1. About $\mathbb{P}(\hat{\mu_1} > \hat{\mu_2})$

Setting:
Best arm $a_1 \sim \mathcal{N}(\mu_1, \sigma^2)$. Pulled m times.
Second best arm $a_2 \sim \mathcal{N}(\mu_2, \sigma^2)$. Pulled n times.

So,
Estimation $\hat{\mu_1} \sim \mathcal{N}(\mu_1, \frac{\sigma^2}{m})$.
Estimation $\hat{\mu_2} \sim \mathcal{N}(\mu_2, \frac{\sigma^2}{n})$

Set $\Delta = \mu_1 - \mu_2 > 0$,
$s^2 = \frac{1}{m} + \frac{1}{n}$,
$X = \hat{\mu_1} - \hat{\mu_2} \sim \mathcal{N}(\mu_1 - \mu_2, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}) = \mathcal{N}(\Delta, (\sigma s)^2)$,
$\Phi(x)$ as the CDF for standard Gaussian distribution.

So,

$$
\begin{aligned}
\mathbb{P}(\hat{\mu_1} > \hat{\mu_2}) &= \mathbb{P}(X > 0) \\
&= 1 - \mathbb{P}(X \leq 0) \\
&= 1 - \mathbb{P}(\frac{X - \Delta}{\sigma s} \leq -\frac{\Delta}{\sigma s}) \qquad \text{where } \frac{X - \Delta}{\sigma s} \sim \mathcal{N}(0, 1) \\
&= 1 - \Phi(-\frac{\Delta}{\sigma s})
\end{aligned}
$$

$$
n \nearrow \Rightarrow s \searrow \Rightarrow -\frac{\Delta}{\sigma s} \searrow \Rightarrow \Phi(-\frac{\Delta}{\sigma s}) \searrow \Rightarrow \mathbb{P}(\hat{\mu_1} > \hat{\mu_2}) \nearrow
$$

When n increases, $\mathbb{P}(\hat{\mu_1} > \hat{\mu_2})$ increases monotonously.

2. Thought

Although more singals increase the probability to choose the best arm at certain round, the events of $\hat{\mu_1} < \hat{\mu_2}$ are more serious. It's harder to recover just by pulling a_2 more.

Instead we should consider how long it will take to achieve $\hat{\mu_1} > \hat{\mu_2}$. For example, given $\hat{\mu_1}$ and $\hat{\mu_2}$. Denote $Y_n(\hat{\mu_1})$ as the estimation of a_2's mean after n rounds. The rounds it takes to recover might be

3. Formulation attempt

Set estimation of $\hat{\mu}_1$ $\hat{\mu}_2$ after t rounds as $\hat{\mu}_1^t$ and $\hat{\mu}_2^t$. They are pulled m^t and n^t times.
So,

$$
\mathbb{P}(\hat{\mu}_1^t > \hat{\mu}_2^t) = \mathbb{P}(\hat{\mu}_1^t > \hat{\mu}_2^t | \hat{\mu}_1^{t-1} > \hat{\mu}_2^{t-1})\mathbb{P}(\hat{\mu}_1^{t-1} > \hat{\mu}_2^{t-1}) + \mathbb{P}(\hat{\mu}_1^t > \hat{\mu}_2^t | \hat{\mu}_1^{t-1} \leq \hat{\mu}_2^{t-1})\mathbb{P}(\hat{\mu}_1^{t-1} \leq \hat{\mu}_2^{t-1})
$$

If $\hat{\mu}_1^{t-1} > \hat{\mu}_2^{t-1}$, arm a_1 is chosen last round. Then $m = m^{t-1} + 1$ in computing $\mathbb{P}(\hat{\mu}_1^t > \hat{\mu}_2^t | \hat{\mu}_1^{t-1} > \hat{\mu}_2^{t-1})$. Similarly, $n = n^{t-1} + 1$ in computing the other condition.

For easier computation, $m^t = \mathbb{E}(m^t) = m^{t-1} + \mathbb{P}(\hat{\mu}_1^t > \hat{\mu}_2^t)$

$\hat{\mu}_{1,i}$ is the estimation of μ_1 at round i.

$$\begin{aligned} R(T) &= (\mu_1 - \mu_2)\mathbb{E}[\text{times of choosing arm 2}] \\ &= (\mu_1 - \mu_2)\Sigma_{i=1}^{T}\mathbb{P}(\hat{\mu}_{1,i} < \hat{\mu}_{2,i}) \end{aligned}$$

$\hat{\mu}_1^m$ is the estimation of μ_1 after pulling m times.

$$\mathbb{P}(\hat{\mu}_1 < \hat{\mu}_2) = \Sigma_{m+n=t}\mathbb{P}(\hat{\mu}_1^m < \hat{\mu}_2^n | m, n)\mathbb{P}(m, n)$$

$m + n = t$

$$\mathbb{P}(m, n) = \mathbb{P}(m-1, n)(1 - \mathbb{P}(\hat{\mu}_1^{t-1} < \hat{\mu}_2^{t-1})) + \mathbb{P}(m, n-1)\mathbb{P}(\hat{\mu}_1^{t-1} < \hat{\mu}_2^{t-1})$$

Denote the sequence of T signals after initialization as $X_1X_2...X_T$. Each X is a random variable that could be A(arm1) or B(arm2). For example, for a sequence of $A_1B_1B_2A_2$, A_i means the i-th received reward of arm1. B_i means of arm2. $A_1B_1B_2A_2$ means a pulling history of arm1, arm2, arm2, arm1. A_0 and B_0 are signals received from initialization of algorithms.

$$\mathbb{E}[\text{times of choosing arm 2}] = \Sigma_{X_1X_2...X_T}\mathbb{P}(X_1X_2...X_T) \cdot (\text{number of } X = B)$$

For example, when round is 2:

$$\begin{aligned} \mathbb{E}[\text{times of choosing arm 2}] &= \Sigma_{X_1X_2...X_T}\mathbb{P}(X_1X_2...X_T) \cdot (\text{number of } X = B) \\ &= \mathbb{P}(A_1B_1) + \mathbb{P}(B_1A_1) + \mathbb{P}(B_1B_2) * 2 \\ &= \mathbb{P}(A_0 > B_0)\mathbb{P}(\frac{A_0 + A_1}{2} < B_0 | A_0 > B_0) \\ &+ \mathbb{P}(A_0 < B_0)\mathbb{P}(A_0 > \frac{B_0 + B_1}{2} | A_0 < B_0) \\ &+ 2 \cdot \mathbb{P}(A_0 < B_0)\mathbb{P}(A_0 < \frac{B_0 + B_1}{2} | A_0 < B_0) \end{aligned}$$

Goal:

$$\begin{aligned} \mathbb{E}[\text{times of choosing arm 2}] &= \Sigma_{i=1}^{T}\mathbb{P}(\text{choose B at round } i) \\ &= \Sigma_{X_1X_2...X_T}\mathbb{P}(X_1X_2...X_T) \cdot (\text{number of } X = B) \end{aligned}$$

Proof:

$$\mathbb{P}(\text{choose B at round } i) = \Sigma_{X_1..X_i..X_T}\mathbb{P}(X_1..X_i..X_T) \text{ where } X_i = B$$