informs
https://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH

Vol. 00, No. 0, Xxxxx 0000, pp. 000–000

Submitted to *Operations Research*

ISSN 0030-364X, EISSN 1526-5463

Module design for pilots training with full flight simulator

Funding: This research was supported by [grant number, funding agency].

1. Background

Full flight simulator (FFS) is a sophisticated simulation system to simulate all aircraft systems that are accessible from the flight deck and are critical to training. For instance, it can simulate the force feedback for the pilot's flight controls, the avionics system, the communication system, the cockpit sounds, the aerodynamics, and the ground handling. It prepares pilots for realistic flight situations and is used for pilots training.

FFS has extremely high fidelity and is typically expensive the run. This will limit the number of sessions each pilot can try with FFS. Apart from initial training, the pilots must carry out recurrent training at regular intervals (such as every six months) in order to retain their qualification. In addition, pilots' reactions to different situations and flight skills are different. Therefore, it is important to design efficient personal training sessions for pilots to detect and improve their inadequacies.

2. Problem Formulation

We aim to design training modules with simulation-based optimization approach. Suppose there are M different training modules from which m are selected in one training session due to the time and cost restriction. Our purpose is to identify the weakness of the pilot with a carefully designed training session. The optimization problem can be formulated as follows:

$$\min_{x_1,...,x_m} \mathbb{E}_\xi[\min\{Y(x_1,\xi),...,Y(x_m,\xi)\}], \tag{1}$$

where $x_1,...,x_m$ represent the m modules in the training session, taken from the module set $\mathcal{X} = \{x_1,...,x_M\}$ and $Y(x_1,\xi)$ is the score the pilot obtain in module x_i. It is evaluated through simulation at FFS. In this model, we assume the score vector in the population $\{Y(x_1),...,Y(x_M)\}$ follows a multivariate normal distribution. For each specific person, his score is a sample from this distribution. In equation (1), the random vector ξ is used to illustrate the randomness in this multivariate normal distribution.

The inner minimization is to find the minimum score one pilot obtain across the given m training modules. We take the minimum value as we need to find the weakness of the pilots considering flight safety. The outer minimization is the find the best set of training sessions in detecting the pilot inadequacy. This formulation is not for 'personal' module design, as we choose the modules that performs the best in expectation over the whole population.

This problem involves two stages. The first stage is to learn the multivariate normal distribution as well as to identify the optimal modules. This is the simulation optimization process. In the two stage, we deploy the recommended session to train the pilots.

3. Methodology

We assume that the score has Gaussian noise: $Y(x) = Z(x) + \mathcal{N}(0,\sigma(x))$ and the joint distribution for $Y(x)_{x\in\{\mathcal{X}\}}$ is a multivariate normal distribution with mean vector $\theta = [Z(x_1),...,Z(x_M)]^T$ and covariance matrix Λ. We use a Bayesian approach to model θ and assume a normal prior:

$$\theta \sim \mathcal{N}(\mu_0,\Sigma_0).$$

As with Xie et al. (2016), we assume Λ, μ_0, Σ_0 are known.

The ith entry of a vector v is denoted as $v(i)$ and the (i,j)th entry of a matrix M is denoted as $M(i,j)$. For an ordered collection of m alternatives $\mathbf{x} = (x_1,...,x_m)$ with element $x_i \in \{1,...,M\}$ for

each i, we use $v(\mathbf{x})$ to denote the subvector of v with the ith entry equal to $v(x_i)$. We further denote by $M(\mathbf{x}, \mathbf{x}')$ the m-by-m submatrix of M with the (i, j)th entry equal to $M(x_i, x'_j)$.

We consider a situation where in each iteration n, one pilot will attend m different modules $\mathbf{x}_n = (x_{n,1}, ..., x_{n,m})^T$ and we obtain his score vector $\mathbf{y}_n = (Y(x_{n,1}), ..., Y(x_{n,m}))^T$. The conditional distribution of $\mathbf{y}_n$ is:

$$\mathbf{y}_n|\theta, \mathbf{x}_n \sim \mathcal{N}(\theta(\mathbf{x}_n), \Lambda(\mathbf{x}_n, \mathbf{x}_n)).$$

Let $\mathbb{X}_n = (\mathbf{x}_1^T, ..., \mathbf{x}_n^T)$ denote the concatenation of the design points of the previous n iteration and similarly $\mathbb{Y}_n = (\mathbf{y}_1^T, ..., \mathbf{y}_n^T)^T$. Then, the posterior distribution for θ is:

$$\theta_n|\mathbb{X}_n, \mathbb{Y}_n \sim \mathcal{N}(\mu_n, \Sigma_n),$$

where for any vector $\mathbf{x} = (x_1, x_2, ..., x_m)$,

$$\mu_n(\mathbf{x}) = \mu_0(\mathbf{x}) + \Sigma_0(\mathbf{x}, \mathbb{X}_\mathbf{n})(\mathbf{\Sigma_0}(\mathbb{X}_\mathbf{n}, \mathbb{X}_\mathbf{n}) + \mathbf{\Gamma_n})^{-1}(\mathbb{Y}_\mathbf{n} - \mu_\mathbf{0}(\mathbb{X}_\mathbf{n})),$$

$$\Sigma_n(\mathbf{x}, \mathbf{x}) = \Sigma_0(\mathbf{x}, \mathbf{x}) - \Sigma_0(\mathbf{x}, \mathbb{X}_\mathbf{n})(\mathbf{\Sigma_0}(\mathbb{X}_\mathbf{n}, \mathbb{X}_\mathbf{n}) + \mathbf{\Gamma_n})^{-1}\mathbf{\Sigma_0}(\mathbb{X}_\mathbf{n}, \mathbf{x}),$$

where Γ_n is the block diagonal matrix with n blocks: $\Lambda(\mathbf{x}_1, \mathbf{x}_1), ..., \Lambda(\mathbf{x}_n, \mathbf{x}_n)$.

We adopt the Expected Improvement (EI) acquisition function to select the m courses for the next iteration $n + 1$. For any candidate vector $\mathbf{x} = (x_1, ..., x_m)$,

$$\text{EI}(\mathbf{x}) = \mathbb{E}_{\theta_n}[(g_c - G(\mathbf{x}))^+|\theta_n],$$

where $(a)^+ = a$ if $a \geq 0$ and $(a)^+ = 0$ otherwise. When θ_n takes a value $\tilde{\theta_n}$, we define $G(\mathbf{x})$ as:

$$G(\mathbf{x}) = \mathbb{E}_\xi[\min\{Y(x_1, \xi), ..., Y(x_m, \xi)\}],$$

where $\{Y(x_1), ..., Y(x_m)\} \sim \mathcal{N}(\tilde{\theta_n}(\mathbf{x}), \Lambda(\mathbf{x}, \mathbf{x}))$ and $\{Y(x_1, \xi), ..., Y(x_m, \xi)\}$ is a random sample from this distribution. In the EI function, g_c is the current best value: $g_c = \min\{G(\mathbf{x}_1), ..., G(\mathbf{x}_n)\}$.

