

CSE 541: Interactive Learning

Lecture 6: Multi-armed bandits, stochastic and adversarial

Kevin Jamieson

Paul G. Allen School of
Computer Science & Engineering
University of Washington
Monday, October 19, 2026



Last time: the multi-armed bandit

Stochastic multi-armed bandit

n arms, unknown 1-sub-Gaussian reward distributions with means θ_i^*
for $t = 1, 2, \dots$: choose $I_t \in [n]$; observe $X_{I_t, t} \sim P_{I_t}$ only

Gaps $\Delta_i = \max_j \theta_j^* - \theta_i^*$; T_i = number of pulls of arm i . Regret:
$$R_T = \max_j \theta_j^* T - \mathbb{E} \left[\sum_{t \leq T} X_{I_t, t} \right] = \sum_{i=1}^n \Delta_i \mathbb{E}[T_i].$$

Today: finish the two-arm warm-up, then an algorithm that adapts to gaps it does not know, then drop the i.i.d. assumption entirely.

Two goals

Regret: pull bad arms rarely.

Best-arm identification: find $\arg \max_i \theta_i^*$ w.p. $\geq 1 - \delta$ with as few pulls as possible.

The tool, again: for an empirical mean of τ pulls, the Chernoff method with sub-Gaussianity gives $\mathbb{P}(|\hat{\theta}_i - \theta_i^*| > \sqrt{2 \log(2/\delta)/\tau}) \leq \delta$.

Warm-up: A/B testing ($n = 2$)

Say $\theta_1^* > \theta_2^*$ and $\Delta = \theta_1^* - \theta_2^*$. The simplest possible algorithm:

Explore then commit

Input: $\tau \in \mathbb{N}$. Pull each arm exactly τ times; compute empirical means $\hat{\theta}_1, \hat{\theta}_2$.
For all $t > 2\tau$ play $\arg \max_i \hat{\theta}_i$.

How large must τ be so that $\hat{\theta}_1 > \hat{\theta}_2$ with probability $\geq 1 - \delta$? The CLT says $\hat{\theta}_i - \theta_i^* \approx \mathcal{N}(0, 1/\tau)$, but we want a non-asymptotic statement: the **Chernoff method** from Lecture 1, with sub-Gaussianity in place of Hoeffding's lemma,

$$\mathbb{P}(\hat{\theta}_i - \theta_i^* > \epsilon) \leq$$

Warm-up: A/B testing ($n = 2$)

Say $\theta_1^* > \theta_2^*$ and $\Delta = \theta_1^* - \theta_2^*$. The simplest possible algorithm:

Explore then commit

Input: $\tau \in \mathbb{N}$. Pull each arm exactly τ times; compute empirical means $\hat{\theta}_1, \hat{\theta}_2$.
For all $t > 2\tau$ play $\arg \max_i \hat{\theta}_i$.

How large must τ be so that $\hat{\theta}_1 > \hat{\theta}_2$ with probability $\geq 1 - \delta$? The CLT says $\hat{\theta}_i - \theta_i^* \approx \mathcal{N}(0, 1/\tau)$, but we want a non-asymptotic statement: the **Chernoff method** from Lecture 1, with sub-Gaussianity in place of Hoeffding's lemma,

$$\mathbb{P}(\hat{\theta}_i - \theta_i^* > \epsilon) \leq \inf_{\lambda > 0} e^{-\lambda \tau \epsilon} \mathbb{E} \left[e^{\lambda \sum_{s=1}^{\tau} (X_{i,s} - \theta_i^*)} \right] \leq \inf_{\lambda > 0} e^{-\lambda \tau \epsilon + \tau \lambda^2 / 2} = e^{-\tau \epsilon^2 / 2}.$$

Warm-up: A/B testing ($n = 2$)

Say $\theta_1^* > \theta_2^*$ and $\Delta = \theta_1^* - \theta_2^*$. The simplest possible algorithm:

Explore then commit

Input: $\tau \in \mathbb{N}$. Pull each arm exactly τ times; compute empirical means $\hat{\theta}_1, \hat{\theta}_2$.
For all $t > 2\tau$ play $\arg \max_i \hat{\theta}_i$.

How large must τ be so that $\hat{\theta}_1 > \hat{\theta}_2$ with probability $\geq 1 - \delta$? The CLT says $\hat{\theta}_i - \theta_i^* \approx \mathcal{N}(0, 1/\tau)$, but we want a non-asymptotic statement: the **Chernoff method** from Lecture 1, with sub-Gaussianity in place of Hoeffding's lemma,

$$\mathbb{P}(\hat{\theta}_i - \theta_i^* > \epsilon) \leq \inf_{\lambda > 0} e^{-\lambda \tau \epsilon} \mathbb{E} \left[e^{\lambda \sum_{s=1}^{\tau} (X_{i,s} - \theta_i^*)} \right] \leq \inf_{\lambda > 0} e^{-\lambda \tau \epsilon + \tau \lambda^2 / 2} = e^{-\tau \epsilon^2 / 2}.$$

So $E_i = \{|\hat{\theta}_i - \theta_i^*| \leq \sqrt{2 \log(4/\delta)/\tau}\}$ fails with probability $\leq \delta/2$, and on $E_1 \cap E_2$ with $\tau = \lceil 8\Delta^{-2} \log(4/\delta) \rceil$ (so that $\sqrt{2 \log(4/\delta)/\tau} \leq \Delta/2$),

$$\hat{\theta}_1 >$$

Warm-up: A/B testing ($n = 2$)

Say $\theta_1^* > \theta_2^*$ and $\Delta = \theta_1^* - \theta_2^*$. The simplest possible algorithm:

Explore then commit

Input: $\tau \in \mathbb{N}$. Pull each arm exactly τ times; compute empirical means $\hat{\theta}_1, \hat{\theta}_2$.
For all $t > 2\tau$ play $\arg \max_i \hat{\theta}_i$.

How large must τ be so that $\hat{\theta}_1 > \hat{\theta}_2$ with probability $\geq 1 - \delta$? The CLT says $\hat{\theta}_i - \theta_i^* \approx \mathcal{N}(0, 1/\tau)$, but we want a non-asymptotic statement: the **Chernoff method** from Lecture 1, with sub-Gaussianity in place of Hoeffding's lemma,

$$\mathbb{P}(\hat{\theta}_i - \theta_i^* > \epsilon) \leq \inf_{\lambda > 0} e^{-\lambda \tau \epsilon} \mathbb{E} \left[e^{\lambda \sum_{s=1}^{\tau} (X_{i,s} - \theta_i^*)} \right] \leq \inf_{\lambda > 0} e^{-\lambda \tau \epsilon + \tau \lambda^2 / 2} = e^{-\tau \epsilon^2 / 2}.$$

So $E_i = \{|\hat{\theta}_i - \theta_i^*| \leq \sqrt{2 \log(4/\delta)/\tau}\}$ fails with probability $\leq \delta/2$, and on $E_1 \cap E_2$ with $\tau = \lceil 8\Delta^{-2} \log(4/\delta) \rceil$ (so that $\sqrt{2 \log(4/\delta)/\tau} \leq \Delta/2$),

$$\hat{\theta}_1 > \theta_1^* - \frac{\Delta}{2} = \theta_2^* + \frac{\Delta}{2} \geq \hat{\theta}_2 - \frac{\Delta}{2} + \frac{\Delta}{2} = \hat{\theta}_2.$$

With probability $\geq 1 - \delta$ we identify the best arm after $2\tau = O(\Delta^{-2} \log(1/\delta))$ pulls, then play it forever.

