At each time $t = 1, 2, 3, \ldots$

- Algorithm chooses an action $I_t \in \{1, \ldots, n\}$
- Observes a reward $X_{I_t,t} \sim P_{I_t}$ where $P_1, \ldots, P_n$ are unknown distributions

$$\theta_i^* = \mathbb{E}_{X \sim P_i}[X]$$

$$\begin{aligned} R_T &= \max_{j=1,\ldots,n} \mathbb{E}\left[\sum_{t=1}^T X_{j,t} - \sum_{t=1}^T X_{I_t,t}\right] \\ &= \max_{j=1,\ldots,n} \theta_j^* T - \underbrace{\mathbb{E}\left[\sum_{t=1}^T X_{I_t,t}\right]} \end{aligned}$$

$$= \sum_{t=1}^T \mathbb{E}\Big[\sum_{i=1}^n \mathbb{1}\{I_t = i\} \cdot \mathbb{E}[X_{I_t,t} \mid I_t = i]\Big]$$

$$= \sum_{t=1}^T \mathbb{E}\Big[\sum_{i=1}^n \mathbb{1}\{I_t = i\}\, \mathbb{E}[X_{i,t}]\Big]$$

$$= \sum_{t=1}^T \mathbb{E}\Big[\sum_{i=1}^n \mathbb{1}\{I_t = i\}\, \theta_i^*\Big]$$

$$= \sum_{i=1}^n \theta_i^*\, \mathbb{E}\Big[\sum_{t=1}^T \mathbb{1}\{I_t = i\}\Big]$$

$$=: \sum_{i=1}^n \theta_i^*\, \mathbb{E}[T_i]$$

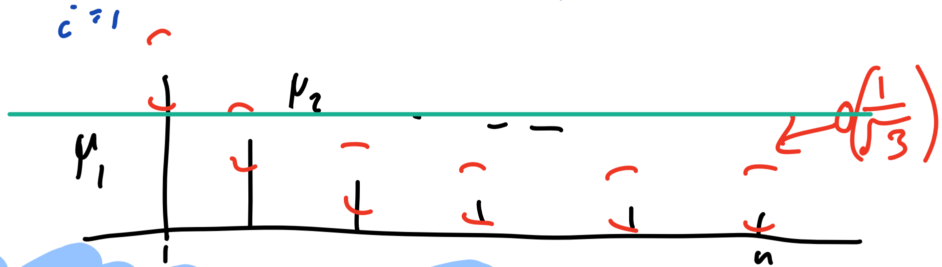
$$T = \sum_{i=1}^n T_i$$

$$= \max_{j=1,\ldots,n} \theta_j^* T - \sum_{i=1}^n \theta_i^* \mathbb{E}[T_i]$$

$$= \sum_{i=1}^n \Big(\max_{j=1,\ldots,n} \theta_j^* - \theta_i^*\Big) \mathbb{E}[T_i]$$

$$=: \sum_{i=1}^n \Delta_i\, \mathbb{E}[T_i], \qquad \Delta_i := \max_{j=1,\ldots,n} \theta_j^* - \theta_i^*$$

If you sample arm i $\approx \Delta_i^{-2}$ then output $\arg\max_i \hat{\mu}_i$



If I sample each arm 3 times

$$\mathbb{P}\left(\frac{1}{\tau}\sum_{t=1}^{\tau} Z_t > \varepsilon\right) \leq \exp(-\tau\varepsilon^2/2) \Rightarrow$$

Corollary 1. *Let $Z_1, Z_2, \ldots$ be independent mean-zero σ^2-sub-Gaussian random variables so that $\psi_Z(\lambda) := \log(\mathbb{E}[\exp(\lambda Z_t)]) \leq \exp(\lambda^2\sigma^2/2)$, then for $\tau = \lceil 2\sigma^2\epsilon^{-2}\log(1/\delta)\rceil$ we have $\mathbb{P}(\frac{1}{\tau}\sum_{t=1}^{\tau} Z_t \leq \epsilon) \geq 1-\delta$.*

Lemma 1 (Hoeffding's Lemma)**.** *Let X be an independent random variable with support in $[a,b]$ almost surely and $\mathbb{E}[X] = 0$. Then $\log(\mathbb{E}[\exp(\lambda X)]) \leq (b-a)^2\lambda^2/8$.*

