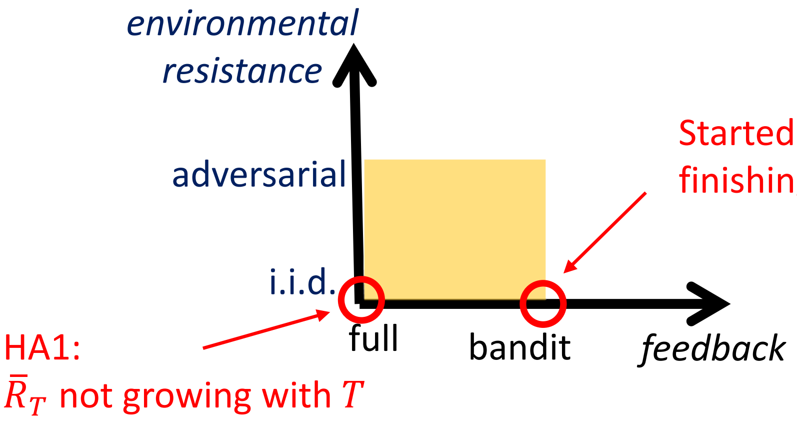
Stochastic Bandits
The UCB algorithm

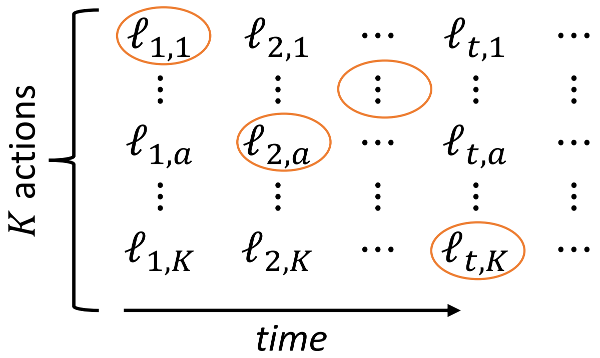
Yevgeny Seldin

Quick recap of the last lecture



Started last time, finishing today

HA1: $\bar{R}_T$ not growing with T



$$\begin{array}{ccccc} \ell_{1,1} & \ell_{2,1} & \cdots & \ell_{t,1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,a} & \ell_{2,a} & \cdots & \ell_{t,a} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,K} & \ell_{2,K} & \cdots & \ell_{t,K} & \cdots \end{array}$$

- Regret: $R_T = \sum_{t=1}^T \ell_{t,A_t} - \min_a \sum_{t=1}^T \ell_{t,a}$
- Expected regret: $\mathbb{E}[R_T] = \mathbb{E}\left[\sum_{t=1}^T \ell_{t,A_t}\right] - \mathbb{E}\left[\min_a \sum_{t=1}^T \ell_{t,a}\right]$
- Pseudo-regret: $\bar{R}_T = \mathbb{E}\left[\sum_{t=1}^T \ell_{t,A_t}\right] - \min_a \mathbb{E}\left[\sum_{t=1}^T \ell_{t,a}\right] = \mathbb{E}\left[\sum_{t=1}^T \ell_{t,A_t}\right] - T\mu^*$

  $= \sum_{a=1}^K \Delta(a)\mathbb{E}[N_T(a)]$

Lower Confidence Bound (LCB) algorithm for losses (Originally Upper Confidence Bound (UCB) for rewards) ("Optimism in the face of uncertainty" approach)

- Define $L_t^{CB}(a) = \hat{\mu}_{t-1}(a) - \sqrt{\frac{3 \ln t}{2N_{t-1}(a)}}$ lower confidence bound
  - (We will show that with high probability $L_t^{CB}(a) \le \mu(a)$ for all t)

- LCB Algorithm:
  - Play each arm once
  - For $t = K+1, K+2, \ldots$:
    - Play $A_t = \arg\min_a L_t^{CB}(a)$

- No knowledge of T
- No knowledge of Δ
- Works for any K

Rewards ↔ Losses

$$\ell_{t,a} = 1 - r_{t,a}$$

$$r_{t,a} = 1 - \ell_{t,a}$$

- Theorem:

$$\bar{R}_T \le 6 \sum_{a:\Delta(a)>0} \frac{\ln T}{\Delta(a)} + \left(1 + \frac{\pi^2}{3}\right) \sum_a \Delta(a)$$

Lower Confidence Bound (LCB) algorithm for losses (Originally Upper Confidence Bound (UCB) for rewards) ("Optimism in the face of uncertainty" approach)