A/B testing: the regret, and a picture we will see again

On $E_1 \cap E_2$ the bad arm is pulled at most τ times; otherwise at most T times.
So

$$R_T = \Delta \mathbb{E}[T_2] \leq \\ \leq$$

A/B testing: the regret, and a picture we will see again

On $E_1 \cap E_2$ the bad arm is pulled at most τ times; otherwise at most T times.
So

$$\begin{aligned} R_T = \Delta \mathbb{E}[T_2] &\leq \tau\Delta + \Delta T \mathbb{P}(E_1^c \cup E_2^c) \\ &\leq \frac{8 \log(4/\delta)}{\Delta} + \Delta + \Delta T \delta \stackrel{\delta=1/T}{\leq} \Delta + \frac{8 \log(4T)}{\Delta}. \end{aligned}$$

A/B testing: the regret, and a picture we will see again

On $E_1 \cap E_2$ the bad arm is pulled at most τ times; otherwise at most T times.
So

$$\begin{aligned} R_T = \Delta \mathbb{E}[T_2] &\leq \tau \Delta + \Delta T \mathbb{P}(E_1^c \cup E_2^c) \\ &\leq \frac{8 \log(4/\delta)}{\Delta} + \Delta + \Delta T \delta \stackrel{\delta=1/T}{\leq} \Delta + \frac{8 \log(4T)}{\Delta}. \end{aligned}$$

Regret can never exceed ΔT either, so (for $\Delta \leq 1$)

$$R_T \leq \min \{ \Delta T, \Delta + 8 \log(4T)/\Delta \} \leq$$

A/B testing: the regret, and a picture we will see again

On $E_1 \cap E_2$ the bad arm is pulled at most τ times; otherwise at most T times.
So

$$\begin{aligned} R_T = \Delta \mathbb{E}[T_2] &\leq \tau \Delta + \Delta T \mathbb{P}(E_1^c \cup E_2^c) \\ &\leq \frac{8 \log(4/\delta)}{\Delta} + \Delta + \Delta T \delta \stackrel{\delta=1/T}{\leq} \Delta + \frac{8 \log(4T)}{\Delta}. \end{aligned}$$

Regret can never exceed ΔT either, so (for $\Delta \leq 1$)

$$R_T \leq \min \{ \Delta T, \Delta + 8 \log(4T)/\Delta \} \leq 1 + 2\sqrt{8T \log(4T)},$$

the worst case being $\Delta \approx \sqrt{8 \log(4T)/T}$.

A/B testing: the regret, and a picture we will see again

On $E_1 \cap E_2$ the bad arm is pulled at most τ times; otherwise at most T times. So

$$\begin{aligned} R_T = \Delta \mathbb{E}[T_2] &\leq \tau \Delta + \Delta T \mathbb{P}(E_1^c \cup E_2^c) \\ &\leq \frac{8 \log(4/\delta)}{\Delta} + \Delta + \Delta T \delta \stackrel{\delta=1/T}{\leq} \Delta + \frac{8 \log(4T)}{\Delta}. \end{aligned}$$

Regret can never exceed ΔT either, so (for $\Delta \leq 1$)

$$R_T \leq \min \{ \Delta T, \Delta + 8 \log(4T)/\Delta \} \leq 1 + 2\sqrt{8T \log(4T)},$$

the worst case being $\Delta \approx \sqrt{8 \log(4T)/T}$.

Takeaways

- Tiny gaps cost little per pull; large gaps are easy to detect; the hard ones are $\approx 1/\sqrt{T}$. This recurs all quarter.
- **The catch:** τ used Δ , which we do not know. Next: elimination, which adapts to the gaps (notes §10.2–10.3).
- Experts: $\sqrt{T \log d}$. Bandit feedback: $\sqrt{nT}$; $\log n \rightarrow n$ is the price of one reward per round.

Successive elimination: adapt to the gaps you do not know

Idea: keep a set of surviving arms; pull them equally; at geometrically increasing precision, throw out any arm that is confidently worse than the current leader. Confidence intervals are the Chernoff bound from the A/B slide.

Successive elimination (notes, §10.2)

Input: arms $\mathcal{X} = [n]$, confidence $\delta \in (0, 1)$. **Initialize** $\mathcal{X}_1 = \mathcal{X}$, $\ell = 1$.

while $|\mathcal{X}_\ell| > 1$:

$\epsilon_\ell = 2^{-\ell}$; pull each arm in $\mathcal{X}_\ell$ exactly $\tau_\ell = \lceil 2\epsilon_\ell^{-2} \log(4\ell^2 n/\delta) \rceil$ times

 compute the empirical means $\hat{\theta}_{i,\ell}$ of these τ_ℓ rewards, $i \in \mathcal{X}_\ell$

$\mathcal{X}_{\ell+1} = \mathcal{X}_\ell \setminus \{i \in \mathcal{X}_\ell : \max_{j \in \mathcal{X}_\ell} \hat{\theta}_{j,\ell} - \hat{\theta}_{i,\ell} > 2\epsilon_\ell\}$; $\ell \leftarrow \ell + 1$

Output the surviving arm (for regret: keep playing it forever).

Successive elimination: adapt to the gaps you do not know

Idea: keep a set of surviving arms; pull them equally; at geometrically increasing precision, throw out any arm that is confidently worse than the current leader. Confidence intervals are the Chernoff bound from the A/B slide.

Successive elimination (notes, §10.2)

Input: arms $\mathcal{X} = [n]$, confidence $\delta \in (0, 1)$. **Initialize** $\mathcal{X}_1 = \mathcal{X}$, $\ell = 1$.

while $|\mathcal{X}_\ell| > 1$:

$\epsilon_\ell = 2^{-\ell}$; pull each arm in $\mathcal{X}_\ell$ exactly $\tau_\ell = \lceil 2\epsilon_\ell^{-2} \log(4\ell^2 n/\delta) \rceil$ times

compute the empirical means $\hat{\theta}_{i,\ell}$ of these τ_ℓ rewards, $i \in \mathcal{X}_\ell$

$\mathcal{X}_{\ell+1} = \mathcal{X}_\ell \setminus \{i \in \mathcal{X}_\ell : \max_{j \in \mathcal{X}_\ell} \hat{\theta}_{j,\ell} - \hat{\theta}_{i,\ell} > 2\epsilon_\ell\}$; $\ell \leftarrow \ell + 1$

Output the surviving arm (for regret: keep playing it forever).

