let there be two arms a,b. Assume total rounds T=5. Assume that the true means (which are unknown to the agent) of the arms be u(a)=4, u(b)=3. If I follow an algorithm that picks arms uniformly at random in all rounds, what is the expected regret performance of the that algorithm?

ChatGPT

The expected regret of an algorithm measures the difference between the total reward obtained by the algorithm and the total reward that could have been obtained by always choosing the best arm. In your case, you're using a simple random selection algorithm.

The expected regret R_T after T rounds can be calculated as follows:

$$R_T = T \cdot \max(u) - \sum_{t=1}^{T} \mathrm{E}[r_t]$$

Where:

- T is the total number of rounds.
- u is the vector of true means of each arm.
- $\mathrm{E}[r_t]$ is the expected reward of round t.

Since you're selecting arms uniformly at random, the expected reward for each round is the average of the true means of all arms. Therefore, $\mathrm{E}[r_t] = \frac{u(a)+u(b)}{2} = \frac{4+3}{2} = \frac{7}{2}$.

Plugging this into the formula for regret:

$$R_5 = 5 \cdot \max(4, 3) - 5 \cdot \frac{7}{2} = 5 \cdot 4 - \frac{35}{2} = 20 - \frac{35}{2} = 20 - 17.5 = 2.5$$

So, the expected regret performance of the algorithm after 5 rounds is 2.5. This means that on average, you'd miss out on 2.5 units of reward compared to always choosing the best arm.