Upper confidence bound

- At time t, sample from the arm with maximum $\text{UCB}_a(t)$.
- Recall that, $\text{UCB}_a(t) = \hat{\mu}_a(t) + \sqrt{\frac{2\log(T)}{N_a(t)}}$.
- **Thus, UCB prioritises arms having:**
  - Higher empirical mean $\hat{\mu}_a(t)$ – (Exploitation)
  - Lower no. of samples $N_a(t)$ – (Exploration)

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:

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