- τ_ℓ is chosen so that $\sqrt{2 \log(4\ell^2 n/\delta)/\tau_\ell} \leq \epsilon_\ell$: each empirical mean in round ℓ is within ϵ_ℓ of its true mean with probability $\geq 1 - \delta/(2n\ell^2)$.
- The ℓ^2 inside the log pays for a union bound over infinitely many rounds: $\sum_\ell 1/\ell^2 < 2$.
- Precision doubles each round and the sample size quadruples, so the *last* round dominates the cost. No Δ anywhere.

The best arm survives, and the survivors' gaps shrink

Assume $\max_i \Delta_i \leq 4$ and let arm 1 be optimal. Define the good event $E = \bigcap_{i \in [n]} \bigcap_{\ell \geq 1} \{|\hat{\theta}_{i,\ell} - \theta_i^*| \leq \epsilon_\ell\}$.

Lemma

$\mathbb{P}(E) \geq 1 - \delta$, and on E : for every ℓ , $1 \in \mathcal{X}_\ell$ and $\max_{i \in \mathcal{X}_\ell} \Delta_i \leq 8\epsilon_\ell$.

The best arm survives, and the survivors' gaps shrink

Assume $\max_i \Delta_i \leq 4$ and let arm 1 be optimal. Define the good event $E = \bigcap_{i \in [n]} \bigcap_{\ell \geq 1} \{|\hat{\theta}_{i,\ell} - \theta_i^*| \leq \epsilon_\ell\}$.

Lemma

$\mathbb{P}(E) \geq 1 - \delta$, and on E : for every ℓ , $1 \in \mathcal{X}_\ell$ and $\max_{i \in \mathcal{X}_\ell} \Delta_i \leq 8\epsilon_\ell$.

Probability. Chernoff with τ_ℓ gives $\mathbb{P}(|\hat{\theta}_{i,\ell} - \theta_i^*| > \epsilon_\ell) \leq 2e^{-\tau_\ell \epsilon_\ell^2/2} \leq \delta/(2n\ell^2)$, so $\mathbb{P}(E^c) \leq$

The best arm survives, and the survivors' gaps shrink

Assume $\max_i \Delta_i \leq 4$ and let arm 1 be optimal. Define the good event $E = \bigcap_{i \in [n]} \bigcap_{\ell \geq 1} \{|\hat{\theta}_{i,\ell} - \theta_i^*| \leq \epsilon_\ell\}$.

Lemma

$\mathbb{P}(E) \geq 1 - \delta$, and on E : for every ℓ , $1 \in \mathcal{X}_\ell$ and $\max_{i \in \mathcal{X}_\ell} \Delta_i \leq 8\epsilon_\ell$.

Probability. Chernoff with τ_ℓ gives $\mathbb{P}(|\hat{\theta}_{i,\ell} - \theta_i^*| > \epsilon_\ell) \leq 2e^{-\tau_\ell \epsilon_\ell^2 / 2} \leq \delta / (2n\ell^2)$, so $\mathbb{P}(E^c) \leq \sum_{i=1}^n \sum_{\ell \geq 1} \frac{\delta}{2n\ell^2} \leq \delta$.

The best arm survives, and the survivors' gaps shrink

Assume $\max_i \Delta_i \leq 4$ and let arm 1 be optimal. Define the good event $E = \bigcap_{i \in [n]} \bigcap_{\ell \geq 1} \{|\hat{\theta}_{i,\ell} - \theta_i^*| \leq \epsilon_\ell\}$.

Lemma

$\mathbb{P}(E) \geq 1 - \delta$, and on E : for every ℓ , $1 \in \mathcal{X}_\ell$ and $\max_{i \in \mathcal{X}_\ell} \Delta_i \leq 8\epsilon_\ell$.

Probability. Chernoff with τ_ℓ gives $\mathbb{P}(|\hat{\theta}_{i,\ell} - \theta_i^*| > \epsilon_\ell) \leq 2e^{-\tau_\ell \epsilon_\ell^2/2} \leq \delta/(2n\ell^2)$, so $\mathbb{P}(E^c) \leq \sum_{i=1}^n \sum_{\ell \geq 1} \frac{\delta}{2n\ell^2} \leq \delta$.

Arm 1 is never eliminated. If $1 \in \mathcal{X}_\ell$, then for any $j \in \mathcal{X}_\ell$, on E ,

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

The best arm survives, and the survivors' gaps shrink

Assume $\max_i \Delta_i \leq 4$ and let arm 1 be optimal. Define the good event $E = \bigcap_{i \in [n]} \bigcap_{\ell \geq 1} \{|\hat{\theta}_{i,\ell} - \theta_i^*| \leq \epsilon_\ell\}$.

Lemma

$\mathbb{P}(E) \geq 1 - \delta$, and on E : for every ℓ , $1 \in \mathcal{X}_\ell$ and $\max_{i \in \mathcal{X}_\ell} \Delta_i \leq 8\epsilon_\ell$.

Probability. Chernoff with τ_ℓ gives $\mathbb{P}(|\hat{\theta}_{i,\ell} - \theta_i^*| > \epsilon_\ell) \leq 2e^{-\tau_\ell \epsilon_\ell^2/2} \leq \delta/(2n\ell^2)$, so $\mathbb{P}(E^c) \leq \sum_{i=1}^n \sum_{\ell \geq 1} \frac{\delta}{2n\ell^2} \leq \delta$.

Arm 1 is never eliminated. If $1 \in \mathcal{X}_\ell$, then for any $j \in \mathcal{X}_\ell$, on E ,

$$\hat{\theta}_{j,\ell} - \hat{\theta}_{1,\ell} = (\hat{\theta}_{j,\ell} - \theta_j^*) - (\hat{\theta}_{1,\ell} - \theta_1^*) - \Delta_j \leq \epsilon_\ell + \epsilon_\ell - \Delta_j \leq 2\epsilon_\ell,$$

so arm 1 does not trail the leader by more than $2\epsilon_\ell$ and stays in $\mathcal{X}_{\ell+1}$. By induction, $1 \in \mathcal{X}_\ell$ for all ℓ .

The best arm survives, and the survivors' gaps shrink

Assume $\max_i \Delta_i \leq 4$ and let arm 1 be optimal. Define the good event $E = \bigcap_{i \in [n]} \bigcap_{\ell \geq 1} \{|\hat{\theta}_{i,\ell} - \theta_i^*| \leq \epsilon_\ell\}$.

Lemma

$\mathbb{P}(E) \geq 1 - \delta$, and on E : for every ℓ , $1 \in \mathcal{X}_\ell$ and $\max_{i \in \mathcal{X}_\ell} \Delta_i \leq 8\epsilon_\ell$.

Probability. Chernoff with τ_ℓ gives $\mathbb{P}(|\hat{\theta}_{i,\ell} - \theta_i^*| > \epsilon_\ell) \leq 2e^{-\tau_\ell \epsilon_\ell^2/2} \leq \delta/(2n\ell^2)$, so $\mathbb{P}(E^c) \leq \sum_{i=1}^n \sum_{\ell \geq 1} \frac{\delta}{2n\ell^2} \leq \delta$.

