Last class: $\{X_n\}_{n\geq 0}$ iid

$$\mathbb{E}[X_1] = \mu \quad , \quad Var[X_1] = \sigma^2$$

$$S_n = X_1 + \cdots + X_n$$

Central Limit Theorem

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{d.} \mathcal{N}(0, \sigma^2) \quad , \quad \text{as } n \to \infty$$

Weak Law of Large Numbers

$$\frac{S_n}{n} \xrightarrow{p.} \mu \quad , \quad \text{as } n \to \infty$$

Sample Mean True Mean

(use of CLT)

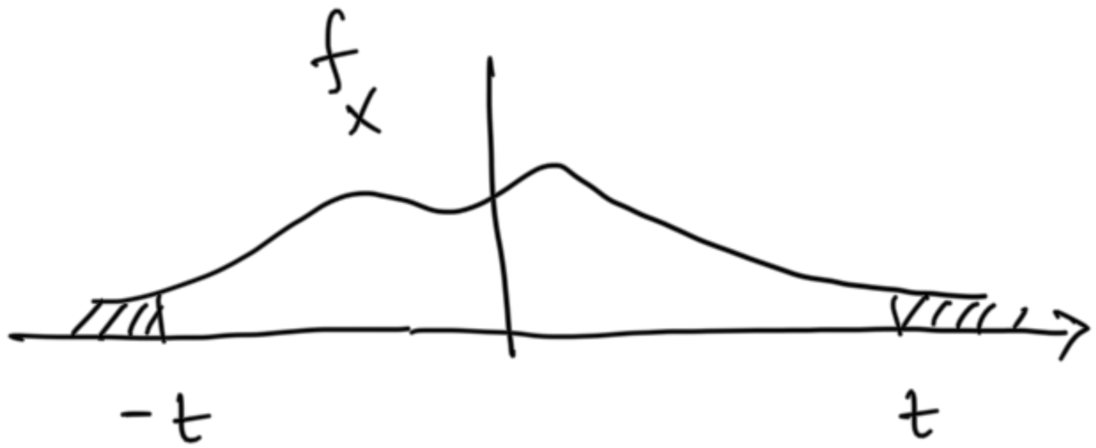
$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \lesssim \sqrt{\frac{2\sigma^2}{\pi n \varepsilon^2}}\ e^{-\left(\frac{n\varepsilon^2}{2\sigma^2}\right)} \quad \text{— (1)}$$

approximation

(chebyshev)

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n \varepsilon^2} \quad — (2)$$

Tail Bounds : $\mathbb{P}(|X| > t)$



R.H.S of (1) is better than R.H.S of (2)
(smaller for large n)

* asymptotic result (as $n \to \infty$)

* no decision to make ; we are simply looking at a sequence S_n
* no algorithmic interaction.

$\Downarrow$

at this stage we really cannot appreciate why R.H.S (2) being better than R.H.S (1) is great.

Goal: Connect Regret in multi-armed bandits to tail bounds.

Multi armed bandit problem

- arms $A = \{1, \ldots, k\}$

- Distributions are $P_1, \ldots, P_k$
- Play an arm A_t and receive rewat

$$X_t \sim P_{A_t}$$

- Means are $\mu_i = \mathbb{E}_{P_i}[X_t]$
- Best reward $\mu_* = \max_i \mu_i$
- Best arm $i_* = \arg\max_i \mu_i$

$$\text{Regret} = R_n = \mathbb{E}\left[\sum_{t=1}^{n} (\mu_* - X_t)\right]$$

Lemma: Let $\Delta_a = \mu_* - \mu_a$ be the sub-optimality gap

random Variable

Let $T_{\cdot}(n) = \sum^{n} \mathbb{I}_{\{A \ldots\}}$

$a \qquad s=1 \quad \{A_s = a\}$

Indicator function $\mathbb{I}_{\{cond\}} = 1$, cond is true

$= 0$, cond is false

number of times arm a was chosen in n rounds.

$$R_n = \sum_{a \in A} \mathbb{E}[T_a(n)] \Delta_a$$

Proof:

$$R_n = \mathbb{E}\left[\sum_{t=1}^{n} (\mu_* - X_t)\right]$$

Since at each time t, only one arm is chosen

only one of $\mathbb{I}_{\{A_s = 1\}}, \ldots, \mathbb{I}_{\{A_s = k\}}$ is

1

$$\mathbb{I}_{\{A_s = 1\}} + \mathbb{I}_{\{A_s = 2\}} + \cdots + \mathbb{I}_{\{A_s = k\}} = 1$$

$$R_n = \mathbb{E}\left[\sum_{t=1}^{n} (\mu_* - X_t) \cdot 1\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n} (\mu_* - X_t)\left(\mathbb{I}_{\{A_t = 1\}} + \mathbb{I}_{\{A_t = 2\}} + \cdots + \mathbb{I}_{\{A_t = k\}}\right)\right]$$

$$= \mathbb{E}\left[\sum_{a \in A} \sum_{t=1}^{n} (\mu_* - X_t)\, \mathbb{I}_{\{A_t = a\}}\right]$$

using Tower property of expectation

$\mathbb{E}[X \mid Y]$ or $\mathbb{E}[Y \mid X]$

$$\mathbb{E}[\underbrace{\mathbb{E}[X \mid Y]}_{\text{function of } Y}] = \mathbb{E}[X]$$

