Regret guarantees of Upper bound confidence

Theorem: UCB algorithm pulls every suboptimal arm a atmost $O\left(\frac{\log(T)}{\Delta_a^2}\right)$

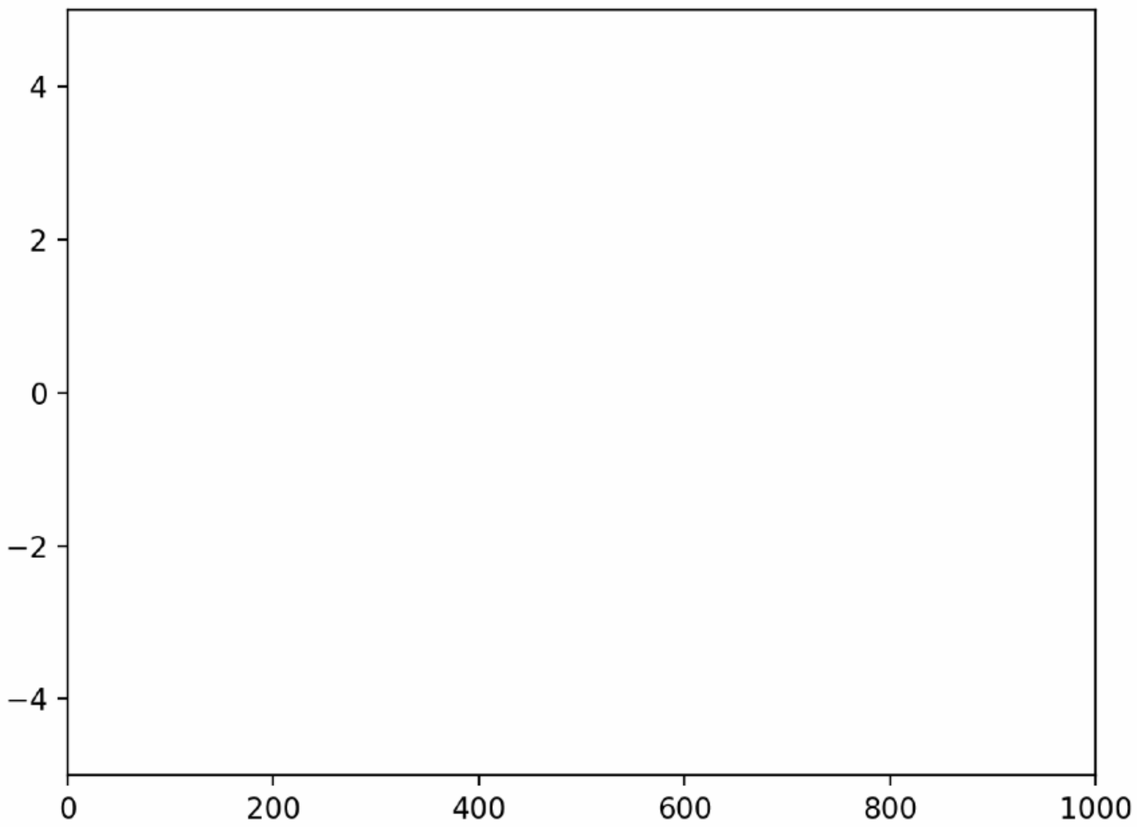
times or more precisely,

$$\mathbb{E}[N_a(T)] \leq 1 + \frac{8\log(T)}{\Delta_a^2}$$

Expected regret is at most:

$$\text{Reg}_T \leq O\left(\sum_{a \neq a^\star} \Delta_a + \sum_{a \neq a^\star} \frac{\log(T)}{\Delta_a}\right)$$

UCB working



UCB working for two coins with probability of head 0.9 (red) and 0.5 (blue)