Arm 1 is never eliminated. If $1 \in \mathcal{X}_\ell$, then for any $j \in \mathcal{X}_\ell$, on E ,

$$\hat{\theta}_{j,\ell} - \hat{\theta}_{1,\ell} = (\hat{\theta}_{j,\ell} - \theta_j^*) - (\hat{\theta}_{1,\ell} - \theta_1^*) - \Delta_j \leq \epsilon_\ell + \epsilon_\ell - \Delta_j \leq 2\epsilon_\ell,$$

so arm 1 does not trail the leader by more than $2\epsilon_\ell$ and stays in $\mathcal{X}_{\ell+1}$. By induction, $1 \in \mathcal{X}_\ell$ for all ℓ .

Bad arms are eliminated. If $i \in \mathcal{X}_\ell$ has $\Delta_i > 4\epsilon_\ell$, then on E
 $\max_j \hat{\theta}_{j,\ell} - \hat{\theta}_{i,\ell} \geq \hat{\theta}_{1,\ell} - \hat{\theta}_{i,\ell} =$

The best arm survives, and the survivors' gaps shrink

Assume $\max_i \Delta_i \leq 4$ and let arm 1 be optimal. Define the good event $E = \bigcap_{i \in [n]} \bigcap_{\ell \geq 1} \{|\hat{\theta}_{i,\ell} - \theta_i^*| \leq \epsilon_\ell\}$.

Lemma

$\mathbb{P}(E) \geq 1 - \delta$, and on E : for every ℓ , $1 \in \mathcal{X}_\ell$ and $\max_{i \in \mathcal{X}_\ell} \Delta_i \leq 8\epsilon_\ell$.

Probability. Chernoff with τ_ℓ gives $\mathbb{P}(|\hat{\theta}_{i,\ell} - \theta_i^*| > \epsilon_\ell) \leq 2e^{-\tau_\ell \epsilon_\ell^2/2} \leq \delta/(2n\ell^2)$, so $\mathbb{P}(E^c) \leq \sum_{i=1}^n \sum_{\ell \geq 1} \frac{\delta}{2n\ell^2} \leq \delta$.

Arm 1 is never eliminated. If $1 \in \mathcal{X}_\ell$, then for any $j \in \mathcal{X}_\ell$, on E ,

$$\hat{\theta}_{j,\ell} - \hat{\theta}_{1,\ell} = (\hat{\theta}_{j,\ell} - \theta_j^*) - (\hat{\theta}_{1,\ell} - \theta_1^*) - \Delta_j \leq \epsilon_\ell + \epsilon_\ell - \Delta_j \leq 2\epsilon_\ell,$$

so arm 1 does not trail the leader by more than $2\epsilon_\ell$ and stays in $\mathcal{X}_{\ell+1}$. By induction, $1 \in \mathcal{X}_\ell$ for all ℓ .

Bad arms are eliminated. If $i \in \mathcal{X}_\ell$ has $\Delta_i > 4\epsilon_\ell$, then on E $\max_j \hat{\theta}_{j,\ell} - \hat{\theta}_{i,\ell} \geq \hat{\theta}_{1,\ell} - \hat{\theta}_{i,\ell} = (\hat{\theta}_{1,\ell} - \theta_1^*) - (\hat{\theta}_{i,\ell} - \theta_i^*) + \Delta_i > -2\epsilon_\ell + 4\epsilon_\ell = 2\epsilon_\ell$, so $i \notin \mathcal{X}_{\ell+1}$. Hence every survivor of round ℓ has $\Delta_i \leq 4\epsilon_\ell = 8\epsilon_{\ell+1}$ (and $\leq 4 = 8\epsilon_1$ at $\ell = 1$). ■

Pure exploration: how many pulls to find the best arm

Theorem (sample complexity)

Assume $\max_i \Delta_i \leq 4$. With probability at least $1 - \delta$, successive elimination outputs arm 1 after at most $c \sum_{i=2}^n \Delta_i^{-2} \log(n \log(\Delta_i^{-2})/\delta)$ pulls, for a universal constant c .

Pure exploration: how many pulls to find the best arm

Theorem (sample complexity)

Assume $\max_i \Delta_i \leq 4$. With probability at least $1 - \delta$, successive elimination outputs arm 1 after at most $c \sum_{i=2}^n \Delta_i^{-2} \log(n \log(\Delta_i^{-2})/\delta)$ pulls, for a universal constant c .

Proof. Work on E . By the lemma, arm $i \neq 1$ is still in $\mathcal{X}_\ell$ only if $\Delta_i \leq 8\epsilon_\ell = 8 \cdot 2^{-\ell}$, i.e. only for rounds $\ell \leq \ell_i := \lceil \log_2(8/\Delta_i) \rceil$. So its number of pulls is

$$T_i \leq \sum_{\ell=1}^{\ell_i} \tau_\ell =$$
$$\leq$$

Pure exploration: how many pulls to find the best arm

Theorem (sample complexity)

Assume $\max_i \Delta_i \leq 4$. With probability at least $1 - \delta$, successive elimination outputs arm 1 after at most $c \sum_{i=2}^n \Delta_i^{-2} \log(n \log(\Delta_i^{-2})/\delta)$ pulls, for a universal constant c .

Proof. Work on E . By the lemma, arm $i \neq 1$ is still in $\mathcal{X}_\ell$ only if $\Delta_i \leq 8\epsilon_\ell = 8 \cdot 2^{-\ell}$, i.e. only for rounds $\ell \leq \ell_i := \lceil \log_2(8/\Delta_i) \rceil$. So its number of pulls is

$$\begin{aligned} T_i &\leq \sum_{\ell=1}^{\ell_i} \tau_\ell = \sum_{\ell=1}^{\ell_i} \left\lceil 2 \cdot 4^\ell \log \frac{4\ell^2 n}{\delta} \right\rceil \leq c' 4^{\ell_i} \log \frac{4\ell_i^2 n}{\delta} \\ &\leq c \Delta_i^{-2} \log \frac{n \log(\Delta_i^{-2})}{\delta}, \end{aligned}$$

(geometric sum, last term dominates; $4^{\ell_i} \leq 64\Delta_i^{-2}$, $\ell_i^2 \lesssim \log \Delta_i^{-2}$). Arm 1 is pulled as often as the longest-surviving other arm: at most a factor 2. ■

Pure exploration: how many pulls to find the best arm

Theorem (sample complexity)

Assume $\max_i \Delta_i \leq 4$. With probability at least $1 - \delta$, successive elimination outputs arm 1 after at most $c \sum_{i=2}^n \Delta_i^{-2} \log(n \log(\Delta_i^{-2})/\delta)$ pulls, for a universal constant c .