$X_t \sim \begin{cases} 1 & \text{w.p. } \mu \\ 0 & \text{w.p. } 1-\mu \end{cases}$ $\quad -\mu \leq X_t - \mu \leq 1-\mu \quad$ $\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}(X_t-\mu) > \varepsilon\right) \leq \exp(-2\varepsilon^2 T)$

$X_1, \ldots, X_T$ sequence of IID R.V.s

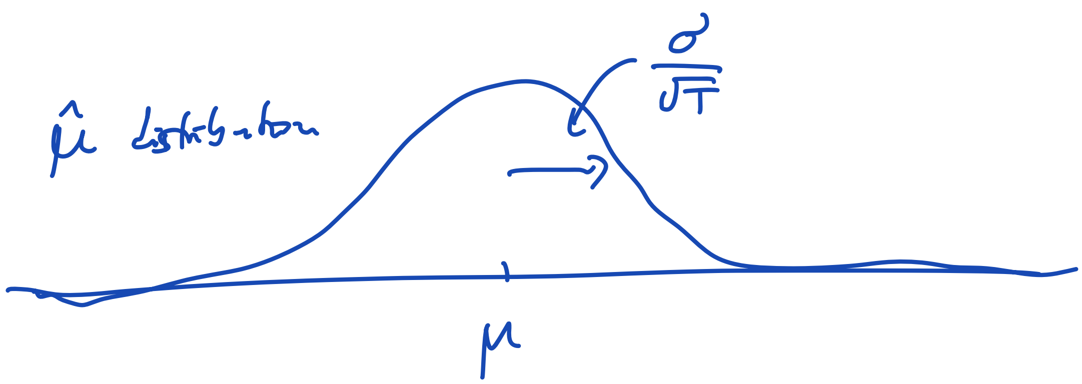
w/ $\mathbb{E}[X_t] = \mu$, $\mathbb{E}[(X_t-\mu)^2] = \sigma^2$

Central Limit Theorem | Under benign conditions

(e.g. $\mathbb{E}[|X_t|^4] < \infty$), if $\hat{\mu}_T = \frac{1}{T}\sum_{t=1}^{T} X_t$ then

$$\lim_{T\to\infty} \sqrt{T}(\hat{\mu}_T - \mu) \sim \mathcal{N}(0, \sigma^2)$$

$$\overset{T\to\infty}{\Longrightarrow} \quad \hat{\mu}_T \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{T}\right)$$

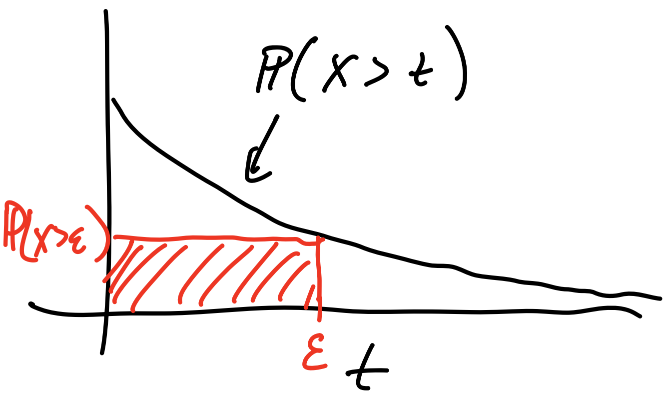


Suggests that if we want $|\hat{\mu}_T - \mu| < \varepsilon$ for some $\varepsilon > 0$, we need $T \geq \frac{\sigma^2}{\varepsilon^2}$

Markov's inequality For $Z \geq 0$ a.s., IID

we have $$\mathbb{P}(Z > \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}.$$

Proof
$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > t)\, dt$$



$$\geq \int_0^\varepsilon \mathbb{P}(X \geq \varepsilon)\, dt$$
$$= \varepsilon\, \mathbb{P}(X \geq \varepsilon) \blacksquare$$

Chebychev Inequality For any X R.V.

we have $$\mathbb{P}(|X-\mu| > \varepsilon) \leq \frac{\mathbb{E}[(X-\mu)^2]}{\varepsilon^2}$$