Tower property: condition on A_t

$$\mathbb{E}[\cdot] = \mathbb{E}[\mathbb{E}[\cdot \mid A_t]]$$

$$\mathbb{E}\Big[\mathbb{E}\Big[\underbrace{\sum_{a \in A} \sum_{t=1}^{n} (\mu_* - X_t)\, \mathbb{I}_{\{A_t = a\}}}\ \Big|\ A_t\Big]\Big]$$

Moving $\mathbb{E}[\cdot \mid A_t]$ inside summation

$$= \mathbb{E}\Big[\sum_{a \in A} \sum_{t=1}^{n} \mathbb{E}\big[(\mu_* - X_t)\, \mathbb{I}_{\{A_t = a\}} \mid A_t\big]\Big]$$

$$\mathbb{E}[g(X) f(Y) \mid Y] = f(Y)\, \mathbb{E}[g(X) \mid Y]$$

$$\mathbb{E}[c\, g(x)] = c\, \mathbb{E}[g(x)]$$

$$= \mathbb{E}\left[\sum_{a \in A} \sum_{t=1}^{n} \mathbb{I}_{\{A_t = a\}} \mathbb{E}[\mu_* - X_t \mid A_t] \right]$$

$$= \mathbb{E}\left[\sum_{a \in A} \sum_{t=1}^{n} \mathbb{I}_{\{A_t = a\}} (\mu_* - \mu_{A_t}) \right]$$

$$= \mathbb{E}\left[\sum_{a \in A} \sum_{t=1}^{n} \mathbb{I}_{\{A_t = a\}} (\mu_* - \mu_a) \right]$$

$$= \mathbb{E}\left[\sum_{a \in A} \sum_{t=1}^{n} \mathbb{I}_{\{A_t = a\}} \cdot \Delta_a \right]$$

$$= \mathbb{E}\left[\sum_{a \in A} \Delta_a \sum_{t=1}^{n} \mathbb{I}_{\{A_t = a\}} \right]$$

moving $\mathbb{E}[\quad]$ inside summation

$$= \sum_{a \in A} \Delta_a \, \mathbb{E}[T_a(n)]$$

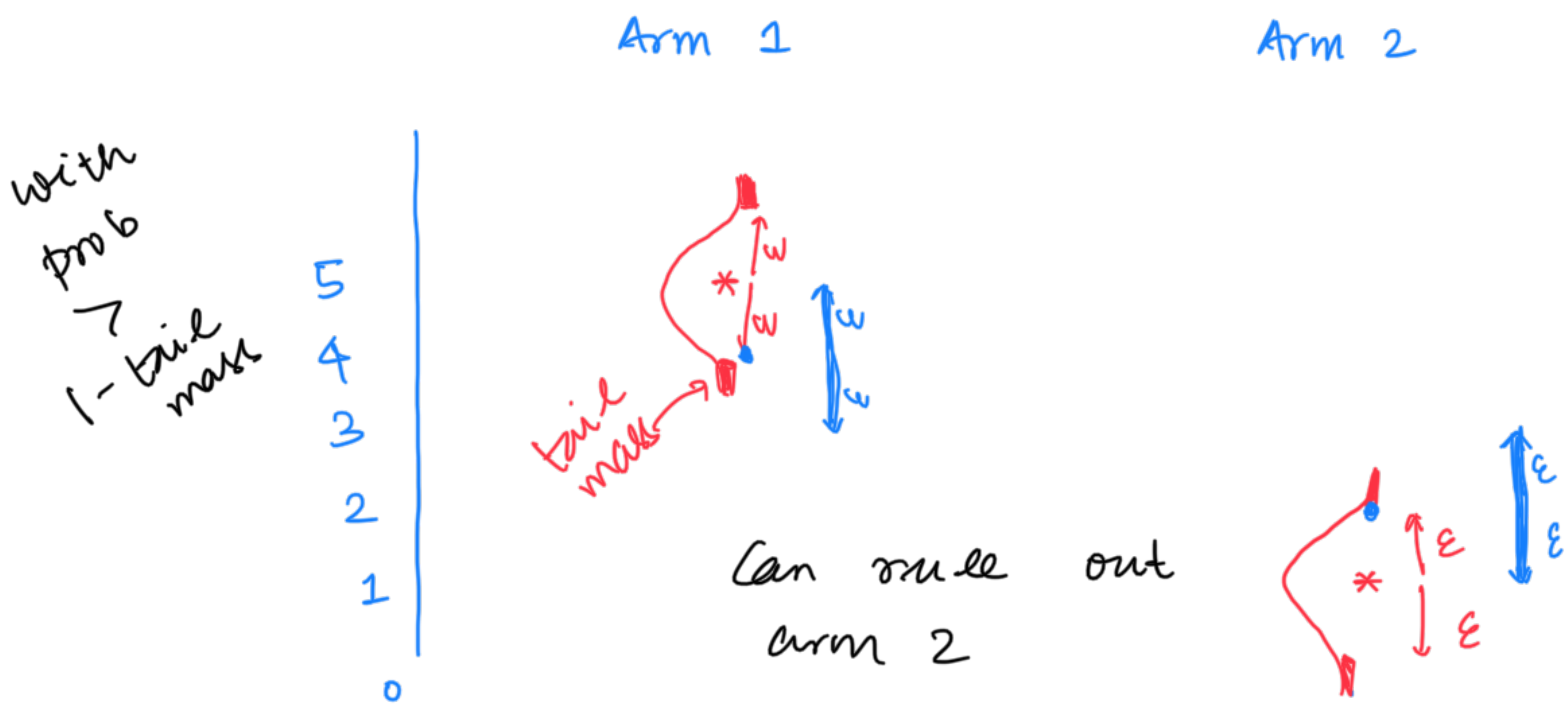
Arm 1 $\quad Y_{11}(\omega) \cdots Y_{12}(\omega) \cdots \cdots$