Proof. Work on E . By the lemma, arm $i \neq 1$ is still in $\mathcal{X}_\ell$ only if $\Delta_i \leq 8\epsilon_\ell = 8 \cdot 2^{-\ell}$, i.e. only for rounds $\ell \leq \ell_i := \lceil \log_2(8/\Delta_i) \rceil$. So its number of pulls is

$$\begin{aligned} T_i &\leq \sum_{\ell=1}^{\ell_i} \tau_\ell = \sum_{\ell=1}^{\ell_i} \left\lceil 2 \cdot 4^\ell \log \frac{4\ell^2 n}{\delta} \right\rceil \leq c' 4^{\ell_i} \log \frac{4\ell_i^2 n}{\delta} \\ &\leq c \Delta_i^{-2} \log \frac{n \log(\Delta_i^{-2})}{\delta}, \end{aligned}$$

(geometric sum, last term dominates; $4^{\ell_i} \leq 64\Delta_i^{-2}$, $\ell_i^2 \lesssim \log \Delta_i^{-2}$). Arm 1 is pulled as often as the longest-surviving other arm: at most a factor 2. ■

Read it: Δ_i^{-2} pulls per arm up to logs, the A/B answer arm by arm, *without knowing any Δ_i* .

Regret of successive elimination

Same algorithm and lemma; now count regret $\sum_i \Delta_i T_i$ after T pulls;
 $L := \log(n \log T / \delta)$.

Theorem (regret)

Assume $\max_i \Delta_i \leq 4$. With probability at least $1 - \delta$, $\Delta_i T_i \lesssim \Delta_i^{-1} L$ for every arm i and every T . Consequently, with $\delta = 1/T$: $R_T \lesssim \sum_{i \geq 2} \log T / \Delta_i$ and $R_T \lesssim \sqrt{n T \log T}$.

Regret of successive elimination

Same algorithm and lemma; now count regret $\sum_i \Delta_i T_i$ after T pulls;
 $L := \log(n \log T / \delta)$.

Theorem (regret)

Assume $\max_i \Delta_i \leq 4$. With probability at least $1 - \delta$, $\Delta_i T_i \lesssim \Delta_i^{-1} L$ for every arm i and every T . Consequently, with $\delta = 1/T$: $R_T \lesssim \sum_{i \geq 2} \log T / \Delta_i$ and $R_T \lesssim \sqrt{n T \log T}$.

Gap-dependent. On E , $T_i \lesssim \Delta_i^{-2} L$ (last slide), so $\Delta_i T_i \lesssim \Delta_i^{-1} L$: each pull of a bad arm costs only Δ_i .

Regret of successive elimination

Same algorithm and lemma; now count regret $\sum_i \Delta_i T_i$ after T pulls;
 $L := \log(n \log T / \delta)$.

Theorem (regret)

Assume $\max_i \Delta_i \leq 4$. With probability at least $1 - \delta$, $\Delta_i T_i \lesssim \Delta_i^{-1} L$ for every arm i and every T . Consequently, with $\delta = 1/T$: $R_T \lesssim \sum_{i \geq 2} \log T / \Delta_i$ and $R_T \lesssim \sqrt{n T \log T}$.

Gap-dependent. On E , $T_i \lesssim \Delta_i^{-2} L$ (last slide), so $\Delta_i T_i \lesssim \Delta_i^{-1} L$: each pull of a bad arm costs only Δ_i .

Gap-free, by the A/B picture. Both bounds hold for each arm, and $\min\{ax, b/x\} \leq \sqrt{ab}$ for every $x > 0$:

$$\Delta_i T_i \leq$$

Regret of successive elimination

Same algorithm and lemma; now count regret $\sum_i \Delta_i T_i$ after T pulls;
 $L := \log(n \log T / \delta)$.

Theorem (regret)

Assume $\max_i \Delta_i \leq 4$. With probability at least $1 - \delta$, $\Delta_i T_i \lesssim \Delta_i^{-1} L$ for every arm i and every T . Consequently, with $\delta = 1/T$: $R_T \lesssim \sum_{i \geq 2} \log T / \Delta_i$ and $R_T \lesssim \sqrt{n T \log T}$.

Gap-dependent. On E , $T_i \lesssim \Delta_i^{-2} L$ (last slide), so $\Delta_i T_i \lesssim \Delta_i^{-1} L$: each pull of a bad arm costs only Δ_i .

Gap-free, by the A/B picture. Both bounds hold for each arm, and $\min\{ax, b/x\} \leq \sqrt{ab}$ for every $x > 0$:

$$\Delta_i T_i \leq \min \left\{ \Delta_i T_i, \frac{cL}{\Delta_i} \right\} \leq \sqrt{cL T_i} \quad (\text{worst case } \Delta_i \approx \sqrt{cL/T_i}).$$

Sum over arms; Cauchy-Schwarz with $\sum_i T_i = T$:

$$R_T \leq \sum_{i=1}^n \sqrt{cL T_i} \leq \sqrt{cL} \sqrt{n \sum_i T_i} = \sqrt{cL n T}. \quad \blacksquare$$

Regret of successive elimination

Same algorithm and lemma; now count regret $\sum_i \Delta_i T_i$ after T pulls;
 $L := \log(n \log T / \delta)$.

Theorem (regret)

Assume $\max_i \Delta_i \leq 4$. With probability at least $1 - \delta$, $\Delta_i T_i \lesssim \Delta_i^{-1} L$ for every arm i and every T . Consequently, with $\delta = 1/T$: $R_T \lesssim \sum_{i \geq 2} \log T / \Delta_i$ and $R_T \lesssim \sqrt{n T \log T}$.

Gap-dependent. On E , $T_i \lesssim \Delta_i^{-2} L$ (last slide), so $\Delta_i T_i \lesssim \Delta_i^{-1} L$: each pull of a bad arm costs only Δ_i .

Gap-free, by the A/B picture. Both bounds hold for each arm, and $\min\{ax, b/x\} \leq \sqrt{ab}$ for every $x > 0$:

$$\Delta_i T_i \leq \min \left\{ \Delta_i T_i, \frac{cL}{\Delta_i} \right\} \leq \sqrt{cL T_i} \quad (\text{worst case } \Delta_i \approx \sqrt{cL/T_i}).$$

Sum over arms; Cauchy-Schwarz with $\sum_i T_i = T$:

$$R_T \leq \sum_{i=1}^n \sqrt{cL T_i} \leq \sqrt{cL} \sqrt{n \sum_i T_i} = \sqrt{cL n T}. \quad \blacksquare$$

Versus experts' $\sqrt{T \log d}$: partial feedback costs $\sqrt{n}$. Next: the same price without i.i.d.

Adversarial bandits

Exponential weights, with one number per round

Adversarial bandits: the setting

Adversarial multi-armed bandit (notes, Ch. 7)

adversary fixes losses $\ell_1, \dots, \ell_T \in [-1, 1]^n$ in advance (*oblivious*)

for $t = 1, 2, \dots, T$:

player picks $q_t \in \Delta_n$ and plays $I_t \sim q_t$

player suffers and observes ℓ_{t, I_t} , and

nothing else

$\text{Regret}_T = \max_{i \in [n]} \mathbb{E} \left[\sum_{t=1}^T \ell_{t, I_t} - \ell_{t, i} \right]$,
the expectation over the player's coins only.

