigma}^2]$, then $m = \sum_{r=0}^{Q-1} \mathbb{1}\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\} = \sum_{r=0}^{Q-1} \mathbb{1}\{\mathcal{U}_p(r)\}$. Note the fact that only the suboptimal solutions yields positive mean gap and that the negative mean gap of the risky solutions can be upper bounded by 0, thus the above equation can be further divided into two parts

$$\begin{aligned}
&\sum_{p=1}^{T'} \left[\Delta_{A_p} + \mu^\star \sum_{r=1}^{Q-1} 2 \cdot \mathbb{1}\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\} \right] \\
&\leq \sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{B}\}\Delta_{A_p} + \sum_{p=1}^{T'} \mu^\star \sum_{r=1}^{Q-1} 2 \cdot \mathbb{1}\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\} \\
&=: \mathrm{R}_1(T') + \mathrm{R}_2(T')
\end{aligned}$$

In conclusion, by summarizing the regret from the initialization stage, the regret due to failure of the good event and the high-probability regret, the expected regret can be bounded by

$$\begin{aligned}
\mathbb{E}[\mathrm{R}(T')] &\leq \mathbb{E}[\mathrm{R}_1(T')|\mathcal{E}]\mathbb{P}[\mathcal{E}] + \mathbb{E}[\mathrm{R}_2(T')|\mathcal{E}]\mathbb{P}[\mathcal{E}] + 2\mu^\star \cdot TL\left(\xi(\omega_\mu) + 2\xi(\omega_v) + 2\xi(\omega'_v)\right) + 2\mu^\star L \\
&\leq \mathbb{E}[\mathrm{R}_1(T')|\mathcal{E}] + \mathbb{E}[\mathrm{R}_2(T')|\mathcal{E}] + 2\mu^\star \cdot TL\left(\xi(\omega_\mu) + 2\xi(\omega_v) + 2\xi(\omega'_v)\right) + 2\mu^\star L \\
&= \mathbb{E}[\mathrm{R}_1(T')|\mathcal{E}] + \mathbb{E}[\mathrm{R}_2(T')|\mathcal{E}] + \mathrm{R}_3(T)
\end{aligned}$$

□

B.3. Regret due to suboptimality

Lemma 6.2. *Conditioned on event $\mathcal{E}$, the regret due to suboptimality $\mathrm{R}_1(T')$ can be bounded by*

$$O\Big(\sum_{i \in E \setminus S^\star} \frac{K}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}} \ln \frac{1}{\omega_\mu} + \sum_{i \in E} \frac{c_i K}{\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min}} \ln \frac{1}{\omega'_{\mathrm{v}}} \Big).$$

Proof of Lemma 6.2. Notice the fact that the suboptimal solution A_p can be safe, i.e. $A_p \in \mathcal{S} \cap \mathcal{B}$, or unsafe, i.e., $A_p \in \mathcal{S}^c \cap \mathcal{B}$, we upper bound the regret under these the two scenarios separately.

Case 1**:** $A_p \in \mathcal{S} \cap \mathcal{B}$

By the definition of the event $\mathcal{G}^\mu_{j,p}$, we have

$$|G^\mu_{j,p}| = \left|\left\{i \in A_p \setminus S^\star : T_i(p-1) \leq m_j(\Delta_{A_p}, \omega_\mu)\right\}\right| \leq b_j K$$

which indicates

$$\mathbb{1}\left\{A_p \in \mathcal{S} \cap \mathcal{B}, \mathcal{G}^\mu_{j,p}\right\} \leq \frac{1}{b_j K} \sum_{i \in A_p \setminus S^\star} \mathbb{1}\left\{A_p \in \mathcal{S} \cap \mathcal{B}, T_i(p-1) \leq m_j(\Delta_{A_p}, \omega_\mu)\right\}.$$

Given an item $i \in E \setminus S^\star$, assume it is included v_i solutions in $\mathcal{S} \cap \mathcal{B}$, we index them according to the decreasing order of their mean gaps, i.e. $\{A^{i,k}\}_{k \in [v_i]}$ with $\Delta_{A^{i,1}} \geq \ldots \geq \Delta_{A^{i,v_i}}$. Therefore,

$$\begin{aligned}
&\sum_{p=1}^{T} \mathbb{1}\left\{A_p \in \mathcal{S} \cap \mathcal{B}\right\} \Delta_{A_p} \cdot \mathbb{1}\left\{\mathcal{E}\right\} \\
&\overset{(a)}{\leq} \sum_{p=1}^{T} \mathbb{1}\left\{A_p \in \mathcal{S} \cap \mathcal{B}, \mathcal{F}^\mu_p\right\} \Delta_{A_p} \\
&\overset{(b)}{\leq} \sum_{p=1}^{T} \sum_{j \in \mathbb{N}} \mathbb{1}\left\{A_p \in \mathcal{S} \cap \mathcal{B}, \mathcal{G}^\mu_{j,p}\right\} \Delta_{A_p} \\
&\leq \sum_{p=1}^{T} \sum_{j \in \mathbb{N}} \frac{1}{b_j K} \sum_{i \in A_p \setminus S^\star} \mathbb{1}\left\{A_p \in \mathcal{S} \cap \mathcal{B}, T_i(p-1) \leq m_j(\Delta_{A_p}, \omega_\mu)\right\} \Delta_{A_p}
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{p=1}^{T}\sum_{j\in\mathbb{N}}\frac{1}{b_jK}\sum_{i\in E\setminus S^\star}\sum_{k\in[v_i]}\mathbb{1}\left\{A^{i,k}=A_p, i\in A_p, T_i(p-1)\leq m_j(\Delta_{A^{i,k}},\omega_\mu)\right\}\Delta_{A^{i,k}}\\
&\leq \sum_{i\in E\setminus S^\star}\sum_{j\in\mathbb{N}}\sum_{p=1}^{T}\sum_{k\in[v_i]}\frac{\Delta_{A^{i,k}}}{b_jK}\mathbb{1}\left\{A^{i,k}=A_p, i\in A^{i,k}, T_i(p-1)\leq\frac{a_j\cdot\gamma K^2}{\Delta^2_{A^{i,k}}}\left(2\ln\frac{1}{\omega_\mu}+\ln\ln_+\frac{1}{\Delta^2_{A^{i,k}}}+D\right)\right\}\\
&\overset{(c)}{\leq} \sum_{i\in E\setminus S^\star}\sum_{j\in\mathbb{N}}\sum_{p=1}^{T}\sum_{k\in[v_i]}\frac{\Delta_{A^{i,k}}}{b_jK}\mathbb{1}\left\{A^{i,k}=A_p, i\in A^{i,k}, T_i(p-1)\leq\frac{a_j\cdot\gamma K^2}{\Delta^2_{A^{i,k}}}\left(2\ln\frac{1}{\omega_\mu}+\ln\ln_+\frac{1}{\Delta^2_{A^{i,v_i}}}+D\right)\right\}\\
&\overset{(d)}{\leq} \sum_{i\in E\setminus S^\star}\sum_{j\in\mathbb{N}}\frac{a_j\cdot\gamma K^2}{b_jK}\left(2\ln\frac{1}{\omega_\mu}+\ln\ln_+\frac{1}{\Delta^2_{A^{i,v_i}}}+D\right)\left(\frac{1}{\Delta_{A^{i,v_i}}}+\sum_{k=2}^{v_i-1}\Delta_{A^{i,k+1}}\left(\frac{1}{\Delta^2_{A^{i,k+1}}}-\frac{1}{\Delta^2_{A^{i,k}}}\right)\right)\\
&\leq \sum_{i\in E\setminus S^\star}\frac{2C\gamma K}{\Delta_{A^{i,v_i}}}\left(2\ln\frac{1}{\omega_\mu}+\ln\ln_+\frac{1}{\Delta^2_{A^{i,v_i}}}+D\right)
\end{aligned}
$$

where (a) and b make use of Lemma B.1 and Lemma B.2, (c) is obtained by relaxing $\ln\ln_+\frac{1}{\Delta_{A^{i,k}}}$ to $\ln\ln_+\frac{1}{\Delta_{A^{i,v_i}}}$, (d) is obtained by solving the optimization problem.

Case 2: $A_p\in\mathcal{S}^c\cap\mathcal{B}$

For the case where $\omega_\mu\leq\omega'_{\mathrm{v}}$, denote $\bar{\omega}:=\sqrt{\frac{\ln\frac{1}{\omega'_{\mathrm{v}}}}{\ln\frac{1}{\omega_\mu}}}$. Given an item $i\in E$, assume it is included v_i solutions in $\mathcal{S}^c\cap\mathcal{B}$, we index them according to the decreasing order of their gaps $\bar{\Delta}_A:=\max\left\{\bar{\omega}\Delta_A,\frac{\Delta^{\mathrm{v}}_A}{3}\right\}$, i.e. $\{A^{i,k}\}_{k\in[v_i]}$ with $\bar{\Delta}_{A^{i,1}}\geq\ldots\geq\bar{\Delta}_{A^{i,v_i}}$. Denote $c_i:=\max_{k\in[v_i]}\left(\frac{\Delta_{A^{i,k}}}{\bar{\Delta}_{A^{i,k}}}\right)^2$, $k_i=\arg\max_{k\in[v_i]}\Delta_{A^{i,k}}$ and $d_i=\min_{k\in[v_i]}\Delta_{A^{i,k}}$, i.e., the minimum mean gap.

We have

$$
\begin{aligned}
&\sum_{p=1}^{T}\mathbb{1}\left\{A_p\in\mathcal{S}^c\cap\mathcal{B}\right\}\Delta_{A_p}\cdot\mathbb{1}\left\{\mathcal{E}\right\}\\
&\overset{(a)}{\leq}\sum_{p=1}^{T}\mathbb{1}\left\{A_p\in\mathcal{S}^c\cap\mathcal{B},\mathcal{F}^\mu_p,\mathcal{F}_p\left(\frac{\Delta^{\mathrm{v}}_{A_p}}{3},\omega'_{\mathrm{v}}\right)\right\}\Delta_{A_p}\\
&\overset{(b)}{\leq}\sum_{p=1}^{T}\mathbb{1}\left\{A_p\in\mathcal{S}^c\cap\mathcal{B},\mathcal{F}_p\left(\max\left\{\bar{\omega}\Delta_{A_p},\frac{\Delta^{\mathrm{v}}_{A_p}}{3}\right\},\omega'_{\mathrm{v}}\right)\right\}\Delta_{A_p}\\
&\overset{(c)}{\leq}\sum_{p=1}^{T}\sum_{j\in\mathbb{N}}\mathbb{1}\left\{A_p\in\mathcal{S}^c\cap\mathcal{B},\mathcal{G}_{j,p}\left(\max\left\{\bar{\omega}\Delta_\mu,\frac{\Delta^{\mathrm{v}}_{A_p}}{3}\right\},\omega'_{\mathrm{v}}\right)\right\}\Delta_{A_p}\\
&\leq\sum_{p=1}^{T}\sum_{j\in\mathbb{N}}\frac{1}{b_jK}\sum_{i\in A_p}\mathbb{1}\left\{A_p\in\mathcal{S}^c\cap\mathcal{B},T_i(p-1)\leq m_j\left(\bar{\Delta}_{A_p},\omega'_{\mathrm{v}}\right)\right\}\Delta_{A_p}\\
&=\sum_{p=1}^{T}\sum_{j\in\mathbb{N}}\frac{1}{b_jK}\sum_{i\in E\setminus S^\star}\sum_{k\in[v_i]}\mathbb{1}\left\{A^{i,k}=A_p,i\in A_p,T_i(p-1)\leq m_j\left(\bar{\Delta}_{A_p},\omega'_{\mathrm{v}}\right)\right\}\Delta_{A^{i,k}}\\
&\leq\sum_{i\in E}\sum_{j\in\mathbb{N}}\sum_{p=1}^{T}\sum_{k\in[v_i]}\frac{\Delta_{A^{i,k}}}{b_jK}\mathbb{1}\left\{A^{i,k}=A_p,i\in A^{i,k},T_i(p-1)\leq\frac{a_j\cdot\gamma K^2}{\bar{\Delta}^2_{A^{i,k}}}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}}+\ln\ln_+\frac{1}{\bar{\Delta}^2_{A^{i,k}}}+D\right)\right\}\\
&\overset{(d)}{\leq}\sum_{i\in E}\sum_{j\in\mathbb{N}}\sum_{p=1}^{T}\sum_{k\in[v_i]}\frac{\Delta_{A^{i,k}}}{b_jK}\mathbb{1}\left\{A^{i,k}=A_p,i\in A^{i,k},T_i(p-1)\leq\frac{a_j\cdot\gamma K^2}{\bar{\Delta}^2_{A^{i,k}}}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}}+\ln\ln_+\frac{1}{\bar{\Delta}^2_{A^{i,v_i}}}+D\right)\right\}
\end{aligned}
\tag{15}
$$

$$\overset{(e)}{\leq} \sum_{i\in E}\sum_{j\in\mathbb{N}} \frac{a_j\cdot\gamma K^2}{b_j K}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}}+\ln\ln_+\frac{1}{\bar{\Delta}^2_{A^{i,v_i}}}+D\right)\left(\frac{c_i}{d_i}+c_i\cdot\left(\frac{1}{d_i}-\frac{1}{\Delta_{A^{i,k_i}}}\right)\right)$$
$$\leq \sum_{i\in E}\frac{2c_i\cdot C\gamma K}{d_i}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}}+\ln\ln_+\frac{1}{\bar{\Delta}^2_{A^{i,v_i}}}+D\right)$$

where (a) comes from Lemma B.1, (b) results from Lemma B.3, (c) is due to Lemma B.2, (d) is obtained by relaxing $\ln\ln_+\frac{1}{\bar{\Delta}^2_{A^{i,k}}}$ to $\ln\ln_+\frac{1}{\bar{\Delta}^2_{A^{i,v_i}}}$ and (e) is achieved by solving the optimization problem in (15):

$$\sum_{p=1}^{T}\sum_{k\in[v_i]}\frac{\Delta_{A^{i,k}}}{b_jK}\mathbb{1}\left\{A^{i,k}=A_p, i\in A^{i,k}, T_i(p-1)\leq\frac{a_j\cdot\gamma K^2}{\bar{\Delta}^2_{A^{i,k}}}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}}+\ln\ln_+\frac{1}{\bar{\Delta}^2_{A^{i,v_i}}}+D\right)\right\}$$
$$\leq\frac{a_j\cdot\gamma K^2}{b_jK}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}}+\ln\ln_+\frac{1}{\bar{\Delta}^2_{A^{i,v_i}}}+D\right)\cdot\left(\frac{c_i}{d_i}+\int_{d_i}^{\Delta_{A^{i,k_i}}}\frac{c_i}{x^2}\,\mathrm{d}x\right)$$
$$=\frac{a_j\cdot\gamma K^2}{b_jK}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}}+\ln\ln_+\frac{1}{\bar{\Delta}^2_{A^{i,v_i}}}+D\right)\left(\frac{c_i}{d_i}+c_i\cdot\left(\frac{1}{d_i}-\frac{1}{\Delta_{A^{i,k_i}}}\right)\right)$$

In conclusion, for $i\in E\setminus S^\star$, denote

$$\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}:=\min_{S\ni i, S\in\mathcal{S}\cap\mathcal{B}}\Delta_S$$

For $i\in E$, denote

$$\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min}:=\min_{S\ni i, S\in\mathcal{S}^c\cap\mathcal{B}}\Delta_S$$
$$\bar{\Delta}'_{i,\mathcal{S}^c\cap\mathcal{B}}:=\min_{S\ni i, S\in\mathcal{S}^c\cap\mathcal{B}}\max\{\bar{\omega}\Delta_S,\Delta^{\mathrm{v}}_S/3\}$$
$$c_i:=\max_{S\ni i, S\in\mathcal{S}^c\cap\mathcal{B}}\left(\frac{\Delta_S}{\max\{\bar{\omega}\Delta_S,\Delta^{\mathrm{v}}_S/3\}}\right)^2$$

The regret due to suboptimality can be upper bounded by

$$\mathrm{R}_1(T)\leq\sum_{i\in E\setminus S^\star}\frac{2C\gamma K}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}}\left(2\ln\frac{1}{\omega_\mu}+\ln\ln_+\frac{1}{\Delta^2_{i,\mathcal{S}\cap\mathcal{B},\min}}+D\right)$$
$$+\sum_{i\in E}\frac{2c_i\cdot C\gamma K}{\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min}}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}}+\ln\ln_+\frac{1}{(\bar{\Delta}'_{i,\mathcal{S}^c\cap\mathcal{B}})^2}+D\right)$$

□

B.4. Regret due to safeness-checking

We firstly introduce two more technical lemmas in order to upper bound the regret. We characterize the number of sub-solutions n_p in Algorithm 2 by the following lemma.

Lemma B.4. *At any phase p, we have $\mathbb{P}[n_p\leq Q]=1$. Furthermore, if $U^{\mathrm{v}}_{A_p}(p-1)\in((m-1)\bar{\sigma}^2,m\bar{\sigma}^2]$ for some $m\in\mathbb{N}$, then $m\leq n_p\leq 2m-1$.*

Proof of Lemma B.4. Recall that an absolutely safe solution S is safe w.p. 1 and $S\in\mathcal{S}_{p-1}$, thus $|A_{p,r}|\geq q,\forall r\in[n_p-1]$. If $n_p>Q$, i.e. $n_p\geq Q+1$, then

$$K\geq|A_p|=\sum_{r=1}^{n_p}|A_{p,r}|>\sum_{r=1}^{Q}|A_{p,r}|\geq Q\cdot q\geq K$$

which constitutes a contradiction. Therefore, we have $\mathbb{P}[n_p \leq Q] = 1$.

If $U^{\mathrm{v}}_{A_p}(p-1) \in ((m-1)\bar{\sigma}^2, m\bar{\sigma}^2]$ for some $m \in \mathbb{N}_+$, we sequentially define $\{j_1, j_2, \ldots\} \subset [|A_p|]$ as follows: denote $j_0 := 0$ and let j_1 be the integer such that

$$\sum_{s=1}^{j_1-1} U^{\mathrm{v}}_{i_s}(p-1) \leq \bar{\sigma}^2 \quad \text{and} \quad \sum_{s=1}^{j_1} U^{\mathrm{v}}_{i_s}(p-1) > \bar{\sigma}^2.$$

Then let $j_2 > j_1$ be the integer such that

$$\sum_{s=j_1+1}^{j_2-1} U^{\mathrm{v}}_{i_s}(p-1) \leq \bar{\sigma}^2 \quad \text{and} \quad \sum_{s=j_1+1}^{j_2} U^{\mathrm{v}}_{i_s}(p-1) > \bar{\sigma}^2.$$

$$\vdots$$

The last integer $j_k = |A_p|$ satisfies

$$0 < \sum_{s=j_{k-1}+1}^{j_k} U^{\mathrm{v}}_{i_s}(p-1) \leq \bar{\sigma}^2.$$

If $k \geq m+1$, we must have

$$\begin{aligned} U^{\mathrm{v}}_{A_t}(p-1) &= \sum_{l=1}^{k} \sum_{s=j_{l-1}+1}^{j_l} U^{\mathrm{v}}_{i_s}(p-1) \\ &\geq \sum_{l=1}^{m} \sum_{s=j_{l-1}+1}^{j_l} U^{\mathrm{v}}_{i_s}(p-1) > m\bar{\sigma}^2 \end{aligned}$$

which contradicts with $U^{\mathrm{v}}_{A_p}(p-1) \in ((m-1)\bar{\sigma}^2, m\bar{\sigma}^2]$. Hence, $k \leq m$. Then we construct the sub-solutions by

$$\begin{aligned} &A_{p,l} = \{i_{j_{l-1}+1}, \ldots, i_{j_l-1}\}, \quad \forall l \in [k-1] \\ &\text{and} \quad A_{p,k} = \{i_{j_{k-1}+1}, \ldots, i_{j_k}\}. \end{aligned}$$

There are $k-1$ items $\{i_{j_l} : l \in [k-1]\}$ left which will compose at most $k-1$ additional sub-solutions. In conclusion, we need at most $2m-1$ sub-solutions, i.e., $n_p \leq 2m-1$. Obviously, we need at least m sub-solutions since $\bar{\sigma}^2 > \sigma^2$. So $m \leq n_p \leq 2m-1$. □

Remark B.5. *Indexing the items in Line* 3 *of* GREEDY-SPLIT *can be done arbitrarily, i.e., it does not require any specific order of the items. As such,* GREEDY-SPLIT *is an efficient greedy algorithm. We note that finding the optimal order that leads to the minimum number of sub-solutions n_p is a combinatorial problem which is generally hard to solve.*

Due to the fact that the upper confidence of any solution S satisfies $U^{\mathrm{v}}_S(p) \leq Q\bar{\sigma}^2$, thus n_p can at most be $2Q-1$.

Lemma 6.1 implies the key to upper bound the regret due to safeness-checking is to upper bound $\sum_{r=1}^{Q-1} \mathbb{1}\{\mathcal{U}_p(r)\}$ over the horizon T. From the definition of $\mathcal{U}_p(r) := \{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\}$, for $r_1, r_2 \in [Q-1]$ with $r_1 > r_2$, event $\mathcal{U}_p(r_1)$ indicates event $\mathcal{U}_p(r_2)$. Thus in order to upper bound $\sum_{r=1}^{Q-1} \mathbb{1}\{\mathcal{U}_p(r)\}$, it suffices to upper bound $\mathbb{1}\{\mathcal{U}_p(r)\}$ for $r \in [Q-1]$. To be more specific, given $l \in [Q-1]$, in order to compute the maximum number of times $\sum_{r=1}^{Q-1} \mathbb{1}\{\mathcal{U}_p(r)\} \geq l$, we only need to compute the maximum number of times event $\mathcal{U}_p(l)$ occurs.

In the following lemma, we show a necessary condition (in terms of event $\mathcal{F}_p(x, \omega)$) for event $\mathcal{U}_p(r)$.

Lemma B.6. *On the event $\mathcal{E}$,*

- *for $A_p \in \mathcal{S}$:*

$$\mathbb{1}\{\mathcal{U}_p(r)\} \leq \mathbb{1}\left\{\mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\}$$

- *for* $A_p \in \mathcal{S}^c$

$$\mathbb{1}\left\{\mathcal{U}_p(r)\right\} \leq \mathbb{1}\left\{\mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\}$$

Proof of Lemma B.6. The proof is straightforward

$$\begin{aligned}
& U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2 \\
\overset{(a)}{\Rightarrow} \quad & \hat{\sigma}^2_{A_p}(p-1) + \sum_{i \in A_p} \beta_{\mathrm{u}}(T_i(p-1)) > r\bar{\sigma}^2 \\
\overset{(b)}{\Rightarrow} \quad & \sigma^2_{A_p} + 2\sum_{i \in A_p} \beta_{\mathrm{u}}(T_i(p-1)) > r\bar{\sigma}^2 \\
\Rightarrow \quad & 2\sum_{i \in A_p} 3 \cdot \mathrm{lil}(T_i(p-1), \omega_{\mathrm{v}}) > (r-1)\bar{\sigma}^2 + (\bar{\sigma}^2 - \sigma^2_{A_p})
\end{aligned}$$

where (a) is due to the definition of the confidence bounds for a solution (4), and (b) utilizes the event $\bigcap_{i \in A_p} \mathcal{E}_{i,T_i(p-1)}$. For $A_p \in \mathcal{S}$, the above event is equivalent to $\mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)$; and for $A_p \in \mathcal{S}^c$, it is equivalent to $\mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)$. □

The above lemma and Lemma B.1 upper bound the components in $\mathrm{R}_2(T')$ by the $\mathcal{F}$ events.

We are now ready to bound the regret due to safeness checking.