Arm 2 $\quad Y_{21}(\omega) \cdots Y_{22}(\omega) \cdots$

arm A_t is played $\quad X_t = Y_{tA_t}$

Moral: Number of times we play the non-optimal arm needs to be kept down.

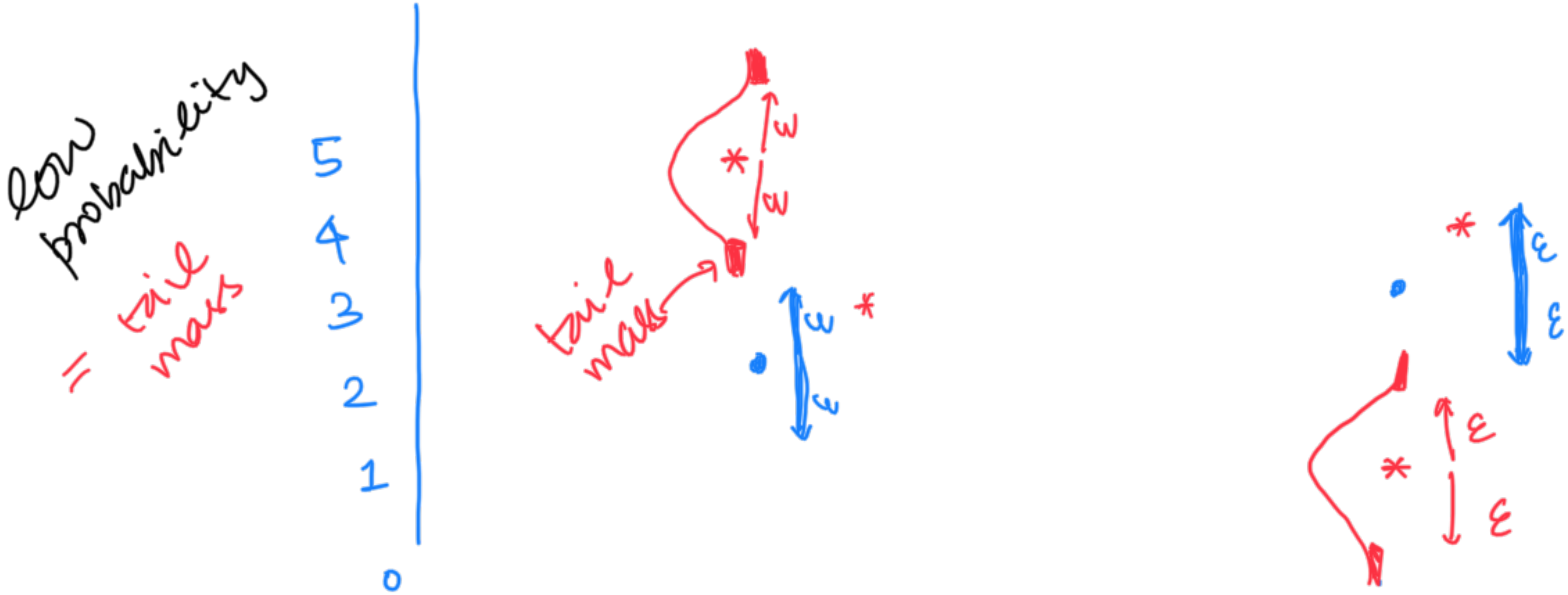
Consider a bandit problem with say 2 arms



$*$: True Mean

$\bullet$: Sample Mean

Tail Bound: $\mathbb{P}(|\underline{\text{Sample Mean}} - \text{True Mean}| > \underline{\varepsilon}) < \text{tail mass}$



Cannot rule out arm 2 and we might end up choosing arm 2 more often than we wish