We next explain how to compute g_c. For any vector $\mathbf{x}_i, 1 \leq i \leq n$, we approximate $G(\mathbf{x}_i)$ through sample average approximation (SAA). Specifically, we generate samples of $\{Y(x_{i,1}, \xi), ..., Y(x_{i,m}, \xi)\}$ from distribution $\mathcal{N}(\mu_n(\mathbf{x}_i), \Lambda(\mathbf{x}_i, \mathbf{x}_i))$. Here, as we already have observations at $\mathbf{x}_i$, we use the posterior mean $\mu_n(\mathbf{x}_i)$ as if it were the true value of $\theta_n(\mathbf{x}_i)$ to compute the current best value. Similar approach has been widely used for ordinary stochastic GP based optimization algorithms. The samples can then be generated as follows. Suppose $\Lambda(\mathbf{x}_i, \mathbf{x}_i) = B_i B_i^T$ and $\mathbf{z}_j \in \mathbb{R}^{m\times 1}$ is a random draw from m iid standard normal distribution. The j-th random samples can

be represented as: $(Y(x_{i,1},\xi_j),...,Y(x_{i,m},\xi_j))^T = \mu_n(\mathbf{x}_i) + B_i\mathbf{z}_j$ and $Y(x_{i,k},\xi_j) = \mu_n(x_{i,k}) + B_i^k\mathbf{z}_j$, where B_i^k is the kth row of B_i. Therefore, we have

$$G(\mathbf{x}_i) \approx \frac{1}{J}\sum_{j=1}^{J} \min \mu_n(\mathbf{x}_i) + B_i\mathbf{z}_j = \frac{1}{J}\sum_{j=1}^{J} \min\{\mu_n(x_{i,1}) + B_i^1\mathbf{z}_j, ..., \mu_n(x_{i,m}) + B_i^m\mathbf{z}_j\}.$$

To compute EI for $\mathbf{x}$, we should further take care of the outer expectation with respect to the posterior distribution of θ. We can use similar approach as above the generate samples from the posterior distribution of $\theta_n(\mathbf{x})$ and obtain the following SAA form:

$$\mathrm{EI}(\mathbf{x}) \approx \frac{1}{K}\sum_{k=1}^{K}(g_c - \frac{1}{J}\sum_{j=1}^{J} \min \theta_n(\mathbf{x}) + A\tilde{\mathbf{z}}_k + B\mathbf{z}_j)^+,$$

where $\Sigma_n(\mathbf{x},\mathbf{x}) = AA^T$, $\Lambda(\mathbf{x},\mathbf{x}) = BB^T$, and both $\tilde{\mathbf{z}}_k$ and $\mathbf{z}_j$ are iid standard normal vectors of length m.

4. Some special cases

4.1. Two normal variables

We need to compare $\mathrm{E}_{\xi_1,\xi_2}[\min\{\mu_1 + A\xi_1, \mu_2 + B_1\xi_1 + B_2\xi_2\}]$ and $\mathrm{E}_{\xi_1,\xi_2}[\min\{\mu_1 + A\xi_1, \mu_2 + B\xi_2\}]$, where ξ_1, ξ_2 are independent standard normal random variable. All the coefficients are positive and that $B_1^2 + B_2^2 = B^2$.

Notice that for any two Gaussian random number $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ with correlation ρ, the expectation $\mathrm{E}[\min\{X_1, X_2\}]$ takes the following form (Clark 1961):

$$\mathrm{E}[\min\{X_1, X_2\}] = \mu_1\Phi(\frac{\mu_2 - \mu_1}{\theta}) + \mu_2\Phi(\frac{\mu_1 - \mu_2}{\theta}) - \theta\phi(\frac{\mu_2 - \mu_1}{\theta}),$$

where ϕ, Φ are cdf and pdf for standard normal distribution, respectively, and $\theta = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} = \sqrt{\mathrm{var}(X_1 - X_2)}$.

We define $X_1 = \mu_1 + A\xi_1 \sim N(\mu_1, A^2)$, $X_2 = \mu_2 + B_1\xi_1 + B_2\xi_2 \sim N(\mu_2, B_1^2 + B_2^2)$, $X_3 = \mu_2 + B\xi_2 \sim N(\mu_2, B^2)$. Thus,

$$\mathrm{cov}(X_1, X_2) = AB_1, \mathrm{cov}(X_1, X_3) = 0.$$

$$\mathrm{var}(X_1 - X_2) = A^2 + B_1^2 + B_2^2 - 2AB_1, \mathrm{var}(X_1 - X_3) = A^2 + B^2.$$

Therefore,

$$\mathrm{E}[\min\{X_1, X_2\}] = \mu_1\Phi(\frac{\mu_2 - \mu_1}{\theta_1}) + \mu_2\Phi(\frac{\mu_1 - \mu_2}{\theta_1}) - \theta_1\phi(\frac{\mu_2 - \mu_1}{\theta_1}),$$
$$\mathrm{E}[\min\{X_1, X_3\}] = \mu_1\Phi(\frac{\mu_2 - \mu_1}{\theta_2}) + \mu_2\Phi(\frac{\mu_1 - \mu_2}{\theta_2}) - \theta_2\phi(\frac{\mu_2 - \mu_1}{\theta_2}),$$

where $\theta_1 = \sqrt{A^2 + B_1^2 + B_2^2 - 2AB_1}$, $\theta_2 = \sqrt{A^2 + B^2}$. Denote:

$$f(\theta) = \mu_1\Phi(\frac{\mu_2-\mu_1}{\theta}) + \mu_2\Phi(\frac{\mu_1-\mu_2}{\theta}) - \theta\phi(\frac{\mu_2-\mu_1}{\theta}) = (\mu_1-\mu_2)\Phi(\frac{\mu_2-\mu_1}{\theta}) + \mu_2 - \theta\phi(\frac{\mu_2-\mu_1}{\theta}).$$

We have

$$\begin{aligned} f'(\theta) &= (\mu_1-\mu_2)\phi(\frac{\mu_2-\mu_1}{\theta})\frac{\mu_2-\mu_1}{-\theta^2} - \phi(\frac{\mu_2-\mu_1}{\theta}) - \theta\phi(\frac{\mu_2-\mu_1}{\theta})\frac{\mu_2-\mu_1}{-\theta}\frac{\mu_2-\mu_1}{-\theta^2} \\ &= -\phi(\frac{\mu_2-\mu_1}{\theta}) < 0 \end{aligned}$$

Therefore, f_θ is a decreasing function w.r.t. θ. We have the following conclusions:

1. When $B_1^2 + B_2^2 = B^2$ and all coefficients are positive, we have $\theta_1 = \sqrt{A^2 + B^2 - 2AB_1} < \theta_2$. Hence, $\mathrm{E}[\min\{X_1, X_2\}] > \mathrm{E}[\min\{X_1, X_3\}]$.

2. When $B_1^2 + B_2^2 = B^2$ and $B_1 < 0$, we have $\theta_1 = \sqrt{A^2 + B^2 - 2AB_1} > \theta_2$. Hence, $\mathrm{E}[\min\{X_1, X_2\}] < \mathrm{E}[\min\{X_1, X_3\}]$.