Markov's inequality $$\mathbb{P}(|X-\mu| > \varepsilon) = \mathbb{P}((X-\mu)^2 > \varepsilon^2)$$
$$\leq \frac{\mathbb{E}(|X-\mu|^2)}{\varepsilon^2}$$

Ex. Let $X_1, \ldots, X_T$ be IID w/ $E[X_t] = \mu$, $E[(X_t - \mu)^2] = \sigma^2$

$$\hat{\mu}_T = \frac{1}{T}\sum_{t=1}^{T} X_t \qquad \mathbb{P}(|\hat{\mu}_T - \mu| > \varepsilon) \leq \frac{\mathbb{E}[(\hat{\mu}_T - \mu)^2]}{\varepsilon^2}$$

$$\mathbb{E}[(\hat{\mu}_T - \mu)^2] = \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T}(X_t - \mu)\right)^2\right]$$

$$= \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T}(X_t - \mu)\right)\left(\frac{1}{T}\sum_{s=1}^{T}(X_s - \mu)\right)\right]$$

$$= \mathbb{E}\left[\frac{1}{T^2}\sum_{t=1}^{T}(X_t - \mu)^2\right]$$

$$= \frac{\sigma^2}{T}$$

$t \neq s$

$$\mathbb{E}[(X_t - \mu)(X_s - \mu)] = \underbrace{\mathbb{E}[(X_t - \mu)]}_{=0}\underbrace{\mathbb{E}[(X_s - \mu)]}_{=0} = 0$$

$$\mathbb{P}(|\hat{\mu}_T - \mu| > \varepsilon) \leq \frac{\mathbb{E}[(\hat{\mu}_T - \mu)^2]}{\varepsilon^2}$$

$$= \frac{\sigma^2}{T\varepsilon^2} = \frac{1}{100}$$

$$\Rightarrow \quad T \geq 100\frac{\sigma^2}{\varepsilon^2}$$

Chernoff Bound Let $X_1, \dots, X_T$ be IID mean-zero

then $\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T} X_t > \varepsilon\right) \leq$

Fix $\lambda \in \mathbb{R}$

$$\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T} X_t > \varepsilon\right) = \mathbb{P}\left(\lambda \cdot \frac{1}{T}\sum_{t} X_t > \lambda \varepsilon\right)$$

$$= \mathbb{P}\left(\exp\left(\lambda \frac{1}{T}\sum_{t=1}^{T} X_t\right) > \exp(\lambda \varepsilon)\right)$$

$$\leq e^{-\lambda \varepsilon}\, \mathbb{E}\left[\exp\left(\lambda \frac{1}{T}\sum_{t=1}^{T} X_t\right)\right]$$

$$= e^{-\lambda \varepsilon}\, \mathbb{E}\left[\prod_{t=1}^{T} \exp\left(\frac{\lambda}{T} X_t\right)\right]$$

$$= e^{-\lambda \varepsilon} \prod_{t=1}^{T} \mathbb{E}\left[\exp\left(\frac{\lambda}{T} X_t\right)\right]$$

$$= e^{-\lambda \varepsilon}\, \mathbb{E}\left[\exp\left(\frac{\lambda}{T} X_1\right)\right]^T \qquad s = \frac{\lambda}{T}$$

$$= e^{-sT\varepsilon}\, \mathbb{E}\left[\exp(s X_1)\right]^T$$

$$\leq \inf_{s} \exp\left(-sT\varepsilon + T\log\left(\mathbb{E}[\exp(sX_1)]\right)\right)$$

$$M_X(s) = \mathbb{E}[\exp(sX)]$$

$$M'_X(s) = \mathbb{E}[X \exp(sX)] \Rightarrow M'_X(0) = \mathbb{E}[X]$$

$$M''_X(s) = \mathbb{E}[X^2 \exp(sX)] \Rightarrow M''_X(0) = \mathbb{E}[X^2]$$

$$Z \sim N(0, \sigma^2) \qquad \mathbb{E}[\exp(sx)] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} e^{sx}\, dx$$

$$= \exp(s^2\sigma^2/2)$$

$$\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t > \varepsilon\right) \leq \inf_s \exp(-sT\varepsilon + Ts^2\sigma^2/2)$$

