BAI D tracking(Prof. Jayakrishnan Nair's version)

Puranjay Datta ,19D070048

02-05-2023

Problem Statement

1. Given n arms, find the arm with highest mean
2. For the sake of simplicity of analysis I took 3 arms with means $1, \mu_1, \mu_2$, $\mu_1 > \mu_2$.
3. define $\Delta_1 = 1 - \mu_1, \Delta_2 = 1 - \mu_2$
4. $H = \frac{2}{\Delta_1^2} + \frac{1}{\Delta_2^2}$
5. $\Delta_i - \sqrt{\frac{T\delta^2}{HT_i(t)}} \leq \hat{\Delta}_i \leq \Delta_i + \sqrt{\frac{T\delta^2}{HT_i(t)}}$

Lemma1

$\Delta_i - \sqrt{\frac{T\delta^2}{HT_i(t)}} \leq \hat{\Delta}_i \leq \Delta_i + \sqrt{\frac{T\delta^2}{HT_i(t)}}$

Proof:

$$|\hat{\mu}_i(t) - \mu_i| \leq \sqrt{\frac{T\delta^2}{HT_i(t)}}.$$

From the reverse triangle inequality

$$\begin{aligned} |\hat{\mu}_i(t) - \mu_i| &= |(\hat{\mu}_i(t) - 1) - (\mu_i - 1)| \\ &\geq ||\hat{\mu}_i(t) - 1| - |\mu_i - 1|| \\ &\geq \left|\widehat{\Delta}_i(t) - \Delta_i\right|. \end{aligned}$$

$$\Delta_i - \sqrt{\frac{T\delta^2}{HT_i(t)}} \leq \widehat{\Delta}_i(t) \leq \Delta_i + \sqrt{\frac{T\delta^2}{HT_i(t)}}$$

Analysis [1]

1. Let the number of pulls of arm 1 be x, arm 2 be y.
2. The number of pulls of arm 3(with mean 1) is T-x-y>x,y
3. In D tracking's variant we are trying to track the optimal number of pulls i.e $\frac{T\Delta_i^{-2}}{H}$
4. we sample $argmin T_i(t) - \frac{t\hat{\Delta}_i^{-2}}{H}$

Analysis

Helpful Arm

Characterization of some helpful arm. At time T, we consider an arm k that has been pulled after the initialization phase and such that $T_k(T) - 1 \geq \frac{(T-K)}{H\Delta_k^2}$. We know that such an arm exists otherwise we get:

$$T - K = \sum_{i=1}^{K} (T_i(T) - 1) < \sum_{i=1}^{K} \frac{T-K}{H\Delta_i^2} = T - K,$$

which is a contradiction. Note that since $T \geq 2K$, we have that $T_k(T) - 1 \geq \frac{T}{2H\Delta_k^2}$ We now consider $t \leq T$ the last time that this arm k was pulled. Using $T_k(t) \geq 2$ (by the initialisation of the algorithm), we know that:

$$T_k(t) \geq T_k(T) - 1 \geq \frac{T}{2H\Delta_k^2}$$

1. My aim was to see how many more pulls does this variant take compared to the optimal number of pulls i.e estimate the constant c in $\frac{T\Delta_1^{-2}(1+c)}{H}$ and how worse the worst arm performs i.e estimate the constant d in $\frac{T\Delta_2^{-2}(1-d)}{H}$
2. Define $T' =$ Instant when the helpful arm(with mean 1 in our case) has reached its last pull.
3. $T' \geq \frac{T\Delta_1^{-2}}{2H}$
4. At T' we have $z - \frac{T'\hat{\Delta}_1^{-2}}{H} \leq y - \frac{T'\hat{\Delta}_2^{-2}}{H}$ and $z > y$.

Simulation

1. For $z > Threshold$ the above equation is not satisfied.
2. Therefore we need to find the maximum value of z above which it is not satisfied.
3. We simulate it for different $H(\Delta_1, \Delta_2)$ and $T = Time\ Horizon$.

Conclusion

1. For a fixed hardness H, as T increases the constant c,d approaches 0.
2. Found the first Time instant using Binary search ,where $c, d < 1$. Call it T_B.
3. T_B have a some relation in H,Δ_1, Δ_2
4. It looks the there is some additive factor i.e the $\log e(T)$ has some additive factor like $\frac{T(1+c(H,T))}{H}$ apart from the multiplicative factor $\frac{T}{H_2 \bar{log}(k)}$.
5. $c(H, T)$ increases as H,T increases.But as T increases c seems to converge.

References

[1] Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. "An optimal algorithm for the thresholding bandit problem". In: *International Conference on Machine Learning*. PMLR. 2016, pp. 1690–1698.

Thank You