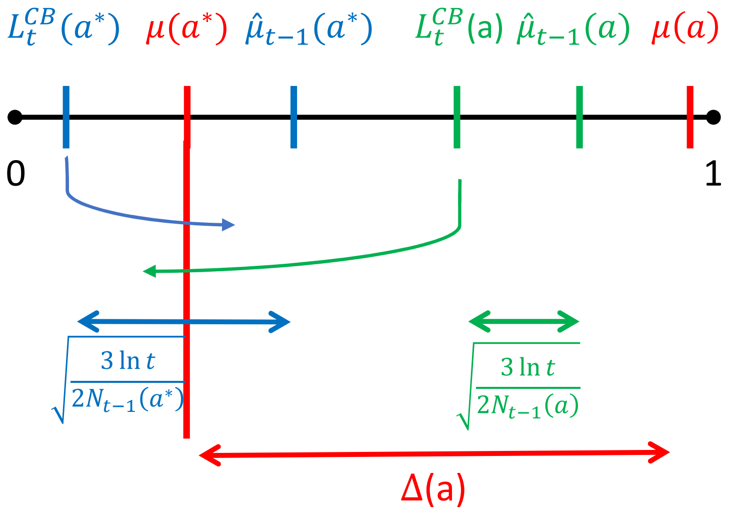
- Define $L_t^{CB}(a) = \hat{\mu}_{t-1}(a) - \sqrt{\frac{3 \ln t}{2N_{t-1}(a)}}$ lower confidence bound
  - (We will show that with high probability $L_t^{CB}(a) \leq \mu(a)$ for all t)

- LCB Algorithm:
  - Play each arm once
  - For $t = K+1, K+2, \ldots$:
    - Play $A_t = \arg\min_a L_t^{CB}(a)$

- Theorem:

$$\bar{R}_T \leq 6 \sum_{a:\Delta(a)>0} \frac{\ln T}{\Delta(a)} + \left(1 + \frac{\pi^2}{3}\right) \sum_a \Delta(a)$$

- Proof:
  - $\bar{R}_T = \sum_{a=1}^{K} \Delta(a) \mathbb{E}[N_T(a)]$
  - When can we play $a \neq a^*$?
  - Bound the expected number of times $L_t^{CB}(a) \leq L_t^{CB}(a^*)$



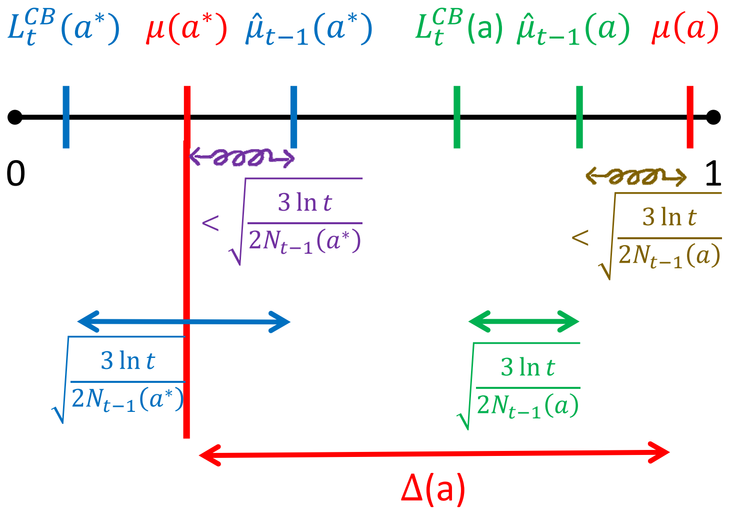
Proof

- $\bar{R}_t(a) = \sum_a \Delta(a) \mathbb{E}[N_T(a)]$
- $L_t^{CB}(a) = \hat{\mu}_{t-1}(a) - \sqrt{\frac{3 \ln t}{2N_{t-1}(a)}}$

- Bound the expected number of times $L_t^{CB}(a) \leq L_t^{CB}(a^*)$

- The expected number of times $L_t^{CB}(a) \leq L_t^{CB}(a^*)$ is bounded by
  1. The expected number of times $L_t^{CB}(a^*) \geq \mu(a^*)$
  2. Plus expected the number of times $L_t^{CB}(a) \leq \mu(a^*)$