Where this sits

Experts (Lecture 2): adversarial losses, but you see all of ℓ_t . EW gets $\sqrt{T \log n / 2}$.

Stochastic bandits (so far today): one number per round, but i.i.d. noise around fixed means.

Now both: one number per round, chosen by an adversary.

Adversarial bandits: the setting

Adversarial multi-armed bandit (notes, Ch. 7)

adversary fixes losses $\ell_1, \dots, \ell_T \in [-1, 1]^n$ in advance (*oblivious*)

for $t = 1, 2, \dots, T$:

player picks $q_t \in \Delta_n$ and plays $I_t \sim q_t$

player suffers and observes ℓ_{t, I_t} , and

nothing else

$\text{Regret}_T = \max_{i \in [n]} \mathbb{E} \left[\sum_{t=1}^T \ell_{t, I_t} - \ell_{t, i} \right]$,
the expectation over the player's coins only.

- Randomization is essential, as in Lecture 2: a deterministic player can be made to pay 1 every round while some arm pays ≤ 0 on average.
- Exponential weights is the natural algorithm, but its update $w_{t+1, i} = w_{t, i} e^{-\eta \ell_{t, i}}$ needs $\ell_{t, i}$ for every i ; we see one coordinate.

Where this sits

Experts (Lecture 2): adversarial losses, but you see all of ℓ_t . EW gets $\sqrt{T \log n / 2}$.

Stochastic bandits (so far today): one number per round, but i.i.d. noise around fixed means.

Now both: one number per round, chosen by an adversary.

The trick: estimate the whole loss vector from one coordinate

Play $I_t \sim q_t$ and define the **importance-weighted** estimate

$$\hat{\ell}_{t,i} = \frac{\mathbf{1}\{I_t = i\}}{q_{t,i}} \ell_{t,i} \quad (i \in [n]),$$

which is computable: it is 0 for every arm except I_t , and $\ell_{t,I_t}/q_{t,I_t}$ there.

The trick: estimate the whole loss vector from one coordinate

Play $I_t \sim q_t$ and define the **importance-weighted** estimate

$$\hat{\ell}_{t,i} = \frac{\mathbf{1}\{I_t = i\}}{q_{t,i}} \ell_{t,i} \quad (i \in [n]),$$

which is computable: it is 0 for every arm except I_t , and $\ell_{t,I_t}/q_{t,I_t}$ there.

Unbiased. Conditionally on q_t , $\mathbb{E}[\hat{\ell}_{t,i} \mid q_t] =$

The trick: estimate the whole loss vector from one coordinate

Play $I_t \sim q_t$ and define the **importance-weighted** estimate

$$\hat{\ell}_{t,i} = \frac{\mathbf{1}\{I_t = i\}}{q_{t,i}} \ell_{t,i} \quad (i \in [n]),$$

which is computable: it is 0 for every arm except I_t , and $\ell_{t,I_t}/q_{t,I_t}$ there.

Unbiased. Conditionally on q_t , $\mathbb{E}[\hat{\ell}_{t,i} \mid q_t] = \sum_{j=1}^n q_{t,j} \frac{\mathbf{1}\{j=i\}}{q_{t,i}} \ell_{t,i} = \ell_{t,i}$.

The trick: estimate the whole loss vector from one coordinate

Play $I_t \sim q_t$ and define the **importance-weighted** estimate

$$\hat{\ell}_{t,i} = \frac{\mathbf{1}\{I_t = i\}}{q_{t,i}} \ell_{t,i} \quad (i \in [n]),$$

which is computable: it is 0 for every arm except I_t , and $\ell_{t,I_t}/q_{t,I_t}$ there.

Unbiased. Conditionally on q_t , $\mathbb{E}[\hat{\ell}_{t,i} \mid q_t] = \sum_{j=1}^n q_{t,j} \frac{\mathbf{1}\{j=i\}}{q_{t,i}} \ell_{t,i} = \ell_{t,i}$.

But noisy. Its second moment is $\mathbb{E}[\hat{\ell}_{t,i}^2 \mid q_t] =$

which blows up for arms we rarely play. So we will never let $q_{t,i}$ get too small: mix in a little exploration.

The trick: estimate the whole loss vector from one coordinate

Play $I_t \sim q_t$ and define the **importance-weighted** estimate

$$\hat{\ell}_{t,i} = \frac{\mathbf{1}\{I_t = i\}}{q_{t,i}} \ell_{t,i} \quad (i \in [n]),$$

which is computable: it is 0 for every arm except I_t , and $\ell_{t,I_t}/q_{t,I_t}$ there.

Unbiased. Conditionally on q_t , $\mathbb{E}[\hat{\ell}_{t,i} \mid q_t] = \sum_{j=1}^n q_{t,j} \frac{\mathbf{1}\{j=i\}}{q_{t,i}} \ell_{t,i} = \ell_{t,i}$.

But noisy. Its second moment is $\mathbb{E}[\hat{\ell}_{t,i}^2 \mid q_t] = \sum_{j=1}^n q_{t,j} \frac{\mathbf{1}\{j=i\}}{q_{t,i}^2} \ell_{t,i}^2 = \frac{\ell_{t,i}^2}{q_{t,i}}$, which blows up for arms we rarely play. So we will never let $q_{t,i}$ get too small: mix in a little exploration.

The trick: estimate the whole loss vector from one coordinate

Play $I_t \sim q_t$ and define the **importance-weighted** estimate

$$\hat{\ell}_{t,i} = \frac{\mathbf{1}\{I_t = i\}}{q_{t,i}} \ell_{t,i} \quad (i \in [n]),$$

which is computable: it is 0 for every arm except I_t , and $\ell_{t,I_t}/q_{t,I_t}$ there.

Unbiased. Conditionally on q_t , $\mathbb{E}[\hat{\ell}_{t,i} \mid q_t] = \sum_{j=1}^n q_{t,j} \frac{\mathbf{1}\{j=i\}}{q_{t,i}} \ell_{t,i} = \ell_{t,i}$.

But noisy. Its second moment is $\mathbb{E}[\hat{\ell}_{t,i}^2 \mid q_t] = \sum_{j=1}^n q_{t,j} \frac{\mathbf{1}\{j=i\}}{q_{t,i}^2} \ell_{t,i}^2 = \frac{\ell_{t,i}^2}{q_{t,i}}$, which blows up for arms we rarely play. So we will never let $q_{t,i}$ get too small: mix in a little exploration.

EXP3(γ): exponential weights for exploration and exploitation

Input: $\eta > 0$, $\gamma \in [0, 1]$, exploration distribution $\lambda \in \Delta_n$ (e.g. uniform). **Initialize** $w_1 \equiv 1$.

for $t = 1, \dots, T$: $p_t = w_t / \sum_j w_{t,j}$; **play** $I_t \sim q_t := (1 - \gamma)p_t + \gamma\lambda$; **observe** ℓ_{t,I_t} ;
form $\hat{\ell}_t$ as above; $w_{t+1,i} = w_{t,i} e^{-\eta \hat{\ell}_{t,i}}$ **for all** i .

