

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

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

$$\theta_i^* = \mathbb{E}_{x \sim p_i} [x]$$

$$R_T = \max_{j=1,\dots,n} \mathbb{E} \left[\sum_{t=1}^T X_{j,t} - \sum_{t=1}^T X_{I_t,t} \right]$$

$$= \max_{j=1,\dots,n} \theta_j^* T - \mathbb{E} \left[\sum_{t=1}^T X_{I_t,t} \right]$$

$$= \sum_{t=1}^T \mathbb{E} \left[\sum_{i=1}^{\hat{I}} \mathbb{1}_{\{I_t=i\}} \cdot \mathbb{E}[X_{I_t,t} \mid I_t=i] \right]$$

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

$$= \sum_{t=1}^T \mathbb{E} \left[\sum_{i=1}^n \mathbb{1}_{\{I_t = i\}} \theta_i^* \right]$$

$$= \sum_{i=1}^n \theta_i^{\text{old}} \mathbb{E} \left[\sum_{t=1}^T \mathbb{1}_{\{I_t = i\}} \right]$$

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

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

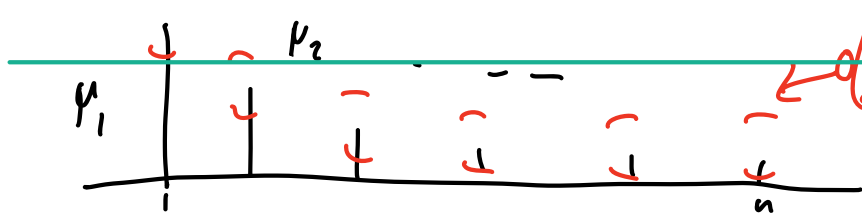
$$= \max_{j=1, \dots, n} \theta_j^* T - \sum_{i=1}^n \theta_i^* E[T_i]$$

If you sample
acc $i \sim \Delta_i^{-2}$ then
output $\arg\min_i \mu_i$

$$= \sum_{i=1}^n (\max_{j=1, \dots, n} \theta_j^A - \theta_i^A) E[T_i]$$

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

$$\Delta_i := \max_{j=1, \dots, n} \theta_j^* - \theta_i^*$$



$\left(\frac{1}{\sqrt{3}} \right)$ If I sample each arm 3 times

Corollary 1. Let $Z_1, Z_2, \dots$ be independent mean-zero σ^2 -sub-Gaussian random variables so that $\psi_Z(\lambda) := \mathbb{E}[\exp(\lambda Z_t)] \leq \exp(\lambda^2 \sigma^2 / 2)$, 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 $\log(\mathbb{E}[\exp(\lambda X)]) \leq (b-a)^2 \lambda^2 / 8$.

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

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

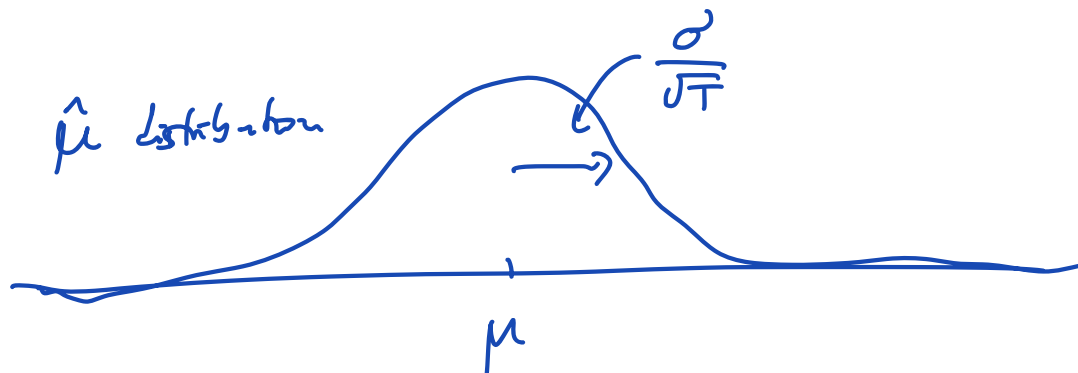
$$\text{w/ } \mathbb{E}[X_t] = \mu, \quad \mathbb{E}[(X_t - \mu)^2] = \sigma^2$$