Proof continued



$L_t^{CB}(a^*)$
$\mu(a^*)$
$\hat{\mu}_{t-1}(a^*)$
$L_t^{CB}(\mathrm{a})$
$\hat{\mu}_{t-1}(a)$
$\mu(a)$
0
1
$< \sqrt{\frac{3\ln t}{2N_{t-1}(a^*)}}$
$< \sqrt{\frac{3\ln t}{2N_{t-1}(a)}}$
$\sqrt{\frac{3\ln t}{2N_{t-1}(a^*)}}$
$\sqrt{\frac{3\ln t}{2N_{t-1}(a)}}$
Δ(a)

1. The expected number of times $L_t^{CB}(a^*) \geq \mu(a^*)$ is bounded by

   The expected number of times $\hat{\mu}_{t-1}(a^*) \geq \mu(a^*) + \sqrt{\frac{3\ln t}{2N_{t-1}(a^*)}}$

   (↭ “breaks” – the confidence bound fails)

2. The expected the number of times $L_t^{CB}(a) \leq \mu(a^*)$ is bounded by

   2.1 The expected number of times $\hat{\mu}_{t-1}(a) \leq \mu(a) - \sqrt{\frac{3\ln t}{2N_{t-1}(a)}}$

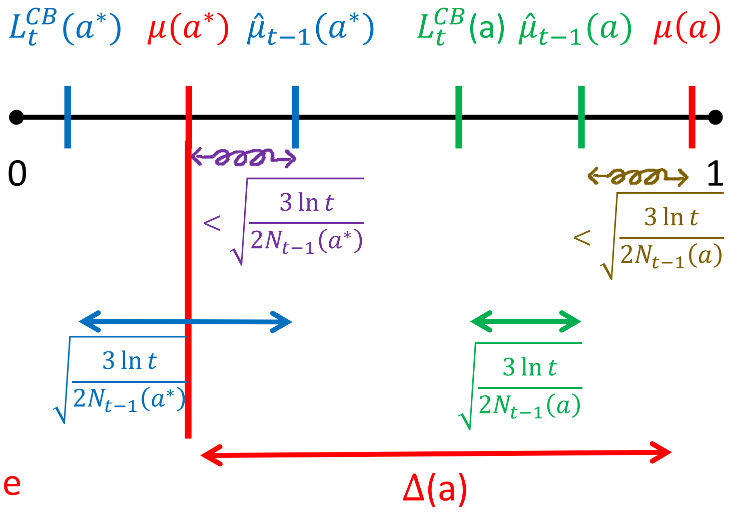
   (↭ “breaks” – the confidence bound fails)

   2.2 Plus the number of times $\sqrt{\frac{3\ln t}{2N_{t-1}(a)}} + \sqrt{\frac{3\ln t}{2N_{t-1}(a)}} = \sqrt{\frac{6\ln t}{N_{t-1}(a)}} \geq \Delta(a)$

   (The confidence bound holds, but it is not tight enough)

   $\Rightarrow \quad N_{t-1}(a) \leq \frac{6\ln t}{\Delta(a)^2} \leq \frac{6\ln T}{\Delta(a)^2}$

- Mid-summary: $\mathbb{E}[N_T(a)] \leq \frac{6\ln T}{\Delta(a)^2} + 1 + \mathbb{E}[1.] + \mathbb{E}[2.1]$


$L_t^{CB}(a^*)$
$\mu(a^*)$
$\hat{\mu}_{t-1}(a^*)$
L_t^{CB}(a)
$\hat{\mu}_{t-1}(a)$
$\mu(a)$
0
1
$< \sqrt{\frac{3\ln t}{2N_{t-1}(a^*)}}$
$< \sqrt{\frac{3\ln t}{2N_{t-1}(a)}}$
$\sqrt{\frac{3\ln t}{2N_{t-1}(a^*)}}$
$\sqrt{\frac{3\ln t}{2N_{t-1}(a)}}$
Δ(a)

Proof continued

- $\mathbb{E}[N_T(a)] \leq \frac{6\ln T}{\Delta(a)^2} + 1 + \mathbb{E}[\cancel{\leftrightsquigarrow}] + \mathbb{E}[\cancel{\leftrightsquigarrow}]$
- Let $F(a^*)$ be the expected number of times $\hat{\mu}_{t-1}(a^*) \geq \mu(a^*) + \sqrt{\frac{3\ln t}{2N_{t-1}(a^*)}}$
- Bound $\mathbb{P}\left(\hat{\mu}_{t-1}(a^*) - \mu(a^*) \geq \sqrt{\frac{3\ln t}{2N_{t-1}(a^*)}}\right)$

$N_{t-1}(a^*)$ is a random variable dependent on $\hat{\mu}_t(a^*)$!