Lemma 6.3. *On the event* $\mathcal{E}$*, if* $T' \in [T'_{r'}, T'_{r'-1})$ *then*

$$\mathrm{R}_2(T') \leq 2\mu^\star[T'(r'-1) + H(r', \Lambda)] \leq 2\mu^\star H(1, \Lambda)$$

Proof of Lemma 6.3. From Lemma 6.1, the high-probability regret due to safeness-checking is

$$\begin{aligned}
\mathrm{R}_2(T') &= \mu^\star \sum_{p=1}^{T'} \left[2\sum_{r=1}^{Q-1} \mathbb{1}\left\{\mathcal{U}_p(r)\right\}\right] \\
&= 2\mu^\star \sum_{r=1}^{Q-1}\sum_{p=1}^{T'} \mathbb{1}\left\{\mathcal{U}_p(r)\right\}
\end{aligned}$$

In the following, given $r \in [Q-1]$, we are going to upper bound $\sum_{p=1}^{T'} \mathbb{1}\{\mathcal{U}_p(r)\}$ conditional on event $\mathcal{E}$. When $U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2$ holds, there are at most $2r+1$ solutions being chosen at phase p, i.e. $n_p \leq 2r+1$, according to Lemma B.4. Therefore, we are also deriving an upper bound for number of phases in which there are at most $2r+1$ sub-solutions being sampled.

The proof scheme is planned as follows: in **Step 1**, we decompose the event $\mathcal{U}_p(r)$ into 4 events according to where A_p lies, i.e. (1) $A_p = S^\star$; (2) $\mathcal{S} \cap \mathcal{B}$; (3) $\mathcal{R}$ and (4) $\mathcal{S}^c \cap \mathcal{B}$. We will upper bound the regret under each of these cases.

In **Step 2**, we apply Lemma B.1 to upper bound the number of times a solution A can be selected via the events $\mathcal{F}^\mu_p$ and $\mathcal{F}_p(x, \omega)$. Because there will be multiple $\mathcal{F}^\mu_p$ and $\mathcal{F}_p(x, \omega)$ events in the indicator function, Lemma B.3 will be adopted to merge them into one event. After that, Lemma B.2 is utilized to bridge the number of times a solution is identified to the number of times of an item is sampled. At the end of this step, we conclude the number of times $\mathbb{1}\{\mathcal{U}_p(r)\}$ occurs under the four cases.

In **Step 3**, we upper bound $\sum_{r=1}^{Q-1}\sum_{p=1}^{T'} \mathbb{1}\left\{\mathcal{U}_p(r)\right\}$ based on the results from Step 2.

Step 1: We decompose $\sum_{p=1}^{T'} \mathbb{1}\{\mathcal{U}_p(r)\}$ conditional on event $\mathcal{E}$ into four parts:

$$
\begin{aligned}
&\sum_{p=1}^{T'} \mathbb{1}\{\mathcal{U}_p(r)\} \\
&\leq \sum_{p=1}^{T'} \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \\
&= \sum_{p=1}^{T'} \Big(\mathbb{1}\{A_p = S^\star\} + \mathbb{1}\{A_p \in \mathcal{S}\cap\mathcal{B}\} + \mathbb{1}\{A_p \in \mathcal{R}\} + \mathbb{1}\{A_p \in \mathcal{S}^c\cap\mathcal{B}\}\Big) \cdot \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \\
&= \sum_{p=1}^{T'} \mathbb{1}\{A_p = S^\star\}\, \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} + \sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{S}\cap\mathcal{B}\}\, \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \\
&\quad + \sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\}\, \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} + \sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{S}^c\cap\mathcal{B}\}\Big) \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\}
\end{aligned}
$$

Step 2: For each of the scenarios, we firstly upper bound the regret by the "$\mathcal{F}$ events" and they can be further bounded in terms of "$\mathcal{G}$ events".

Case 1: $A_p = S^\star$

$$
\begin{aligned}
&\sum_{p=1}^{T'} \mathbb{1}\{A_p = S^\star\}\, \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \\
&\overset{(a)}{\leq} \sum_{p=1}^{T'} \mathbb{1}\{A_p = S^\star\}\, \mathbb{1}\left\{\mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&\overset{(b)}{\leq} \sum_{p=1}^{T'} \sum_{j\in\mathbb{N}} \mathbb{1}\{A_p = S^\star\}\, \mathbb{1}\left\{\mathcal{G}_{j,p}\left(\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&\overset{(c)}{\leq} \sum_{p=1}^{T'} \sum_{j\in\mathbb{N}} \mathbb{1}\{A_p = S^\star\} \frac{1}{b_j K} \sum_{i\in A_p} \mathbb{1}\left\{T_i(p-1) \leq m_j\left(\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&= \sum_{i\in S^\star} \sum_{j\in\mathbb{N}} \sum_{p=1}^{T'} \frac{1}{b_j K} \mathbb{1}\left\{T_i(p-1) \leq m_j\left(\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{S^\star}}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&\leq \sum_{i\in S^\star} \sum_{j\in\mathbb{N}} \frac{1}{b_j K} m_j\left(\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{S^\star}}{3}, \omega_{\mathrm{v}}\right) \\
&= \sum_{i\in S^\star} \sum_{j\in\mathbb{N}} \frac{a_j \cdot \gamma K^2}{b_j K} \frac{9}{((r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{S^\star})^2}\left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+ \frac{1}{((r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{S^\star})^2} + D\right) \\
&\leq \sum_{i\in S^\star} C \cdot \frac{9\cdot\gamma K}{((r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{S^\star})^2}\left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+ \frac{1}{{\Delta^{\mathrm{v}}_{S^\star}}^2} + D\right)
\end{aligned}
$$

where (a) utilizes Lemma B.6, (b) makes use of Lemma B.2 and (c) follows the definition of $\mathcal{G}_{j,p}$. For simplicity, we denote

$$
g_{S^\star}(r, \Delta^{\mathrm{v}}_{S^\star}) = C \cdot \frac{9\cdot\gamma K}{((r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{S^\star})^2}\left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+ \frac{1}{{\Delta^{\mathrm{v}}_{S^\star}}^2} + D\right).
$$

Thus

$$\sum_{p=1}^{T'} \mathbb{1}\left\{A_p = S^\star\right\} \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \leq \sum_{i \in S^\star} g_{S^\star}(r, \Delta^{\mathrm{v}}_{S^\star})$$

Case 2: $A_p \in \mathcal{S} \cap \mathcal{B}$

Under this case, there will be a comparison between ω_μ and ω_{v} thus we denote $\tilde{\omega} = \sqrt{\frac{\ln \frac{1}{\omega_\mu}}{\ln \frac{1}{\omega_{\mathrm{v}}}}}$. For $i \in E$, denote $\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}} := \min_{S \ni i, S \in \mathcal{S}\cap\mathcal{B}} \max\left\{\frac{\Delta_S}{\sqrt{\ln(1/\omega_\mu)}}, \frac{\Delta^{\mathrm{v}}_S}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}\right\}$ which is achieved by solution $S_{i,\mathcal{S}\cap\mathcal{B}}$, and assume $\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}} \in (\frac{r_i\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}, \frac{(r_i+1)\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}]$ for some $r_i \in \mathbb{N}$.

Scenario 1: $\omega_\mu \geq \omega_{\mathrm{v}}$

We firstly deal with the case where $\omega_\mu \geq \omega_{\mathrm{v}}$, i.e., $\tilde{\omega} \leq 1$. We have

$$\begin{aligned}
&\sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \\
&\overset{(a)}{\leq} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{\mathcal{F}^\mu_p, \mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&\overset{(b)}{\leq} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{\mathcal{F}_p\left(\Delta_{A_p}, \omega_\mu\right), \mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&\overset{(c)}{\leq} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{\mathcal{F}_p\left(\max\left\{\Delta_{A_p}, \tilde{\omega}\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_\mu\right)\right\} \\
&\overset{(d)}{\leq} \sum_{p=1}^{T'}\sum_{j\in\mathbb{N}} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{\mathcal{G}_{j,p}\left(\max\left\{\Delta_{A_p}, \tilde{\omega}\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_\mu\right)\right\} \\
&\overset{(e)}{\leq} \sum_{p=1}^{T'}\sum_{j\in\mathbb{N}} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \frac{1}{b_j K} \sum_{i\in A_p} \mathbb{1}\left\{T_i(p-1) \leq m_j\left(\max\left\{\Delta_{A_p}, \tilde{\omega}\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_\mu\right)\right\} \\
&= \sum_{i\in E}\sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \leq m_j\left(\max\left\{\Delta_{A_p}, \tilde{\omega}\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_\mu\right)\right\} \quad (16)
\end{aligned}$$

where (a) utilizes Lemma B.6, (b) and (c) make use of Lemma B.3, (d) is due to Lemma B.2 and (e) follows the definition of $\mathcal{G}_{j,p}$.

Given $i \in E$,

(1) if $r \geq r_i + 2$, for any $S \in \mathcal{S}\cap\mathcal{B}$ that contains item i,

$$\max\left\{\Delta_S, \tilde{\omega}\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_S}{3}\right\} \geq \tilde{\omega}\frac{(r-1)\bar{\sigma}^2}{3} \geq \tilde{\omega}\frac{(r_i+1)\bar{\sigma}^2}{3} \geq \sqrt{\ln\frac{1}{\omega_\mu}} \cdot \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}.$$

Thus,

$$\begin{aligned}
&\sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \leq m_j\left(\max\left\{\Delta_{A_p}, \tilde{\omega}\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_\mu\right)\right\} \\
&\sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \leq m_j\left(\tilde{\omega}\frac{(r-1)\bar{\sigma}^2}{3}, \omega_\mu\right)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j\in\mathbb{N}} \frac{1}{b_j K} m_j \left(\tilde{\omega}\frac{(r-1)\bar{\sigma}^2}{3}, \omega_\mu\right) \\
&= \sum_{j\in\mathbb{N}} \frac{a_j\cdot\gamma K^2}{b_j K}\frac{9}{((r-1)\bar{\sigma}^2)^2}\left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \frac{1}{\tilde{\omega}^2}\ln\ln_+\frac{9}{(\tilde{\omega}(r-1)\bar{\sigma}^2)^2} + \frac{1}{\tilde{\omega}^2}D\right) \\
&\le C\cdot\frac{9\cdot\gamma K}{((r-1)\bar{\sigma}^2)^2}\left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \frac{1}{\tilde{\omega}^2}\ln\ln_+\frac{1}{\ln(1/\omega_\mu)\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + \frac{1}{\tilde{\omega}^2}D\right) \\
&= C\cdot\frac{\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}})^2}\left(2 + \frac{1}{\ln(1/\omega_\mu)}\ln\ln_+\frac{1}{\ln(1/\omega_\mu)\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + \frac{1}{\ln(1/\omega_\mu)}D\right)
\end{aligned}
$$

(2) if $r \le r_i + 1$, for any $S \in \mathcal{S}\cap\mathcal{B}$ that contains item i

$$
\max\left\{\Delta_S, \tilde{\omega}\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_S}{3}\right\} \ge \max\left\{\Delta_S, \tilde{\omega}\frac{\Delta^{\mathrm{v}}_S}{3}\right\} \ge \sqrt{\ln\frac{1}{\omega_\mu}}\cdot\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}.
$$

Thus,

$$
\begin{aligned}
&\sum_{j\in\mathbb{N}}\frac{1}{b_j K}\sum_{p=1}^{T'}\mathbb{1}\{A_p\in\mathcal{S}\cap\mathcal{B}\}\,\mathbb{1}\left\{i\in A_p, T_i(p-1)\le m_j\left(\max\left\{\Delta_{A_p}, \tilde{\omega}\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_\mu\right)\right\} \\
&\le \sum_{j\in\mathbb{N}}\frac{1}{b_j K}\sum_{p=1}^{T'}\mathbb{1}\{A_p\in\mathcal{S}\cap\mathcal{B}\}\,\mathbb{1}\left\{i\in A_p, T_i(p-1)\le m_j\left(\sqrt{\ln\frac{1}{\omega_\mu}}\cdot\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}, \omega_\mu\right)\right\} \\
&\overset{(a)}{\le} \sum_{j\in\mathbb{N}}\frac{1}{b_j K} m_j\left(\sqrt{\ln\frac{1}{\omega_\mu}}\cdot\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}, \omega_\mu\right) \\
&= \sum_{j\in\mathbb{N}}\frac{a_j\cdot\gamma K^2}{b_j K}\frac{1}{(\sqrt{\ln\frac{1}{\omega_\mu}}\cdot\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}})^2}\left(2\ln\frac{1}{\omega_\mu} + \ln\ln_+\frac{1}{(\sqrt{\ln\frac{1}{\omega_\mu}}\cdot\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}})^2} + D\right) \\
&\le C\cdot\frac{\gamma K}{\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}\left(2 + \frac{1}{\ln(1/\omega_\mu)}\ln\ln_+\frac{1}{\ln(1/\omega_\mu)\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + \frac{1}{\ln(1/\omega_\mu)}D\right)
\end{aligned}
$$

where (a) is achieved by sampling $A_{i,\mathcal{S}\cap\mathcal{B}}$. Therefore, if we denote $g_{\mathcal{S}\cap\mathcal{B},1}(r, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) :=$

$$
\begin{cases}
\dfrac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}})^2}\left(2 + \dfrac{1}{\ln(1/\omega_\mu)}\ln\ln_+\dfrac{1}{\ln(1/\omega_\mu)\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + \dfrac{1}{\ln(1/\omega_\mu)}D\right), r \ge \left\lfloor\dfrac{3\sqrt{\ln(1/\omega_{\mathrm{v}})}\cdot\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\bar{\sigma}^2}\right\rfloor + 2 \\
\dfrac{C\gamma K}{\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}\left(2 + \dfrac{1}{\ln(1/\omega_\mu)}\ln\ln_+\dfrac{1}{\ln(1/\omega_\mu)\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + \dfrac{1}{\ln(1/\omega_\mu)}D\right), r \le \left\lfloor\dfrac{3\sqrt{\ln(1/\omega_{\mathrm{v}})}\cdot\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\bar{\sigma}^2}\right\rfloor + 1
\end{cases}
$$

then (16) can be upper bounded by

$$
\sum_{i\in E} g_{\mathcal{S}\cap\mathcal{B},1}(r, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}})
$$

Scenario 2: $\omega_\mu \le \omega_{\mathrm{v}}$

For the case where $\omega_\mu \le \omega_{\mathrm{v}}$, i.e., $\tilde{\omega} \ge 1$. We have

$$
\begin{aligned}
&\sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \\
&\overset{(a)}{\le} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{\mathcal{F}^{\mu}_p, \mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&\overset{(b)}{\le} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{\mathcal{F}_p\left(\Delta_{A_p}, \omega_\mu\right), \mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&\overset{(c)}{\le} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{\mathcal{F}_p\left(\max\left\{\frac{\Delta_{A_p}}{\tilde{\omega}}, \frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_{\mathrm{v}}\right)\right\} \\
&\overset{(d)}{\le} \sum_{p=1}^{T'}\sum_{j\in\mathbb{N}} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{\mathcal{G}_{j,p}\left(\max\left\{\frac{\Delta_{A_p}}{\tilde{\omega}}, \frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_{\mathrm{v}}\right)\right\} \\
&\overset{(e)}{\le} \sum_{p=1}^{T'}\sum_{j\in\mathbb{N}} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \frac{1}{b_j K} \sum_{i\in A_p} \mathbb{1}\left\{T_i(p-1) \le m_j\left(\max\left\{\frac{\Delta_{A_p}}{\tilde{\omega}}, \frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_{\mathrm{v}}\right)\right\} \\
&= \sum_{i\in E}\sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \le m_j\left(\max\left\{\frac{\Delta_{A_p}}{\tilde{\omega}}, \frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_{\mathrm{v}}\right)\right\} \quad (17)
\end{aligned}
$$

Given $i \in E$,

(1) if $r \ge r_i + 2$, for any $S \in \mathcal{S}\cap\mathcal{B}$ that contains item i,

$$
\max\left\{\frac{\Delta_S}{\tilde{\omega}}, \frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_S}{3}\right\} \ge \frac{(r-1)\bar{\sigma}^2}{3} \ge \frac{(r_i+1)\bar{\sigma}^2}{3} \ge \sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}} \cdot \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}.
$$

Thus,

$$
\begin{aligned}
&\sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \le m_j\left(\max\left\{\frac{\Delta_{A_p}}{\tilde{\omega}}, \frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_{\mathrm{v}}\right)\right\} \\
&\le \sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \le m_j\left(\frac{(r-1)\bar{\sigma}^2}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&= \sum_{j\in\mathbb{N}} \frac{1}{b_j K} m_j\left(\frac{(r-1)\bar{\sigma}^2}{3}, \omega_{\mathrm{v}}\right) \\
&= \sum_{j\in\mathbb{N}} \frac{a_j \cdot \gamma K^2}{b_j K} \frac{9}{((r-1)\bar{\sigma}^2)^2}\left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+ \frac{9}{((r-1)\bar{\sigma}^2)^2} + D\right) \\
&\le C \cdot \frac{9\cdot\gamma K}{((r-1)\bar{\sigma}^2)^2}\left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+ \frac{1}{\ln(1/\omega_{\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + D\right) \\
&= C \cdot \frac{\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}})^2}\left(2 + \frac{1}{\ln(1/\omega_{\mathrm{v}})}\ln\ln_+ \frac{1}{\ln(1/\omega_{\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + \frac{1}{\ln(1/\omega_{\mathrm{v}})} D\right)
\end{aligned}
$$

(2) if $r \le r_i + 1$, for any $S \in \mathcal{S}\cap\mathcal{B}$ that contains item i,

$$
\max\left\{\frac{\Delta_S}{\tilde{\omega}}, \frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_S}{3}\right\} \ge \max\left\{\frac{\Delta_S}{\tilde{\omega}}, \frac{\Delta^{\mathrm{v}}_S}{3}\right\} \ge \sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}} \cdot \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}.
$$

Thus,

$$\begin{aligned}
&\sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{ i \in A_p, T_i(p-1) \le m_j \left(\max\left\{ \frac{\Delta_{A_p}}{\tilde{\omega}}, \frac{(r-1)\bar{\sigma}^2 + \Delta^{\mathrm{v}}_{A_p}}{3} \right\}, \omega_{\mathrm{v}} \right) \right\} \\
&\le \sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{ i \in A_p, T_i(p-1) \le m_j \left(\sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}} \cdot \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}, \omega_{\mathrm{v}} \right) \right\} \\
&= \sum_{j\in\mathbb{N}} \frac{1}{b_j K} m_j \left(\sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}} \cdot \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}, \omega_{\mathrm{v}} \right) \\
&= \sum_{j\in\mathbb{N}} \frac{a_j \cdot \gamma K^2}{b_j K} \frac{1}{(\sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}} \cdot \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}})^2} \left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+ \frac{1}{(\sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}} \cdot \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}})^2} + D \right) \\
&\le C \cdot \frac{\gamma K}{\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} \left(2 + \frac{1}{\ln(1/\omega_{\mathrm{v}})} \ln\ln_+ \frac{1}{\ln(1/\omega_{\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + \frac{1}{\ln(1/\omega_{\mathrm{v}})} D \right)
\end{aligned}$$

Therefore, if we denote $g_{\mathcal{S}\cap\mathcal{B},2}(r, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) :=$

$$\begin{cases}
\dfrac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}})^2} \left(2 + \dfrac{1}{\ln(1/\omega_{\mathrm{v}})} \ln\ln_+ \dfrac{1}{\ln(1/\omega_{\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + \dfrac{1}{\ln(1/\omega_{\mathrm{v}})} D \right), r \ge \left\lfloor \dfrac{3\sqrt{\ln(1/\omega_{\mathrm{v}})} \cdot \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\bar{\sigma}^2} \right\rfloor + 2 \\
\dfrac{C\gamma K}{\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} \left(2 + \dfrac{1}{\ln(1/\omega_{\mathrm{v}})} \ln\ln_+ \dfrac{1}{\ln(1/\omega_{\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + \dfrac{1}{\ln(1/\omega_{\mathrm{v}})} D \right), r \le \left\lfloor \dfrac{3\sqrt{\ln(1/\omega_{\mathrm{v}})} \cdot \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\bar{\sigma}^2} \right\rfloor + 1
\end{cases}$$

then (17) can be upper bounded by

$$\sum_{i\in E} g_{\mathcal{S}\cap\mathcal{B},2}(r, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}).$$

In conclusion, for $i \in E$, we denote $\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}} := \min_{S\ni i, S\in\mathcal{S}\cap\mathcal{B}} \max\left\{ \frac{\Delta_S}{\sqrt{\ln(1/\omega_\mu)}}, \frac{\Delta^{\mathrm{v}}_A}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}} \right\}$, $\omega_{\mu\mathrm{v}} := \max\{\omega_\mu, \omega_{\mathrm{v}}\}$ and $g_{\mathcal{S}\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) :=$

$$\begin{cases}
\dfrac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}})^2} \left(2 + \dfrac{1}{\ln(1/\omega_{\mu\mathrm{v}})} \left(\ln\ln_+ \dfrac{1}{\ln(1/\omega_{\mu\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + D \right) \right), r \ge \left\lfloor \dfrac{3\sqrt{\ln(1/\omega_{\mathrm{v}})} \cdot \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\bar{\sigma}^2} \right\rfloor + 2 \\
\dfrac{C\gamma K}{\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} \left(2 + \dfrac{1}{\ln(1/\omega_{\mu\mathrm{v}})} \left(\ln\ln_+ \dfrac{1}{\ln(1/\omega_{\mu\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}} + D \right) \right), r \le \left\lfloor \dfrac{3\sqrt{\ln(1/\omega_{\mathrm{v}})} \cdot \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\bar{\sigma}^2} \right\rfloor + 1
\end{cases}$$

then

$$\sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}\cap\mathcal{B}\right\} \mathbb{1}\left\{ U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2 \right\} \le \sum_{i\in E} g_{\mathcal{S}\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}})$$

Case 3: $A_p \in \mathcal{R}$

Under this case, there will be a comparison between ω_{v} and ω'_{v} thus we denote $\bar{\omega} = \sqrt{\frac{\ln\frac{1}{\omega'_{\mathrm{v}}}}{\ln\frac{1}{\omega_{\mathrm{v}}}}}$ and $\omega_{\mathrm{sum}} := \sqrt{\ln\frac{1}{\omega'_{\mathrm{v}}}} + \sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}}$.

For $i \in E$, denote $\Delta^{\mathrm{v}}_{i,\mathcal{R}} := \min_{S\ni i, S\in\mathcal{R}} \Delta^{\mathrm{v}}_S$ and assume $\Delta^{\mathrm{v}}_{i,\mathcal{R}} \in (r_i \frac{\bar{\omega}}{\bar{\omega}+1}\bar{\sigma}^2, (r_i+1)\frac{\bar{\omega}}{\bar{\omega}+1}\bar{\sigma}^2]$.

Scenario 1: $\omega_{\mathrm{v}} \le \omega'_{\mathrm{v}}$ For the case $\omega_{\mathrm{v}} \le \omega'_{\mathrm{v}}$, we have $\bar{\omega} \le 1$.

$$
\begin{aligned}
&\sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{R}\right\} \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \\
&\overset{(a)}{\leq} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{R}\right\} \mathbb{1}\left\{\mathcal{F}_p\left(\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \omega'_{\mathrm{v}}\right), \mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&\overset{(b)}{\leq} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{R}\right\} \mathbb{1}\left\{\mathcal{F}_p\left(\max\left\{\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \bar{\omega}\cdot\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega'_{\mathrm{v}}\right)\right\} \\
&\overset{(c)}{\leq} \sum_{p=1}^{T'}\sum_{j\in\mathbb{N}} \mathbb{1}\left\{A_p \in \mathcal{R}\right\} \mathbb{1}\left\{\mathcal{G}_{j,p}\left(\max\left\{\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \bar{\omega}\cdot\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega'_{\mathrm{v}}\right)\right\} \\
&\overset{(d)}{\leq} \sum_{p=1}^{T'}\sum_{j\in\mathbb{N}} \mathbb{1}\left\{A_p \in \mathcal{R}\right\} \frac{1}{b_j K}\sum_{i\in A_p} \mathbb{1}\left\{T_i(p-1) \leq m_j\left(\max\left\{\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \bar{\omega}\cdot\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega'_{\mathrm{v}}\right)\right\} \\
&= \sum_{i\in E}\sum_{j\in\mathbb{N}} \frac{1}{b_j K}\sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{R}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \leq m_j\left(\max\left\{\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \bar{\omega}\cdot\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega'_{\mathrm{v}}\right)\right\} \quad (18)
\end{aligned}
$$

where (a) utilizes Lemma B.6, (b) makes use of Lemma B.3, (c) is due to Lemma B.2 and (d) follows the definition of $\mathcal{G}_{j,p}$.

