Fixed-Budget Best-Arm Identification with Heterogeneous Reward Variances (Supplementary Material)

Anusha Lalitha **Kousha Kalantari** **Yifei Ma** **Anoop Deoras** **Branislav Kveton**

AWS AI Labs
{anlalith,kkalant,yifeim,adeoras,bkveton}@amazon.com

A PROOF OF THEOREM 3

First, we decompose the probability of choosing a suboptimal arm. For any $s \in [m]$, let $E_s = \{1 \in \mathcal{A}_{s+1}\}$ be the event that the best arm is not eliminated in stage s and $\bar{E}_s$ be its complement. Then by the law of total probability,

$$\mathbb{P}\left(\hat{I} \neq 1\right) = \mathbb{P}\left(\bar{E}_m\right) = \sum_{s=1}^{m} \mathbb{P}\left(\bar{E}_s, E_{s-1} \dots, E_1\right) \leq \sum_{s=1}^{m} \mathbb{P}\left(\bar{E}_s \,\middle|\, E_{s-1} \dots, E_1\right) .$$

We bound $\mathbb{P}\left(\bar{E}_s \,\middle|\, E_{s-1} \dots, E_1\right)$ based on the observation that the best arm can be eliminated only if the estimated mean rewards of at least a half of the arms in $\mathcal{A}_s$ are at least as high as that of the best arm. Specifically, let $\mathcal{A}'_s = \mathcal{A}_s \setminus \{1\}$ be the set of all arms in stage s but the best arm and

$$N'_s = \sum_{i \in \mathcal{A}'_s} \mathbb{1}\{\hat{\mu}_{s,i} \geq \hat{\mu}_{s,1}\} .$$

Then by the Markov's inequality,

$$\mathbb{P}\left(\bar{E}_s \,\middle|\, E_{s-1} \dots, E_1\right) \leq \mathbb{P}\left(N'_s \geq \frac{n_s}{2} \,\middle|\, E_{s-1} \dots, E_1\right) \leq \frac{2\, \mathbb{E}\left[N'_s \,\middle|\, E_{s-1} \dots, E_1\right]}{n_s} .$$

The key step in bounding the above expectation is understanding the probability that any arm has a higher estimated mean reward than the best one. We bound this probability next.

Lemma 1. *For any stage $s \in [m]$ with the best arm, $1 \in \mathcal{A}_s$, and any suboptimal arm $i \in \mathcal{A}_s$, we have*

$$\mathbb{P}\left(\hat{\mu}_{s,i} \geq \hat{\mu}_{s,1}\right) \leq \exp\left[-\frac{n_s \Delta_i^2}{4 \sum_{j \in \mathcal{A}_s} \sigma_j^2}\right] .$$

Proof. The proof is based on concentration inequalities for sub-Gaussian random variables [Boucheron et al., 2013]. In particular, since $\hat{\mu}_{s,i} - \mu_i$ and $\hat{\mu}_{s,1} - \mu_1$ are sub-Gaussian with variance proxies $\sigma_i^2 / N_{s,i}$ and $\sigma_1^2 / N_{s,1}$, respectively; their difference is sub-Gaussian with a variance proxy $\sigma_i^2 / N_{s,i} + \sigma_1^2 / N_{s,1}$. It follows that

$$\begin{aligned} \mathbb{P}\left(\hat{\mu}_{s,i} \geq \hat{\mu}_{s,1}\right) &= \mathbb{P}\left(\hat{\mu}_{s,i} - \hat{\mu}_{s,1} \geq 0\right) = \mathbb{P}\left((\hat{\mu}_{s,i} - \mu_i) - (\hat{\mu}_{s,1} - \mu_1) > \Delta_i\right) \\ &\leq \exp\left[-\frac{\Delta_i^2}{2\left(\frac{\sigma_i^2}{N_{s,i}} + \frac{\sigma_1^2}{N_{s,1}}\right)}\right] = \exp\left[-\frac{n_s \Delta_i^2}{4 \sum_{j \in \mathcal{A}_s} \sigma_j^2}\right] , \end{aligned}$$

where the last step follows from the definitions of $N_{s,i}$ and $N_{s,1}$ in Lemma 1. □