- Idea: break dependent events into independent events and take a union bound
- Introduce $X_1, \ldots, X_T$ r.v. with the same distribution as ℓ_{t,a^*}
- Let $\bar{\mu}_s = \frac{1}{s}\sum_{i=1}^{s} X_i$
- $\mathbb{P}\left(\hat{\mu}_{t-1}(a^*) - \mu(a^*) \geq \sqrt{\frac{3\ln t}{2N_{t-1}(a^*)}}\right) \leq \mathbb{P}\left(\exists s \in \{1,\ldots,t\}: \bar{\mu}_s - \mu(a^*) \geq \sqrt{\frac{\ln t^3}{2s}}\right)$

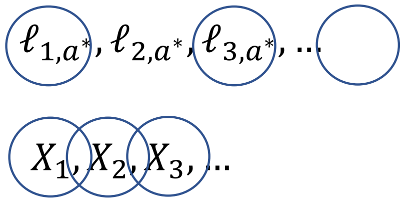
$$\underbrace{\leq}_{\text{union}} \sum_{s=1}^{t} \mathbb{P}\left(\bar{\mu}_s - \mu(a^*) \geq \sqrt{\frac{\ln t^3}{2s}}\right)$$

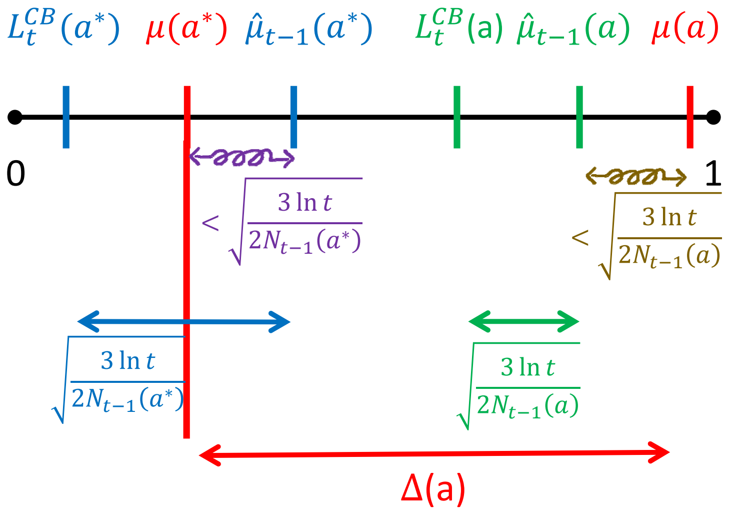
$$\underbrace{\leq}_{\text{Hoeffding}} \sum_{s=1}^{t} \frac{1}{t^3} = \frac{1}{t^2}$$

- $\mathbb{E}[F(a^*)] = \sum_{t=1}^{\infty} \mathbb{P}\left(L_t^{CB}(a^*) \geq \mu(a^*)\right) \leq \sum_{t=1}^{\infty} \frac{1}{t^2} \leq \frac{\pi^2}{6}$

Hoeffding:

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - \mu \geq \sqrt{\frac{\ln\frac{1}{\delta}}{2n}}\right) \leq \delta$$




$L_t^{CB}(a^*)$
$\mu(a^*)$
$\hat{\mu}_{t-1}(a^*)$
L_t^{CB}(a)
$\hat{\mu}_{t-1}(a)$
$\mu(a)$
0
1
$< \sqrt{\frac{3\ln t}{2N_{t-1}(a^*)}}$
$< \sqrt{\frac{3\ln t}{2N_{t-1}(a)}}$
$\sqrt{\frac{3\ln t}{2N_{t-1}(a^*)}}$
$\sqrt{\frac{3\ln t}{2N_{t-1}(a)}}$
Δ(a)

