Regret

$$R_r(T) = T\mu_{a^*} - E\left[\sum_{t=1}^{T} R_t\right]$$

$$R_r(T) = \sum_{a=1}^{K} \Delta_a \, E\left[N_a(T)\right]$$ → Regret decomposition.

$$\Delta_a = \mu^* - \mu_a$$

↳ best mean reward amongst all actions.

$N_a(T)$: no. of times arm a is selected in T rounds.

<u>ε- greedy strategy</u>: ε- fraction of T rounds are exploratory.

During exploitation, suboptimal arms are choosen only when their mean appears better than the optimal arm's mean.

For exploration: $E\left[N_a^{explore}(T)\right] = \dfrac{\varepsilon T}{K}$

In $(1-\varepsilon)T$ rounds, the agent exploits. Suboptimal arms are rarely choosen during exploitation, so

$$E\left[N_a^{exploit}(T)\right] \simeq 0$$

$$\therefore E\left[N_a(T)\right] = E\left[N_a^{explore}(T)\right] + E\left[N_a^{exploit}(T)\right]$$

$$\geq E\left[N_a^{explore}(T)\right] = \frac{\varepsilon T}{K}$$

Summing over all $K-1$ suboptimal arms :

$$\sum_{a \neq a^*} E\left[N_a(T)\right] \geqslant \frac{\varepsilon T}{K}(K-1)$$

↳ best arm.

The regret is : $R_\nu(T) = \sum_{a \neq a^*} \Delta_a E\left[N_a(T)\right]$

$\Delta_a \geqslant \Delta_{min} \quad \forall\, a \neq a^*$, we have :

$$R_\nu(T) \geqslant \Delta_{min} \sum_{a \neq a^*} E\left[N_a(T)\right]$$

$$\geqslant \Delta_{min} \frac{\varepsilon \cdot T}{K}(K-1)$$