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

The last major step is bounding $\mathbb{E}\left[N'_s \,|\, E_{s-1} \ldots, E_1\right]$ with the help of Lemma 1. Starting with the union bound, we get

$$\begin{aligned}\mathbb{E}\left[N'_s \,|\, E_{s-1} \ldots, E_1\right] &\leq \sum_{i \in \mathcal{A}'_s} \mathbb{P}\left(\hat{\mu}_{s,i} \geq \hat{\mu}_{s,1}\right) \leq \sum_{i \in \mathcal{A}'_s} \exp\left[-\frac{n_s \Delta_i^2}{4 \sum_{j \in \mathcal{A}_s} \sigma_j^2}\right] \\ &\leq n_s \max_{i \in \mathcal{A}'_s} \exp\left[-\frac{n_s \Delta_i^2}{4 \sum_{j \in \mathcal{A}_s} \sigma_j^2}\right] = n_s \exp\left[-\frac{n_s \min_{i \in \mathcal{A}'_s} \Delta_i^2}{4 \sum_{j \in \mathcal{A}_s} \sigma_j^2}\right] .\end{aligned}$$

Now we chain all inequalities and get

$$\mathbb{P}\left(\hat{I} \neq 1\right) \leq 2 \sum_{s=1}^{m} \exp\left[-\frac{n_s \min_{i \in \mathcal{A}'_s} \Delta_i^2}{4 \sum_{j \in \mathcal{A}_s} \sigma_j^2}\right] .$$

To get the final claim, we use that

$$m = \log_2 K \,, \quad n_s = \frac{n}{\log_2 K} \,, \quad \min_{i \in \mathcal{A}'_s} \Delta_i^2 \geq \Delta_{\min}^2 \,, \quad \sum_{j \in \mathcal{A}_s} \sigma_j^2 \leq \sum_{j \in \mathcal{A}} \sigma_j^2 \,.$$

This concludes the proof.

B PROOF OF THEOREM 4

This proof has the same steps as that in Appendix A. The only difference is that $N_{s,i}$ and $N_{s,1}$ in Lemma 1 are replaced with their lower bounds, based on the following lemma.

Lemma 2. *Fix stage s and arm $i \in \mathcal{A}_s$ in* `SHVar`. *Then*

$$N_{s,i} \geq \frac{\sigma_i^2}{\sigma_{\max}^2} \left(\frac{n_s}{|\mathcal{A}_s|} - 1\right) ,$$

where $\sigma_{\max} = \max_{i \in \mathcal{A}} \sigma_i$ is the maximum reward noise and n_s is the budget in stage s.

Proof. Let J be the most pulled arm in stage s and $\ell \in [n_s]$ be the round where arm J is pulled the last time. By the design of `SHVar`, since arm J is pulled in round ℓ,

$$\frac{\sigma_J^2}{N_{s,\ell,J}} \geq \frac{\sigma_i^2}{N_{s,\ell,i}}$$

holds for any arm $i \in \mathcal{A}_s$. This can be further rearranged as

$$N_{s,\ell,i} \geq \frac{\sigma_i^2}{\sigma_J^2} N_{s,\ell,J} \,.$$

Since arm J is the most pulled arm in stage s and ℓ is the round of its last pull,

$$N_{s,\ell,J} = N_{s,J} - 1 \geq \frac{n_s}{|\mathcal{A}_s|} - 1 \,.$$

Moreover, $N_{s,i} \geq N_{s,\ell,i}$. Now we combine all inequalities and get

$$N_{s,i} \geq \frac{\sigma_i^2}{\sigma_J^2} \left(\frac{n_s}{|\mathcal{A}_s|} - 1\right) . \tag{1}$$

To eliminate dependence on random J, we use $\sigma_J \leq \sigma_{\max}$. This concludes the proof. □

When plugged into Lemma 1, we get

$$\mathbb{P}\left(\hat{\mu}_{s,i} \geq \hat{\mu}_{s,1}\right) \leq \exp\left[-\frac{\Delta_i^2}{2\left(\frac{\sigma_i^2}{N_{s,i}} + \frac{\sigma_1^2}{N_{s,1}}\right)}\right] \leq \exp\left[-\frac{\left(\frac{n_s}{|\mathcal{A}_s|} - 1\right) \Delta_i^2}{4\sigma_{\max}^2}\right] .$$

This completes the proof.

C PROOF OF THEOREM 5

This proof has the same steps as that in Appendix A. The main difference is that $N_{s,i}$ and $N_{s,1}$ in Lemma 1 are replaced with their lower bounds, based on the following lemma.