Another way of presenting these results. Suppose $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ with correlation ρ and $\theta = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} = \sqrt{\mathrm{var}(X_1 - X_2)}$. Denote $f = \mathrm{E}[\min\{X_1, X_2\}]$. We have:

$$\frac{\partial f}{\partial \theta} = -\phi(\frac{\mu_2-\mu_1}{\theta}) < 0, \tag{2}$$

$$\frac{\partial f}{\partial \rho} = \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial \rho} = \frac{\sigma_1\sigma_2}{\theta}\phi(\frac{\mu_2-\mu_1}{\theta}) > 0, \tag{3}$$

$$\frac{\partial f}{\partial \sigma_1} = \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial \sigma_1} = -\frac{\sigma_1 - \rho\sigma_2}{\theta}\phi(\frac{\mu_2-\mu_1}{\theta}), \tag{4}$$

$$\frac{\partial f}{\partial \mu_1} = \Phi(\frac{\mu_2-\mu_1}{\theta}) > 0. \tag{5}$$

We summarize the conclusions as follows:

1. From (3) and (5), we observe that variables with smaller correlations and expectations are preferred.

2. The relation between the objective function and the variance for each individual variable is not monotone. We can, however, derive some special cases:

(a) When $\rho < \frac{\sigma_1}{\sigma_2}$ (specifically when $\rho < 0$), a larger value of σ_1 is preferred.

(b) When $\rho > \frac{\sigma_1}{\sigma_2}$, smaller value of σ_1 is preferred. This seems a little bit counter-intuitive. We can consider some special situation to understand this. Consider when $\rho = 1$ and $\sigma_1 < \sigma_2$. In this situation, X_1 and X_2 are positively linear dependent: $X_1 = \frac{\sigma_1}{\sigma_2}X_2$. We can find that they always have the same sign and when X_1 and $X_2 > 0$, $\min\{X_1, X_2\} = X_1 > 0$, which increases with σ_1. When X_1

and $X_2 < 0$, $\min\{X_1, X_2\} = X_2 < 0$, which does not change with σ_1. Therefore, in this case, a larger value of σ_1 will increase the objective value. Although a larger value of σ_1 will reduce X_1 when it is negative, these negative values, however, are not counted in the objective function as long as $\sigma_1 < \sigma_2$. This situation will slightly gets better when ρ deviates from 1, but still makes less negative values of X_1 to be counted in the objective function until $\rho < \frac{\sigma_1}{\sigma_2}$.

4.2. Several normal variables

We consider the following situations: m normal random numbers $X_1, ..., X_m$ with equal mean μ and equal variance σ^2. The correlation between each pair of variables is $\rho \geq 0$. We would like to explore the relationship between the objective function $f = \mathrm{E}[\min\{X_1, ..., X_m\}]$ and ρ, μ, and σ.

In this case, each random variable can be represented as $X_i = \mu + \sigma(\sqrt{\rho}W + \sqrt{1-\rho}N_i)$, where W and $N_1, ..., N_m$ are i.i.d. standard normal random variables. Then,

$$\begin{aligned} f = \mathrm{E}[\min\{X_1, ..., X_m\}] &= \mathrm{E}[\min\{\mu + \sigma(\sqrt{\rho}W + \sqrt{1-\rho}N_1), ..., \mu + \sigma(\sqrt{\rho}W + \sqrt{1-\rho}N_m)\}] \\ &= \mu + \sigma\mathrm{E}[\min\{\sqrt{\rho}W + \sqrt{1-\rho}N_1, ..., \sqrt{\rho}W + \sqrt{1-\rho}N_m\}] \\ &= \mu + \sigma\mathrm{E}[\sqrt{\rho}W + \sqrt{1-\rho}\min\{N_1, ..., N_m\}] = \mu + \sigma\sqrt{1-\rho}\mathrm{E}[\min\{N_1, ..., N_m\}]. \end{aligned}$$

This had no analytical solution. Some special cases:

1. When $\rho = 1$, $f = \mu$.
2. When $\rho = 0$, $f = \mu + \sigma\mathrm{E}[\min\{N_1, ..., N_m\}]$.

$\mathrm{E}[\min\{N_1, ..., N_m\}]$ is a constant for a given m. It is negative and can be approximated as

$$-\sqrt{2\ln m} + \frac{\ln\ln m + \ln 4\pi}{2\sqrt{2\ln m}}.$$

Hence, $f \approx \mu - \sigma\sqrt{1-\rho}\sqrt{2\ln m}$.

We can see that in this case, smaller μ, larger σ and smaller ρ are preferred. Intuitively, when ρ decreases, the m variables are less correlated and thus each of them has larger probability to take smaller values more independently. Therefore, the objective value increases as ρ.

5. Simulations with known θ and Λ

5.1. SAA approach

We first demonstrate how to identify the optimal module set $\mathbf{x} = (x_1, ..., x_m)$ when the parameters θ and Λ are known. Consider selecting 8 modules from 50 candidates to minimize the expected minimum score. Building upon our earlier discussion of Sample Average Approximation (SAA), we propose a sequential selection procedure:

1. Let $x_1 = \arg\min_{x\in\mathcal{X}} \theta(x)$, get the collection of selected modules $S = \{x_1\}$
2. Let $x_2 = \arg\min_{x\in\mathcal{X}/\{x_1\}} \mathrm{E}[\min\{x_1, x\}]$, update the collection $S = \{x_1, x_2\}$
3. Let $x_3 = \arg\min_{x\in\mathcal{X}/S} \mathrm{E}[\min\{x_1, x_2, x\}]$, update the collection $S = \{x_1, x_2, x_3\}$
4. …keep doing this until we get the collection $S = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$
5. Output $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ as the result

While this approach provides accurate results, the computational cost of repeated SAA evaluations becomes prohibitive for large-scale problems. This motivates our alternative approximation method presented next.

5.2. Improvement-based approach

We define the improvement of module x_2 over module x_1 as $\mathrm{Ipr}(x_2, x_1)$:

$$\mathrm{Ipr}(x_2, x_1) = \mathrm{E}[\min\{x_1, x_2\}] - \mathrm{E}[x_1]$$

Obviously, when x_1 and x_2 are exactly the same, $\mathrm{Ipr}(x_2, x_1) = 0$. Next, suppose we have a set of modules $S = \{x_1, x_3, x_4\}$, we can samely define the improvement of module x_2 over set S as $\mathrm{Ipr}(x_2, S)$:

$$\mathrm{Ipr}(x_2, S) = \mathrm{E}[\min\{S, x_2\}] - \mathrm{E}[\min\{S\}]$$

Now let's consider an extreme case, assuming x_1 and x_2 are exactly the same (i.e. $\mathrm{Ipr}(x_2, x_1) = 0$), regardless of the value of $\mathrm{Ipr}(x_2, x_3)$ and $\mathrm{Ipr}(x_2, x_4)$, the value of $\mathrm{Ipr}(x_2, S)$ is always 0. This inspires us to speculate that the value of the improvement of module x_2 over set S is dominated by pairwise minima:: $\min_{x\in S} \mathrm{Ipr}(x_2, x)$. So we can define the contribution of module x_2 to set S:

$$\mathrm{Ctr}(x_2, S) = \min_{x\in S} \mathrm{Ipr}(x_2, x)$$

Here, the contribution measure $\mathrm{Ctr}(x_2, S)$ represents quantifies the reduction potential in the minimum score when adding module x to set S.

Based on this we can get our improvement-based approach as follows:

1. Let $x_1 = \arg\min_{x\in\mathcal{X}} \theta(x)$, get the collection of selected modules $S = \{x_1\}$
2. Let $x_2 = \arg\min_{x\in\mathcal{X}/\{x_1\}} \mathrm{E}[\min\{x_1, x\}]$, update the collection $S = \{x_1, x_2\}$ (Step 1.and 2. are exactly the same as SAA approach)
3. Let $x_3 = \arg\max_{x\in\mathcal{X}/S} \mathrm{Ctr}(x, S)$, update the collection $S = \{x_1, x_2, x_3\}$
4. Let $x_4 = \arg\max_{x\in\mathcal{X}/S} \mathrm{Ctr}(x, S)$, update the collection $S = \{x_1, x_2, x_3, x_4\}$

5. …keep doing this until we get the collection $S = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$
6. Output $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ as the result.