Given $i \in E$,

(1) if $r \geq r_i + 2$, for any $S \in \mathcal{R}$ that contains item i,

$$
\max\left\{\frac{\Delta^{\mathrm{v}}_S}{3}, \bar{\omega}\cdot\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_S}{3}\right\} \geq \max\left\{\frac{\Delta^{\mathrm{v}}_S}{3}, \frac{\bar{\omega}}{1+\bar{\omega}}\cdot\frac{(r-1)\bar{\sigma}^2}{3}\right\} \geq \frac{\bar{\omega}}{1+\bar{\omega}}\cdot\frac{(r-1)\bar{\sigma}^2}{3} \geq \frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3}
$$

where the first inequality uses the fact that for $x, y \in \mathbb{R}_+$ and $z \in [0,1]$, , $\max\{x, y\} \geq \max\{x, xz + y(1-z)\}$. We take $x = \frac{\Delta^{\mathrm{v}}_{A_p}}{3}, y = \bar{\omega}\cdot\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}$ and $z = \frac{\bar{\omega}}{1+\bar{\omega}}$. Thus,

$$
\begin{aligned}
&\sum_{j\in\mathbb{N}} \frac{1}{b_j K}\sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{R}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \leq m_j\left(\max\left\{\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \bar{\omega}\cdot\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega'_{\mathrm{v}}\right)\right\} \\
&\leq \sum_{j\in\mathbb{N}} \frac{1}{b_j K}\sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{R}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \leq m_j\left(\frac{\bar{\omega}}{1+\bar{\omega}}\cdot\frac{(r-1)\bar{\sigma}^2}{3}, \omega'_{\mathrm{v}}\right)\right\} \\
&= \sum_{j\in\mathbb{N}} \frac{1}{b_j K} m_j\left(\frac{\bar{\omega}}{1+\bar{\omega}}\cdot\frac{(r-1)\bar{\sigma}^2}{3}, \omega'_{\mathrm{v}}\right) \\
&= \sum_{j\in\mathbb{N}} \frac{a_j\cdot\gamma K^2}{b_j K}\frac{1}{(\frac{\bar{\omega}}{1+\bar{\omega}}\cdot\frac{(r-1)\bar{\sigma}^2}{3})^2}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}} + \ln\ln_+\frac{1}{(\frac{\bar{\omega}}{1+\bar{\omega}}\cdot\frac{(r-1)\bar{\sigma}^2}{3})^2} + D\right) \\
&\leq \frac{C\gamma K}{(\frac{\bar{\omega}}{1+\bar{\omega}}\cdot\frac{(r-1)\bar{\sigma}^2}{3})^2}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}} + \ln\ln_+\frac{1}{(\frac{\bar{\omega}}{1+\bar{\omega}}\cdot\frac{(r-1)\bar{\sigma}^2}{3})^2} + D\right) \\
&\leq \frac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\omega_{\mathrm{sum}}})^2}\left(2 + \frac{1}{\ln(1/\omega'_{\mathrm{v}})}\left(\ln\ln_+\frac{1}{\ln(1/\omega'_{\mathrm{v}})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2} + D\right)\right)
\end{aligned}
$$

(2) if $r \leq r_i + 1$, for any $S \in \mathcal{R}$ that contains item i,

$$
\max\left\{\frac{\Delta^{\mathrm{v}}_S}{3}, \bar{\omega}\cdot\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_S}{3}\right\} \geq \max\left\{\frac{\Delta^{\mathrm{v}}_S}{3}, \frac{\bar{\omega}}{1+\bar{\omega}}\cdot\frac{(r-1)\bar{\sigma}^2}{3}\right\} \geq \frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3}
$$

Thus,

$$
\begin{aligned}
&\sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\} \mathbb{1}\left\{ i \in A_p, T_i(p-1) \le m_j \left(\max\left\{ \frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \bar{\omega} \cdot \frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3} \right\}, \omega'_{\mathrm{v}} \right) \right\} \\
&\le \sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\} \mathbb{1}\left\{ i \in A_p, T_i(p-1) \le m_j \left(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3}, \omega'_{\mathrm{v}} \right) \right\} \\
&= \sum_{j\in\mathbb{N}} \frac{1}{b_j K} m_j \left(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3}, \omega'_{\mathrm{v}} \right) \\
&= \sum_{j\in\mathbb{N}} \frac{a_j \cdot \gamma K^2}{b_j K} \frac{1}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3})^2} \left(2 \ln \frac{1}{\omega'_{\mathrm{v}}} + \ln\ln_+ \frac{1}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3})^2} + D \right) \\
&\le \frac{C\gamma K}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3})^2} \left(2 \ln \frac{1}{\omega'_{\mathrm{v}}} + \ln\ln_+ \frac{1}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3})^2} + D \right) \\
&= \frac{C\gamma K}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2} \left(2 + \frac{1}{\ln(1/\omega'_{\mathrm{v}})} \left(\ln\ln_+ \frac{1}{\ln(1/\omega'_{\mathrm{v}})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2} + D \right) \right)
\end{aligned}
$$

Therefore, if we denote

$$
g_{\mathcal{R},1}(r, \Delta^{\mathrm{v}}_{i,\mathcal{R}}) := \begin{cases} \frac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\omega_{\mathrm{sum}}})^2} \left(2 + \frac{1}{\ln(1/\omega'_{\mathrm{v}})} \left(\ln\ln_+ \frac{1}{\ln(1/\omega'_{\mathrm{v}})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2} + D \right) \right), r \ge \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2} \right\rfloor + 2 \\ \frac{C\gamma K}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2} \left(2 + \frac{1}{\ln(1/\omega'_{\mathrm{v}})} \left(\ln\ln_+ \frac{1}{\ln(1/\omega'_{\mathrm{v}})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2} + D \right) \right), r \le \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2} \right\rfloor + 1 \end{cases}
$$

then (18) can be upper bounded by

$$
\sum_{i\in E} g_{\mathcal{R},1}(r, \Delta^{\mathrm{v}}_{i,\mathcal{R}}).
$$

Scenario 2: $\omega_{\mathrm{v}} \ge \omega'_{\mathrm{v}}$ For the case $\omega_{\mathrm{v}} \ge \omega'_{\mathrm{v}}$, we have $\bar{\omega} \ge 1$.

$$
\begin{aligned}
&\sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\} \mathbb{1}\left\{ U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2 \right\} \\
&\overset{(a)}{\le} \sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\} \mathbb{1}\left\{ \mathcal{F}_p\left(\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \omega'_{\mathrm{v}} \right), \mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}} \right) \right\} \\
&\overset{(b)}{\le} \sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\} \mathbb{1}\left\{ \mathcal{F}_p\left(\max\left\{ \frac{1}{\bar{\omega}} \frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3} \right\}, \omega_{\mathrm{v}} \right) \right\} \\
&\overset{(c)}{\le} \sum_{p=1}^{T'} \sum_{j\in\mathbb{N}} \mathbb{1}\{A_p \in \mathcal{R}\} \mathbb{1}\left\{ \mathcal{G}_{j,p}\left(\max\left\{ \frac{1}{\bar{\omega}} \frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3} \right\}, \omega_{\mathrm{v}} \right) \right\} \\
&\overset{(d)}{\le} \sum_{p=1}^{T'} \sum_{j\in\mathbb{N}} \mathbb{1}\{A_p \in \mathcal{R}\} \frac{1}{b_j K} \sum_{i\in A_p} \mathbb{1}\left\{ T_i(p-1) \le m_j \left(\max\left\{ \frac{1}{\bar{\omega}} \frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3} \right\}, \omega_{\mathrm{v}} \right) \right\}
\end{aligned}
$$

$$= \sum_{i\in E}\sum_{j\in\mathbb{N}} \frac{1}{b_j K}\sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\}\, \mathbb{1}\left\{ i \in A_p, T_i(p-1) \le m_j \left(\max\left\{ \frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3} \right\}, \omega_{\mathrm{v}} \right) \right\} \quad (19)$$

where (a) utilizes Lemma B.6, (b) makes use of Lemma B.3, (c) is due to Lemma B.2 and (d) follows the definition of $\mathcal{G}_{j,p}$.

Given $i \in E$,

(1) if $r \ge r_i + 2$, for any $S \in \mathcal{R}$ that contains item i,

$$\max\left\{ \frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{S}}{3}, \frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{S}}{3} \right\} \ge \max\left\{ \frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{S}}{3}, \frac{\bar{\omega}}{1+\bar{\omega}} \cdot \frac{(r-1)\bar{\sigma}^2}{3} \right\} \ge \frac{1}{1+\bar{\omega}} \cdot \frac{(r-1)\bar{\sigma}^2}{3} \ge \frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3}$$

where the first inequality uses the fact that for $x, y \in \mathbb{R}_+$ and $z \in [0,1]$, , $\max\{x, y\} \ge \max\{x, xz + y(1-z)\}$. We take $x = \frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{A_p}}{3}$, $y = \frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}$ and $z = \frac{\bar{\omega}}{1+\bar{\omega}}$. Thus,

$$\begin{aligned}
&\sum_{j\in\mathbb{N}} \frac{1}{b_j K}\sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\}\, \mathbb{1}\left\{ i \in A_p, T_i(p-1) \le m_j \left(\max\left\{ \frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3} \right\}, \omega_{\mathrm{v}} \right) \right\} \\
&\le \sum_{j\in\mathbb{N}} \frac{1}{b_j K}\sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\}\, \mathbb{1}\left\{ i \in A_p, T_i(p-1) \le m_j \left(\frac{1}{1+\bar{\omega}} \cdot \frac{(r-1)\bar{\sigma}^2}{3}, \omega_{\mathrm{v}} \right) \right\} \\
&= \sum_{j\in\mathbb{N}} \frac{1}{b_j K} m_j \left(\frac{1}{1+\bar{\omega}} \cdot \frac{(r-1)\bar{\sigma}^2}{3}, \omega_{\mathrm{v}} \right) \\
&= \sum_{j\in\mathbb{N}} \frac{a_j \cdot \gamma K^2}{b_j K} \frac{1}{(\frac{1}{1+\bar{\omega}} \cdot \frac{(r-1)\bar{\sigma}^2}{3})^2} \left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+ \frac{1}{(\frac{1}{1+\bar{\omega}} \cdot \frac{(r-1)\bar{\sigma}^2}{3})^2} + D \right) \\
&\le \frac{C\gamma K}{(\frac{1}{1+\bar{\omega}} \cdot \frac{(r-1)\bar{\sigma}^2}{3})^2} \left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+ \frac{1}{(\frac{1}{1+\bar{\omega}} \cdot \frac{(r-1)\bar{\sigma}^2}{3})^2} + D \right) \\
&\le \frac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\omega_{\mathrm{sum}}})^2} \left(2 + \frac{1}{\ln(1/\omega_{\mathrm{v}})} \left(\ln\ln_+ \frac{1}{\ln(1/\omega_{\mathrm{v}})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2} + D \right) \right)
\end{aligned}$$

(2) if $r \le r_i + 1$, for any $S \in \mathcal{R}$ that contains item i,

$$\max\left\{ \frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{S}}{3}, \frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{S}}{3} \right\} \ge \max\left\{ \frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{A}}{3}, \frac{\bar{\omega}}{1+\bar{\omega}} \cdot \frac{(r-1)\bar{\sigma}^2}{3} \right\} \ge \frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3}$$

Thus,

$$\begin{aligned}
&\sum_{j\in\mathbb{N}} \frac{1}{b_j K}\sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\}\, \mathbb{1}\left\{ i \in A_p, T_i(p-1) \le m_j \left(\max\left\{ \frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3} \right\}, \omega_{\mathrm{v}} \right) \right\} \\
&\le \sum_{j\in\mathbb{N}} \frac{1}{b_j K}\sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\}\, \mathbb{1}\left\{ i \in A_p, T_i(p-1) \le m_j \left(\frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3}, \omega_{\mathrm{v}} \right) \right\} \\
&= \sum_{j\in\mathbb{N}} \frac{1}{b_j K} m_j \left(\frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3}, \omega_{\mathrm{v}} \right) \\
&= \sum_{j\in\mathbb{N}} \frac{a_j \cdot \gamma K^2}{b_j K} \frac{1}{(\frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3})^2} \left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+ \frac{1}{(\frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3})^2} + D \right) \\
&\le \frac{C\gamma K}{(\frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3})^2} \left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+ \frac{1}{(\frac{1}{\bar{\omega}}\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3})^2} + D \right)
\end{aligned}$$

$$= \frac{C\gamma K}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mathrm{v}})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mathrm{v}})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}+D\right)\right)$$

Therefore, if we denote

$$g_{\mathcal{R},2}(r,\Delta^{\mathrm{v}}_{i,\mathcal{R}}) := \begin{cases} \frac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\omega_{\mathrm{sum}}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mathrm{v}})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mathrm{v}})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}+D\right)\right), r \geq \left\lfloor \frac{\omega_{\mathrm{sum}}\cdot\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2}\right\rfloor + 2 \\ \frac{C\gamma K}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mathrm{v}})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mathrm{v}})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}+D\right)\right), r \leq \left\lfloor \frac{\omega_{\mathrm{sum}}\cdot\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2}\right\rfloor + 1 \end{cases}$$

then (19) can be upper bounded by

$$\sum_{i\in E} g_{\mathcal{R},2}(r,\Delta^{\mathrm{v}}_{i,\mathcal{R}}).$$

In conclusion, for $i \in E$, we denote $\Delta^{\mathrm{v}}_{i,\mathcal{R}} := \min_{S\ni i, S\in\mathcal{R}} \Delta^{\mathrm{v}}_S$, $\omega_{\mathrm{vv}'} := \max\{\omega_{\mathrm{v}}, \omega'_{\mathrm{v}}\}$, $\omega_{\mathrm{sum}} := \sqrt{\ln\frac{1}{\omega'_{\mathrm{v}}}} + \sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}}$and

$$g_{\mathcal{R}}(r,\Delta^{\mathrm{v}}_{i,\mathcal{R}}) := \begin{cases} \frac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\omega_{\mathrm{sum}}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}+D\right)\right), r \geq \left\lfloor \frac{\omega_{\mathrm{sum}}\cdot\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2}\right\rfloor + 2 \\ \frac{C\gamma K}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}+D\right)\right), r \leq \left\lfloor \frac{\omega_{\mathrm{sum}}\cdot\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2}\right\rfloor + 1 \end{cases}$$

then

$$\sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{R}\}\, \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \leq \sum_{i\in E} g_{\mathcal{R}}(r,\Delta^{\mathrm{v}}_{i,\mathcal{R}})$$

Case 4: $A_p \in \mathcal{S}^c \cap \mathcal{B}$

Under this case, there will be a comparison among $\omega_\mu, \omega_{\mathrm{v}}$ and ω'_{v} thus we denote $\omega_{\max} = \max\{\omega_\mu, \omega_{\mathrm{v}}, \omega'_{\mathrm{v}}\}$ and $\omega_1 = \sqrt{\frac{\ln\frac{1}{\omega_{\max}}}{\ln\frac{1}{\omega_\mu}}}, \omega_2 = \sqrt{\frac{\ln\frac{1}{\omega_{\max}}}{\ln\frac{1}{\omega_{\mathrm{v}}}}}, \omega_3 = \sqrt{\frac{\ln\frac{1}{\omega_{\max}}}{\ln\frac{1}{\omega'_{\mathrm{v}}}}}$. For $S \in \mathcal{S}^c \cap \mathcal{B}$, we denote $\bar{\Delta}_S := \max\left\{\omega_1\Delta_S, \omega_3\frac{\Delta^{\mathrm{v}}_S}{3}\right\}$.

For $i \in E$, denote

$$\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}} := \min_{S\ni i, S\in\mathcal{S}^c\cap\mathcal{B}} \max\left\{\frac{\Delta_S}{\sqrt{\ln(1/\omega_\mu)}}, \frac{\Delta^{\mathrm{v}}_S}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}}\right\} = \sqrt{\frac{1}{\ln(1/\omega_{\max})}} \min_{S\ni i, S\in\mathcal{S}^c\cap\mathcal{B}} \bar{\Delta}_S.$$

and assume $\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}} \in (r_i\frac{\bar{\sigma}^2/3}{\sqrt{\ln(1/\omega_{\mathrm{v}})}+\sqrt{\ln(1/\omega'_{\mathrm{v}})}}, (r_i+1)\frac{\bar{\sigma}^2/3}{\sqrt{\ln(1/\omega_{\mathrm{v}})}+\sqrt{\ln(1/\omega'_{\mathrm{v}})}}]$

$$
\begin{aligned}
&\sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\}\Big) \mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \cdot \mathbb{1}\left\{\mathcal{E}\right\} \\
&\overset{(a)}{\leq} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \mathbb{1}\left\{\mathcal{F}^{\mu}_p, \mathcal{F}_p\left(\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \omega'_{\mathrm{v}}\right), \mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&\overset{(b)}{\leq} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \mathbb{1}\left\{\mathcal{F}_p\left(\Delta_{A_p}, \omega_{\mu}\right), \mathcal{F}_p\left(\frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \omega'_{\mathrm{v}}\right), \mathcal{F}_p\left(\frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}, \omega_{\mathrm{v}}\right)\right\} \\
&\overset{(c)}{\leq} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \mathbb{1}\left\{\mathcal{F}_p\left(\max\left\{\omega_1 \Delta_{A_p}, \omega_3 \frac{\Delta^{\mathrm{v}}_{A_p}}{3}, \omega_2 \frac{(r-1)\bar{\sigma}^2 - \Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_{\max}\right)\right\} \\
&\overset{(d)}{\leq} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \mathbb{1}\left\{\mathcal{F}_p\left(\max\left\{\max\left\{\omega_1 \Delta_{A_p}, \omega_3 \frac{\Delta^{\mathrm{v}}_{A_p}}{3}\right\}, \omega_2 \frac{(r-1)\bar{\sigma}^2}{3} - \frac{\omega_2}{\omega_3} \max\left\{\omega_1 \Delta_{A_p}, \omega_3 \frac{\Delta^{\mathrm{v}}_{A_p}}{3}\right\}\right\}\right.\right. \\
&\qquad\qquad \left.\left., \omega_{\max}\right)\right\} \\
&\overset{(e)}{\leq} \sum_{p=1}^{T'} \sum_{j \in \mathbb{N}} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \mathbb{1}\left\{\mathcal{G}_{j,p}\left(\max\left\{\bar{\Delta}_{A_p}, \omega_2 \frac{(r-1)\bar{\sigma}^2}{3} - \frac{\omega_2}{\omega_3} \bar{\Delta}_{A_p}\right\}, \omega_{\max}\right)\right\} \\
&\overset{(f)}{\leq} \sum_{p=1}^{T'} \sum_{j \in \mathbb{N}} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \frac{1}{b_j K} \sum_{i \in A_p} \mathbb{1}\left\{T_i(p-1) \leq m_j\left(\max\left\{\bar{\Delta}_{A_p}, \omega_2 \frac{(r-1)\bar{\sigma}^2}{3} - \frac{\omega_2}{\omega_3} \bar{\Delta}_{A_p}\right\}, \omega_{\max}\right)\right\} \\
&= \sum_{i \in E} \sum_{j \in \mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \leq m_j\left(\max\left\{\bar{\Delta}_{A_p}, \omega_2 \frac{(r-1)\bar{\sigma}^2}{3} - \frac{\omega_2}{\omega_3} \bar{\Delta}_{A_p}\right\}, \omega_{\max}\right)\right\}
\end{aligned}
$$

Given $i \in E$

(1) if $r \geq r_i + 2$, for any $S \in \mathcal{S}^c \cap \mathcal{B}$ that contains item i,

$$
\max\left\{\bar{\Delta}_S, \omega_2 \frac{(r-1)\bar{\sigma}^2}{3} - \frac{\omega_2}{\omega_3} \bar{\Delta}_S\right\} \geq \max\left\{\bar{\Delta}_S, \frac{\omega_2 \omega_3}{\omega_2 + \omega_3} \frac{(r-1)\bar{\sigma}^2}{3}\right\} \geq \frac{\omega_2 \omega_3}{\omega_2 + \omega_3} \frac{(r-1)\bar{\sigma}^2}{3} \geq \sqrt{\ln(1/\omega_{\max})} \bar{\Delta}_{i, \mathcal{S}^c \cap \mathcal{R}}
$$

where the first inequality uses the fact that for $x, y \in \mathbb{R}_+$ and $z \in [0,1]$, , $\max\{x, y\} \geq \max\{x, xz + y(1-z)\}$. We take $x = \bar{\Delta}_{A_p}, y = \omega_2 \frac{(r-1)\bar{\sigma}^2}{3} - \frac{\omega_2}{\omega_3} \bar{\Delta}_{A_p}$ and $z = \frac{\omega_2}{\omega_2 + \omega_3}$.

Similar to the computations for the case $A_p \in \mathcal{R}$, we have

$$
\begin{aligned}
&\sum_{j \in \mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \leq m_j\left(\max\left\{\bar{\Delta}_{A_p}, \omega_2 \frac{(r-1)\bar{\sigma}^2}{3} - \frac{\omega_2}{\omega_3} \bar{\Delta}_{A_p}\right\}, \omega_{\max}\right)\right\} \\
&\leq \sum_{j \in \mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \leq m_j\left(\frac{\omega_2 \omega_3}{\omega_2 + \omega_3} \frac{(r-1)\bar{\sigma}^2}{3}, \omega_{\max}\right)\right\} \\
&\leq \frac{C\gamma K}{\left(\frac{\omega_2 \omega_3}{\omega_2 + \omega_3} \frac{(r-1)\bar{\sigma}^2}{3}\right)^2} \left(2 \ln \frac{1}{\omega_{\max}} + \ln \ln_+ \frac{1}{\left(\frac{\omega_2 \omega_3}{\omega_2 + \omega_3} \frac{(r-1)\bar{\sigma}^2}{3}\right)^2} + D\right) \\
&\leq \frac{C\gamma K}{\left(\frac{\omega_2 \omega_3}{\omega_2 + \omega_3} \frac{(r-1)\bar{\sigma}^2}{3}\right)^2} \left(2 \ln \frac{1}{\omega_{\max}} + \ln \ln_+ \frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}_{i, \mathcal{S}^c \cap \mathcal{R}})^2} + D\right) \\
&= \frac{C\gamma K}{\left(\frac{(r-1)\bar{\sigma}^2/3}{\sqrt{\ln(1/\omega_{\mathrm{v}})} + \sqrt{\ln(1/\omega'_{\mathrm{v}})}}\right)^2} \left(2 + \frac{1}{\ln(1/\omega_{\max})} \left(\ln \ln_+ \frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}_{i, \mathcal{S}^c \cap \mathcal{R}})^2} + D\right)\right)
\end{aligned}
$$

(2) if $r \leq r_i + 1$, for any $S \in \mathcal{S}^c \cap \mathcal{B}$ that contains item i,

$$\max\left\{\bar{\Delta}_S, \omega_2 \frac{(r-1)\bar{\sigma}^2}{3} - \frac{\omega_2}{\omega_3}\bar{\Delta}_S\right\} \geq \max\left\{\bar{\Delta}_S, \frac{\omega_2\omega_3}{\omega_2+\omega_3}\frac{(r-1)\bar{\sigma}^2}{3}\right\} \geq \bar{\Delta}_S \geq \sqrt{\ln(1/\omega_{\max})}\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{R}}$$

Similar to the computations for the case $A_p \in \mathcal{R}$, we have

$$\begin{aligned}
&\sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \leq m_j\left(\max\left\{\bar{\Delta}_{A_p}, \omega_2\frac{(r-1)\bar{\sigma}^2}{3} - \frac{\omega_2}{\omega_3}\bar{\Delta}_{A_p}\right\}, \omega_{\max}\right)\right\} \\
&\leq \sum_{j\in\mathbb{N}} \frac{1}{b_j K} \sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \mathbb{1}\left\{i \in A_p, T_i(p-1) \leq m_j\left(\sqrt{\ln(1/\omega_{\max})}\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{R}}, \omega_{\max}\right)\right\} \\
&\leq \frac{C\gamma K}{(\sqrt{\ln(1/\omega_{\max})}\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}\left(2\ln\frac{1}{\omega_{\max}} + \ln\ln_+\frac{1}{(\sqrt{\ln(1/\omega_{\max})}\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2} + D\right) \\
&= \frac{C\gamma K}{(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}\left(2 + \frac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{R}})^2} + D\right)\right)
\end{aligned}$$

