Limitations of explore then commit (contd.)

- **Question:** What if we don't know the minimum gap $\Delta_{\min}$?
- If $\Delta_{\min} \to 0$, $\quad m^\star = O\left(\frac{K\log(T)}{\Delta_{\min}}\right) \to \infty$!!!
- We need infinitely many samples to seperate arbitrarily close arms !!!
- Trick: For arms which are arbitarily close ($\Delta_a = O(1/T)$), we will suffer constant regret for pulling them. We can ignore such arms !!

Theorem: Upon having $m = O(N^{2/3}\log(N)^{1/3})$ explore then commit suffers a regret of at most $O(N^{2/3}(K\log(N))^{1/3})$ over all instances !!!

Question: Can there be an algorithm that learns the gap $\Delta_{\min}$ and performs the optimal amount of exploration.

Empirical reward vs true reward

- **Strong Law of large number:** Empirical means coverge to the true mean with probability 1. **But at what rate?**

- **Hoeffding's inequality :**

Theorem: Let $X_1, X_2, \cdots, X_n$ be n i.i.d. samples from a distribution over $[0,1]$ with mean μ, then

$$\mathbb{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) \leq 2\exp(-2n\epsilon^2)\,.$$