The $\min_{y \in S} \mathrm{Ipr}(x, y)$ in the above process can be quickly calculated using the formula in 4.1, which saves a lot of computing resources compared to SAA.

5.3. Mixed approach

Although the improvement-based approach offers superior computational efficiency, its solution quality may be suboptimal compared to the SAA method. To leverage the strengths of both approaches, we propose a mixed selection strategy that combines their advantages. The algorithm proceeds iteratively as follows:

1. Let $x_1 = \arg\min_{x \in \mathcal{X}} \theta(x)$, get the collection of selected modules $S = \{x_1\}$
2. Let $x_2 = \arg\min_{x \in \mathcal{X}/\{x_1\}} \mathrm{E}[\min\{x_1, x\}]$, update the collection $S = \{x_1, x_2\}$ (Step 1.and 2. are exactly the same as SAA approach)
3. Find set $C \subseteq S$ s.t. $C = \arg\max_{|C|=10} \sum_{x \in C} \mathrm{Ctr}(x, S)$. Using SAA approach to get $x_3 = \arg\min_{x \in C} \mathrm{E}[\min\{x_1, x_2, x\}]$, update set $S = \{x_1, x_2, x_3\}$.
4. Find set $C \subseteq S$ s.t. $C = \arg\max_{|C|=10} \sum_{x \in C} \mathrm{Ctr}(x, S)$. Using SAA approach to get $x_4 = \arg\min_{x \in C} \mathrm{E}[\min\{x_1, x_2, x_3, x\}]$, update set $S = \{x_1, x_2, x_3, x_4\}$.
5. …keep doing this until we get the collection $S = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$
6. Output $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ as the result.

The proposed mixed approach achieves superior solution quality compared to the pure improvement-based approach, while maintaining significantly lower computational complexity than the full SAA implementation. Our numerical experiments in Section 5.5 demonstrate its effectiveness across various test scenarios, showing consistent performance improvements over both baseline methods. The mixed method's balanced trade-off between accuracy and efficiency makes it particularly suitable for large-scale training module selection problems.

5.4. Heristical approaches

Here we give three other heristical approaches:

1. **Smallest** : Directly select the 8 modules with the smallest mean as the output result
2. **Cluster** : First divide the modules into 8 categories according to the correlation matrix by hierarchical clustering, and then the module with the smallest mean value in each category was selected as the output result.

3. **LCB_max** : Similar to Improvement-based approach, multi-objective optimization and LCB are used to filter and then use SAA approach to choose from the remaining modules.

5.5. simulation experiment

In our experimental setup, we maintain the selection of 8 modules from a pool of 50 candidates. We evaluate and compare the performance and computational efficiency of each approach across three distinct scenarios:

1. **Regulargrouping** : The 50 modules can be divided into 5 groups of 10 modules each. the correlation between modules within a group is high, while there is no correlation between groups.

2. **Centralizedminimum** : Similar to Regular grouping, the difference is that the modules with the smallest mean in this case are clustered in the same group.

3. **Random** : The correlation between individual modules is completely randomized.

For each scenario, we generate 100 random instances and report average performance metrics and computation times in the following tables:

Table 1 Regular grouping

Total modules	Tested modules	Algorithm Name	Avg Rank	Avg Result	Avg Time (s)
50	8	smallest	3.720000	64.067880	0.000032
50	8	cluster	4.800000	64.313566	0.000456
50	8	LCB_max	6.000000	66.164651	1.511647
50	8	improvement	2.540000	63.899411	0.053409
50	8	SAA	2.060000	63.860563	7.538640
50	8	mixed	1.880000	63.849019	1.714283

Table 2 Centralized minimum

Total modules	Tested modules	Algorithm Name	Avg Rank	Avg Result	Avg Time (s)
50	8	smallest	5.750000	68.599959	0.000026
50	8	cluster	4.060000	66.970837	0.000332
50	8	LCB_max	5.130000	67.868257	1.601738
50	8	improvement	2.590000	66.439135	0.075206
50	8	SAA	1.790000	66.348565	7.688974
50	8	mixed	1.680000	66.343133	1.774847

Table 3 Random

Total modules	Tested modules	Algorithm Name	Avg Rank	Avg Result	Avg Time (s)
50	8	smallest	3.780000	52.630880	0.000025
50	8	cluster	4.880000	53.525543	0.000328
50	8	LCB_max	5.980000	56.080551	1.519013
50	8	improvement	3.030000	52.415651	0.076352
50	8	SAA	1.630000	52.145339	7.808298
50	8	mixed	1.700000	52.139484	1.794801

Our experimental results demonstrate that both the SAA and mixed approaches achieve nearly identical performance, with their results being substantially superior to all other methods considered. While we present detailed comparisons for the case of selecting 8 modules from 50 due to space limitations, these findings consistently hold across various selection sizes.

6. Convergence analysis

LEMMA 1. *For any $\epsilon > 0$ and any $n = T_{i,t-1} \in \mathbb{Z}_+$, we have :*

$$\Pr[|\mu_{t-1}(i) - \mu(i)| \geq \epsilon_i] \leq \exp(-\frac{n\epsilon_i^2}{2\sigma^2(i)})$$

Proof. Fix $t \geq 1$ and $i \in [m]$. Conditioned on $(\mathbf{y}_1, ..., \mathbf{y}_{t-1}), \{S_1, ..., S_{t-1}\}$ are deterministic, and $\mu(i) \sim N(\mu_{t-1}(i), \sigma_{t-1}^2(i))$. Now, if $r \sim N(0,1)$,then:

$$\begin{aligned}\Pr\{r > c\} &= e^{-c^2/2}(2\pi)^{-1/2} \int e^{-(r-c)^2/2 - c(r-c)} dr \\ &\leq e^{-c^2/2} \Pr\{r > 0\} = (1/2)e^{-c^2/2}\end{aligned}$$

for $c > 0$, since $e^{-c(r-c)} \leq 1$ for $r \geq c$. Therefore, $\Pr\{|\mu_{t-1}(i) - \mu(i)| > \lambda_i \sigma_{t-1}(i)\} \leq e^{-\lambda_i/2}$

Assume that arm i has been observed n times, yielding n observations. Further assume that these n observations are our only observational data. Using GP modeling, we obtain:

$$\sigma_n^2(i) = k(i,i) - \mathbf{k}_n^\top(i)(\mathbf{K}_n + \sigma^2(i)\mathbf{I}_n)^{-1}\mathbf{k}_n(i)$$

where:

- $\mathbf{k}_n(i) = [k(i,i), \ldots, k(i,i)]^\top \in \mathbb{R}^n$
- $\mathbf{K}_n = \sigma_0^2 \mathbf{1}_n \mathbf{1}_n^\top$ $(\sigma_0^2 = k(i,i))$

For $\mathbf{A} = \sigma^2(i)\mathbf{I}_n$, $\mathbf{u} = \sigma_0^2 \mathbf{1}_n$, $\mathbf{v} = \mathbf{1}_n$, by **Sherman-Morrison** formula:

$$\begin{aligned}(\mathbf{K}_n + \sigma^2(i)\mathbf{I}_n)^{-1} &= \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^\top\mathbf{A}^{-1}}{1 + \mathbf{v}^\top\mathbf{A}^{-1}\mathbf{u}} \\ &= \frac{1}{\sigma^2(i)}\mathbf{I}_n - \frac{\sigma_0^2/\sigma(i)^4 \cdot \mathbf{1}_n\mathbf{1}_n^\top}{1 + n\sigma_0^2/\sigma^2(i)}\end{aligned}$$