$$= \exp(-\varepsilon^2 T/\sigma^2 + T\varepsilon^2/2\sigma^2)$$

$$= \exp(-T\varepsilon^2/2\sigma^2)$$

$$= \delta$$

$$\Rightarrow \quad T \geq \frac{2\sigma^2 \log(1/\delta)}{\varepsilon^2}$$

We say a mean-zero R.V. is sub-Gaussian-σ^2 if

$$\mathbb{E}[\exp(\lambda X)] \leq \exp(\lambda^2\sigma^2/2).$$

Assume R.V.s are 1-sub-Gaussian

Input: n arms $\mathcal{X} = \{1, \dots, n\}$, confidence level $\delta \in (0, 1)$.
Let $\mathcal{X}_1 \leftarrow \mathcal{X}, \ell \leftarrow 1$
while $|\mathcal{X}_\ell| > 1$ **do**
  $\epsilon_\ell = 2^{-\ell}$
  Pull each arm in $\mathcal{X}_\ell$ exactly $\tau_\ell = \lceil 2\epsilon_\ell^{-2} \log(\frac{4\ell^2|\mathcal{X}|}{\delta}) \rceil$ times
  Compute the empirical mean of these rewards $\widehat{\theta}_{i,\ell}$ for all $i \in \mathcal{X}_\ell$
  $\mathcal{X}_{\ell+1} \leftarrow \mathcal{X}_\ell \setminus \{i \in \mathcal{X}_\ell : \max_{j \in \mathcal{X}_\ell} \widehat{\theta}_{j,\ell} - \widehat{\theta}_{i,\ell} > 2\epsilon_\ell\}$
  $\ell \leftarrow \ell + 1$
Output: $\mathcal{X}_{\ell+1}$ (or play the last arm forever in the regret setting)

Lemma 2. *Assume that* $\max_{i \in \mathcal{X}} \Delta_i \le 4$. *With probability at least* $1 - \delta$, *we have* $1 \in \mathcal{X}_\ell$ *and* $\max_{i \in \mathcal{X}_\ell} \Delta_i \le 8\epsilon_\ell$ *for all* $\ell \in \mathbb{N}$.

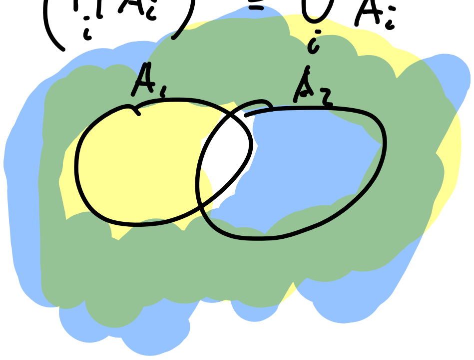
Proof. For any $\ell \in \mathbb{N}$ and $i \in [n]$ define

$$\mathcal{E}_{i,\ell} = \left\{ |\widehat{\theta}_{i,\ell} - \theta_i^*| \le \epsilon_\ell \right\}$$

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \bigcap_{\ell=1}^{\infty} \mathcal{E}_{i,\ell} \right) \ge 1 - \delta$$

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$$

$$\left(\bigcap_i A_i \right)^c = \bigcup_i A_i^c$$



$$\mathbb{P}\left(\left(\bigcap_i \bigcap_\ell \mathcal{E}_{i,\ell} \right)^c \right)$$

$$= \mathbb{P}\left(\bigcup_{i=1}^{n} \bigcup_{\ell=1}^{\infty} \mathcal{E}_{i,\ell}^c \right)$$

$$\le \sum_{i=1}^{n} \sum_{\ell=1}^{\infty} \mathbb{P}(\mathcal{E}_{i,\ell}^c)$$

$$= \sum_{i=1}^{n} \sum_{\ell=1}^{\infty} \mathbb{P}\left(\hat{\theta}_{i,\ell} - \theta_i \ge \varepsilon_\ell \text{ or } \hat{\theta}_{i,\ell} - \theta_i \le -\varepsilon_\ell \right)$$

$$\le \sum_{i=1}^{n} \sum_{\ell=1}^{\infty} \left(\mathbb{P}(\hat{\theta}_{i,\ell} - \theta_i \ge \varepsilon_\ell) + \mathbb{P}(\hat{\theta}_{i,\ell} - \theta_i \le -\varepsilon_\ell) \right)$$