In conclusion, for $i \in E$, we denote

$$\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}} := \min_{S\ni i, S\in\mathcal{S}^c\cap\mathcal{B}} \max\left\{\frac{\Delta_S}{\sqrt{\ln(1/\omega_\mu)}}, \frac{\Delta_S^{\mathrm{v}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}}\right\}.$$

and $\omega_{\max} := \max\{\omega_\mu, \omega_{\mathrm{v}}, \omega'_{\mathrm{v}}\}$ and $\omega_{\mathrm{sum}} := \sqrt{\ln\frac{1}{\omega'_{\mathrm{v}}}} + \sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}}$ and $g_{\mathcal{S}^c\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}) :=$

$$\begin{cases}
\dfrac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\omega_{\mathrm{sum}}})^2}\left(2 + \dfrac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+\dfrac{1}{\ln(1/\omega_{\max})(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{R}})^2} + D\right)\right), r \geq \left\lfloor\dfrac{\omega_{\mathrm{sum}}\cdot\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}}{\bar{\sigma}^2/3}\right\rfloor + 2 \\
\dfrac{C\gamma K}{(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{R}})^2}\left(2 + \dfrac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+\dfrac{1}{\ln(1/\omega_{\max})(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{R}})^2} + D\right)\right), r \leq \left\lfloor\dfrac{\omega_{\mathrm{sum}}\cdot\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}}{\bar{\sigma}^2/3}\right\rfloor + 1
\end{cases}$$

then

$$\sum_{p=1}^{T'} \mathbb{1}\left\{A_p \in \mathcal{S}^c \cap \mathcal{B}\right\} \mathbb{1}\left\{U_{A_p}^{\mathrm{v}}(p-1) > r\bar{\sigma}^2\right\} \leq \sum_{i\in E} g_{\mathcal{S}^c\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})$$

Conclusion of Step 2:

<u>For $S^\star$</u>: denote

$$g_{S^\star}(r, \Delta_{S^\star}^{\mathrm{v}}) := \frac{9\cdot C\gamma K}{((r-1)\bar{\sigma}^2 + \Delta_{S^\star}^{\mathrm{v}})^2}\left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+\frac{1}{{\Delta_{S^\star}^{\mathrm{v}}}^2} + D\right) \tag{20}$$

we have

$$\sum_{p=1}^{T'} \mathbb{1}\left\{A_p = S^\star\right\} \mathbb{1}\left\{U_{A_p}^{\mathrm{v}}(p-1) > r\bar{\sigma}^2\right\} \leq \sum_{i\in S^\star} g_{S^\star}(r, \Delta_{S^\star}^{\mathrm{v}})$$

<u>For $\mathcal{S}\cap\mathcal{B}$</u>: for $i\in E$, denote $\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}} := \min_{S\ni i, S\in\mathcal{S}\cap\mathcal{B}} \max\left\{\frac{\Delta_S}{\sqrt{\ln(1/\omega_\mu)}}, \frac{\Delta^{\mathrm{v}}_S}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}\right\}$, $\omega_{\mu\mathrm{v}} := \max\{\omega_\mu, \omega_{\mathrm{v}}\}$ and

$$
\begin{aligned}
&g_{\mathcal{S}\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) \\
&:= \begin{cases}
\frac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}+D\right)\right), r \geq \left\lfloor \frac{\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}}\right\rfloor + 2 \\
\frac{C\gamma K}{\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}\left(2+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}+D\right)\right), r \leq \left\lfloor \frac{\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}}\right\rfloor + 1
\end{cases}
\end{aligned}
\tag{21}
$$

then

$$
\sum_{p=1}^{T'} \mathbb{1}\left\{A_p\in\mathcal{S}\cap\mathcal{B}\right\}\mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \leq \sum_{i\in E} g_{\mathcal{S}\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}})
$$

<u>For $\mathcal{R}$</u>: for $i\in E$, denote $\Delta^{\mathrm{v}}_{i,\mathcal{R}} := \min_{S\ni i, S\in\mathcal{R}} \Delta^{\mathrm{v}}_S$, $\omega_{\mathrm{vv}'} := \max\{\omega_{\mathrm{v}}, \omega'_{\mathrm{v}}\}$, $\omega_{\mathrm{sum}} := \sqrt{\ln\frac{1}{\omega'_{\mathrm{v}}}} + \sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}}$ and

$$
\begin{aligned}
&g_{\mathcal{R}}(r, \Delta^{\mathrm{v}}_{i,\mathcal{R}}) \\
&:= \begin{cases}
\frac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\omega_{\mathrm{sum}}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}+D\right)\right), r \geq \left\lfloor \frac{\omega_{\mathrm{sum}}\cdot\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2}\right\rfloor + 2 \\
\frac{C\gamma K}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}+D\right)\right), r \leq \left\lfloor \frac{\omega_{\mathrm{sum}}\cdot\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2}\right\rfloor + 1
\end{cases}
\end{aligned}
\tag{22}
$$

then

$$
\sum_{p=1}^{T'} \mathbb{1}\left\{A_p\in\mathcal{R}\right\}\mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \leq \sum_{i\in E} g_{\mathcal{R}}(r, \Delta^{\mathrm{v}}_{i,\mathcal{R}})
$$

<u>For $\mathcal{S}^c\cap\mathcal{B}$</u>: for $i\in E$, denote

$$
\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}} := \min_{S\ni i, S\in\mathcal{S}^c\cap\mathcal{B}} \max\left\{\frac{\Delta_S}{\sqrt{\ln(1/\omega_\mu)}}, \frac{\Delta^{\mathrm{v}}_S}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}}\right\}.
$$

and $\omega_{\max} := \max\{\omega_\mu, \omega_{\mathrm{v}}, \omega'_{\mathrm{v}}\}$ and $\omega_{\mathrm{sum}} := \sqrt{\ln\frac{1}{\omega'_{\mathrm{v}}}} + \sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}}$ and

$$
\begin{aligned}
&g_{\mathcal{S}^c\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}) \\
&:= \begin{cases}
\frac{C\gamma K}{(\frac{(r-1)\bar{\sigma}^2}{3\omega_{\mathrm{sum}}})^2}\left(2+\frac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}+D\right)\right), r \geq \left\lfloor \frac{\omega_{\mathrm{sum}}\cdot\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}}{\bar{\sigma}^2/3}\right\rfloor + 2 \\
\frac{C\gamma K}{(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}\left(2+\frac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}+D\right)\right), r \leq \left\lfloor \frac{\omega_{\mathrm{sum}}\cdot\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}}{\bar{\sigma}^2/3}\right\rfloor + 1
\end{cases}
\end{aligned}
\tag{23}
$$

then

$$
\sum_{p=1}^{T'} \mathbb{1}\left\{A_p\in\mathcal{S}^c\cap\mathcal{B}\right\}\mathbb{1}\left\{U^{\mathrm{v}}_{A_p}(p-1) > r\bar{\sigma}^2\right\} \leq \sum_{i\in E} g_{\mathcal{S}^c\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})
$$

Step 3:

According to the results in Step 2, given $r \in [Q-1]$, the event $\mathcal{U}_p(r)$ can happen at most

$$\sum_{p=1}^{T'} \mathbb{1}\{\mathcal{U}_p(r)\} \tag{24}$$
$$\leq \min\left\{T', \sum_{i\in S^\star} g_{S^\star}(r, \Delta^{\mathrm{v}}_{S^\star}) + \sum_{i\in E} g_{\mathcal{S}\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) + \sum_{i\in E} g_{\mathcal{R}}(r, \Delta^{\mathrm{v}}_{i,\mathcal{R}}) + \sum_{i\in E} g_{\mathcal{S}^c\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})\right\}$$

phases, where the g functions are defined in (20), (21), (22) and (23). By Lemma B.4, (1) event $\mathcal{U}_p(r)$ indicates $r+1 \leq n_p$, thus (24) also indicates at most in this number of phases there are at least $r+1$ being pulled at each phase. (2) event $\mathcal{U}_p(r) \cap \mathcal{U}_p(r+1)^c$ indicates $r+1 \leq n_p \leq 2r+1$, i.e., there are at least $r+1$ and at most $2r+1$ solutions being pulled at each phase. For $r \in [Q]$, we denote

$$T'_r := \sum_{i\in S^\star} g_{S^\star}(r, \Delta^{\mathrm{v}}_{S^\star}) + \sum_{i\in E} g_{\mathcal{S}\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) + \sum_{i\in E} g_{\mathcal{R}}(r, \Delta^{\mathrm{v}}_{i,\mathcal{R}}) + \sum_{i\in E} g_{\mathcal{S}^c\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}), \quad r \in [Q-1] \tag{25}$$

and $T'_Q := 0, T'_0 = \infty$. Note that the g functions are increasing as r decreases, so if there exists an $r' \in [Q]$, such that $T' \in [T'_{r'}, T'_{r'-1})$ (or $T' \leq T'_{r'-1}$) , then

$$\begin{aligned}
&\sum_{r=1}^{Q-1}\sum_{p=1}^{T'} \mathbb{1}\{\mathcal{U}_p(r)\} \\
&= \sum_{r=1}^{r'-1}\sum_{p=1}^{T'} \mathbb{1}\{\mathcal{U}_p(r)\} + \sum_{r=r'}^{Q-1}\sum_{p=1}^{T'} \mathbb{1}\{\mathcal{U}_p(r)\} \\
&\leq T'\cdot(r'-1) + \sum_{r=r'}^{Q-1}\sum_{i\in S^\star} g_{S^\star}(r, \Delta^{\mathrm{v}}_{S^\star}) + \sum_{i\in E} g_{\mathcal{S}\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) + \sum_{i\in E} g_{\mathcal{R}}(r, \Delta^{\mathrm{v}}_{i,\mathcal{R}}) + \sum_{i\in E} g_{\mathcal{S}^c\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}) \\
&= T'\cdot(r'-1) + \sum_{r=r'}^{Q-1} T'_r \\
&\leq T'\cdot(r'-1) + \sum_{i\in S^\star} h_{S^\star}(r', \Delta^{\mathrm{v}}_{S^\star}) + \sum_{i\in E} h_{\mathcal{S}\cap\mathcal{B}}(r', \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) + \sum_{i\in E} h_{\mathcal{R}}(r', \Delta^{\mathrm{v}}_{i,\mathcal{R}}) + \sum_{i\in E} h_{\mathcal{S}^c\cap\mathcal{B}}(r', \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}) \\
&= T'\cdot(r'-1) + H(r', \Lambda)
\end{aligned}$$

where

$$H(r', \Lambda) := \sum_{i\in S^\star} h_{S^\star}(r', \Delta^{\mathrm{v}}_{S^\star}) + \sum_{i\in E} h_{\mathcal{S}\cap\mathcal{B}}(r', \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) + \sum_{i\in E} h_{\mathcal{R}}(r', \Delta^{\mathrm{v}}_{i,\mathcal{R}}) + \sum_{i\in E} h_{\mathcal{S}^c\cap\mathcal{B}}(r', \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}) \tag{26}$$

and the h functions are defined at the end of the proof. This indicates, when $T' \leq T'_{r'-1}$, the upper bound of the high-probability regret due to safeness-checking $\mathrm{R}_2(T')$ (hence the the total regret) is lower bounded by a linear function with slope $r'-1$. In particular, the upper bound of $\mathrm{R}_2(T')$ is lower bounded by a linear function when $T' \leq T'_1$ and it remains a constant when $T' > T'_1$. We can compute an upper bound for the number of solutions being pulled during these T' phases when $T' \in [T'_{r'}, T'_{r'-1})$:

$$\begin{aligned}
&T'_{Q-1}\cdot(2Q-1) + (T'_{Q-2} - T'_{Q-1})\cdot(2Q-3) + \cdots + (T'_{r'} - T'_{r'+1})\cdot(2r'+1) + (T' - T'_{r'})(2r'-1) \\
&= T' + 2\left(T'_{Q-1}\cdot(Q-1) + (T'_{Q-2} - T'_{Q-1})\cdot(Q-2) + \cdots + (T'_{r'} - T'_{r'+1})\cdot r' + (T' - T'_{r'})(r'-1)\right) \\
&= (2r'-1)T' + 2\sum_{r=r'}^{Q-1} T'_r \\
&\leq (2r'-1)T' + 2H(r', \Lambda)
\end{aligned}$$

Thus, the number of pulled solutions is at most $2H(1,\Lambda)$.[1]

In conclusion, the regret due to safeness-checking can be upper bounded by

$$\begin{aligned}
\mathrm{R}_2(T') &= 2\mu^\star \sum_{r=1}^{Q-1}\sum_{p=1}^{T'} \mathbb{1}\left\{\mathcal{U}_p(r)\right\} \\
&\le 2\mu^\star T'\cdot(r'-1) \\
&\quad + 2\mu^\star\left(\sum_{i\in S^\star} h_{S^\star}(r',\Delta^{\mathrm{v}}_{S^\star}) + \sum_{i\in E} h_{\mathcal{S}\cap\mathcal{B}}(r',\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) + \sum_{i\in E} h_{\mathcal{R}}(r',\Delta^{\mathrm{v}}_{i,\mathcal{R}}) + \sum_{i\in E} h_{\mathcal{S}^c\cap\mathcal{B}}(r',\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})\right) \\
&= 2\mu^\star T'\cdot(r'-1) + 2\mu^\star\cdot H(r',\Lambda)
\end{aligned}$$

where $T'\in[T'_{r'},T'_{r'-1})$ with $T'_{r'}$ defined in (25), for $i\in E$, $\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}} := \min_{S\ni i, S\in\mathcal{S}\cap\mathcal{B}} \max\left\{\frac{\Delta_S}{\sqrt{\ln(1/\omega_\mu)}}, \frac{\Delta^{\mathrm{v}}_S}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}\right\}$, $\Delta^{\mathrm{v}}_{i,\mathcal{R}} := \min_{S\ni i, S\in\mathcal{R}} \Delta^{\mathrm{v}}_S$ and $\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}} := \min_{S\ni i, S\in\mathcal{S}^c\cap\mathcal{B}} \max\left\{\frac{\Delta_S}{\sqrt{\ln(1/\omega_\mu)}}, \frac{\Delta^{\mathrm{v}}_S}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}}\right\}$.

The h functions:

(For convenience, we restate the notations: $\omega_{\mu\mathrm{v}} := \max\{\omega_\mu,\omega_{\mathrm{v}}\}$, $\omega_{\mathrm{vv}'} := \max\{\omega_{\mathrm{v}},\omega'_{\mathrm{v}}\}$, $\omega_{\max} := \max\{\omega_\mu,\omega_{\mathrm{v}},\omega'_{\mathrm{v}}\}$ and $\omega_{\mathrm{sum}} := \sqrt{\ln\frac{1}{\omega'_{\mathrm{v}}}} + \sqrt{\ln\frac{1}{\omega_{\mathrm{v}}}}$.)

For $S^\star$: for each $i\in S^\star$

$$\begin{aligned}
h_{S^\star}(r',\Delta^{\mathrm{v}}_{S^\star}) &:= \sum_{r=r'}^{Q-1} g_{S^\star}(r,\Delta^{\mathrm{v}}_{S^\star}) \qquad (27)\\
&= \sum_{r=r'}^{Q-1} \frac{9\cdot C\gamma K}{((r-1)\bar{\sigma}^2+\Delta^{\mathrm{v}}_{S^\star})^2}\left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+\frac{1}{{\Delta^{\mathrm{v}}_{S^\star}}^2} + D\right) \\
&\le \begin{cases} \dfrac{18\cdot C\gamma K}{(r'-1)\bar{\sigma}^4}\left(2\ln\dfrac{1}{\omega_{\mathrm{v}}} + \ln\ln_+\dfrac{1}{(\Delta^{\mathrm{v}}_{S^\star})^2} + D\right), & r'\ge 2 \\ \dfrac{18\cdot C\gamma K}{(\Delta^{\mathrm{v}}_{S^\star})^2}\left(2\ln\dfrac{1}{\omega_{\mathrm{v}}} + \ln\ln_+\dfrac{1}{(\Delta^{\mathrm{v}}_{S^\star})^2} + D\right), & r'=1 \end{cases}
\end{aligned}$$

For $\mathcal{S}\cap\mathcal{B}$: for each $i\in E$, there is a changing point $\left\lfloor \frac{\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}}\right\rfloor + 2$ in $g_{\mathcal{S}\cap\mathcal{B}}(r,\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}})$, thus

$$h_{\mathcal{S}\cap\mathcal{B}}(r',\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) := \sum_{r=r'}^{Q-1} g_{\mathcal{S}\cap\mathcal{B}}(r,\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) \qquad (28)$$

[1] Note that the upper bound for the regret is $2\mu^\star T'\cdot(r'-1)+2\mu^\star\cdot H(r',\Lambda)$ and the upper bound for the number of solutions is $(2r'-1)T'+2H(r',\Lambda)$, which indicates we can roughly use $\min\{T\mu^\star, 2\mu^\star H(r',\Lambda)\}$ to bound the regret due to safeness-checking with T time steps.

$$\leq \begin{cases} 0, r' = Q \\ (Q-r')\frac{C\gamma K}{\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}\left(2+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}+D\right)\right), \left\lfloor\frac{\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}}\right\rfloor \geq Q-3 \\ \frac{2\cdot C\gamma K}{(r'-1)(\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}+D\right)\right), \left\lfloor\frac{\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}}\right\rfloor + 2 \leq r' < Q-1 \\ \frac{3\cdot C\gamma K}{\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}\left(2+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}+D\right)\right), \left\lfloor\frac{\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}}\right\rfloor = 0, r' = 1 \\ C\gamma K\left(\frac{4}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}\cdot\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}}-\frac{r'-1}{\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}\right)\left(2+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mu\mathrm{v}})\bar{\Delta}^2_{i,\mathcal{S}\cap\mathcal{B}}}+D\right)\right), \text{otherwise} \end{cases}$$

For $\mathcal{R}$: for each $i \in E$, there is a changing point $\left\lfloor\frac{\omega_{\mathrm{sum}}\cdot\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2}\right\rfloor + 2$, thus

$$h_{\mathcal{R}}(r', \Delta^{\mathrm{v}}_{i,\mathcal{R}}) := \sum_{r=r'}^{Q-1} g_{\mathcal{R}}(r, \Delta^{\mathrm{v}}_{i,\mathcal{R}}) \leq \tag{29}$$

$$\begin{cases} 0, r' = Q \\ (Q-r')\frac{C\gamma K}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}+D\right)\right), \left\lfloor\frac{\omega_{\mathrm{sum}}\cdot\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2}\right\rfloor \geq Q-3 \\ \frac{2C\gamma K}{(r'-1)(\frac{\bar{\sigma}^2}{3\omega_{\mathrm{sum}}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}+D\right)\right), \left\lfloor\frac{\omega_{\mathrm{sum}}\cdot\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln\frac{1}{\omega'_{\mathrm{v}}}}\bar{\sigma}^2}\right\rfloor + 2 \leq r' < Q-1 \\ \frac{3C\gamma K}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}\left(2+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}+D\right)\right), \left\lfloor\frac{\omega_{\mathrm{sum}}\cdot\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2}\right\rfloor = 0, r' = 1 \\ C\gamma K\left(\frac{3}{\frac{\bar{\sigma}^2}{3\omega_{\mathrm{sum}}}\cdot\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}}}-\frac{r'-1}{(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}\right)\left(2+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\mathrm{vv}'})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}+D\right)\right), \text{else} \end{cases}$$

For $\mathcal{S}^c \cap \mathcal{B}$: for each $i \in E$, there is a changing point $\left\lfloor\frac{\omega_{\mathrm{sum}}\cdot\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}}{\bar{\sigma}^2/3}\right\rfloor + 2$, thus

$$h_{\mathcal{S}^c\cap\mathcal{B}}(r', \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}) := \sum_{r=r'}^{Q-1} g_{\mathcal{S}^c\cap\mathcal{B}}(r, \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}) \tag{30}$$

$$\leq \begin{cases} 0, r' = Q \\ (Q-r')\frac{C\gamma K}{(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}\left(2+\frac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}+D\right)\right), \left\lfloor\frac{\omega_{\text{sum}}\cdot\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}}{\bar{\sigma}^2/3}\right\rfloor \geq Q-3 \\ \frac{2C\gamma K}{(r'-1)(\frac{\bar{\sigma}^2}{3\omega_{\text{sum}}})^2}\left(2+\frac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}+D\right)\right), \left\lfloor\frac{\omega_{\text{sum}}\cdot\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}}{\bar{\sigma}^2/3}\right\rfloor + 2 \leq r' < Q-1 \\ \frac{3C\gamma K}{(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}\left(2+\frac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}+D\right)\right), \left\lfloor\frac{\omega_{\text{sum}}\cdot\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}}{\bar{\sigma}^2/3}\right\rfloor = 0, r' = 1 \\ C\gamma K\left(\frac{3}{\frac{\bar{\sigma}^2}{3\omega_{\text{sum}}}\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}}-\frac{r'-1}{(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}\right)\left(2+\frac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+\frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}})^2}+D\right)\right), \text{otherwise} \end{cases}$$

□

B.5. Proofs of Theorem 4.1 and Theorem 5.1

Theorem 4.1 (Problem-dependent upper bound)**.** *Let $\Lambda = (E, \mathcal{A}_K, \nu, \bar{\sigma}^2)$ be an instance and let $\{\delta_T\}_{T=1}^{\infty} \in o(1)$ be a sequence that satisfies $\ln(1/\delta_T) = o(T^b)$ for all $b > 0$ (i.e., $\{\delta_T\}$ is not exponentially decaying). Then,* PASCOMBUCB *is a $\{\delta_T\}_{T=1}^{\infty}$-variance-constrained consistent algorithm. More precisely, given a time budget T, the probably anytime-safe constraint is satisfied and the regret of* PASCOMBUCB $\mathrm{Reg}(T)$ *is upper bounded by*

$$\min\{T\mu^\star, \mathrm{Reg}_1(T) + \mathrm{Reg}_2(T)\} + \mathrm{Reg}_3(T),$$

where

$$\mathrm{Reg}_1(T) = O\Bigg(\sum_{i\in E\setminus S^\star}\frac{K\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}} + \sum_{i\in E}\frac{c_i K\ln T}{\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min}}\Bigg)$$

$$\mathrm{Reg}_2(T) = 2\mu^\star H\left(\Delta(\Lambda)\right), \quad \mathrm{Reg}_3(T) = 2\mu^\star(L+1)$$

where $\Delta(\Lambda) = \{\Delta^{\mathrm{v}}_{S^\star}\} \cup \{\Delta^{\mathrm{v}}_{i,\mathcal{R}}, \Psi_{i,\mathcal{S}\cap\mathcal{B}}, \Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\}_{i\in E}$ and $H\left(\Delta(\Lambda)\right) := H(1,\Lambda)$ is defined in (26) in App. B.4.

Proof of Theorem 4.1. According to Lemma 6.1, Lemma 6.2 and Lemma 6.3, and take $\omega_\mu = \omega'_{\mathrm{v}} = \frac{1}{T^2}$ and $\omega_{\mathrm{v}} = \frac{\delta_T}{T^2}$, the expected regret of T' phases $\mathbb{E}[\mathrm{R}(T')]$ can be upper bounded as

$$\begin{aligned}
\mathbb{E}[\mathrm{R}(T')] &\leq \mathbb{E}[\mathrm{R}_1(T')|\mathcal{E}] + \mathbb{E}[\mathrm{R}_2(T')|\mathcal{E}] + \mathrm{R}_3(T) \qquad (31)\\
&\leq O\left(\sum_{i\in E\setminus S^\star}\frac{K}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}}\ln\frac{1}{\omega_\mu} + \sum_{i\in E}\frac{c_i K}{\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min}}\ln\frac{1}{\omega'_{\mathrm{v}}}\right) + 2\mu^\star\left[T'\cdot(r'-1) + H(r',\Lambda)\right] \\
&\quad + 2\mu^\star L + 2\mu^\star TL\left(\xi(\omega_\mu) + 2\xi(\omega_v) + 2\xi(\omega'_v)\right) \\
&\leq O\left(\sum_{i\in E\setminus S^\star}\frac{K}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}}\ln T + \sum_{i\in E}\frac{c_i K}{\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min}}\ln T\right) + 2\mu^\star H(1,\Lambda) \\
&\quad + 2\mu^\star L + 2\mu^\star TL\left(3\xi(1/T^2) + 2\xi(\delta_T/T^2)\right) \\
&= \mathrm{Reg}_1(T) + \mathrm{Reg}_2(T) + \mathrm{Reg}_3(T)
\end{aligned}$$

On the other hand, the high-probability regret (14) can be naïve bounded as

$$\begin{aligned}
\hat{\mathrm{R}}(T') &= \sum_{p=1}^{T'}\sum_{r=1}^{n_p}(\mu^\star - \mu_{A_{p,r}}) \\
&\leq \sum_{p=1}^{T'}\sum_{r=1}^{n_p}\mu^\star \\
&= T\mu^\star.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}\mathbb{E}[\mathrm{R}(T')] &\leq \mathbb{E}[\hat{\mathrm{R}}(T')|\mathcal{E}] + \mathrm{Reg}_3(T) \\ &\leq T\mu^\star + \mathrm{Reg}_3(T)\end{aligned} \tag{32}$$

(31) and (32) give the final upper bound with T time steps.