So we have:

$$\mathbf{k}_n^\top(i)(\mathbf{K}_n + \sigma^2(i)\mathbf{I}_n)^{-1}\mathbf{k}_n(i) = \frac{\sigma_0^4 n}{\sigma^2(i) + n\sigma_0^2}$$

Substitute the results into the posterior variance formula:

$$\begin{aligned}\sigma_n^2(i) &= \sigma_0^2 - \frac{\sigma_0^4 n}{\sigma^2(i) + n\sigma_0^2} = \frac{\sigma_0^2\sigma^2(i)}{\sigma^2(i) + n\sigma_0^2} \\ &= \frac{n\sigma_0^2}{\sigma^2(i) + n\sigma_0^2} \cdot \frac{\sigma^2(i)}{n} \leq \frac{\sigma^2(i)}{n}\end{aligned}$$

Considering that adding other observation points will not increase the uncertainty of arm i, so we have $\sigma_{t-1}(i) \leq \sigma_n(i) \leq \frac{\sigma(i)}{\sqrt{n}}$. Therefore:

$$\Pr\left\{|\mu_{t-1}(i)-\mu(i)|>\lambda_i\frac{\sigma(i)}{\sqrt{n}}\right\}\leq \Pr\{|\mu_{t-1}(i)-\mu(i)|>\lambda_i\sigma_{t-1}(i)\}\leq e^{-\lambda_i/2}$$

Denote $\epsilon_i=\lambda_i\frac{\sigma(i)}{\sqrt{n}}$, then we have $\Pr[|\mu_{t-1}(i)-\mu(i)|\geq\epsilon_i]\leq\exp(-\frac{n\epsilon_i^2}{2\sigma^2(i)})$ ■

LEMMA 2. *If for any $i\in[m]$ we have $\mu_1(i)\geq\mu_2(i)$ (Abbreviated as $\boldsymbol{\mu}_1\geq\boldsymbol{\mu}_2$), then for any super arm $S\in\mathcal{F}$,we have:*

$$r_{P_1}(S)\geq r_{P_2}(S)$$

LEMMA 3. *If for any $i\in[m]$ we have $\mu_1(i)-\mu_2(i)\leq\Lambda_i(\Lambda_i>0)$ ($\boldsymbol{\mu}_1\geq\boldsymbol{\mu}_2$). For any super arm $S\in\mathcal{F}$, let $\Lambda_S=\max_{i\in S}\Lambda_i$, then we have:*

$$r_{P_1}(S)-r_{P_2}(S)\leq\Lambda_S$$

Proof. For any super arm $S=\{i_1,...,i_n\}$, denote the corresponding mean vectors in distributions P_1 and P_2 as $\boldsymbol{\mu}_{1,S}=[\mu_1(i_1),...,\mu_1(i_n)]^T$, $\boldsymbol{\mu}_{2,S}=[\mu_1(i_2),...,\mu_2(i_n)]^T$. Note that P_1 and P_2 are both multivariate normal distributions, and have same covariance matrices Σ_S.

Now, let $\mathbf{z}=[z_1,...,z_n]^T\sim N(0,\Sigma_S)$. We have:

$$r_{P_1}(S)=\int\cdots\int_{\mathbf{z}}\min\{\mu_1(i_1)+z_1,\mu_1(i_2)+z_2,\ldots,\mu_1(i_n)+z_n\}\,dz_1\ldots dz_k$$

$$r_{P_2}(S)=\int\cdots\int_{\mathbf{z}}\min\{\mu_2(i_1)+z_1,\mu_2(i_2)+z_2,\ldots,\mu_2(i_n)+z_n\}\,dz_1\ldots dz_k$$

Denote $f(\boldsymbol{\mu}_{1,S}+\mathbf{z})=\min\{\mu_1(i_1)+z_1,\mu_1(i_2)+z_2,\ldots,\mu_1(i_n)+z_n\}$ and $f(\boldsymbol{\mu}_{2,S}+\mathbf{z})=\min\{\mu_2(i_1)+z_1,\mu_2(i_2)+z_2,\ldots,\mu_2(i_n)+z_n\}$. We want to show that $f(\mu_{1,S}+\mathbf{z})-f(\mu_{2,S}+\mathbf{z})\leq\Lambda_S$.

If there exists $i,j\in S=\{i_1,...,i_n\}$ s.t. $\mu_1(i)+z_i=f(\mu_{1,S}+\mathbf{z})$, $\mu_2(j)+z_j=f(\mu_{2,S}+\mathbf{z})$, $f(\mu_{1,S}+\mathbf{z})-f(\mu_{2,S}+\mathbf{z})>\Lambda_S$. Since $\mu_1(i)+z_i=f(\mu_{1,S}+\mathbf{z})=\min\{\mu_1(i_1)+z_1,\mu_1(i_2)+z_2,\ldots,\mu_1(i_n)+z_n\}$, we have: $\mu_1(j)+z_j>\mu_1(i)+z_i$.

Thus $\mu_1(j)-\mu_2(j)=\mu_1(j)+z_j-(\mu_2(j)+z_j)>\mu_1(i)+z_i-(\mu_2(j)+z_j)>\Lambda_S$, which is contradict with $\boldsymbol{\mu}_1\geq\boldsymbol{\mu}_2$.

Therefore

$$\begin{aligned}r_{P_1}(S)-r_{P_2}(S)&=\int\cdots\int_{\mathbf{z}}f(\boldsymbol{\mu}_{1,S}+\mathbf{z})-f(\boldsymbol{\mu}_{2,S}+\mathbf{z})\,dz_1\ldots dz_k\\&\leq\int\cdots\int_{\mathbf{z}}\Lambda_S\,dz_1\ldots dz_k=\Lambda_S\end{aligned}$$

■

LEMMA 4. *Define an event in each round $t(m+1) \le t \le T$:*

$$\mathcal{H}_t = \left\{ 0 < \Delta_{S_t} \le 2 \max_{i \in S_t} \sigma(i) \sqrt{\frac{6\ln(t)}{T_{i,t-1}}} \right\}$$

Then the α-approximation regret in T rounds is at most

$$\mathbb{E}[\sum_{t=m+1}^{T} \mathbb{I}\{\mathcal{H}_t\}\Delta_{S_t}] + (\frac{\pi^2}{6}+1)\alpha M m.$$

Proof. For each super arm S, we define:

$$\Delta_S = \max\{r_P(S) - \alpha \cdot r_P(S^*), 0\}$$

From $m+1 \le t \le T$, Define an event:

$$\varepsilon_t = \left\{ \text{there exists } i \in [m] \text{ such that } |\hat{\mu}^i_{T_{i,t-1}} - \mu^i| \ge \sigma(i) \sqrt{\frac{6\ln(t)}{T_{i,t-1}}} \right\}$$