$$\le 2 \sum_{i=1}^{n} \sum_{\ell=1}^{\infty} \exp(-\tau_\ell \varepsilon_\ell^2 / 2)$$

$$\leq 2 \sum_{i=1}^{n} \sum_{\ell=1}^{\infty} \exp\left(- \left(2 \varepsilon_\ell^{-2} \log\left(\frac{4n\ell^2}{\delta}\right)\right) \varepsilon_\ell^2 / 2 \right)$$

$$= 2 \sum_{i=1}^{n} \sum_{\ell=1}^{\infty} \frac{\delta}{4n\ell^2} \qquad \sum_{\ell=1}^{\infty} \frac{1}{\ell^2} = \frac{\pi^2}{6} \leq 2$$

$$= \sum_{\ell=1}^{\infty} \frac{\delta}{2\ell^2} \leq \delta.$$

$1 \in \mathcal{X}_{\ell+1}$ $\quad \max_j \hat{\theta}_{j,\ell} - \hat{\theta}_{1,\ell} \leq 2\varepsilon_\ell$

$$\max_j \hat{\theta}_{j,\ell} - \hat{\theta}_{1,\ell} =: \hat{\theta}_{\hat{j},\ell} - \hat{\theta}_{1,\ell}$$

$$= \underbrace{\hat{\theta}_{\hat{j},\ell} - \theta_{\hat{j}}}_{\leq \varepsilon_\ell \text{ on } \mathcal{E}_{\hat{j},\ell}} + \underbrace{\theta_1 - \hat{\theta}_{1,\ell}}_{\leq \varepsilon_\ell \text{ on } \mathcal{E}_{1,\ell}} + \underbrace{\theta_{\hat{j}} - \theta_1}_{\leq 0}$$

$$\leq 2\varepsilon_\ell$$

$$\Rightarrow \quad 1 \in \mathcal{X}_{\ell+1} \quad \forall \ell$$

Fix some arm i: $\Delta_i > 4\varepsilon_\ell$.

$$\max_j \hat{\theta}_{j,\ell} - \hat{\theta}_{i,\ell} \geq \hat{\theta}_{1,\ell} - \hat{\theta}_{i,\ell}$$

$$= \underbrace{\hat{\theta}_{1,\ell} - \theta_1}_{\geq -\varepsilon_\ell} + \underbrace{\theta_i - \hat{\theta}_{i,\ell}}_{\geq -\varepsilon_\ell} + \underbrace{\theta_1 - \theta_i}_{= \Delta_i > 4\varepsilon_\ell}$$

$$> 2\varepsilon_\ell$$

$$\max_{j \in \mathcal{X}_{\ell+1}} \Delta_j \leq 4\varepsilon_\ell = 8\varepsilon_{\ell+1}$$

Fix $\nu > 0$ $$\sum_{i=1}^{n} T_i = T$$

$$\text{Regret} = \sum_{i=2}^{n} \Delta_i T_i = \sum_{i:\Delta_i \le \nu} \Delta_i T_i + \sum_{i:\Delta_i > \nu} \Delta_i T_i$$

$$\le \nu T + \sum_{i:\Delta_i > \nu} \Delta_i T_i$$

$$a \vee b = \max\{a,b\}$$

$$= \nu T + \sum_{i:\Delta_i > \nu} \sum_{\ell=1}^{\infty} \Delta_i \mathcal{T}_\ell \mathbb{1}\{i \in \mathcal{X}_\ell\}$$

$$\le \nu T + \sum_{i:\Delta_i > \nu} \sum_{\ell=1}^{\infty} \Delta_i \mathcal{T}_\ell \mathbb{1}\{\Delta_i \vee \nu \le 8\varepsilon_\ell\}$$

$$\le \nu T + \sum_{i:\Delta_j > \nu} \sum_{\ell=1}^{\lceil \log_2(\frac{8}{\Delta_i \vee \nu}) \rceil} 8\varepsilon_\ell \mathcal{T}_\ell$$