□

Theorem 5.1 (Problem-independent upper bound)**.** *Let $\{\delta_T\}_{T=1}^\infty \in o(1)$ be a sequence that satisfies $\ln(1/\delta_T) = o(T^b)$ for all $b > 0$. If $T > L$, for any instance Λ with variance gaps lower bounded by $\Delta^{\mathrm{v}} \leq \min_{S\in\mathcal{A}_K} \Delta^{\mathrm{v}}_S$, the regret of* PASCombUCB *is upper bounded by*

$$O\left(\sqrt{KLT\ln T} + \frac{LK^2}{(\Delta^{\mathrm{v}})^2}\ln\left(\frac{1}{\delta_T}\right)\right).$$

Proof of Theorem 5.1. We firstly deal with the regret due to suboptimality $\mathrm{R}_1(T')$. Let Δ^μ be a constant that is to be chosen.

$$\begin{aligned}\mathrm{R}_1(T') &= \sum_{p=1}^{T'} \mathbb{1}\{A_p \in \mathcal{B}\}\Delta_{A_p} \\ &= \sum_{p=1}^{T'} \mathbb{1}\{\Delta_{A_p} \geq \Delta^\mu\}\mathbb{1}\{A_p \in \mathcal{B}\}\Delta_{A_p} + \sum_{p=1}^{T'} \mathbb{1}\{\Delta_{A_p} < \Delta^\mu\}\mathbb{1}\{A_p \in \mathcal{B}\}\Delta_{A_p} \\ &\leq \sum_{p=1}^{T'} \mathbb{1}\{\Delta_{A_p} \geq \Delta^\mu\}\mathbb{1}\{A_p \in \mathcal{B}\}\Delta_{A_p} + T\cdot\Delta^\mu\end{aligned}$$

where the second term makes use of the fact that $T' \leq T$ w.p. 1.

The first term can be upper bounded by adopting the proof of Lemma 6.3 with the constraint that $\Delta_{A_p} \geq \Delta^\mu$. Thus, for $i \in E \setminus S^\star$, $\Delta_{i,\mathcal{S}\cap\mathcal{B},\min} \geq \Delta^\mu$. For $i \in E$,

$$\begin{aligned}\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min} &\geq \Delta^\mu \\ \bar{\Delta}'_{i,\mathcal{S}^c\cap\mathcal{B}} &\geq \max\{\bar{\omega}\Delta^\mu, \Delta^{\mathrm{v}}/3\} = \max\{\Delta^\mu, \Delta^{\mathrm{v}}/3\} =: \Delta^{\mu\mathrm{v}} \\ c_i &\leq \frac{1}{\bar{\omega}^2} = 1\end{aligned}$$

where $\bar{\omega} := \sqrt{\frac{\ln\frac{1}{\omega'_{\mathrm{v}}}}{\ln\frac{1}{\omega_\mu}}} = 1$. The regret due to suboptimality can be upper bounded by

$$\begin{aligned}\sum_{p=1}^{T'} \mathbb{1}\{\Delta_{A_p} \geq \Delta^\mu\}\mathbb{1}\{A_p \in \mathcal{B}\}\Delta_{A_p} \leq& \sum_{i\in E\setminus S^\star} \frac{2C\gamma K}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}}\left(2\ln\frac{1}{\omega_\mu} + \ln\ln_+\frac{1}{\Delta^2_{i,\mathcal{S}\cap\mathcal{B},\min}} + D\right) \\ &+ \sum_{i\in E} \frac{2c_i\cdot C\gamma K}{\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min}}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}} + \ln\ln_+\frac{1}{(\bar{\Delta}'_{i,\mathcal{S}^c\cap\mathcal{B}})^2} + D\right) \\ \leq& \sum_{i\in E\setminus S^\star} \frac{2C\gamma K}{\Delta^\mu}\left(2\ln\frac{1}{\omega_\mu} + \ln\ln_+\frac{1}{(\Delta^\mu)^2} + D\right) \\ &+ \sum_{i\in E} \frac{2C\gamma K}{\Delta^\mu}\left(2\ln\frac{1}{\omega'_{\mathrm{v}}} + \ln\ln_+\frac{1}{(\Delta^{\mu\mathrm{v}})^2} + D\right) \\ \leq& L\cdot\frac{8C\gamma K}{\Delta^\mu}\left(2\ln T + \ln\ln_+\frac{1}{(\Delta^\mu)^2} + D\right)\end{aligned}$$

By taking $\Delta^{\mu} = \sqrt{\frac{KL\ln(TL)}{T}}$, the regret due to suboptimality is bounded by

$$\begin{aligned}
\mathrm{R}_1(T') &\leq \sum_{p=1}^{T'} \mathbb{1}\{\Delta_{A_p} \geq \Delta^{\mu}\}\mathbb{1}\{A_p \in \mathcal{B}\}\Delta_{A_p} + T \cdot \Delta^{\mu} \\
&\leq 16C\gamma\sqrt{KLT\ln(TL)} + 8C\gamma\sqrt{\frac{KLT}{\ln T}}\ln\ln_+ \frac{T}{KL\ln T} + 8C\gamma\sqrt{\frac{KLT}{\ln T}}D + \sqrt{KLT\ln(TL)} \\
&= O(\sqrt{KLT\ln(TL)}) \\
&= O(\sqrt{KLT\ln T}).
\end{aligned}$$

where we utilize $T \geq L$.

We then cope with the regret due to safeness-checking $\mathrm{R}_2(T')$. According to Lemma 6.3, we only need to upper bound $H(1, \Lambda)$, i.e., the h functions defined in (27), (28), (29) and (30).

For $h_{S^\star}(1, \Delta^{\mathrm{v}}_{S^\star})$:

$$\begin{aligned}
h_{S^\star}(1, \Delta^{\mathrm{v}}_{S^\star}) &\leq h_{S^\star}(1, \Delta^{\mathrm{v}}) \\
&= \frac{18 \cdot C\gamma K}{(\Delta^{\mathrm{v}})^2}\left(2\ln\frac{1}{\omega_{\mathrm{v}}} + \ln\ln_+ \frac{1}{(\Delta^{\mathrm{v}})^2} + D\right) \\
&= O\left(\frac{K}{(\Delta^{\mathrm{v}})^2}\ln\frac{1}{\omega_{\mathrm{v}}}\right) \\
&= O\left(\frac{K}{(\Delta^{\mathrm{v}})^2}\ln\frac{T}{\delta}\right)
\end{aligned}$$

For $h_{\mathcal{S}\cap\mathcal{B}}(1, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}})$, define $\Delta^{\mu\mathrm{v}} := \max\left\{\frac{\Delta^{\mu}}{\sqrt{\ln(1/\omega_{\mu})}}, \frac{\Delta^{\mathrm{v}}}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}\right\} = \max\left\{\sqrt{\frac{KL}{T}}, \frac{\Delta^{\mathrm{v}}}{3\sqrt{\ln(TL/\delta)}}\right\}$. We have $\bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}} \geq \Delta^{\mu\mathrm{v}}$. The threshold $\left\lfloor \frac{\Delta^{\mu\mathrm{v}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}} \right\rfloor = \left\lfloor \frac{\max\left\{3\sqrt{\frac{KL\ln(TL/\delta)}{T}}, \Delta^{\mathrm{v}}\right\}}{\bar{\sigma}^2} \right\rfloor$.

$$\begin{aligned}
&h_{\mathcal{S}\cap\mathcal{B}}(1, \bar{\Delta}_{i,\mathcal{S}\cap\mathcal{B}}) \leq h_{\mathcal{S}\cap\mathcal{B}}(1, \Delta^{\mu\mathrm{v}}) \qquad (33) \\
&= \begin{cases}
(Q-1)\frac{C\gamma K}{(\Delta^{\mu\mathrm{v}})^2}\left(2 + \frac{1}{\ln(1/\omega_{\mu\mathrm{v}})}\left(\ln\ln_+ \frac{1}{\ln(1/\omega_{\mu\mathrm{v}})(\Delta^{\mu\mathrm{v}})^2} + D\right)\right), & \left\lfloor \frac{\Delta^{\mu\mathrm{v}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}} \right\rfloor \geq Q-3 \\
C\gamma K\frac{4}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}} \cdot \Delta^{\mu\mathrm{v}}}\left(2 + \frac{1}{\ln(1/\omega_{\mu\mathrm{v}})}\left(\ln\ln_+ \frac{1}{\ln(1/\omega_{\mu\mathrm{v}})(\Delta^{\mu\mathrm{v}})^2} + D\right)\right), & 0 < \left\lfloor \frac{\Delta^{\mu\mathrm{v}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}} \right\rfloor < Q-3 \\
\frac{3 \cdot C\gamma K}{(\Delta^{\mu\mathrm{v}})^2}\left(2 + \frac{1}{\ln(1/\omega_{\mu\mathrm{v}})}\left(\ln\ln_+ \frac{1}{\ln(1/\omega_{\mu\mathrm{v}})(\Delta^{\mu\mathrm{v}})^2} + D\right)\right), & \left\lfloor \frac{\Delta^{\mu\mathrm{v}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}} \right\rfloor = 0
\end{cases} \\
&= \begin{cases}
O\left((Q-1)\frac{K}{(\Delta^{\mu\mathrm{v}})^2}\right), & \left\lfloor \frac{\Delta^{\mu\mathrm{v}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}} \right\rfloor \geq Q-3 \\
O\left(\frac{K}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}} \cdot \Delta^{\mu\mathrm{v}}}\right), & 0 < \left\lfloor \frac{\Delta^{\mu\mathrm{v}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}} \right\rfloor < Q-3 \\
O\left(\frac{K}{(\Delta^{\mu\mathrm{v}})^2}\right), & \left\lfloor \frac{\Delta^{\mu\mathrm{v}}}{\frac{\bar{\sigma}^2}{3\sqrt{\ln(1/\omega_{\mathrm{v}})}}} \right\rfloor = 0
\end{cases}
\end{aligned}$$

In particular, in the asymptotic case where $T \to \infty$ and $\ln \frac{1}{\delta} = \ln \frac{1}{\delta_T} = o(T^b), \forall b > 0$ (this includes the scenario where δ is fixed with respect to T), we have

$$h_{\mathcal{S}\cap\mathcal{B}}(1, \Delta^{\mu\mathrm{v}}) = O\left(\frac{K}{(\Delta^{\mu\mathrm{v}})^2}\right) = O\left(\frac{K}{(\Delta^{\mathrm{v}})^2} \ln \frac{T}{\delta}\right)$$

For $h_{\mathcal{R}}(1, \Delta^{\mathrm{v}}_{i,\mathcal{R}})$, we have $\Delta^{\mathrm{v}}_{i,\mathcal{R}} \geq \Delta^{\mathrm{v}}$. Furthermore, $\omega_{\mathrm{sum}} = \sqrt{\ln(TL)} + \sqrt{\ln(TL/\delta)}$ and the changing point $\left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \Delta^{\mathrm{v}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2} \right\rfloor = \left\lfloor \frac{(\sqrt{\ln(TL)} + \sqrt{\ln(TL/\delta)}) \cdot \Delta^{\mathrm{v}}}{\sqrt{\ln(TL)}\bar{\sigma}^2} \right\rfloor$.

Thus

$$\begin{aligned}
&h_{\mathcal{R}}(1, \Delta^{\mathrm{v}}_{i,\mathcal{R}}) \leq h_{\mathcal{R}}(r', \Delta^{\mathrm{v}}) \qquad (34)\\
&= \begin{cases}
(Q-1)\frac{C\gamma K}{(\frac{\Delta^{\mathrm{v}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}\left(2 + \frac{1}{\ln(1/\omega_{\mathrm{vv}'})}\left(\ln\ln_+ \frac{1}{\ln(1/\omega_{\mathrm{vv}'})(\frac{\Delta^{\mathrm{v}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2} + D\right)\right), & \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \Delta^{\mathrm{v}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2} \right\rfloor \geq Q-3 \\
C\gamma K \frac{3}{\frac{\bar{\sigma}^2}{3\omega_{\mathrm{sum}}} \cdot \frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}}}\left(2 + \frac{1}{\ln(1/\omega_{\mathrm{vv}'})}\left(\ln\ln_+ \frac{1}{\ln(1/\omega_{\mathrm{vv}'})(\frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2} + D\right)\right), & 0 < \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \Delta^{\mathrm{v}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2} \right\rfloor \leq Q-3 \\
\frac{3C\gamma K}{(\frac{\Delta^{\mathrm{v}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}\left(2 + \frac{1}{\ln(1/\omega_{\mathrm{vv}'})}\left(\ln\ln_+ \frac{1}{\ln(1/\omega_{\mathrm{vv}'})(\frac{\Delta^{\mathrm{v}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2} + D\right)\right), & \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \Delta^{\mathrm{v}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2} \right\rfloor = 0
\end{cases}\\
&= \begin{cases}
O\left((Q-1)\frac{K}{(\frac{\Delta^{\mathrm{v}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}\right), & \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \Delta^{\mathrm{v}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2} \right\rfloor \geq Q-3 \\
O\left(\frac{K}{\frac{\bar{\sigma}^2}{3\omega_{\mathrm{sum}}} \cdot \frac{\Delta^{\mathrm{v}}_{i,\mathcal{R}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}}}\right), & 0 < \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \Delta^{\mathrm{v}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2} \right\rfloor \leq Q-3 \\
= O\left(\frac{K}{(\frac{\Delta^{\mathrm{v}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}})^2}\right), & \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \Delta^{\mathrm{v}}}{\sqrt{\ln(1/\omega'_{\mathrm{v}})}\bar{\sigma}^2} \right\rfloor = 0
\end{cases}
\end{aligned}$$

In particular, in the asymptotic case where $T \to \infty$ and $\ln \frac{1}{\delta} = \ln \frac{1}{\delta_T} = o(T^b), \forall b > 0$, we have

$$h_{\mathcal{R}}(1, \Delta^{\mathrm{v}}) = O\left(\frac{QK}{(\Delta^{\mathrm{v}})^2} \ln \frac{1}{\omega'_{\mathrm{v}}}\right) = O\left(\frac{QK}{(\Delta^{\mathrm{v}})^2} \ln T\right)$$

For $\mathcal{S}^c \cap \mathcal{B}$: define $\bar{\Delta}^{\mu\mathrm{v}} := \max\left\{\frac{\Delta^{\mu}}{\sqrt{\ln(1/\omega_{\mu})}}, \frac{\Delta^{\mathrm{v}}}{3\sqrt{\ln(1/\omega'_{\mathrm{v}})}}\right\} = \max\left\{\sqrt{\frac{KL}{T}}, \frac{\Delta^{\mathrm{v}}}{3\sqrt{\ln(TL)}}\right\}$. We have $\bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}} \geq \bar{\Delta}^{\mu\mathrm{v}}$ and the changing point $\left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \bar{\Delta}^{\mu\mathrm{v}}}{\bar{\sigma}^2/3} \right\rfloor = \left\lfloor \frac{(\sqrt{\ln(TL)} + \sqrt{\ln(TL/\delta)}) \cdot \Delta^{\mu\mathrm{v}}}{\bar{\sigma}^2/3} \right\rfloor$ thus

$$\begin{aligned}
&h_{\mathcal{S}^c\cap\mathcal{B}}(1, \bar{\Delta}_{i,\mathcal{S}^c\cap\mathcal{B}}) \leq h_{\mathcal{S}^c\cap\mathcal{B}}(1, \bar{\Delta}^{\mu\mathrm{v}}) \qquad (35)\\
&= \begin{cases}
(Q-1)\frac{C\gamma K}{(\bar{\Delta}^{\mu\mathrm{v}})^2}\left(2 + \frac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+ \frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}^{\mu\mathrm{v}})^2} + D\right)\right), & \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \bar{\Delta}^{\mu\mathrm{v}}}{\bar{\sigma}^2/3} \right\rfloor \geq Q-3 \\
C\gamma K\left(\frac{3}{\frac{\bar{\sigma}^2}{3\omega_{\mathrm{sum}}}\bar{\Delta}^{\mu\mathrm{v}}}\right)\left(2 + \frac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+ \frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}^{\mu\mathrm{v}})^2} + D\right)\right), & 0 < \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \bar{\Delta}^{\mu\mathrm{v}}}{\bar{\sigma}^2/3} \right\rfloor < Q-3 \\
\frac{3C\gamma K}{(\bar{\Delta}^{\mu\mathrm{v}})^2}\left(2 + \frac{1}{\ln(1/\omega_{\max})}\left(\ln\ln_+ \frac{1}{\ln(1/\omega_{\max})(\bar{\Delta}^{\mu\mathrm{v}})^2} + D\right)\right), & \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \bar{\Delta}^{\mu\mathrm{v}}}{\bar{\sigma}^2/3} \right\rfloor = 0, r' = 1
\end{cases}
\end{aligned}$$

$$= \begin{cases} O\left((Q-1)\frac{K}{(\bar{\Delta}^{\mu_{\mathrm{v}}})^2}\right), & \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \bar{\Delta}^{\mu_{\mathrm{v}}}}{\bar{\sigma}^2/3} \right\rfloor \geq Q-3 \\ O\left(\frac{K}{\frac{\bar{\sigma}^2}{3\omega_{\mathrm{sum}}} \cdot \bar{\Delta}^{\mu_{\mathrm{v}}}}\right), & 0 < \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \bar{\Delta}^{\mu_{\mathrm{v}}}}{\bar{\sigma}^2/3} \right\rfloor < Q-3 \\ O\left(\frac{K}{(\bar{\Delta}^{\mu_{\mathrm{v}}})^2}\right), & \left\lfloor \frac{\omega_{\mathrm{sum}} \cdot \bar{\Delta}^{\mu_{\mathrm{v}}}}{\bar{\sigma}^2/3} \right\rfloor = 0, r' = 1 \end{cases}$$

In particular, in the asymptotic case where $T \to \infty$ and $\ln \frac{1}{\delta} = \ln \frac{1}{\delta_T} = o(T^b), \forall b > 0$, we have

$$h_{\mathcal{S}^c \cap \mathcal{B}}(1, \bar{\Delta}^{\mu_{\mathrm{v}}}) = O\left(\frac{QK}{(\Delta^{\mathrm{v}})^2} \ln T\right)$$

Lastly,

$$\begin{aligned} \mathrm{R}_3(T) &= 2\mu^\star L + 2\mu^\star TL\left(\xi(\omega_\mu) + 2\xi(\omega_v) + 2\xi(\omega_v')\right) \\ &\leq 2KL + K\delta + 2KTL \cdot 4 \cdot \frac{2+\epsilon}{\epsilon}\left(\frac{\frac{1}{T^2}}{\ln(1+\epsilon)}\right)^{1+\epsilon} \\ &\leq 2KL + K\delta + 4K \cdot \frac{2+\epsilon}{\epsilon}\left(\frac{1}{\ln(1+\epsilon)}\right)^{1+\epsilon} \\ &= O(1) \end{aligned}$$

where we utilize $T > L$ and the O notation refers to the fact that the preceding term is bounded as a function of T.

Note that $\mu^\star \leq K$, so $T\mu^\star + K\delta \leq TK + K\delta$.

In conclusion, according to Theorem 4.1, for any $T > L$, the problem-independent upper bound is the minimum of $TK + K\delta$ and

$$\begin{aligned} &O(\sqrt{KLT\ln T}) + K\left(\sum_{i\in S^\star} h_{S^\star}(1, \Delta^{\mathrm{v}}) + \sum_{i\in E} h_{\mathcal{S}\cap\mathcal{B}}(1, \Delta^{\mu_{\mathrm{v}}}) + \sum_{i\in E} h_{\mathcal{R}}(1, \Delta^{\mathrm{v}}) + \sum_{i\in E} h_{\mathcal{S}^c\cap\mathcal{B}}(1, \bar{\Delta}^{\mu_{\mathrm{v}}})\right) \\ &\leq O(\sqrt{KLT\ln T}) + K\left(Kh_{S^\star}(1, \Delta^{\mathrm{v}}) + Lh_{\mathcal{S}\cap\mathcal{B}}(1, \Delta^{\mu_{\mathrm{v}}}) + Lh_{\mathcal{R}}(1, \Delta^{\mathrm{v}}) + Lh_{\mathcal{S}^c\cap\mathcal{B}}(1, \bar{\Delta}^{\mu_{\mathrm{v}}})\right) \\ &= O(\sqrt{KLT\ln T}) + O\left(\frac{K^3}{(\Delta^{\mathrm{v}})^2}\ln\frac{T}{\delta}\right) + KL\left(h_{\mathcal{S}\cap\mathcal{B}}(1, \Delta^{\mu_{\mathrm{v}}}) + h_{\mathcal{R}}(1, \Delta^{\mathrm{v}}) + h_{\mathcal{S}^c\cap\mathcal{B}}(1, \bar{\Delta}^{\mu_{\mathrm{v}}})\right) \end{aligned}$$

where the h functions are defined in (33), (34) and (35). In the asymptotic case where $T \to \infty$ and $\ln \frac{1}{\delta} = \ln \frac{1}{\delta_T} = o(T^b), \forall b > 0$ (this includes the scenario where δ is fixed with respect to T), the asymptotic problem-independent upper bound is

$$\begin{aligned} &O(\sqrt{KLT\ln T}) + K\left(\sum_{i\in E} O\left(\frac{K}{(\Delta^{\mathrm{v}})^2}\ln\frac{T}{\delta}\right) + O\sum_{i\in E}\left(\frac{QK}{(\Delta^{\mathrm{v}})^2}\ln T\right)\right) \\ &= O(\sqrt{KLT\ln T}) + O\left(\frac{LK^2}{(\Delta^{\mathrm{v}})^2}\ln\frac{1}{\delta}\right) \end{aligned}$$

where we utilize $\sqrt{T\ln T} \geq \frac{QK^2}{(\Delta^{\mathrm{v}})^2}\ln T$ when T is sufficiently large. □

C. Proofs of the Lower Bounds

C.1. Preliminaries and the Impossibility Result

Let $\mathrm{KL}(\nu, \nu')$ denote the KL divergence between distributions ν and ν', and

$$d(x, y) := x\ln\left(\frac{x}{y}\right) + (1-x)\ln\left(\frac{1-x}{1-y}\right)$$

denote the Kullback–Leibler (KL) divergence between the Bernoulli distributions $\mathrm{Bern}(x)$ and $\mathrm{Bern}(y)$.

Lemma C.1 (Pinsker's and reverse Pinsker's Inequality)**.** *Consider two probability mass functions P_X, P_Y defined on the same discrete probability space $\mathcal{A} \subset [0,1]$. The following inequalities hold:*

$$|\mathbb{E}_{X\sim P_X}[X] - \mathbb{E}_{Y\sim P_Y}[Y]| \le \delta(P_X, P_Y) \le \sqrt{\frac{1}{2}\mathrm{KL}(P_X, P_Y)} \le \frac{1}{\sqrt{\alpha_Y}} \cdot \delta(P_X, P_Y),.$$

where $\delta(P_X, P_Y) := \sup_{A\subseteq\mathcal{A}} \left\{\sum_{a\in A} P_X(a) - \sum_{a\in A} P_Y(a)\right\} = \frac{1}{2}\sum_{a\in\mathcal{A}} |P_X(a) - P_Y(a)|$ is the total variational distance, and $\alpha_Y := \min_{a\in\mathcal{A}:P_Y(a)>0} Q(a)$.