We bound the α-approximation regret as:

$$\begin{aligned} \mathrm{Reg}_{P,\alpha}(T) &= \sum_{t=1}^{T} \mathbb{E}\left[\alpha r_P(S^*) - r_P(S_t)\right] \\ &\le \sum_{t=1}^{T} \mathbb{E}[\Delta_{S_t}] \\ &= \mathbb{E}[\sum_{t=1}^{m} \Delta_{S_t}] + \mathbb{E}[\sum_{t=m+1}^{T} \mathbb{I}\{\varepsilon_t\}\Delta_{S_t}] \\ &+ \mathbb{E}[\sum_{t=m+1}^{T} \mathbb{I}\{\neg\varepsilon_t\}\Delta_{S_t}] \end{aligned} \tag{6}$$

(a) the first term

Let $M = 100$, The first term can be trivially bounded as:

$$\mathbb{E}[\sum_{t=1}^{m} \Delta_{S_t}] \le \sum_{t=1}^{m} \alpha \cdot r_P(S^*) \le m \cdot \alpha M \tag{7}$$

(b) the second term

By Lemma 1 we know that for any $i \in [m], t \ge m+1$, denote $c_i = \sigma(i)\sqrt{\frac{6\ln(t)}{T_{i,t-1}}}$, we have:

$$\Pr[|\hat{\mu}_{t-1}(i) - \mu(i)| \ge c_i] \le \exp(-\frac{n c_i^2}{2\sigma^2(i)}) = t^{-3}$$

Therefore

$$\begin{aligned}\mathbb{E}[\sum_{t=m+1}^{T}\mathbb{I}\{\varepsilon_t\}] &\leq \sum_{t=m+1}^{T}\sum_{i=1}^{m}\sum_{l=1}^{t-1}t^{-3}\\ &\leq m\sum_{t=m+1}^{T}t^{-2}\\ &\leq \frac{\pi^2}{6}m\end{aligned} \tag{8}$$

and then the second term in (6) can be bounded as

$$\mathbb{E}[\sum_{t=m+1}^{T}\mathbb{I}\{\varepsilon_t\}\Delta_{S_t}] \leq \frac{\pi^2}{6}m\cdot(\alpha\cdot r_P(S^*)) \leq \frac{\pi^2}{6}\alpha Mm \tag{9}$$

(c) the third term

We fix $t > m$ and first asuume $\neg\varepsilon_t$ happens. For each $i \in [m]$, since $\neg\varepsilon_t$ happens, we have:

$$|\hat{\mu}_{T_{i,t-1}}(i) - \mu(i)| < c_i \quad \forall i \in [m] \tag{10}$$

Recall that in round t, the input to the oracle is $\underline{P} = \mathcal{N}(\underline{\boldsymbol{\mu}}, \Sigma)$, where the mean vector of $\underline{P}$ is:

$$\underline{\mu}(i) = \hat{\mu}(i) - c_i \quad \forall i \in [m] \tag{11}$$

From (10) and (11) we konw that $\mu(i) > \underline{\mu}(i) > \mu(i) - 2c_i$ for all $i \in [m]$. Thus, from Lemma 2 we have:

$$r_{\underline{P}}(S) \leq r_P(S) \quad \forall S \in \mathcal{F}. \tag{12}$$

and from Lemma 3 we have:

$$r_{\underline{P}}(S) \geq r_P(S) - 2\Lambda_S \quad \forall S \in \mathcal{F}. \tag{13}$$

where $\Lambda = \max_i c_i$

Also, from the fact that the algorithm chooses S_t in the t-th round, we have:

$$r_{\underline{P}}(S_t) \leq \alpha \cdot \min_{S\in\mathcal{F}} r_{\underline{P}}(S) \leq \alpha \cdot r_{\underline{P}}(S^*). \tag{14}$$

From (12),(13) and (14) we have:

$$r_P(S) - 2\Lambda_S \leq r_{\underline{P}}(S_t) \leq \alpha \cdot r_{\underline{P}}(S^*) \leq \alpha \cdot r_P(S^*) \tag{15}$$

which implies:

$$\Delta_{S_t} \leq 2\Lambda_S$$

Therefore, when $\neg\varepsilon_t$ happens, we always have $\Delta_{S_t} \leq 2 \max_{i \in S_t} c_i$.

This implies:

$$\{\neg\varepsilon_t, \Delta_{S_t} > 0\} \Longrightarrow \{0 < \Delta_{S_t} \leq 2 \max_{i \in S_t} c_i\} = \mathcal{H}_t.$$

Hence, the third term in (6) can be bounded as:

$$\begin{aligned} \mathbb{E}[\sum_{t=m+1}^{T} \mathbb{I}\{\neg\varepsilon_t\}\Delta_{S_t}] &= \mathbb{E}[\sum_{t=m+1}^{T} \mathbb{I}\{\neg\varepsilon_t.\Delta_{S_t} > 0\}\Delta_{S_t}] \\ &\leq \mathbb{E}[\sum_{t=m+1}^{T} \mathbb{I}\{\mathcal{H}_t\}\Delta_{S_t}] \end{aligned} \tag{16}$$

Finally, by combining (6),(7),(9),(16) we have:

$$\mathrm{Reg}_{P,\alpha}(T) \leq \mathbb{E}[\sum_{t=m+1}^{T} \mathbb{I}\{\mathcal{H}_t\}\Delta_{S_t}] + (\frac{\pi^2}{6} + 1)\alpha M m$$

■

Now it remains to bound $\mathbb{E}[\sum_{t=m+1}^{T} \mathbb{I}\{\mathcal{H}_t\}\Delta_{S_t}]$.

Define two decreasing sequences of positive contants:

$$1 = \beta_0 > \beta_1 > \beta_2 > \dots$$

$$\alpha_1 > \alpha_2 > \dots$$

such that $\lim_{k\to\infty} \alpha_k = \lim_{k\to\infty} \beta_k = 0$. We choose $\{\alpha_k\}$ and $\{\beta_k\}$ as in Theorem 4 of (Kveton et al. 2014), which satisfy:

$$\sqrt{6} \sum_{k=1}^{\infty} \frac{\beta_{k-1} - \beta_k}{\sqrt{\alpha_k}} \leq 1 \tag{17}$$

and

$$\sum_{k=1}^{\infty} \frac{\alpha_k}{\beta_k} < 267. \tag{18}$$

For $t \in \{m+1, ..., T\}$ and $k \in \mathbb{Z}_+$, let

$$m_{k,t} = \begin{cases} \alpha_k \left(\frac{2\sigma K}{\Delta_{S_t}}\right)^2 \cdot \ln(T) & \Delta_{S_t} > 0, \\ +\infty & \Delta_{S_t} = 0, \end{cases}$$

and

$$A_{k,t} = \{i \in S_t | T_{i,t-1} \leq m_{k,t}\}.$$

Then we define an event:

$$\mathcal{G}_{k,t} = \{|A_{k,t}| \geq \beta_k K\},$$

which means "in the $t-th$ round, at least $\beta_k K$ arms in S_t had been observed at most $m_{k,t}$ times."

LEMMA 5. *In the $t-th$ round, at least $\beta_k K$ arms in S_t had been observed at most $m_{k,t}$ times.*

Proof. Assume that $\mathcal{H}_t$ happens and that none of $\mathcal{G}_{1,t}, \mathcal{G}_{2,t}, \ldots$ happens. Then $|A_{k,t}| < \beta_k K$ for all $k \in \mathbb{Z}_+$.