Lemma 3. *Fix stage s and arm $i \in \mathcal{A}_s$ in* `SHAdaVar`. *Then*

$$N_{s,i} \geq \frac{\sigma_i^2}{\sigma_{\max}^2} \alpha(|\mathcal{A}_s|, n_s, \delta) \left(\frac{n_s}{|\mathcal{A}_s|} - 1\right),$$

where $\sigma_{\max} = \max_{i \in \mathcal{A}} \sigma_i$ is the maximum reward noise, n_s is the budget in stage s, and

$$\alpha(k, n, \delta) = \frac{1 - 2\sqrt{\frac{\log(1/\delta)}{n/k-2}}}{1 + 2\sqrt{\frac{\log(1/\delta)}{n/k-2}} + \frac{2\log(1/\delta)}{n/k-2}}$$

is an arm-independent constant.

Proof. Let J be the most pulled arm in stage s and $\ell \in [n_s]$ be the round where arm J is pulled the last time. By the design of `SHAdaVar`, since arm J is pulled in round ℓ,

$$\frac{U_{s,\ell,J}}{N_{s,\ell,J}} \geq \frac{U_{s,\ell,i}}{N_{s,\ell,i}}$$

holds for any arm $i \in \mathcal{A}_s$. Analogously to (1), this inequality can be rearranged and loosened as

$$N_{s,i} \geq \frac{U_{s,\ell,i}}{U_{s,\ell,J}} \left(\frac{n_s}{|\mathcal{A}_s|} - 1\right). \tag{2}$$

We bound $U_{s,\ell,i}$ from below using the fact that $U_{s,\ell,i} \geq \sigma_i^2$ holds with probability at least $1 - \delta$, based on the first claim in Lemma 2. To bound $U_{s,\ell,J}$, we apply the second claim in Lemma 2 to bound $\hat{\sigma}_{s,\ell,J}^2$ in $U_{s,\ell,J}$, and get that

$$U_{s,\ell,J} \leq \sigma_J^2 \frac{1 + 2\sqrt{\frac{\log(1/\delta)}{N_{s,\ell,J}-1}} + \frac{2\log(1/\delta)}{N_{s,\ell,J}-1}}{1 - 2\sqrt{\frac{\log(1/\delta)}{N_{s,\ell,J}-1}}}$$

holds with probability at least $1 - \delta$. Finally, we plug both bounds into (2) and get

$$N_{s,i} \geq \frac{\sigma_i^2}{\sigma_J^2} \frac{1 - 2\sqrt{\frac{\log(1/\delta)}{N_{s,\ell,J}-1}}}{1 + 2\sqrt{\frac{\log(1/\delta)}{N_{s,\ell,J}-1}} + \frac{2\log(1/\delta)}{N_{s,\ell,J}-1}} \left(\frac{n_s}{|\mathcal{A}_s|} - 1\right).$$

To eliminate dependence on random J, we use that $\sigma_J \leq \sigma_{\max}$ and $N_{s,\ell,J} \geq n_s/|\mathcal{A}_s| - 1$. This yields our claim and concludes the proof of Lemma 3. □

Similarly to Lemma 2, this bound is asymptotically tight when all reward variances are identical. Also $\alpha(|\mathcal{A}_s|, n_s, \delta) \to 1$ as $n_s \to \infty$. Therefore, the bound has the same shape as that in Lemma 2.

The application of Lemma 3 requires more care. Specifically, it relies on high-probability confidence intervals derived in Lemma 2, which need $N_{s,t,i} > 4\log(1/\delta) + 1$. This is guaranteed whenever $n \geq K \log_2 K(4\log(1/\delta) + 1)$. Moreover, since the confidence intervals need to hold in any stage s and round t, and for any arm i, we need a union bound over Kn events. This leads to the following claim.

Suppose that $n \geq K \log_2 K(4\log(1/\delta) + 1)$. Then, when Lemma 3 is plugged into Lemma 1, we get that

$$\mathbb{P}(\hat{\mu}_{s,i} \geq \hat{\mu}_{s,1}) \leq \exp\left[-\frac{\Delta_i^2}{2\left(\frac{\sigma_i^2}{N_{s,i}} + \frac{\sigma_1^2}{N_{s,1}}\right)}\right] \leq \exp\left[-\frac{\alpha(|\mathcal{A}_s|, n_s, Kn\delta)\left(\frac{n_s}{|\mathcal{A}_s|} - 1\right)\Delta_i^2}{4\sigma_{\max}^2}\right].$$

This completes the proof.

References

Stephane Boucheron, Gabor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.