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

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

$$\xRightarrow{T \rightarrow \infty} \hat{\mu}_T \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{T}\right)$$



Suggests that if we want $|\hat{\mu}_T - \mu| < \varepsilon$

for some $\varepsilon > 0$, we need $T \geq \frac{\sigma^2}{\varepsilon^2}$

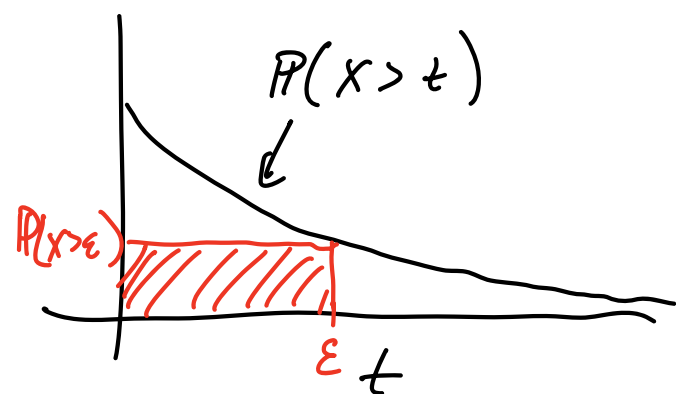
Markov's inequality | For $Z \geq 0$ a.s., IID

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

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

$$\geq \int_0^{\varepsilon} P(X \geq \varepsilon) dt$$

$$= \varepsilon P(X \geq \varepsilon)$$



Chebyshev Inequality | For any X R.V.

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

Markov's inequality $P(|X - \mu| > \varepsilon) = P((X - \mu)^2 > \varepsilon^2)$

$$\leq \frac{E[(X - \mu)^2]}{\varepsilon^2}$$

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

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

$$\begin{aligned} E[(\hat{\mu}_T - \mu)^2] &= E\left[\left(\frac{1}{T} \sum_{t=1}^T (X_t - \mu)\right)^2\right] \\ &= E\left[\left(\frac{1}{T} \sum_{t=1}^T (X_t - \mu)\right) \left(\frac{1}{T} \sum_{s=1}^T (X_s - \mu)\right)\right] \\ &= E\left[\frac{1}{T^2} \sum_{t=1}^T (X_t - \mu)^2\right] \\ &= \frac{\sigma^2}{T} \end{aligned}$$

4.5

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

$$\begin{aligned} P(|\hat{\mu}_T - \mu| > \varepsilon) &\leq \frac{E[(\hat{\mu}_T - \mu)^2]}{\varepsilon^2} \\ &= \frac{\sigma^2}{T \varepsilon^2} = \frac{1}{100} \end{aligned}$$

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

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

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

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

$$\begin{aligned} P(\frac{1}{T} \sum_{t=1}^T X_t > \varepsilon) &= P(\lambda \frac{1}{T} \sum_{t=1}^T X_t > \lambda \varepsilon) \\ &= P(\exp(\lambda \frac{1}{T} \sum_{t=1}^T X_t) > \exp(\lambda \varepsilon)) \\ &\leq e^{-\lambda \varepsilon} \mathbb{E}[\exp(\lambda \frac{1}{T} \sum_{t=1}^T X_t)] \\ &= e^{-\lambda \varepsilon} \mathbb{E}[\prod_{t=1}^T \exp(\frac{\lambda}{T} X_t)] \\ &= e^{-\lambda \varepsilon} \prod_{t=1}^T \mathbb{E}[\exp(\frac{\lambda}{T} X_t)] \\ &= e^{-\lambda \varepsilon} \mathbb{E}[\exp(\frac{\lambda}{T} X_1)]^T \quad s = \frac{\lambda}{T} \\ &= e^{-s T \varepsilon} \mathbb{E}[\exp(s X_1)]^T \\ &\leq \inf_s \exp(-s T \varepsilon + T \log(\mathbb{E}[\exp(s X_1)])) \end{aligned}$$

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

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

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

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

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

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

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

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

$$= \delta$$

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

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

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

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

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

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