Exponential weights, run on the *estimated* losses, with its play smoothed toward λ .

The regret of EXP3: a general theorem

Theorem (notes, Thm. 12)

Fix any losses $\ell_t \in [-1, 1]^n$. For any estimates with $\mathbb{E}[\hat{\ell}_{t,i} \mid \mathcal{F}_{t-1}] = \ell_{t,i}$ and $-\eta \hat{\ell}_{t,i} \leq 1$ for all i, t ,

$$\max_i \mathbb{E} \left[\sum_{t=1}^T \ell_{t,I_t} - \ell_{t,i} \right] \leq 2\gamma T + \frac{\log n}{\eta} + (1 - \gamma) \eta \mathbb{E} \left[\sum_{t=1}^T \sum_{j=1}^n p_{t,j} \hat{\ell}_{t,j}^2 \right].$$

The regret of EXP3: a general theorem

Theorem (notes, Thm. 12)

Fix any losses $\ell_t \in [-1, 1]^n$. For any estimates with $\mathbb{E}[\hat{\ell}_{t,i} \mid \mathcal{F}_{t-1}] = \ell_{t,i}$ and $-\eta \hat{\ell}_{t,i} \leq 1$ for all i, t ,

$$\max_i \mathbb{E} \left[\sum_{t=1}^T \ell_{t,I_t} - \ell_{t,i} \right] \leq 2\gamma T + \frac{\log n}{\eta} + (1 - \gamma) \eta \mathbb{E} \left[\sum_{t=1}^T \sum_{j=1}^n p_{t,j} \hat{\ell}_{t,j}^2 \right].$$

Step A: pass to the estimated losses. Fix i . Since $q_t = (1 - \gamma)p_t + \gamma\lambda$ and $|\ell_{t,j}| \leq 1$,

$$\begin{aligned} \mathbb{E} \left[\sum_t \ell_{t,I_t} - \ell_{t,i} \right] &= \mathbb{E} \left[\sum_t \langle q_t, \ell_t \rangle - \ell_{t,i} \right] \leq \\ &\leq \end{aligned}$$

The regret of EXP3: a general theorem

Theorem (notes, Thm. 12)

Fix any losses $\ell_t \in [-1, 1]^n$. For any estimates with $\mathbb{E}[\hat{\ell}_{t,i} \mid \mathcal{F}_{t-1}] = \ell_{t,i}$ and $-\eta \hat{\ell}_{t,i} \leq 1$ for all i, t ,

$$\max_i \mathbb{E} \left[\sum_{t=1}^T \ell_{t,I_t} - \ell_{t,i} \right] \leq 2\gamma T + \frac{\log n}{\eta} + (1 - \gamma) \eta \mathbb{E} \left[\sum_{t=1}^T \sum_{j=1}^n p_{t,j} \hat{\ell}_{t,j}^2 \right].$$

Step A: pass to the estimated losses. Fix i . Since $q_t = (1 - \gamma)p_t + \gamma\lambda$ and $|\ell_{t,j}| \leq 1$,

$$\begin{aligned} \mathbb{E} \left[\sum_t \ell_{t,I_t} - \ell_{t,i} \right] &= \mathbb{E} \left[\sum_t \langle q_t, \ell_t \rangle - \ell_{t,i} \right] \leq \gamma T + \mathbb{E} \left[\sum_t (1 - \gamma) \langle p_t, \ell_t \rangle - \ell_{t,i} \right] \\ &\leq 2\gamma T + (1 - \gamma) \mathbb{E} \left[\sum_t \langle p_t, \ell_t \rangle - \ell_{t,i} \right], \end{aligned}$$

using $\gamma \langle \lambda, \ell_t \rangle \leq \gamma$ and $-\ell_{t,i} \leq -(1 - \gamma)\ell_{t,i} + \gamma$. As p_t depends only on the past and $\mathbb{E}[\hat{\ell}_t \mid \mathcal{F}_{t-1}] = \ell_t$, the last term is $\mathbb{E}[\sum_t \langle p_t, \hat{\ell}_t \rangle - \hat{\ell}_{t,i}]$: EW's regret *on the estimates*.

Step B: the potential argument

$W_t = \sum_i w_{t,i}$, $W_1 = n$. **Lower bound**, as always:
 $\log \frac{W_{T+1}}{W_1} \geq \log w_{T+1,i} - \log n = -\eta \sum_t \hat{\ell}_{t,i} - \log n.$

Step B: the potential argument

$W_t = \sum_i w_{t,i}$, $W_1 = n$. **Lower bound**, as always:

$$\log \frac{W_{T+1}}{W_1} \geq \log w_{T+1,i} - \log n = -\eta \sum_t \hat{\ell}_{t,i} - \log n.$$

Upper bound. Hoeffding's lemma (slack $\eta^2/8$ per round) needed bounded losses; $\hat{\ell}_{t,i}$ can be huge. What we have is $-\eta \hat{\ell}_{t,i} \leq 1$, and $e^x \leq 1 + x + x^2$ for $x \leq 1$. One round:

$$\begin{aligned} \log \frac{W_{t+1}}{W_t} &= \log \left(\sum_i p_{t,i} e^{-\eta \hat{\ell}_{t,i}} \right) \leq \\ &\leq \end{aligned}$$

Step B: the potential argument

$W_t = \sum_i w_{t,i}$, $W_1 = n$. **Lower bound**, as always:

$$\log \frac{W_{T+1}}{W_1} \geq \log w_{T+1,i} - \log n = -\eta \sum_t \hat{\ell}_{t,i} - \log n.$$

Upper bound. Hoeffding's lemma (slack $\eta^2/8$ per round) needed bounded losses; $\hat{\ell}_{t,i}$ can be huge. What we have is $-\eta \hat{\ell}_{t,i} \leq 1$, and $e^x \leq 1 + x + x^2$ for $x \leq 1$. One round:

$$\begin{aligned} \log \frac{W_{t+1}}{W_t} &= \log \left(\sum_i p_{t,i} e^{-\eta \hat{\ell}_{t,i}} \right) \leq \log \left(1 - \eta \langle p_t, \hat{\ell}_t \rangle + \eta^2 \sum_i p_{t,i} \hat{\ell}_{t,i}^2 \right) \\ &\leq -\eta \langle p_t, \hat{\ell}_t \rangle + \eta^2 \sum_i p_{t,i} \hat{\ell}_{t,i}^2 \end{aligned}$$

by $e^x \leq 1 + x + x^2$, then $\log(1 + x) \leq x$. Sum over t and combine with the lower bound:

$$\sum_{t=1}^T \langle p_t, \hat{\ell}_t \rangle - \sum_{t=1}^T \hat{\ell}_{t,i} \leq \frac{\log n}{\eta} + \eta \sum_{t=1}^T \sum_{j=1}^n p_{t,j} \hat{\ell}_{t,j}^2.$$