Proof summary

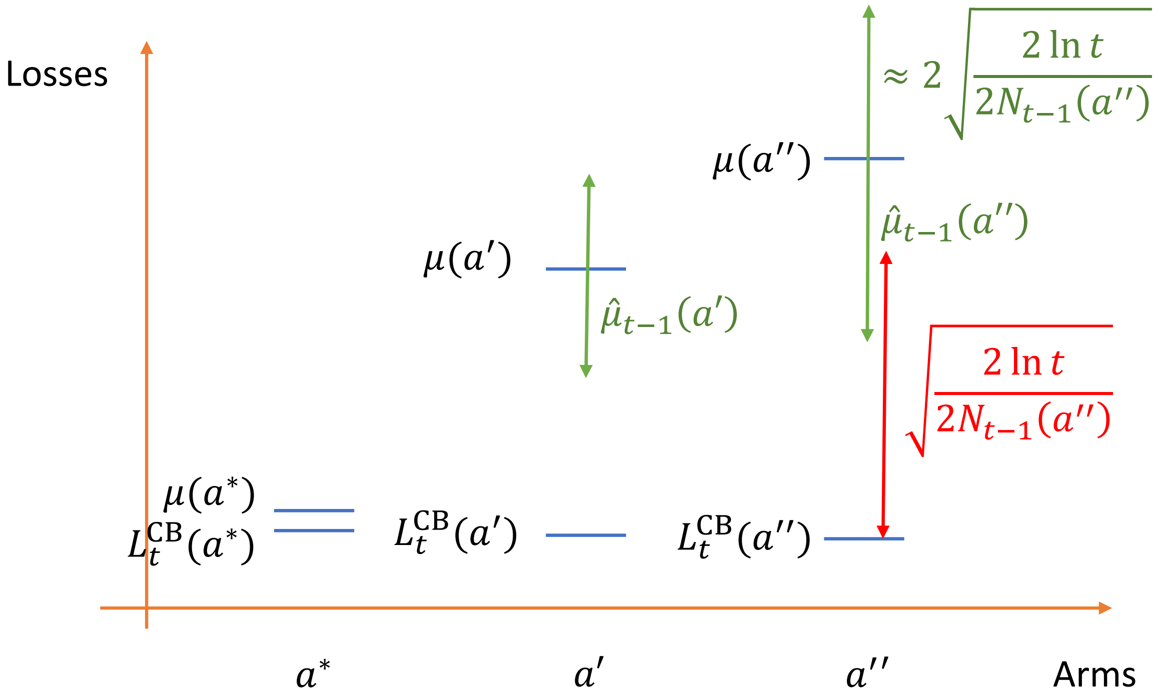
- $\bar{R}_t(a) = \sum_a \Delta(a)\mathbb{E}[N_T(a)]$

- $\mathbb{E}[N_T(a)] \le \underbrace{\frac{6\ln T}{\Delta(a)^2}}_{\text{The time it takes for confidence intervals to start working}} + 1 + \underbrace{\frac{\pi^2}{6} + \frac{\pi^2}{6}}_{\text{The expected number of times confidence intervals fail}}$

- $\bar{R}_T \le 6\sum_{a:\Delta(a)>0}\frac{\ln T}{\Delta(a)} + \left(1+\frac{\pi^2}{3}\right)\sum_a \Delta(a)$

- Home assignment:
  - Take $L_t^{CB}(a) = \hat{\mu}_{t-1}(a) - \sqrt{\frac{2\ln t}{2N_{t-1}(a)}}$ (instead of $L_t^{CB}(a) = \hat{\mu}_{t-1}(a) - \sqrt{\frac{3\ln t}{2N_{t-1}(a)}}$; i.e. confidence $\frac{1}{t^2}$ instead $\frac{1}{t^3}$)
  - Show $\bar{R}_T \le 4\sum_{a:\Delta(a)>0}\frac{\ln T}{\Delta(a)} + (2\ln T + 3)\sum_a \Delta(a)$

LCB algorithm dynamics (with $L_t^{CB}(a) = \hat{\mu}_{t-1}(a) - \sqrt{\frac{2\ln t}{2N_{t-1}(a)}}$)



- Confidence interval of the played arm shrinks ($N_{t-1}(a)$ grows)
- Confidence intervals of all other arms grow ($\ln t$ grows)
- $\Rightarrow$ all LCBs are roughly at the same level
- Most of the time $L_t^{\text{CB}}(a^*) \leq \mu(a^*)$
- a^* is played a lot, so $L_t^{\text{CB}}(a^*)$ is very close to $\mu(a^*)$
- All other arms are played just enough to keep $\sqrt{\frac{2\ln t}{2N_{t-1}(a)}} = \theta(\Delta(a))$, i.e. $N_t(a) = \theta\left(\frac{\ln t}{\Delta(a)^2}\right)$