Let $A_{0,t} = S_t$ and $\bar{A}_{k,t} = S_t \setminus A_{k,t}$ for $k \in \mathbb{Z}_+ \cup \{0\}$. It is easy to see $\bar{A}_{k-1,t} \subseteq \bar{A}_{k,t}$ for all $k \in \mathbb{Z}_+$. Note that $\lim_{k\to\infty} m_{k,t} = 0$. Thus there exists $N \in \mathbb{Z}_+$ such that $\bar{A}_{k,t} = S_t$ for all $k \geq N$, and then we have $S_t = \bigcup_{k=1}^{\infty} \left(\bar{A}_{k,t} \setminus \bar{A}_{k-1,t}\right)$. Finally, note that for all $i \in \bar{A}_{k,t}$, we have $T_{i,t-1} > m_{k,t}$. Therefore

$$\begin{aligned}
\sum_{i\in S_t} \frac{1}{\sqrt{T_{i,t-1}}} &= \sum_{k=1}^{\infty} \sum_{i\in \bar{A}_{k,t}\setminus\bar{A}_{k-1,t}} \frac{1}{\sqrt{T_{i,t-1}}} \leq \sum_{k=1}^{\infty} \sum_{i\in \bar{A}_{k,t}\setminus\bar{A}_{k-1,t}} \frac{1}{\sqrt{m_{k,t}}} \\
&= \sum_{k=1}^{\infty} \frac{\left|\bar{A}_{k,t} \setminus \bar{A}_{k-1,t}\right|}{\sqrt{m_{k,t}}} = \sum_{k=1}^{\infty} \frac{\left|A_{k-1,t} \setminus A_{k,t}\right|}{\sqrt{m_{k,t}}} = \sum_{k=1}^{\infty} \frac{\left|A_{k-1,t}\right| - \left|A_{k,t}\right|}{\sqrt{m_{k,t}}} \\
&= \frac{|S_t|}{\sqrt{m_{1,t}}} + \sum_{k=1}^{\infty} |A_{k,t}| \left(\frac{1}{\sqrt{m_{k+1,t}}} - \frac{1}{\sqrt{m_{k,t}}}\right) \\
&< \frac{K}{\sqrt{m_{1,t}}} + \sum_{k=1}^{\infty} \beta_k K \left(\frac{1}{\sqrt{m_{k+1,t}}} - \frac{1}{\sqrt{m_{k,t}}}\right) \\
&= \sum_{k=1}^{\infty} \frac{(\beta_{k-1} - \beta_k)K}{\sqrt{m_{k,t}}}.
\end{aligned}$$

Note that we asuume $\mathcal{H}_t$ happens. Denote $\sigma = \max_{i\in S_t} \sigma(i)$, then we have:

$$\begin{aligned}
\Delta_{S_t} \leq 2\Lambda_S &= 2 \max_{i\in S_t} \sigma(i) \sqrt{\frac{6\ln(t)}{T_{i,t-1}}} \\
&\leq 2\sigma \cdot \sqrt{6\ln(T)} \cdot \max_{i\in S_t} \frac{1}{\sqrt{T_{i,t-1}}} \\
&\leq 2\sigma \cdot \sqrt{6\ln(T)} \cdot \sum_{i\in S_t} \frac{1}{\sqrt{T_{i,t-1}}} \\
&< 2\sigma \cdot \sqrt{6\ln(T)} \cdot \sum_{k=1}^{\infty} \frac{(\beta_{k-1} - \beta_k)K}{\sqrt{m_{k,t}}} \\
&= \sqrt{6} \sum_{k=1}^{\infty} \frac{\beta_{k-1} - \beta_k}{\sqrt{\alpha_k}} \cdot \Delta_{S_t} \leq \Delta_{S_t},
\end{aligned}$$

We reach a contradiction here. The proof of lemma 5 is completed. ∎

By Lemma 5 we have

$$\sum_{t=m+1}^{T} \mathbb{I}\{\mathcal{H}_t\}\Delta_{S_t} \leq \sum_{k=1}^{\infty}\sum_{t=m+1}^{T} \mathbb{I}\{\mathcal{G}_{k,t}, \Delta_{S_t} > 0\}\Delta_{S_t}.$$

For $i \in [m], k \in \mathbb{Z}_+, t \in \{m+1, ..., T\}$, define an event

$$\mathcal{G}_{i,k,t} = \mathcal{G}_{k,t} \wedge \{i \in S_t, T_{i,t-1} \leq m_{k,t}\}.$$

Then by the definitions of $\mathcal{G}_{k,t}$ and $\mathcal{G}_{i,k,t}$ we have

$$\mathbb{I}\{\mathcal{G}_{k,t}, \Delta_{S_t} > 0\} \leq \frac{1}{\beta_k K}\sum_{i\in E_{\mathrm{B}}} \mathbb{I}\{\mathcal{G}_{i,k,t}, \Delta_{S_t} > 0\}.$$

Therefore

$$\sum_{t=m+1}^{T} \mathbb{I}\{\mathcal{H}_t\}\Delta_{S_t} \leq \sum_{i\in E_{\mathrm{B}}}\sum_{k=1}^{\infty}\sum_{t=m+1}^{T} \mathbb{I}\{\mathcal{G}_{i,k,t}, \Delta_{S_t} > 0\}\frac{\Delta_{S_t}}{\beta_k K}.$$

For each arm $i \in E_B$, suppose i is contained in N_i bad super arms $S^{\mathrm{B}}_{i,1}, S^{\mathrm{B}}_{i,2}, \ldots, S^{\mathrm{B}}_{i,N_i}$. Let $\Delta_{i,l} = \Delta_{S^{\mathrm{B}}_{i,l}} (l \in [N_i])$. Without loss of generality, we assume $\Delta_{i,1} \geq \Delta_{i,2} \geq \ldots \geq \Delta_{i,N_i}$. Note that $\Delta_{i,N_i} = \Delta_{i,\min}$. For convenience, we also define $\Delta_{i,0} = +\infty$, i.e., $\alpha_k \left(\frac{2\sigma K}{\Delta_{i,0}}\right)^2 = 0$. Then we have