Take expectations and plug into Step A. The new slack, $\eta^2 \sum_j p_{t,j} \hat{\ell}_{t,j}^2$, is what exploration keeps in check. ■

EXP3 for n -armed bandits: $\sqrt{12 nT \log n}$

Corollary (notes, Cor. 3)

With λ uniform, $\gamma = \eta n$ and $\eta = \sqrt{\log(n)/(3nT)}$:

$$\max_i \mathbb{E} \left[\sum_{t=1}^T \ell_{t, I_t} - \ell_{t, i} \right] \leq \sqrt{12 nT \log n}.$$

EXP3 for n -armed bandits: $\sqrt{12 n T \log n}$

Corollary (notes, Cor. 3)

With λ uniform, $\gamma = \eta n$ and $\eta = \sqrt{\log(n)/(3nT)}$:

$$\max_i \mathbb{E} \left[\sum_{t=1}^T \ell_{t, I_t} - \ell_{t, i} \right] \leq \sqrt{12 n T \log n}.$$

The condition. Since $q_{t,i} \geq \gamma \lambda_i = \gamma/n$ and $\gamma = \eta n$,
 $|\eta \hat{\ell}_{t,i}| = \eta \frac{\mathbf{1}_{\{I_t=i\}}}{q_{t,i}} |\ell_{t,i}| \leq$

EXP3 for n -armed bandits: $\sqrt{12 n T \log n}$

Corollary (notes, Cor. 3)

With λ uniform, $\gamma = \eta n$ and $\eta = \sqrt{\log(n)/(3nT)}$:

$$\max_i \mathbb{E} \left[\sum_{t=1}^T \ell_{t, I_t} - \ell_{t, i} \right] \leq \sqrt{12 n T \log n}.$$

The condition. Since $q_{t,i} \geq \gamma \lambda_i = \gamma/n$ and $\gamma = \eta n$,
 $|\eta \hat{\ell}_{t,i}| = \eta \frac{\mathbf{1}_{\{I_t=i\}}}{q_{t,i}} |\ell_{t,i}| \leq \frac{\eta n}{\gamma} = 1.$

EXP3 for n -armed bandits: $\sqrt{12 n T \log n}$

Corollary (notes, Cor. 3)

With λ uniform, $\gamma = \eta n$ and $\eta = \sqrt{\log(n)/(3nT)}$:

$$\max_i \mathbb{E} \left[\sum_{t=1}^T \ell_{t, I_t} - \ell_{t, i} \right] \leq \sqrt{12 n T \log n}.$$

The condition. Since $q_{t,i} \geq \gamma \lambda_i = \gamma/n$ and $\gamma = \eta n$,
 $|\eta \hat{\ell}_{t,i}| = \eta \frac{\mathbf{1}_{\{I_t=i\}}}{q_{t,i}} |\ell_{t,i}| \leq \frac{\eta n}{\gamma} = 1.$

The second moment. Only $j = I_t$ contributes, and $p_{t,j} \leq q_{t,j}/(1 - \gamma)$:

$$\mathbb{E} \left[\sum_{j=1}^n p_{t,j} \hat{\ell}_{t,j}^2 \right] = \mathbb{E} \left[p_{t, I_t} \frac{\ell_{t, I_t}^2}{q_{t, I_t}^2} \right] \leq$$

EXP3 for n -armed bandits: $\sqrt{12 n T \log n}$

Corollary (notes, Cor. 3)

With λ uniform, $\gamma = \eta n$ and $\eta = \sqrt{\log(n)/(3nT)}$:

$$\max_i \mathbb{E} \left[\sum_{t=1}^T \ell_{t,I_t} - \ell_{t,i} \right] \leq \sqrt{12 n T \log n}.$$

The condition. Since $q_{t,i} \geq \gamma \lambda_i = \gamma/n$ and $\gamma = \eta n$,
 $|\eta \hat{\ell}_{t,i}| = \eta \frac{\mathbf{1}\{I_t=i\}}{q_{t,i}} |\ell_{t,i}| \leq \frac{\eta n}{\gamma} = 1.$

The second moment. Only $j = I_t$ contributes, and $p_{t,j} \leq q_{t,j}/(1 - \gamma)$:

$$\mathbb{E} \left[\sum_{j=1}^n p_{t,j} \hat{\ell}_{t,j}^2 \right] = \mathbb{E} \left[p_{t,I_t} \frac{\ell_{t,I_t}^2}{q_{t,I_t}^2} \right] \leq \frac{1}{1 - \gamma} \mathbb{E} \left[\frac{\ell_{t,I_t}^2}{q_{t,I_t}} \right] = \frac{1}{1 - \gamma} \sum_{j=1}^n \ell_{t,j}^2 \leq \frac{n}{1 - \gamma}.$$

So $\text{Regret}_T \leq 3\eta n T + \log(n)/\eta = \sqrt{12 n T \log n}$ at $\eta = \sqrt{\log(n)/(3nT)}$. ■

EXP3 for n -armed bandits: $\sqrt{12 n T \log n}$

Corollary (notes, Cor. 3)

With λ uniform, $\gamma = \eta n$ and $\eta = \sqrt{\log(n)/(3nT)}$:

$$\max_i \mathbb{E} \left[\sum_{t=1}^T \ell_{t,I_t} - \ell_{t,i} \right] \leq \sqrt{12 n T \log n}.$$

The condition. Since $q_{t,i} \geq \gamma \lambda_i = \gamma/n$ and $\gamma = \eta n$,
 $|\eta \hat{\ell}_{t,i}| = \eta \frac{\mathbf{1}_{\{I_t=i\}}}{q_{t,i}} |\ell_{t,i}| \leq \frac{\eta n}{\gamma} = 1.$

The second moment. Only $j = I_t$ contributes, and $p_{t,j} \leq q_{t,j}/(1 - \gamma)$:

$$\mathbb{E} \left[\sum_{j=1}^n p_{t,j} \hat{\ell}_{t,j}^2 \right] = \mathbb{E} \left[p_{t,I_t} \frac{\ell_{t,I_t}^2}{q_{t,I_t}^2} \right] \leq \frac{1}{1 - \gamma} \mathbb{E} \left[\frac{\ell_{t,I_t}^2}{q_{t,I_t}} \right] = \frac{1}{1 - \gamma} \sum_{j=1}^n \ell_{t,j}^2 \leq \frac{n}{1 - \gamma}.$$

So $\text{Regret}_T \leq 3\eta n T + \log(n)/\eta = \sqrt{12 n T \log n}$ at $\eta = \sqrt{\log(n)/(3nT)}$. ■

The price of bandit feedback, again

Full information $\sqrt{T \log n}$; stochastic $\sqrt{n T \log T}$; adversarial $\sqrt{n T \log n}$: $\sqrt{n}$ is the price of one coordinate per round. Next: lower bounds, and UCB.