Lemma C.2 (Lemma 1 in Kaufmann et al. (2016))**.** *Assume the distributions under instance $\Lambda_1 = (E, \mathcal{A}_K, \nu^{(1)}, \bar{\sigma}^2)$ and instance $\Lambda_2 = (E, \mathcal{A}_K, \nu^{(2)}, \bar{\sigma}^2)$ are mutually absolutely continuous. Given time budget T,*

$$\sum_{i=1}^{L} \mathbb{E}_{\Lambda_1}[N_i(T)] \cdot \mathrm{KL}(\nu_i^{(1)}, \nu_i^{(2)}) \ge \sup_{\mathcal{E}\in\mathcal{H}_T^{\Lambda_1}} d\big(\mathbb{P}_{\Lambda_1}(\mathcal{E}),\ \mathbb{P}_{\Lambda_2}(\mathcal{E})\big).$$

where $N_i(t)$ denotes the number of time steps item i is selected up to and including time step t and $\mathcal{H}_T^{\Lambda_1}$ is all the possible events generated by instance Λ_1 and algorithm π with T time steps.

Lemma C.3. *Let solution S containing $|S| = m(q < m \le K)$ items be a safe solution under instance $\Lambda_1 = (E, \mathcal{A}_K, \nu^{(1)}, \bar{\sigma}^2)$. Each item in S is i.i.d. with reward distribution ν_1 , mean μ_1 and variance $\sigma_1^2 < \bar{\sigma}^2$. Define event $\mathcal{E}_{(t,1)} = \{S$ is identified as safe after time step $t\}$, $\mathcal{E}_{(t,2)} = \{S$ is chosen at least once after time step $t\}$, and $\mathcal{E}_{(t)} = \mathcal{E}_{(t,1)} \cap \mathcal{E}_{(t,2)}$. Assume there exists $\tau \le T$ such that $\mathbb{P}_{\Lambda_1}[\mathcal{E}_{(\tau)}] \ge 1-\delta$ and $\mathbb{P}_{\Lambda_1}[\mathcal{E}_{(\tau-1,1)}] < 1-\delta$. If τ exists, we have*

$$\sum_{i\in S} \mathbb{E}_{\Lambda_1}[N_i(\tau)] \ge \sup_{\nu_2\in E(\nu_1)} \frac{d(\delta, 1-\delta)}{\mathrm{KL}(\nu_1, \nu_2)}.$$

Furthermore,

$$\mathbb{E}_{\Lambda_1}[M(\tau)] \ge \sup_{\nu_2\in E(\nu_1)} \frac{1}{|S|-1} \cdot \frac{d(\delta, 1-\delta)}{\mathrm{KL}(\nu_1, \nu_2)} := T(\nu^{(1)}),$$

where $M(t)$ is the number of times that a solution $S' \subset S$ is sampled up to and include time step t and $E(\nu_1) = \{\nu_2 :$ the variance associated to ν_2 is larger than $\bar{\sigma}^2/|S|\}$.

Proof. With $\sigma_2^2 > \bar{\sigma}^2/|S|$, we construct an alternative instance $\Lambda_2 = (E, \mathcal{A}_K, \nu_2, \bar{\sigma}^2)$, under which each item in S is with reward distribution ν_2 , mean μ_2 and variance σ_2^2, while the distributions of other items remain unchanged.

Define event $\mathcal{E}_{(t,1)} = \{S$ is identified as safe after time step $t\}$, $\mathcal{E}_{(t,2)} = \{S$ is chosen at least once after time step $t\}$, and $\mathcal{E}_{(t)} = \mathcal{E}_{(t,1)} \cap \mathcal{E}_{(t,2)}$. Assume there exists $\tau \le T$ such that $\mathbb{P}_{\Lambda_1}[\mathcal{E}_{(\tau)}] \ge 1-\delta$ and $\mathbb{P}_{\Lambda_1}[\mathcal{E}_{(\tau-1,1)}] < 1-\delta$. Since S is unsafe under instance Λ_2 and all the solutions chosen $\{S_t\}_{t=1}^T \subset \mathcal{A}_K$ are safe with probability at least $1-\delta$, we have $\mathbb{P}_{\Lambda_2}[\mathcal{E}_{(t)}] < \delta$ for all $t \le T$.

We now apply Lemma C.2 to obtain that

$$\sum_{i\in S} \mathbb{E}_{\Lambda_1}[N_i(\tau)] \cdot \mathrm{KL}(\nu_1, \nu_2) \ge d(\mathbb{P}_1(\mathcal{E}_{(\tau)}^c), \mathbb{P}_2(\mathcal{E}_{(\tau)}^c)) \ge d(\delta, 1-\delta) \ \Rightarrow\ \sum_{i\in S} \mathbb{E}_{\Lambda_1}[N_i(\tau)] \ge \frac{d(\delta, 1-\delta)}{\mathrm{KL}(\nu_1, \nu_2)}.$$

Since $\mathbb{P}_{\Lambda_1}[\mathcal{E}_{(\tau-1,1)}] < 1-\delta$, we can select at most $m-1$ items at one time step among the first τ time steps. Therefore,

$$\mathbb{E}_{\Lambda_1}[M(\tau)] \ge \frac{1}{m-1} \sum_{i\in P} \mathbb{E}_{\Lambda_1}[N_i(\tau)] \ge \frac{1}{m-1} \cdot \frac{d(\delta, 1-\delta)}{\mathrm{KL}(\nu_1, \nu_2)}.$$

□

Theorem 4.5 (Impossibility result)**.** *Let $\{\delta_T\}_{T=1}^{\infty} \in o(1)$ be a sequence that satisfies that there exists $b \in (0, 1]$ such that $\ln(1/\delta_T) = \Omega(T^b)$. For any instance Λ, the regret of any algorithm is lower bounded by $\Omega(T^b)$.*

Proof. The proof is similar to the proof of Lemma C.3. We consider an alternative instance $\Lambda_2 = (E, \mathcal{A}_K, \nu^{(2)}, \bar{\sigma}^2)$ with the distributions of the items in the optimal safe solution $S^\star$ changed such that (assume the variances of all the items in $S^\star$ are changed (increased))

$$\sum_{i \in S^\star} (\sigma_i^{(2)})^2 \geq \bar{\sigma}^2$$

i.e., under instance Λ_2, this solution $S^\star$ is unsafe (thus not optimal safe). The other items remain unchanged.

By a similar argument as the proof of Lemma C.3, we have

$$\sum_{i \in S^\star} \mathbb{E}_{\Lambda_1}[N_i(\tau)] \cdot \mathrm{KL}(\nu_i^{(1)}, \nu_i^{(2)}) \geq d(\delta, 1-\delta)$$
$$\Rightarrow \sum_{i \in S^\star} \mathbb{E}_{\Lambda_1}[N_i(\tau)] \geq \frac{d(\delta, 1-\delta)}{\min_{i \in S^\star} \mathrm{KL}(\nu_i^{(1)}, \nu_i^{(2)})}$$

So the safeness checking of $S^\star$ will take $\Omega(\ln \frac{1}{2.4\delta_T}) = \Omega(T^b)$

Recall the probably anytime-safe constraint (1):

$$\mathbb{P}\big[\forall\, t \in [T], S_t \in \mathcal{S}\big] \geq 1 - \delta_T.$$

This indicates at time step t, for any solution S, if $\mathbb{P}_{\mathcal{H}_t^{(0)}}[S \in \mathcal{S}] < 1 - \delta_T$, S will not be selected at this time step. Otherwise, (1) is violated. Therefore, before the safeness of the optimal safe solution S is ascertained, it is not going to be sampled and the instantaneous regret will be lower bounded by $\min_{S \in \mathcal{S} \cap \mathcal{B}} \Delta_S$.

In conclusion, the regret is at least $\ln \frac{1}{2.4\delta_T} \cdot \min_{S \in \mathcal{S} \cap \mathcal{B}} \Delta_S = \Omega(T^b)$. □

We derive both the problem-dependent and problem independent lower bounds on the K*-path semi-bandit* problem. The items in the ground set are divided into L_0 paths: $P_1, \ldots, P_{L_0}$, each of which contains K unique items. Path P_j contains items $(j-1)K+1, \ldots, jK$. Without loss of generality, we assume that L/K is an integer. A set S is a solution if and only if $S \subset P_j$ for some j. In other words, the solution set $\mathcal{A}_K = \{S : S \subset P_j,\ \exists j = 1, \ldots, L_0\}$. We let $N_i(t)$ denote the number of time steps item i is selected up to and including time step t, $M_j(t)$ denote the number of time steps a safe subset in path j is selected up to and including time step t, and $S_j(t)$ denote the time steps when a safe subset in path j is selected, i.e.,

$$N_i(t) = \sum_{s=1}^{t} \mathbb{1}\{i \in S_s\}, \quad M_j(t) = \sum_{s=1}^{t} \mathbb{1}\{S_s \subset P_j, S_s \text{ is safe }\}, \quad S_j(t) = \{1 \leq s \leq t : S_s \subset P_j, S_s \text{ is safe }\}.$$

We let $\mathrm{Reg}[j](t)$ denote the regret accumulated in $S_j(t)$. Since all the chosen solutions are safe with probability at least $1-\delta$, we have $\mathrm{Reg}(T) \geq (1-\delta) \cdot \sum_{j=1}^{L_0} \mathrm{Reg}[j](T)$. In the following, we lower bound the regret accumulated in time steps where safe solutions are chosen.

We will construct several instances to prove each of the lower bounds (which will be specified in the proof) such that under instance k ($\Lambda_k = (E, \mathcal{A}_K, \nu^{(k)}, \bar{\sigma}^2)$), the stochastic reward of items in path j ($1 \leq j \leq L_0$) are *i.i.d.*, i.e. , $\nu_i^{(k)} = \nu_{P_j}^{(k)}, \forall i \in P_j$, which will be specified in each case. Under instance k, we define several other notations as follows:

- Let $W_i(t)^{(k)}$ be the random reward of arm i at time step t.
- Let $S_t^{(k)}$ be the pulled solution at time step t, and $\mathcal{H}_t^{(j)} = \{(S_s^{(j)}, \{W_i(s)^{(k)}\}_{s \in S_s^{(k)}})\}_{s=1}^{t}$ be the sequence of selected solutions and observed rewards up to and including time step t.

For simplicity, we abbreviate $\mathbb{E}_{\mathcal{H}_T^{(k)}}, \mathbb{P}_{\mathcal{H}_T^{(k)}}$, as $\mathbb{E}_k, \mathbb{P}_k$ respectively.

C.2. Problem-dependent Lower bound

In order to provide a better understanding of the analysis, we first derive a lower bound with Gaussian distributions (unbounded) in Theorem C.4. With the same technique, we derive a lower bound with bounded Bernoulli distributions in Theorem 4.3, which corroborates with our problem setup.

Theorem C.4 (Problem-dependent lower bound for sub-Gaussian instances)**.** *Let $\{\delta_T\}_{T=1}^{\infty} \in o(1)$ be a sequence that satisfies $\ln(1/\delta_T) = o(T^b)$ for all $b > 0$. There exists an instance Λ, for any $\{\delta_T\}_{T\in\mathbb{N}}$-variance-constrained consistent algorithm π, the regret is lower bounded by*

$$\Omega\left(\sum_{i\in E}\frac{\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}}\right) + \frac{\mu^\star}{K}\cdot\Omega\left(\frac{K\cdot\ln(1/\delta_T)}{(\Delta_{S^\star})^2} + \sum_{i\in E}\left(\Psi'_{i,\mathcal{S}\cap\mathcal{B}} + \frac{\ln T}{(\Delta^{\mathrm{v}}_{i,\mathcal{R}})^2} + \Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\right)\right).$$

Proof. Given a vanishing sequence $\{\delta_T\}_{T\in\mathbb{N}}$, we consider a fixed $\{\delta_T\}_{T\in\mathbb{N}}$-variance-constrained consistent algorithm π on the K-path semi-bandit problem. For simplicity, the distributions of the items are assumed to be Gaussian in the proof, but the techniques can be applied to the Bernoulli case and we provide the instance design and the corresponding bound at the end of the proof.

Under instance Λ_0 (the base instance), $\sigma^2 = \frac{2\bar{\sigma}^2}{K}$ (so the absolutely safe solutions contain at most $K/2$ items) and the distributions of the items are

$$\nu_i^{(0)} = N(\mu_j^{(0)}, (\sigma_i^2)^{(0)}) = \begin{cases} \nu_{P_1}^{(0)} = N(\Delta, \dfrac{\bar{\sigma}^2 - \epsilon^{\mathrm{v}}}{K}), & i \in P_1 \\ \nu_{P_j}^{(0)} = N(\Delta - \dfrac{\epsilon^\mu}{K}, \dfrac{\bar{\sigma}^2 - \epsilon^{\mathrm{v}}}{K}), & i \in P_j, 2 \le j \le L_1 + 1 \\ \nu_{P_j}^{(0)} = N(\Delta + \dfrac{\epsilon^\mu}{K}, \dfrac{\bar{\sigma}^2 + \epsilon^{\mathrm{v}}}{K}), & i \in P_j, L_1 + 2 \le j \le L_2 + 1 \\ \nu_{P_j}^{(0)} = N(\Delta - \dfrac{\epsilon^\mu}{K}, \dfrac{\bar{\sigma}^2 + \epsilon^{\mathrm{v}}}{K}), & i \in P_j, L_2 + 2 \le j \le L_0 \end{cases}$$

where $\epsilon^\mu < \frac{\Delta}{K}$ and ϵ^{v} are small positive constants (e.g. $\epsilon^\mu = \frac{\Delta}{2K}, \epsilon^{\mathrm{v}} \le \frac{\bar{\sigma}^2}{K^2}$), and $L_1 = L_2 - L_1 = L_0 - L_2 - 1 = \frac{L_0 - 1}{3}$ (assume it is an integer). In this case,

- path 1 is an optimal safe path,
- path 2 to path $L_1 + 1$ are the safe and suboptimal paths,
- path $L_1 + 2$ to path $L_2 + 1$ are the risky paths
- path $L_2 + 2$ to path $L_3 + 1 = L_0$ are the unsafe and suboptimal paths.

In the following, we will compute the minimum regret yielded from each of the paths.

Case 1: the optimal safe path P_1

In order to achieve $o(T^a), \forall a > 0$ regret, any algorithm has to identify the safeness of the optimal safe solution P_1 and sample P_1 $\Omega(T)$ times, otherwise, the regret is linear. According to Lemma C.3, the expected number of time steps needed for the safeness identification of P_1 is lower bounded by

$$\mathbb{E}_0[M_1(\tau)] \ge \sup_{\nu_{P_1}^{(1)} \in E(\nu_{P_1}^{(0)})} \frac{1}{K-1}\cdot\frac{d(\delta_T, 1-\delta_T)}{\mathrm{KL}(\nu_{P_1}^{(0)}, \nu_{P_1}^{(1)})} := T(\nu_{P_1}^{(0)})$$

where $E(\nu_{P_1}^{(0)}) = \{\nu_{P_1}^{(1)} :$ the variance associated to $\nu_{P_1}^{(1)}$ is larger than $\bar{\sigma}^2/K\}$. In particular, we let instance $\Lambda_1 = (E, \mathcal{A}_K, \nu^{(1)}, \bar{\sigma}^2)$ with

$$\nu_i^{(1)} = \begin{cases} \nu_{P_1}^{(1)} = N\left(\Delta, \dfrac{\bar{\sigma}^2 + \epsilon_1^{\mathrm{v}}}{K}\right), & i \in P_1 \\ \nu_i^{(0)}, & i \notin P_1 \end{cases}$$

where ϵ_1^{v} is an arbitrarily small constant. Thus, we can take supremum over $\epsilon_1^{\mathrm{v}} > 0$ and have

$$T(\nu_{P_1}^{(0)}) \geq \frac{1}{K-1} \cdot \frac{\ln \frac{1}{2.4\delta_T}}{\mathrm{KL}(N(\Delta, \frac{\bar{\sigma}^2 - \epsilon^{\mathrm{v}}}{K}), N(\Delta, \frac{\bar{\sigma}^2}{K}))} \geq \frac{1}{K-1} \cdot \frac{4K^2(\bar{\sigma}^2/K)^2}{(\epsilon^{\mathrm{v}})^2} \ln \frac{1}{2.4\delta_T} = \frac{K(\sigma^2)^2}{(\epsilon^{\mathrm{v}})^2} \ln \frac{1}{2.4\delta_T}$$

When a solution $S \subset P_1$ is chosen before this time step, the instantaneous regret is lower bounded by Δ. The accumulative regret from P_1 is lower bounded by

$$\mathrm{Reg}[1](T) \geq \Delta \cdot T(\nu_{P_1}^{(0)}) = \Omega(\frac{K\Delta}{(\Delta_{P_1}^{\mathrm{v}})^2} \ln \frac{1}{\delta_T})$$

Case 2: the safe and suboptimal paths

For any safe and suboptimal path $P_j (2 \leq j \leq L_1 + 1)$, we define instance $\Lambda_j = (E, \mathcal{A}_K, \nu^{(j)}, \bar{\sigma}^2)$ with

$$\nu_i^{(j)} = \begin{cases} \nu_{P_j}^{(j)} = N(\Delta + \frac{\epsilon_j^{\mu}}{K}, \frac{\bar{\sigma}^2 - \epsilon^{\mathrm{v}}}{K}), & i \in P_j \\ \nu_i^{(0)}, & i \notin P_j \end{cases}$$

where $\epsilon_j^{\mu} < \Delta$ is an arbitrarily small constant. So P_j is the optimal safe solution under instance j.

Fix any item $i \in P_j$, consider the event $\mathcal{E}_j = \{N_i(T) \geq \frac{T}{2}\}$, under instance 1, $\mathcal{E}_j$ indicates the optimal safe solution P_1 is sampled less than $\frac{T}{2}$ times; under instance j, $\mathcal{E}_j^c$ indicates the optimal safe solution P_j is sampled less than $\frac{T}{2}$ times. Therefore,

$$\begin{aligned} \mathrm{Reg}_{\Lambda_0}(T) + \mathrm{Reg}_{\Lambda_j}(T) &\geq \frac{T}{2}\epsilon^{\mu}\mathbb{P}_0[\mathcal{E}_j] + \frac{T}{2}\epsilon_j^{\mu}\mathbb{P}_j[\mathcal{E}_j^c] \\ &\geq \frac{T}{2}\min\{\epsilon^{\mu}, \epsilon_j^{\mu}\}\left(\mathbb{P}_0[\mathcal{E}_j] + \mathbb{P}_j[\mathcal{E}_j^c]\right) \end{aligned}$$

According to Lemma C.1 and Lemma C.2, we have

$$\begin{aligned} &\mathbb{P}_0[\mathcal{E}_j] + \mathbb{P}_j[\mathcal{E}_j^c] \geq \frac{1}{2}\exp\{-\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)\} \\ &\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j) = \sum_{i=1}^{L} \mathbb{E}_0[N_i(T)] \cdot \mathrm{KL}\big(\nu_i^{(0)}, \nu_i^{(j)}\big) = \sum_{i \in P_j} \mathbb{E}_0[N_i(T)] \cdot \mathrm{KL}\big(\nu_i^{(0)}, \nu_i^{(j)}\big) \end{aligned}$$

Thus

$$\begin{aligned} & \mathrm{Reg}_{\Lambda_0}(T) + \mathrm{Reg}_{\Lambda_j}(T) \geq \frac{T}{2}\min\{\epsilon^{\mu}, \epsilon_j^{\mu}\} \cdot \frac{1}{2}\exp\{-\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)\} \\ \iff & \frac{\ln\Big(\mathrm{Reg}_{\Lambda_0}(T) + \mathrm{Reg}_{\Lambda_j}(T)\Big)}{\ln T} \geq 1 + \frac{\min\{\epsilon^{\mu}, \epsilon_j^{\mu}\}/4}{\ln T} - \frac{\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)}{\ln T} \\ \overset{(*)}{\implies} & \frac{\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)}{\ln T} \geq 1 \\ \iff & \frac{\sum_{i \in P_j} \mathbb{E}_0[N_i(T)]}{\ln T} \geq \frac{1}{\mathrm{KL}\big(\nu_{P_j}^{(0)}, \nu_{P_j}^{(j)}\big)} = \frac{2\frac{\bar{\sigma}^2 - \epsilon^{\mathrm{v}}}{K}}{(\frac{\epsilon^{\mu} + \epsilon_j^{\mu}}{K})^2} \\ \overset{(**)}{\implies} & \frac{\sum_{i \in P_j} \mathbb{E}_0[N_i(T)]}{\ln T} \geq \frac{2\frac{\bar{\sigma}^2 - \epsilon^{\mathrm{v}}}{K}}{(\frac{\epsilon^{\mu}}{K})^2} =: T(\nu_{P_j}^{(0)}) \end{aligned}$$

where in $(*)$ we let $T \to \infty$ and note that both $\mathrm{Reg}_{\Lambda_0}(T)$ and $\mathrm{Reg}_{\Lambda_j}(T)$ are of order $o(T^a), \forall a > 0$; in $(**)$ we take supremum over $\epsilon_j^{\mu} > 0$.

We also have to take the safeness constraint into consideration. According to Lemma C.3, in order to check the safeness of P_j, the items in P_j have to be sampled

$$\begin{aligned}
&\sum_{i\in P_j}\mathbb{E}_{\Lambda_0}[N_i(\tau)] \geq \sup_{\nu'_{P_j}\in E(\nu^{(0)}_{P_j})} \frac{d(\delta_T, 1-\delta_T)}{\mathrm{KL}(\nu^{(0)}_{P_j}, \nu'_{P_j})}. \\
\iff\quad & \frac{\sum_{i\in P_j}\mathbb{E}_{\Lambda_0}[N_i(\tau)]}{\ln T} \geq \sup_{\nu'_{P_j}\in E(\nu^{(0)}_{P_j})} \frac{1}{\mathrm{KL}(\nu^{(0)}_{P_j}, \nu'_{P_j})}\cdot\frac{d(\delta_T, 1-\delta_T)}{\ln T} \geq \frac{4K^2(\bar{\sigma}^2/K)^2}{(\epsilon^{\mathrm{v}})^2}\frac{\ln\frac{1}{2.4\delta_T}}{\ln T} := T_{\mathrm{safe}}(\nu^{(0)}_{P_j}).
\end{aligned}$$

If $T(\nu^{(0)}_{P_j}) \leq T_{\mathrm{safe}}(\nu^{(0)}_{P_j})$, it indicates the suboptimality of P_j is identified before the safeness. Furthermore, whenever a solution $S \subset P_j$ is sampled, S can have most $K-1$ items. So the regret is lower bounded by

$$\begin{aligned}
&\frac{\mathrm{Reg}[j](T)}{\ln T} \geq \frac{T(\nu^{(0)}_{P_j})}{K-1}\cdot(\epsilon^\mu + \Delta - \frac{\epsilon^\mu}{K}) \geq \frac{2K^2\frac{\bar{\sigma}^2-\epsilon^{\mathrm{v}}}{K}}{(\epsilon^\mu)^2}\cdot(\frac{\Delta}{K-1}+\frac{\epsilon^\mu}{K}) \\
\implies\quad & \frac{\mathrm{Reg}[j](T)}{\ln T} \geq \frac{2K\frac{\bar{\sigma}^2-\epsilon^{\mathrm{v}}}{K}}{(\epsilon^\mu)^2}\cdot(\Delta+\epsilon^\mu) \geq \frac{K\sigma^2/2}{(\epsilon^\mu)^2}\cdot(\Delta+\epsilon^\mu) \\
\implies\quad & \mathrm{Reg}[j](T) = \Omega\left(\frac{K}{\Delta^2_{P_j}}(\Delta_{P_j}+\Delta)\ln T\right)
\end{aligned} \tag{36}$$

If $T(\nu^{(0)}_{P_j}) \geq T_{\mathrm{safe}}(\nu^{(0)}_{P_j})$, it indicates the suboptimality of P_j is identified after the safeness. Thus, whenever a solution $S \subset P_j$ is sampled, S can have most $K-1$ items before the safeness checking is finished. So the regret is lower bounded by

$$\begin{aligned}
\frac{\mathrm{Reg}[j](T)}{\ln T} &\geq \frac{T_{\mathrm{safe}}(\nu^{(0)}_{P_j})}{K-1}\cdot(\epsilon^\mu+\Delta-\frac{\epsilon^\mu}{K}) + \frac{T(\nu^{(0)}_{P_j}) - T_{\mathrm{safe}}(\nu^{(0)}_{P_j})}{K}\cdot\epsilon^\mu \\
&\geq \frac{T(\nu^{(0)}_{P_j})}{K}\cdot\epsilon^\mu + (\Delta-\frac{\epsilon^\mu}{K})\cdot\frac{T_{\mathrm{safe}}(\nu^{(0)}_{P_j})}{K-1} \\
&\geq \frac{2K\frac{\bar{\sigma}^2-\epsilon^{\mathrm{v}}}{K}}{(\epsilon^\mu)^2}\cdot\epsilon^\mu + (\Delta-\frac{\epsilon^\mu}{K})\cdot\frac{1}{K-1}\cdot\frac{4K^2(\bar{\sigma}^2/K)^2}{(\epsilon^{\mathrm{v}})^2}\frac{\ln\frac{1}{2.4\delta_T}}{\ln T} \\
&\geq \frac{K\sigma^2/2}{\epsilon^\mu} + (\Delta-\frac{\epsilon^\mu}{K})\cdot\frac{1}{K-1}\cdot\frac{K^2(\sigma^2)^2}{(\epsilon^{\mathrm{v}})^2}\frac{\ln\frac{1}{2.4\delta_T}}{\ln T} \\
\implies\quad \mathrm{Reg}[j](T) &= \Omega\left(\frac{K}{\Delta_{P_j}}\ln T + (\Delta-\frac{\Delta_{P_j}}{K})\frac{K}{(\Delta^{\mathrm{v}}_{P_j})^2}\ln\frac{1}{\delta_T}\right)
\end{aligned} \tag{37}$$

Based on (36) and (37), the regret is

$$\mathrm{Reg}[j](T) = \Omega\left(\frac{K}{\Delta_{P_j}}\ln T + \Delta\cdot\min\left\{\frac{K}{\Delta^2_{P_j}}\ln T, \frac{K}{(\Delta^{\mathrm{v}}_{P_j})^2}\ln\frac{1}{\delta_T}\right\}\right)$$

Case 3: the risky paths

For any risky path P_j ($L_1+2 \leq j \leq L_2+1$, we define instance $\Lambda_j = (E, \mathcal{A}_K, \nu^{(j)}, \bar{\sigma}^2)$ with

$$\nu^{(j)}_i = \begin{cases} \nu^{(j)}_{P_j} = N(\Delta+\frac{\epsilon^\mu}{K}, \frac{\bar{\sigma}^2-\epsilon^{\mathrm{v}}_j}{K}), & i\in P_j \\ \nu^{(0)}_i, & i\notin P_j \end{cases}$$

where ϵ^{v}_j is an arbitrarily small constant. So P_j is the optimal safe solution under instance j.