$$\begin{aligned}
&\sum_{t=m+1}^{T} \mathbb{I}\{\mathcal{H}_t\}\Delta_{S_t} \\
&\leq \sum_{i\in E_{\mathrm{B}}}\sum_{k=1}^{\infty}\sum_{t=m+1}^{T}\sum_{l=1}^{N_i} 1\{\mathcal{G}_{i,k,t}, S_t = S^{\mathrm{B}}_{i,l}\}\frac{\Delta_{S_t}}{\beta_k K} \\
&\leq \sum_{i\in E_{\mathrm{B}}}\sum_{k=1}^{\infty}\sum_{t=m+1}^{T}\sum_{l=1}^{N_i} \mathbb{I}\{T_{i,t-1} \leq m_{k,t}, S_t = S^{\mathrm{B}}_{i,l}\}\frac{\Delta_{i,l}}{\beta_k K} \\
&= \sum_{i\in E_{\mathrm{B}}}\sum_{k=1}^{\infty}\sum_{t=m+1}^{T}\sum_{l=1}^{N_i} \mathbb{I}\left\{T_{i,t-1} \leq \alpha_k \left(\frac{2\sigma K}{\Delta_{i,l}}\right)^2 \ln T, S_t = S^{\mathrm{B}}_{i,l}\right\}\frac{\Delta_{i,l}}{\beta_k K} \\
&= \sum_{i\in E_{\mathrm{B}}}\sum_{k=1}^{\infty}\sum_{t=m+1}^{T}\sum_{l=1}^{N_i}\sum_{j=1}^{l} \mathbb{I}\left\{\alpha_k \left(\frac{2\sigma K}{\Delta_{i,j-1}}\right)^2 \ln T < T_{i,t-1} \leq \alpha_k \left(\frac{2\sigma K}{\Delta_{i,j}}\right)^2 \ln T, S_t = S^{\mathrm{B}}_{i,l}\right\}\frac{\Delta_{i,l}}{\beta_k K} \\
&\leq \sum_{i\in E_{\mathrm{B}}}\sum_{k=1}^{\infty}\sum_{t=m+1}^{T}\sum_{l=1}^{N_i}\sum_{j=1}^{l} \mathbb{I}\left\{\alpha_k \left(\frac{2\sigma K}{\Delta_{i,j-1}}\right)^2 \ln T < T_{i,t-1} \leq \alpha_k \left(\frac{2\sigma K}{\Delta_{i,j}}\right)^2 \ln T, S_t = S^{\mathrm{B}}_{i,l}\right\}\frac{\Delta_{i,j}}{\beta_k K} \\
&\leq \sum_{i\in E_{\mathrm{B}}}\sum_{k=1}^{\infty}\sum_{t=m+1}^{T}\sum_{l=1}^{N_i}\sum_{j=1}^{N_i} \mathbb{I}\left\{\alpha_k \left(\frac{2\sigma K}{\Delta_{i,j-1}}\right)^2 \ln T < T_{i,t-1} \leq \alpha_k \left(\frac{2\sigma K}{\Delta_{i,j}}\right)^2 \ln T, S_t = S^{\mathrm{B}}_{i,l}\right\}\frac{\Delta_{i,j}}{\beta_k K} \\
&\leq \sum_{i\in E_{\mathrm{B}}}\sum_{k=1}^{\infty}\sum_{t=m+1}^{T}\sum_{j=1}^{N_i} 1\left\{\alpha_k \left(\frac{2\sigma K}{\Delta_{i,j-1}}\right)^2 \ln T < T_{i,t-1} \leq \alpha_k \left(\frac{2\sigma K}{\Delta_{i,j}}\right)^2 \ln T\right\}\frac{\Delta_{i,j}}{\beta_k K}
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{i\in E_{\mathrm{B}}}\sum_{k=1}^{\infty}\sum_{j=1}^{N_i}\left(\alpha_k\left(\frac{2\sigma K}{\Delta_{i,j}}\right)^2\ln T-\alpha_k\left(\frac{2\sigma K}{\Delta_{i,j-1}}\right)^2\ln T\right)\frac{\Delta_{i,j}}{\beta_k K}\\
&=4\sigma^2K\left(\sum_{k=1}^{\infty}\frac{\alpha_k}{\beta_k}\right)\ln T\cdot\sum_{i\in E_{\mathrm{B}}}\sum_{j=1}^{N_i}\left(\frac{1}{\Delta_{i,j}^2}-\frac{1}{\Delta_{i,j-1}^2}\right)\Delta_{i,j}\\
&\leq 1068\sigma^2K\ln T\cdot\sum_{i\in E_{\mathrm{B}}}\sum_{j=1}^{N_i}\left(\frac{1}{\Delta_{i,j}^2}-\frac{1}{\Delta_{i,j-1}^2}\right)\Delta_{i,j},
\end{aligned}
$$

where the last inequality is due to (18).

Finally, for each $i\in E_B$ we have

$$
\begin{aligned}
\sum_{j=1}^{N_i}\left(\frac{1}{\Delta_{i,j}^2}-\frac{1}{\Delta_{i,j-1}^2}\right)\Delta_{i,j}&=\frac{1}{\Delta_{i,N_i}}+\sum_{j=1}^{N_i-1}\frac{1}{\Delta_{i,j}^2}(\Delta_{i,j}-\Delta_{i,j+1})\\
&\leq\frac{1}{\Delta_{i,N_i}}+\int_{\Delta_{i,N_i}}^{\Delta_{i,1}}\frac{1}{x^2}\mathrm{d}x\\
&=\frac{2}{\Delta_{i,N_i}}-\frac{1}{\Delta_{i,1}}\\
&<\frac{2}{\Delta_{i,\min}}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\sum_{t=m+1}^{T}\mathbb{I}\{\mathcal{H}_t\}\Delta_{S_t}&\leq 1068\sigma^2K\ln T\cdot\sum_{i\in E_{\mathrm{B}}}\frac{2}{\Delta_{i,\min}}\\
&=\sigma^2K\sum_{i\in E_{\mathrm{B}}}\frac{2136}{\Delta_{i,\min}}\ln T.
\end{aligned}
\tag{19}
$$

Combining (19) with Lemma 4, the distribution-dependent regret bound in Theorem 1 is proved.

To prove the distribution-independent bound, we decompose $\sum_{t=m+1}^{T}\mathbb{I}\{\mathcal{H}_t\}\Delta_{S_t}$ into two parts:

$$
\begin{aligned}
\sum_{t=m+1}^{T}\mathbb{I}\{\mathcal{H}_t\}\Delta_{S_t}&=\sum_{t=m+1}^{T}\mathbb{I}\{\mathcal{H}_t,\Delta_{S_t}\leq\epsilon\}\Delta_{S_t}\\
&\quad+\sum_{t=m+1}^{T}\mathbb{I}\{\mathcal{H}_t,\Delta_{S_t}>\epsilon\}\Delta_{S_t}\\
&\leq\epsilon T+\sum_{t=m+1}^{T}1\{\mathcal{H}_t,\Delta_{S_t}>\epsilon\}\Delta_{S_t},
\end{aligned}
\tag{20}
$$

where $\epsilon>0$ is a constant to be determined. The second term can be bounded in the same way as in the proof of the distribution-dependent regret bound, except that we only consider the case

$\Delta_{S_t} > \epsilon$. Thus we can replace (19) by

$$\begin{aligned}
&\sum_{t=m+1}^{T} \mathbb{I}\{\mathcal{H}_t, \Delta_{S_t} > \epsilon\}\Delta_{S_t} \\
&\leq \sigma^2 K \sum_{i \in E_{\mathrm{B}}, \Delta_{i,\min} > \epsilon} \frac{2136}{\Delta_{i,\min}} \ln T \\
&\leq \sigma^2 K m \frac{2136}{\epsilon} \ln T.
\end{aligned} \tag{21}$$

It follows that

$$\sum_{t=m+1}^{T} \mathbb{I}\{\mathcal{H}_t\}\Delta_{S_t} \leq \epsilon T + \sigma^2 K m \frac{2136}{\epsilon} \ln T.$$

Finally, letting $\epsilon = \sqrt{\frac{2136\sigma^2 K m \ln T}{T}}$, we get

$$\sum_{t=m+1}^{T} 1\{\mathcal{H}_t\}\Delta_{S_t} \leq 2\sqrt{2136\sigma^2 K m T \ln T} < 93\sigma\sqrt{mKT \ln T}.$$

Combining this with Lemma 4, we conclude the proof of the distribution-independent regret bound in Theorem 1. ■

References

Clark CE (1961) The greatest of a finite set of random variables. *Operations Research* 9(2):145–162.

Kveton B, Wen Z, Ashkan A, Szepesvari C (2014) Tight regret bounds for stochastic combinatorial semi-bandits. *ArXiv* abs/1410.0949, URL `https://api.semanticscholar.org/CorpusID:6152788`.

Xie J, Frazier PI, Chick SE (2016) Bayesian optimization via simulation with pairwise sampling and correlated prior beliefs. *Operations Research* 64(2):542–559.