$$\sum_{\ell=1}^{k} 2^\ell = 2^{k+1} - 1$$

$$\le \nu T + \sum_{i:\Delta_i > \nu} \sum_{\ell=1}^{\lceil \log_2(\frac{8}{\Delta_i \vee \nu}) \rceil} 8\varepsilon_\ell \cdot 2\varepsilon_\ell^{-2} \log\left(\frac{4n\ell^2}{\delta}\right)$$

$$\le \nu T + \sum_{i:\Delta_i > \nu} 16 \log\left(\frac{4n \lceil \log_2(\frac{8}{\Delta_i \vee \nu}) \rceil}{\delta}\right) \sum_{\ell=1}^{\lceil \log_2(\frac{8}{\Delta_i \vee \nu}) \rceil} 2^\ell$$

$$\le \nu T + c \sum_{i:\Delta_i > \nu} (\Delta_i \vee \nu)^{-1} \log\left(\frac{4n \log((\Delta_i \vee \nu)^{-1})}{\delta}\right)$$

$$\nu = 0 \Rightarrow \text{Regret} \le c \sum_i \Delta_i^{-1} \log\left(\frac{n \log \Delta_i^{-1}}{\delta}\right)$$

$$\le \nu T + c n \nu^{-1} \log\left(\frac{n \log \nu^{-1}}{\delta}\right) \le c' \sqrt{n T \log(nT/\delta)}$$

$$\mathbb{E}[\text{Regret}] = \mathbb{E}[\text{Regret} \cdot \mathbb{1}\{\mathcal{E}\}] + \mathbb{E}[\underbrace{\text{Regret}}_{\leq T} \cdot \mathbb{1}\{\mathcal{E}^c\}]$$

$$= \sqrt{nT\log\left(\frac{nT}{\delta}\right)} + \underbrace{\qquad\qquad}_{\leq \delta T}$$

If $\delta = 1/T$ then $\mathbb{E}[\text{Regret}] \leq c\sqrt{nT\log(nT)}$

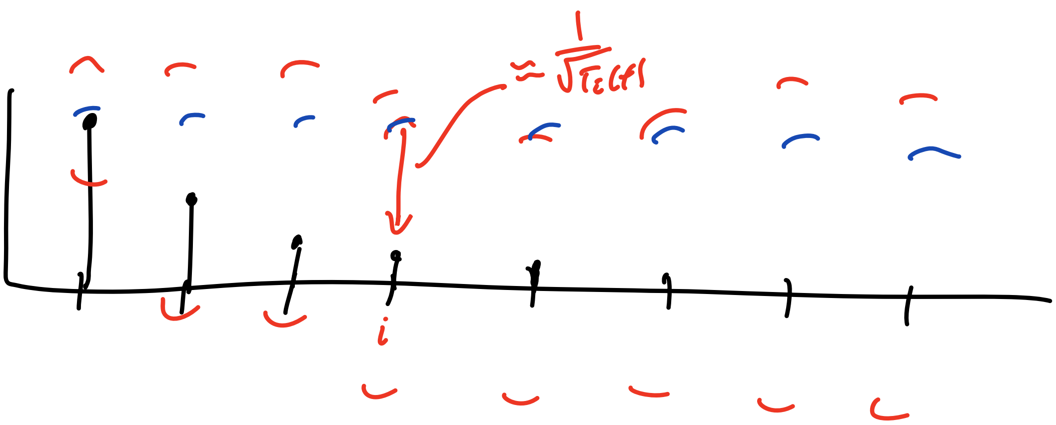
$$\mathbb{E}[\text{Regret}] \leq c\sum_{i \neq i_*} \Delta_i^{-1} \log(nT\Delta_i^{-1})$$

UCB - upper confidence bound algorithm

Pull every arm once

while True

$$I_t = \arg\max_i \hat{\theta}_{i,T_i(t)} + \sqrt{\frac{c\log(T)}{T_i(t)}}$$



Thompson Sampling / Posterior Sampling / Ensemble sampling

Let $P_{t,i}$ be the posterior distribution of θ_i after t pulls of arm i.

At each time t draw

$$\tilde{\theta}_{t,i} \sim P_{T_i(t),i}$$

and play $I_t = \arg\max_i \tilde{\theta}_{t,i}$. Update posterior

$$\mathbb{P}(I_t = i) = \mathbb{P}_t(\text{arm } i \text{ is the best arm})$$