Fix any item $i \in P_j$, consider the event $\mathcal{E}_j = \{N_i(T) \geq \frac{T}{2}\}$, under instance 1, $\mathcal{E}_j$ indicates the optimal safe solution P_1 is sampled less than $\frac{T}{2}$ times; under instance j, $\mathcal{E}_j^c$ indicates the optimal safe solution P_j is sampled less than $\frac{T}{2}$ times. Therefore,

$$\begin{aligned}\mathrm{Reg}_{\Lambda_0}(T) + \mathrm{Reg}_{\Lambda_j}(T) &\geq \frac{T}{2}\left(\Delta - \frac{K-1}{K}\epsilon^\mu\right)\mathbb{P}_0[\mathcal{E}_j] + \frac{T}{2}\epsilon^\mu \mathbb{P}_j[\mathcal{E}_j^c] \\ &\geq \frac{T}{2}\epsilon^\mu \cdot \left(\mathbb{P}_0[\mathcal{E}_j] + \mathbb{P}_j[\mathcal{E}_j^c]\right)\end{aligned}$$

According to Lemma C.1 and Lemma C.2, we have

$$\begin{aligned}&\mathbb{P}_0[\mathcal{E}_j] + \mathbb{P}_j[\mathcal{E}_j^c] \geq \frac{1}{2}\exp\{-\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)\} \\ &\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j) = \sum_{i=1}^{L} \mathbb{E}_0[N_i(T)] \cdot \mathrm{KL}\big(\nu_i^{(0)}, \nu_i^{(j)}\big) = \sum_{i \in P_j} \mathbb{E}_0[N_i(T)] \cdot \mathrm{KL}\big(\nu_i^{(0)}, \nu_i^{(j)}\big)\end{aligned}$$

Thus

$$\begin{aligned}& \mathrm{Reg}_{\Lambda_0}(T) + \mathrm{Reg}_{\Lambda_j}(T) \geq \frac{T}{2}\epsilon^\mu \cdot \frac{1}{2}\exp\{-\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)\} \\ \iff\ & \frac{\ln\Big(\mathrm{Reg}_{\Lambda_0}(T) + \mathrm{Reg}_{\Lambda_j}(T)\Big)}{\ln T} \geq 1 + \frac{\min\{\epsilon^\mu, \epsilon_j^\mu\}/4}{\ln T} - \frac{\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)}{\ln T} \\ \overset{(*)}{\Longrightarrow}\ & \frac{\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)}{\ln T} \geq 1 \\ \iff\ & \frac{\sum_{i \in P_j} \mathbb{E}_0[N_i(T)]}{\ln T} \geq \frac{1}{\mathrm{KL}\big(\nu_{P_j}^{(0)}, \nu_{P_j}^{(j)}\big)} = \frac{K^2(\sigma^2)^2}{(\epsilon^{\mathrm{v}})^2} =: T(\nu_{P_j}^{(0)})\end{aligned}$$

where in $(*)$ we let $T \to \infty$ and note that both $\mathrm{Reg}_{\Lambda_0}(T)$ and $\mathrm{Reg}_{\Lambda_j}(T)$ are of order $o(T^a), \forall a > 0$.

Note that P_j is a risky path and if any solution $S \subset P_j$ is sampled, $|S| \leq K-1$. Thus, $\mathbb{E}[M_j(T)] \geq \frac{T(\nu_{P_j}^{(0)})}{K-1}\ln T$. The regret is lower bounded by

$$\begin{aligned}\mathrm{Reg}[j](T) &= \Omega\left(\left(\Delta - \frac{K-1}{K}\epsilon^\mu\right)\frac{1}{K-1}\frac{K^2(\sigma^2)^2}{(\epsilon^{\mathrm{v}})^2}\ln T\right) \\ &= \Omega\left(\frac{K\Delta}{(\Delta_{P_j}^{\mathrm{v}})^2}\ln T\right)\end{aligned}$$

Case 4: the unsafe and suboptimal paths

For any unsafe and suboptimal path $P_j (L_2 + 2 \leq j \leq L_0)$, we define instance $\Lambda_j = (E, \mathcal{A}_K, \nu^{(j)}, \bar{\sigma}^2)$ with

$$\nu_i^{(j)} = \begin{cases} \nu_{P_j}^{(j)} = N(\Delta + \frac{\epsilon_j^\mu}{K}, \frac{\bar{\sigma}^2 - \epsilon_j^{\mathrm{v}}}{K}), & i \in P_j \\ \nu_i^{(0)}, & i \notin P_j \end{cases}$$

where $\epsilon_j^\mu < \Delta, \epsilon_j^{\mathrm{v}} < \frac{\bar{\sigma}^2}{K}$ are arbitrarily small constants. So P_j is the optimal safe solution under instance j.

Fix any item $i \in P_j$, consider the event $\mathcal{E}_j = \{N_i(T) \geq \frac{T}{2}\}$, under instance 1, $\mathcal{E}_j$ indicates the optimal safe solution P_1 is sampled less than $\frac{T}{2}$ times; under instance j, $\mathcal{E}_j^c$ indicates the optimal safe solution P_j is sampled less than $\frac{T}{2}$ times. Therefore,

$$\begin{aligned}\mathrm{Reg}_{\Lambda_0}(T) + \mathrm{Reg}_{\Lambda_j}(T) &\geq \frac{T}{2}\epsilon^\mu \mathbb{P}_0[\mathcal{E}_j] + \frac{T}{2}\epsilon_j^\mu \mathbb{P}_j[\mathcal{E}_j^c] \\ &\geq \frac{T}{2}\min\{\epsilon^\mu, \epsilon_j^\mu\}\left(\mathbb{P}_0[\mathcal{E}_j] + \mathbb{P}_j[\mathcal{E}_j^c]\right)\end{aligned}$$

According to Lemma C.1 and Lemma C.2, we have

$$\mathbb{P}_0[\mathcal{E}_j] + \mathbb{P}_j[\mathcal{E}_j^c] \geq \frac{1}{2}\exp\{-\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)\}$$
$$\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j) = \sum_{i=1}^{L} \mathbb{E}_0[N_i(T)] \cdot \mathrm{KL}\big(\nu_i^{(0)}, \nu_i^{(j)}\big) = \sum_{i \in P_j} \mathbb{E}_0[N_i(T)] \cdot \mathrm{KL}\big(\nu_i^{(0)}, \nu_i^{(j)}\big)$$

Thus

$$\begin{aligned}
& \mathrm{Reg}_{\Lambda_0}(T) + \mathrm{Reg}_{\Lambda_j}(T) \geq \frac{T}{2}\min\{\epsilon^\mu, \epsilon_j^\mu\} \cdot \frac{1}{2}\exp\{-\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)\} \\
\iff \quad & \frac{\ln\Big(\mathrm{Reg}_{\Lambda_0}(T) + \mathrm{Reg}_{\Lambda_j}(T)\Big)}{\ln T} \geq 1 + \frac{\min\{\epsilon^\mu, \epsilon_j^\mu\}/4}{\ln T} - \frac{\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)}{\ln T} \\
\overset{(*)}{\implies} \quad & \frac{\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_j)}{\ln T} \geq 1 \\
\iff \quad & \frac{\sum_{i \in P_j} \mathbb{E}_0[N_i(T)]}{\ln T} \geq \frac{1}{\mathrm{KL}\big(\nu_{P_j}^{(0)}, \nu_{P_j}^{(j)}\big)} = 4K^2 \left(\left(\frac{\epsilon_j^{\mathrm{v}} + \epsilon^{\mathrm{v}}}{(\bar{\sigma}^2 - \epsilon_j^{\mathrm{v}})/K} \right)^2 + \frac{2(\epsilon_j^\mu + \epsilon^\mu)^2}{(\bar{\sigma}^2 - \epsilon_j^{\mathrm{v}})/K} \right)^{-1} \\
\overset{(**)}{\implies} \quad & \frac{\sum_{i \in P_j} \mathbb{E}_0[N_i(T)]}{\ln T} \geq 4K^2 \left(\left(\frac{\epsilon^{\mathrm{v}}}{\sigma^2/2} \right)^2 + \frac{2(\epsilon^\mu)^2}{\sigma^2/2} \right)^{-1} =: T(\nu_{P_j}^{(0)})
\end{aligned}$$

where in $(*)$ we let $T \to \infty$ and note that both $\mathrm{Reg}_{\Lambda_0}(T)$ and $\mathrm{Reg}_{\Lambda_j}(T)$ are of order $o(T^a), \forall a > 0$; in $(**)$ we take the supremum over $\epsilon_j^\mu > 0, \epsilon_j^{\mathrm{v}} > 0$.

Note that P_j is an unsafe path and if any solution $S \subset P_j$ is sampled, $|S| \leq K - 1$. Thus, $\mathbb{E}[M_j(T)] \geq \frac{T(\nu_{P_j}^{(0)})}{K-1} \ln T$. The regret is lower bounded by

$$\begin{aligned}
\mathrm{Reg}[j](T) &\geq \left(\Delta + \frac{K-1}{K}\epsilon^\mu\right) \frac{4K^2}{K-1} \left(\left(\frac{\epsilon^{\mathrm{v}}}{\sigma^2/2} \right)^2 + \frac{2(\epsilon^\mu)^2}{\sigma^2/2} \right)^{-1} \ln T \\
&= \Omega\left(\frac{K\Delta + K\epsilon^\mu}{\max\{\epsilon^\mu, \epsilon^{\mathrm{v}}\}^2} \ln T \right) \\
&= \Omega\left(\min\left\{ \frac{K\Delta}{\Delta_{P_j}^2} \ln T, \frac{K\Delta}{(\Delta_{P_j}^{\mathrm{v}})^2} \ln T \right\} \right)
\end{aligned}$$

In conclusion, the regret yielded from these paths is lower bounded by

$$\begin{aligned}
\frac{\mathrm{Reg}(T)}{1 - \delta_T} &\geq \mathrm{Reg}[1](T) + \sum_{j=2}^{L_1+1} \mathrm{Reg}[j](T) + \sum_{j=L_1+2}^{L_2+1} \mathrm{Reg}[j](T) + \sum_{j=L_2+2}^{L_0} \mathrm{Reg}[j](T) \\
&\geq \Omega\left(\frac{K\Delta}{(\epsilon^{\mathrm{v}})^2} \ln \frac{1}{\delta_T} \right) + \sum_{j=2}^{L_1+1} \Omega\left(\frac{K}{\epsilon^\mu} \ln T + \Delta \cdot \min\left\{ \frac{K}{(\epsilon^\mu)^2} \ln T, \frac{K}{(\epsilon^{\mathrm{v}})^2} \ln \frac{1}{\delta_T} \right\} \right) \\
&\quad + \sum_{j=L_1+2}^{L_2+1} \Omega\left(\frac{K\Delta}{(\epsilon^{\mathrm{v}})^2} \ln T \right) + \sum_{j=L_2+2}^{L_0} \Omega\left(\min\left\{ \frac{K\Delta}{(\epsilon^\mu)^2} \ln T, \frac{K\Delta}{(\epsilon^{\mathrm{v}})^2} \ln T \right\} \right).
\end{aligned}$$

Note

- $L_1 = L_2 - L_1 = L_0 - L_2 - 1 = \frac{L_0 - 1}{3}$, $L_0 = \frac{L}{K}$ and $\mu^\star = K\Delta$.
- for $S^\star$, $\epsilon^{\mathrm{v}} = \Delta_{S^\star}^{\mathrm{v}}$.

- for $S \in \mathcal{S} \cap \mathcal{B}$ and $i \in S$, we check that if $S \subset P_j, 2 \leq j \leq L_1 + 1$, $\Delta_{i,\mathcal{S}\cap\mathcal{B},\min} = \epsilon^\mu$ and

$$\Psi'_{i,\mathcal{S}\cap\mathcal{B}} \geq \min\left\{\frac{\ln T}{\Delta_S^2}, \frac{9\ln(1/\delta_T)}{(\Delta_S^{\mathrm{v}})^2}\right\}.$$

  where the equality holds when $S = P_j, 2 \leq j \leq L_1 + 1$.
- for $S \in \mathcal{R}$ and $i \in S$, $\epsilon^{\mathrm{v}} = \Delta^{\mathrm{v}}_{i,\mathcal{R}}$.
- $S \in \mathcal{S}^c \cap \mathcal{B}$ and $i \in S$, we can easily check $S = P_j$ for some $L_2 + 2 \leq j \leq L_0$, thus $\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min} = \epsilon^\mu$ and

$$\Phi_{i,\mathcal{S}^c\cap\mathcal{B}} = \min\left\{\frac{\ln T}{\Delta_S^2}, \frac{9\ln T}{(\Delta_S^{\mathrm{v}})^2}\right\}.$$

Therefore,

$$\mathrm{Reg}(T) \geq \Omega\left(\frac{L\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}}\right) + \frac{\mu^\star}{K}\cdot\Omega\left(\frac{K\cdot\ln(1/\delta_T)}{(\Delta_{S^\star})^2} + \sum_{i\in E}\Psi'_{i,\mathcal{S}\cap\mathcal{B}} + \sum_{i\in E}\frac{\ln T}{(\Delta^{\mathrm{v}}_{i,\mathcal{R}})^2} + \sum_{i\in E}\Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\right).$$

□

Theorem 4.3 (Problem-dependent lower bound)**.** *Let $\{\delta_T\}_{T=1}^\infty \in o(1)$ be a sequence that satisfies $\ln(1/\delta_T) = o(T^b)$ for all $b > 0$. There exists an instance Λ such that for any $\{\delta_T\}_{T\in\mathbb{N}}$-variance-constrained consistent algorithm π, the regret is lower bounded by*

$$\Omega\Bigg(\sum_{i\in E}\frac{\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}}\Bigg) + \frac{\mu^\star}{K}\cdot\Omega\Bigg(\frac{K\ \ln(1/\delta_T)}{(\Delta^{\mathrm{v}}_{S^\star})^2} + \sum_{i\in E}\Big(\Psi'_{i,\mathcal{S}\cap\mathcal{B}} + \frac{\ln T}{(\Delta^{\mathrm{v}}_{i,\mathcal{R}})^2} + \Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\Big)\Bigg).$$

Proof. We divide the whole ground set into several paths. Meanwhile, we define four sets $E_1, \ldots, E_4$ such that $E_i \cap E_j = \emptyset$ for any $i \neq j$. $P_1 = E_1$, $P_j \subset E_j$ for $j = 2, 3, 4$. For any $j > 4$, there exists $k \neq 1$ such that $P_j \subset E_k$. For any path P_j, if $P_j \subset E_4$, $|P_j| = K_2$; otherwise, $|P_j| = K_1$. P_j consists of arms

$$\sum_{i=1}^{j-1}|P_i| + 1, \ldots, \sum_{i=1}^{j-1}|P_i| + K_1\cdot\mathbb{1}\{P_j \not\subset E_4\} + K_2\cdot\mathbb{1}\{P_j \subset E_4\}.$$

The feasible solution set $\mathcal{A}_K$ consists of all subsets of each path.

We let $\mathrm{Bern}(a)$ denote the Bernoulli distribution with parameter $a (a \in (0,1))$. Note that the variance of $\mathrm{Bern}(a)$ is $a(1-a)$. We construct an instance with $\nu_i = \mathrm{Bern}(\mu_i)$. With $\varepsilon_1, \varepsilon_2 > 0$, we set

$$\begin{aligned}
\mu_i &= \Delta \quad \forall i \in E_1,\\
\mu_i &= \Delta - \varepsilon_1 \quad \forall i \in E_2,\\
\mu_i &= \Delta + \varepsilon_2 \quad \forall i \in E_3,\\
\mu_i &= \Delta - \varepsilon_3 \quad \forall i \in E_4.
\end{aligned}$$

We let $2 \leq K_1 < K_2 \leq K$, $\Delta < 1/2$ and

$K_1\Delta(1-\Delta) < \bar{\sigma}^2$ (paths in E_1 and E_2 are safe),
$K_1(\Delta+\varepsilon_2)(1-\Delta-\varepsilon_2) > \bar{\sigma}^2$ (paths in E_3 are unsafe),
$K_2(\Delta-\varepsilon_3)(1-\Delta+\varepsilon_3) > \bar{\sigma}$ (paths in E_4 are unsafe),
$K_2(\Delta-\varepsilon_3) < K_1\Delta$ (paths in E_4 are suboptimal),
$(K_1-1)\cdot(\Delta+\varepsilon_2) < K_1\cdot\Delta \Leftrightarrow (K_1-1)\cdot\varepsilon_2 < \Delta \Leftrightarrow \varepsilon_2 < \dfrac{\Delta}{K_1-1}$ (paths in E_3 are suboptimal or unsafe).

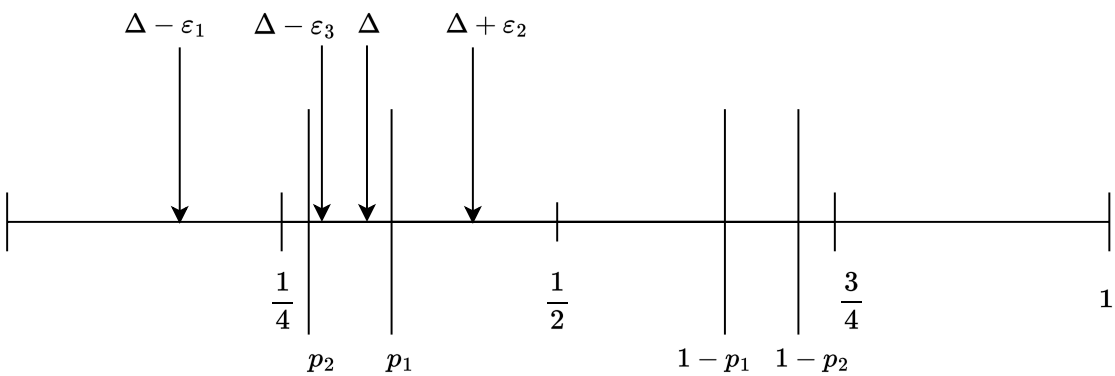
The conditions above indicate that

- P_1 is the unique optimal safe set;
- if $P_j \subset E_2$, P_j and its subsets are safe but suboptimal;
- if $P_j \subset E_3$, P_j is risky, and its proper subsets are suboptimal;
- if $P_j \subset E_4$, P_j and its subsets are suboptimal, and P_j is unsafe.

Let

$$p_1 := \frac{1-\sqrt{1-\bar{\sigma}^2/K_1}}{2} \text{ and } p_2 := \frac{1-\sqrt{1-\bar{\sigma}^2/K_2}}{2},$$

The relations between μ_i's are as in the following figure:



With a similar proof to that of Theorem C.4, we have, the accumulative regret from the optimal P_1 is lower bounded by

$$\mathrm{Reg}[1](T) \geq \Omega(\frac{K_1\Delta}{(\Delta^{\mathrm{v}}_{P_1})^2}\ln\frac{1}{\delta_T});$$

the accumulative regret from a safe and suboptimal path in E_2 is lower bounded by

$$\mathrm{Reg}[j](T) \geq \Omega\left(\frac{K_1}{\Delta_{P_j}}\ln T + \Delta\cdot\min\left\{\frac{K_1}{\Delta^2_{P_j}}\ln T, \frac{K_1}{(\Delta^{\mathrm{v}}_{P_j})^2}\ln\frac{1}{\delta_T}\right\}\right);$$

the accumulative regret from a risky path in E_3 is lower bounded by

$$\mathrm{Reg}[j](T) \geq \Omega\left(\frac{K_1\Delta}{(\Delta^{\mathrm{v}}_{P_j})^2}\ln T\right);$$

the accumulative regret from a unsafe and suboptimal path in E_4 is lower bounded by

$$\mathrm{Reg}[j](T) \geq \Omega\left(\min\left\{\frac{K_2\Delta}{\Delta^2_{P_j}}\ln T, \frac{K_2\Delta}{(\Delta^{\mathrm{v}}_{P_j})^2}\ln T\right\}\right).$$

We let

$$\epsilon^{\mu} = \min_j\{\Delta_{P_j}\}, \quad \epsilon^{\mathrm{v}} = \min_j\{\Delta^{\mathrm{v}}_{P_j}\};$$

and set K_1, K_2, ε_1, ε_2, ε_3 such that

$$K_1 > \frac{3}{4}\cdot K_2, \quad K_2 = K, \quad \min_j\{\Delta_{P_j}\} < 2\epsilon^{\mu}, \quad \min_j\{\Delta^{\mathrm{v}}_{P_j}\} < 2\epsilon^{\mathrm{v}}.$$

In conclusion, the regret yielded from these paths is lower bounded by

$$\begin{aligned}\frac{\mathrm{Reg}(T)}{1-\delta_T} &\geq \mathrm{Reg}[1](T) + \sum_{P_j \in E_2} \mathrm{Reg}[j](T) + \sum_{P_j \in E_3} \mathrm{Reg}[j](T) + \sum_{P_j \in E_4} \mathrm{Reg}[j](T) \\ &\geq \Omega\left(\frac{K\Delta}{(\epsilon^{\mathrm{v}})^2} \ln \frac{1}{\delta_T}\right) + |E_2| \cdot \Omega\left(\frac{K}{\epsilon^{\mu}} \ln T + \Delta \cdot \min\left\{\frac{K}{(\epsilon^{\mu})^2} \ln T, \frac{K}{(\epsilon^{\mathrm{v}})^2} \ln \frac{1}{\delta_T}\right\}\right) \\ &\quad + |E_3| \cdot \Omega\left(\frac{K\Delta}{(\epsilon^{\mathrm{v}})^2} \ln T\right) + |E_4| \cdot \Omega\left(\min\left\{\frac{K\Delta}{(\epsilon^{\mu})^2} \ln T, \frac{K\Delta}{(\epsilon^{\mathrm{v}})^2} \ln T\right\}\right).\end{aligned}$$

Note that

- $\mu^\star = K_1 \Delta$.
- for $S^\star$, $\epsilon^{\mathrm{v}} = \Delta^{\mathrm{v}}_{S^\star}$.
- for $S \in \mathcal{S} \cap \mathcal{B}$ and $i \in S$, we check that if $S \subset P_j, j \in E_2$, $\Delta_{i,\mathcal{S}\cap\mathcal{B},\min} = \epsilon^{\mu}$ and

$$\Psi'_{i,\mathcal{S}\cap\mathcal{B}} \geq \min\left\{\frac{\ln T}{\Delta_S^2}, \frac{9\ln(1/\delta_T)}{(\Delta^{\mathrm{v}}_S)^2}\right\}.$$

  where the equality holds when $S = P_j, P_j \in E_2$.
- for $S \in \mathcal{R}$ and $i \in S$, $\epsilon^{\mathrm{v}} = \Delta^{\mathrm{v}}_{i,\mathcal{R}}$.
- $S \in \mathcal{S}^c \cap \mathcal{B}$ and $i \in S$, we can easily check $S = P_j$ for some $P_j \in E_4$, thus $\Delta_{i,\mathcal{S}^c\cap\mathcal{B},\min} = \epsilon^{\mu}$ and

$$\Phi_{i,\mathcal{S}^c\cap\mathcal{B}} = \min\left\{\frac{\ln T}{\Delta_S^2}, \frac{9\ln T}{(\Delta^{\mathrm{v}}_S)^2}\right\}.$$

Therefore,

$$\mathrm{Reg}(T) \geq \Omega\left(\frac{L\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}}\right) + \frac{\mu^\star}{K} \cdot \Omega\left(\frac{K \cdot \ln(1/\delta_T)}{(\Delta_{S^\star})^2} + \sum_{i\in E} \Psi'_{i,\mathcal{S}\cap\mathcal{B}} + \sum_{i\in E} \frac{\ln T}{(\Delta^{\mathrm{v}}_{i,\mathcal{R}})^2} + \sum_{i\in E} \Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\right).$$

□

C.3. Problem-independent Lower bound

Theorem 5.2 (Problem-independent lower bound)**.** *Let the minimum variance gap be* $\Delta^{\mathrm{v}} := \min_{S\in\mathcal{A}_K} \Delta^{\mathrm{v}}_S$. *When* $K^3 \geq L^2$, *we have*

$$\mathrm{Reg}(T) = \Omega\left(\sqrt{KLT} + \min\left\{\frac{L}{(\Delta^{\mathrm{v}})^2} \ln\left(\frac{1}{\delta_T}\right), T\right\}\right).$$

Proof. Since the rewards of items are bounded in $[0,1]$, the variance of each arm is at most $1/4$. Therefore, when $\bar{\sigma}^2 \in [K/4, \infty)$, any solution in $\mathcal{A}_K$ is safe and there exists a generic lower bound (Kveton et al., 2015). When $\bar{\sigma}^2 \in (0, K/4)$, let

$$\bar{a} := \frac{1+\sqrt{1-\bar{\sigma}^2/K}}{2} \text{ and } \underline{a} := \frac{1-\sqrt{1-\bar{\sigma}^2/K}}{2}.$$

We consider the instances containing items with Bernoulli reward distributions. We let $\mathrm{Bern}(a)$ denote the Bernoulli distribution with mean a. An item $i \in [L]$ with reward distribution $\mathrm{Bern}(\mu_i)$ is with variance $\mu_i(1-\mu_i)$, which is smaller than $\bar{\sigma}^2/K$ if and only if $\mu_i \in (0, \underline{a}) \cup (\bar{a}, 1)$.

We will construct 3 instances such that under instance k ($0 \leq k \leq 2$), the stochastic reward of an item in path j ($1 \leq j \leq L_0$) is drawn from distribution $\nu_j^{(k)} = \mathrm{Bern}(\mu_j^{(k)})$, where $\mu_j^{(k)}$ will be specified in each case.

Under instance 0, with $\mu_0 < \underline{a}$, we let $\mu_j^{(0)} = \mu_0$ for all j, i.e., the reward distribution of each item is $\mathrm{Bern}(\mu_0)$. Since $\mu_{P_j} = K\mu_0$ and $\sigma^2_{P_j} = K\mu_0(1-\mu_0) < \bar{\sigma}^2$ for all j, each path is an identical safe and optimal solution. Since all paths

are equivalent under instance 0, we have $\mathbb{E}_0[M_j(t)] = T/L_0$ for all $j \in 1, \ldots, L_0$, where $\mathbb{E}_0$ denote the expectation under instance 0.

We next construct instance 1 such that

$$\mu_1^{(1)} = \mu_1, \qquad \mu_j^{(1)} = \mu_0 \quad j \neq 1,$$

where $\mu_0 < \mu_1 < \underline{a}$.[2] Hence, P_1 is the unique optimal safe solution under instance 1 while all other solutions are safe but suboptimal.

With an analysis similar to that of Lemma 6.4 in Zhong et al. (2021), we can show that

Lemma C.5. *Let the reward distribution of item i be $\nu_i^{(j)}$ under instance $j (j = 1, 2)$, then*

$$\mathrm{KL}\big(\mathcal{H}_T^{(1)}, \mathcal{H}_T^{(2)}\big) = \sum_{i=1}^{L} \mathbb{E}_0[N_i(T)] \cdot \mathrm{KL}\big(\nu_i^{(1)}, \nu_i^{(2)}\big).$$

Hence, we have

$$\mathrm{KL}\big(\mathcal{H}_T^{(0)}, \mathcal{H}_T^{(1)}\big) = \sum_{i \in P_1} \mathbb{E}_0[N_i(T)] \cdot \big(\nu_i^{(0)}, \nu_i^{(1)}\big) \overset{(a)}{\leq} K \cdot \mathbb{E}_0[M_1(t)] \cdot d(\mu_0, \mu_1) = K \cdot \frac{T}{L_0} \cdot d(\mu_0, \mu_1),$$

where (a) follows from the fact that at most K items are selected at one time step and the definition of instances.

Next, we apply Pinsker's Inequality to bound $\mathbb{E}_1[M_1(T)]$. Lemma C.1 indicates that

$$\begin{aligned} &|\mathbb{E}_0[M_1(T)] - \mathbb{E}_1[M_1(T)]| \leq \sqrt{\frac{1}{2}\mathrm{KL}\big(\mathcal{H}_T^{(0)}, \mathcal{H}_T^{(1)}\big)}, \\ \Rightarrow \quad &\left|\mathbb{E}_1[M_1(T)] - \frac{T}{L_0}\right| \leq \sqrt{\frac{KT}{2L_0} \cdot d(\mu_0, \mu_1)}. \end{aligned}$$

Moreover, since the paths P_j for $j \neq 1$ are identical under instance 1, we have

$$\mathbb{E}_1[M_j(T)] = \frac{1}{L_0 - 1}\big(T - \mathbb{E}_1[M_1(T)]\big) \geq \frac{1}{L_0 - 1}\left(T - \frac{T}{L_0} - \sqrt{\frac{KT}{2L_0} \cdot d(\mu_0, \mu_1)}\right) := \underline{M}.$$

To learn the regret incurred by P_2 under instance 1, we need to take the effects of the safety constraint into consideration. Lemma C.3 indicates that

- if $\underline{M} < T(\nu_0)$, at each of the first $\underline{M}$ time steps in $S_2(T)$, at most $K - 1$ items are pulled and regret of at least $[K(\mu_1 - \mu_2) + \mu_2] \cdot \underline{M}$ is incurred, i.e.,

$$\mathrm{Reg}[2] \geq [K(\mu_1 - \mu_2) + \mu_2] \cdot \underline{M}.$$

- if $\underline{M} \geq T(\nu_0)$, at each of the first $T(\nu_0)$ time steps in $S_2(T)$, at most $K - 1$ items are pulled and regret of at least $[K(\mu_1 - \mu_2) + \mu_2] \cdot T(\nu_0)$ is incurred; besides, at the subsequent time steps in $S_2(T)$, regret of at least $K(\mu_1 - \mu_2) \cdot [\underline{M} - T(\nu_0)]$ is incurred, i.e.,

$$\mathrm{Reg}[2] \geq [K(\mu_1 - \mu_2) + \mu_2] \cdot T(\nu_0) + K(\mu_1 - \mu_2) \cdot [\underline{M} - T(\nu_0)] = K(\mu_1 - \mu_2) \cdot \underline{M} + \mu_2 \cdot T(\nu_0).$$

In short, we have

$$\mathrm{Reg}[2] \geq K(\mu_1 - \mu_2) \cdot \underline{M} + \mu_2 \cdot \min\{T(\nu_0), \underline{M}\}.$$

We can lower bound $\mathrm{Reg}(j)$ for all $j = 3, \ldots, L_0$ with the same method.

[2] In this proof, μ_1 are μ_2 are redefined to minimize clutter; the previous definitions of them not used.

Besides, since

$$\mathbb{E}_1[M_1(T)] \geq T - \sqrt{\frac{KT}{2L_0} \cdot d(\mu_0, \mu_1)}$$

and at most $K-1$ items are selected at each of the first $T(\nu_1)$ time steps in $S_1(T)$, we have

$$\mathrm{Reg}[1] \geq \mu_1 \cdot \min\left\{T(\nu_1), T - \sqrt{\frac{KT}{2L_0} \cdot d(\mu_0, \mu_1)}\right\}.$$

Therefore, under instance 1, we have

$$\begin{aligned}
\frac{\mathrm{Reg}(T)}{1-\delta} &\geq \sum_{j=1}^{L_0} \mathrm{Reg}[j] \\
&\geq \mu_1 \cdot \min\left\{T(\nu_1), T - \sqrt{\frac{KT}{2L_0} \cdot d(\mu_0, \mu_1)}\right\} + (L_0 - 1) \cdot [K(\mu_1 - \mu_2) \cdot \underline{M} + \mu_2 \cdot \min\{T(\nu_0), \underline{M}\}] \\
&= \mu_1 \cdot \min\left\{\sup_{\nu_1' \in E(\nu_1)} \frac{1}{K-1} \cdot \frac{d(\delta, 1-\delta)}{\mathrm{KL}(\nu_1, \nu_1')}, T - \sqrt{\frac{KT}{2L_0} \cdot d(\mu_0, \mu_1)}\right\} \\
&\quad + (L_0 - 1) \cdot \left[K(\mu_1 - \mu_2) \cdot \underline{M} + \mu_2 \cdot \min\left\{\sup_{\nu_0' \in E(\nu_0)} \frac{1}{K-1} \cdot \frac{d(\delta, 1-\delta)}{\mathrm{KL}(\nu_0, \nu_0')}, \underline{M}\right\}\right]
\end{aligned}$$

where

$$\begin{aligned}
E(\nu_0) &= \{\nu(0') : \text{ the variance related to } \nu(0') \text{ is larger than } \bar{\sigma}^2/K\}, \\
E(\nu_1) &= \{\nu(1') : \text{ the variance related to } \nu(1') \text{ is larger than } \bar{\sigma}^2/K\}, \\
\underline{M} &= \frac{1}{L_0 - 1}\left(T - \frac{T}{L_0} - \sqrt{\frac{KT}{2L_0} \cdot d(\mu_0, \mu_1)}\right).
\end{aligned}$$

By Pinsker's inequality (see Lemma C.1), for $\mu_i \geq 1/2$,

$$d(\mu_i, \bar{a}) \leq \frac{(\mu_i - \bar{a})^2 \cdot (1 - \mu_i - \bar{a})^2}{\underline{a}(1 - \mu_i - \bar{a})^2} = \frac{[\mu_i(1-\mu_i) - \bar{\sigma}^2/K]^2}{\underline{a}(1 - \mu_i - \bar{a})^2} \leq \frac{[\mu_i(1-\mu_i) - \bar{\sigma}^2/K]^2}{\underline{a}(\bar{a} - 1/2)^2} = \frac{[\mu_i(1-\mu_i) - \bar{\sigma}^2/K]^2}{\underline{a}(1/2 - \underline{a})^2};$$

for $\mu_i < 1/2$,

$$d(\mu_i, \underline{a}) \leq \frac{(\mu_i - \underline{a})^2 \cdot (1 - \mu_i - \underline{a})^2}{\underline{a}(1 - \mu_i - \underline{a})^2} = \frac{[\mu_i(1-\mu_i) - \bar{\sigma}^2/K]^2}{\underline{a}(1 - \mu_i - \underline{a})^2} \leq \frac{[\mu_i(1-\mu_i) - \bar{\sigma}^2/K]^2}{\underline{a}(1/2 - \underline{a})^2}.$$

Since $\nu_0 = \mathrm{Bern}(\mu_0)$, $\nu_1 = \mathrm{Bern}(\mu_1)$, and $0 < \mu_0 < \mu_1 < \underline{a} < 1/2$, we have

$$\begin{aligned}
\sup_{\nu_0' \in E(\nu_0)} \frac{1}{K-1} \cdot \frac{d(\delta, 1-\delta)}{\mathrm{KL}(\nu_0, \nu_0')} &= \frac{d(\delta, 1-\delta)}{K-1} \cdot \frac{(1/2 - \underline{a})^2}{[\mu_0(0 - \mu_0) - \bar{\sigma}^2/K]^2}, \\
\sup_{\nu_1' \in E(\nu_1)} \frac{1}{K-1} \cdot \frac{d(\delta, 1-\delta)}{\mathrm{KL}(\nu_1, \nu_1')} &= \frac{d(\delta, 1-\delta)}{K-1} \cdot \frac{(1/2 - \underline{a})^2}{[\mu_1(1 - \mu_1) - \bar{\sigma}^2/K]^2}.
\end{aligned}$$

We define the minimum variance gap $\Delta^v := \min_{S \in \mathcal{A}} \Delta_S^v$. When $K^3 \leq L^2$ and $LK/T \leq \underline{a}^2$, we can let $\mu_1 - \mu_2 = \sqrt{L/KT}$. Then we have

$$\begin{aligned}
\mathrm{Reg}(T) &= \Omega\left(\min\left\{\frac{1}{K-1} \cdot \frac{d(\delta, 1-\delta)}{(\Delta^v/K)^2}, T\right\} + \sqrt{KLT} + L_0 \cdot \min\left\{\frac{1}{K-1} \cdot \frac{d(\delta, 1-\delta)}{(\Delta^v/K)^2}, \frac{TK}{L}\right\}\right) \\
&= \Omega\left(\min\left\{K \cdot \frac{d(\delta, 1-\delta)}{(\Delta^v/)^2}, T\right\} + \sqrt{KLT} + L \cdot \min\left\{\frac{d(\delta, 1-\delta)}{(\Delta^v)^2}, \frac{T}{L}\right\}\right) \\
&= \Omega\left(\sqrt{KLT} + \min\left\{L \cdot \frac{d(\delta, 1-\delta)}{(\Delta^v)^2}, T\right\}\right).
\end{aligned}$$

Moreover, we complete the proof with $d(\delta, 1-\delta) \geq -\ln(2.4\delta)$ and $\delta_T = \delta$.

□

D. Tightness of the Upper bound

Corollary D.1 (Tightness of problem-dependent bounds)**.** *Let* $\{\delta_T\}_{T=1}^{\infty} \in o(1)$ *be a sequence that satisfies* $\ln(1/\delta_T) = o(T^b)$ *for all* $b > 0$*,*

- *if* $\ln(1/\delta_T) \in o(\ln T)$*, in particular, if* $\delta_T = \delta_0 > 0$ *for all* T*, the regret* $\mathrm{Reg}(T)$ *is*

$$\Omega\left(\sum_{i\in E}\frac{\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}} + \frac{\mu^\star \ln T/K}{(\Delta^{\mathrm{v}}_{i,\mathcal{R}})^2} + \frac{\mu^\star}{K}\Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\right)$$
$$\cap\, O\left(\sum_{i\in E}\frac{K\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}} + \frac{\mu^\star K\ln T}{(\Delta^{\mathrm{v}}_{i,\mathcal{R}})^2} + \mu^\star K\Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\right)$$

- *if* $\ln(1/\delta_T) = \lambda \ln T$*, i.e.,* $\delta_T = T^{-\lambda}$ *with a fixed* $\lambda > 0$*, the regret* $\mathrm{Reg}(T)$ *is*

$$\Omega\left(\sum_{i\in E}\frac{\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}} + \frac{\lambda\mu^\star \ln T}{(\Delta^{\mathrm{v}}_{S^\star})^2} + \frac{\mu^\star}{K}\sum_{i\in E}\left(\Psi'_{i,\mathcal{S}\cap\mathcal{B}} + \frac{\ln T}{(\Delta^{\mathrm{v}}_{i,\mathcal{R}})^2} + \Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\right)\right)$$
$$\cap\, O\left(\sum_{i\in E}\frac{K\ln T}{\Delta_{i,\mathcal{S}\cap\mathcal{B},\min}} + \frac{\lambda\mu^\star K^2 \ln T}{(\Delta^{\mathrm{v}}_{S^\star})^2} + K\mu^\star\sum_{i\in E}\left(\Psi_{i,\mathcal{S}\cap\mathcal{B}} + \frac{\ln T}{(\Delta^{\mathrm{v}}_{i,\mathcal{R}})^2} + \Phi_{i,\mathcal{S}^c\cap\mathcal{B}}\right)\right)$$

  where $\ln(1/\delta_T)$ *in the* Ψ *and* Ψ' *functions should be replaced by* $\lambda \ln T$*.*
- *if* $\ln(1/\delta_T) \in \omega(\ln T)$*, the regret* $\mathrm{Reg}(T)$ *is*

$$\mathrm{Reg}(T) \in \Omega\left(\frac{\mu^\star \ln(1/\delta_T)}{(\Delta^{\mathrm{v}}_{S^\star})^2}\right) \cap O\left(\frac{\mu^\star K^2 \ln(1/\delta_T)}{(\Delta^{\mathrm{v}}_{S^\star})^2}\right)$$

The upper bounds above are achieved by PASCOMBUCB.

E. Additional Discussions and Future Research

E.1. Discussions on the Tightness Results

In terms of the problem-dependent bounds in Corollary 4.4, we consider general instances where the rewards from the items are independent and the gap in $\mathrm{Reg}_1(T)$ can be closed when that the rewards from the items are correlated, as in the lower bound for the unconstrained combinatorial bandits in Kveton et al. (2015). This assumption also allows us to remove a factor of K from the gap of $\mathrm{Reg}_2(T)$.

In terms of the problem-independent bounds in Corollary 5.4, the regret due to suboptimality is almost tight as in the unconstrained case (Kveton et al., 2015). The regret due to safeness checking is tight up to a factor of K^2. During each phase, PASCOMBUCB selects and samples solutions which are disjoint subsets of A_p, and hence one item is sampled at most once during one phase. However, it is empirically feasible to sample some items more than once during one phase, which will help reduce the regret but requires more delicate analysis.

For future directions, it is of interest to close the gap (the factor K) in the regret due to safeness checking with improved analyses or additional assumptions on the instance.

E.2. Discussions on the Problem Formulation

Anytime safety is important in safety-critical applications where *at each point in time* the risk cannot exceed a certain threshold. For example, in a self-driving car that is scheduled to move from start point x_0 to end point x_n via $(x_1, x_2, \ldots, x_{n-1})$ (the choice of these waypoints is a combinatorial problem), it is necessary that the car stay in its designated lane at all points in time, and not just "on average", otherwise a catastrophic accident might result at some point in time with non-negligible probability. In this example, we might want to choose a route that is safe at all times w.h.p. (in the sense that the car stays in its designated lane) at the cost of a longer travel time.

This work studies the anytime-safe constraint at the *super-arm*[3] level and provides a first step to understanding risk in combinatorial semi-bandits. The *sum* $\sum_{i\in A}\sigma_i^2$ of a set of items (base arms) in A is adopted as the risk measure, which is a

[3] In this discussion, the terms "super arm" and "base arm" refer to "solution" and "item" respectively.

certain function of $\sigma_i^2, i \in A$. It is also of interest to study the anytime safety at the *base arm level*, where the risk function is $\max_{i \in A} \sigma_i^2$. From a technical point of view, if the safeness of any single base arm has not been ascertained (as we need to learn the safeness/risk of the base arms), then pulling any base arm can be risky, in the sense that with non-vanishing probability, the anytime-safe constraint (or even the less stringent stagewise safety constraint (Khezeli & Bitar, 2020)) is violated when we do the exploration (by pulling any base arm) *at the beginning*. Thus, this seems infeasible from a technical standpoint. Nevertheless, we believe additional proper assumptions can be made to formulate a practical and feasible model that leads to future researches.

E.3. Comparisons

Comparison with Wu et al. (2016); Kazerouni et al. (2017): While the conservative (linear) constraint

$$\mathbb{P}\left[\bigcap_{t \geq 1}\left\{\sum_{k=1}^{t}\langle X_k, \theta^*\rangle \geq (1-\alpha)tb_0\right\}\right] \geq 1-\delta$$

requires the constraint should be satisfied w.p. $1-\delta$ over the whole horizon, which is similar to our probably anytime-safe constraint, the constraint is in terms of the *cumulative* expected reward (up to time step t). The cumulative nature of the conservative constraint maintains a "budget reservoir" that makes this constraint less stringent than the stagewise safety constraint (Khezeli & Bitar, 2020), in the sense that one algorithm may satisfy the conservative constraint but violate the stagewise safety constraint. Both the stagewise safety constraint and the probably anytime-safe constraint consider the reward/risk that is incurred at each *single* time step.

Comparison with Khezeli & Bitar (2020); Moradipari et al. (2020): To the best of our knowledge, the *stagewise safety* (or the *stagewise conservative)* constraint (Khezeli & Bitar, 2020; Moradipari et al., 2020) is the most related risk-aware constraint to our anytime-safe constraint.
(1) The stagewise conservative constraint is *a constraint on the mean reward* (hence, only one statistics is involved in the problem), which originates from the conservative constraint (Wu et al., 2016). In our setup, we post the anytime-safe constraint on the risk while minimizing the regret, which requires us to consider *two statistics and the interaction between them*.
(2) The stagewise safety constraint has only been utilized under the linear bandit setup in the literature, where the arm set is assumed to be a *convex and compact* set in $\mathbb{R}^n$ (Moradipari et al., 2020), and thus, the arms constitute an *uncountable continuous* set. If an arm A is known to be safe, then it is safe to pull any arm "near" A. However, in the combinatorial bandit setup, such a continuity property of the arm set does not hold since it is "discrete". Specifically, given that super arm A is safe (but not absolutely safe), even the safeness of a nearby arm $\tilde{A}$, which is obtained by replacing one single base arm in A with another base arm, cannot be guaranteed by the safeness of A.
(3) In terms of the safety level, consider the stagewise safety constraint with a constant confidence parameter δ (independent of T); intuitively, the safety constraint can be violated approximately δT times, which is *linear* in T. In addition, consider an algorithm which does safeness checking first, followed by exploration-and-exploitation on the safe super arms, it takes $\Theta(\frac{1}{(\Delta_A^{\mathrm{Y}})^2}\ln\frac{1}{\delta})$ pulls to identify the unsafeness of an unsafe super arm A. Note that the time required is independent of T, so the regret due to safeness checking is $o(T^a)$ for all $a > 0$. From this perspective, the stagewise safety constraint is not stringent enough and can be easily satisfied by such a naive algorithm. A more direct intuition (yet not completely rigorous) is, if the algorithm ignores the stagewise safety constraint, it only takes $o(T^a)$ for all $a > 0$ to rule out the unsafe super arms, which indicates it satisfies the stagewise safety constraint (since the unsafe super arms are pulled $o(T^a) < \delta T$ times). From another point of view, given a confidence parameter δ, if the stagewise safety constraint is satisfied w.p. $1-\delta_t$ at time step t with $\sum_{t=1}^{T}\delta_t = \delta$, then the probably anytime-safe constraint is met w.p. $1-\delta$.
(4) Besides the constant confidence parameter δ, we have investigated the *whole spectrum* of δ in terms of the time horizon T. The tightness result (Corollary D.1) indicates our algorithm is capable of dealing with *an even stricter constraint* (in the sense that δ decreases with respect to T) and we provide a sharp threshold on the achievability of $o(T^a)$ (for all $a > 0$) regret (see Theorem 4.5).