Lecture #6: Lower bound

Recall bandit setting

- To each arm k is associated a probability distribution $\nu_k \in \mathcal{D}$.

- $\mathcal{D}$ is the bandit model ($\mathcal{D} \subset \mathcal{P}_1(\mathbb{R})$)
- A bandit instance is denoted by $\nu = (\nu_k)_{k \in [K]}$
- Goal: minimise the regret, which can be rewritten as:

$$R_T = \sum_{k=1}^{K} \Delta_k \, \mathbb{E}[N_k(T)]$$

Bounding the regret $\iff$ bounding $\mathbb{E}[N_k(T)]$.

What are the best possible (by an algorithm) bounds?

- what is a randomized strategy π?

a sequence of measurable functions $(\pi_t)_{t \geq 1}$ with

$$\pi_{t+1}: \underbrace{H_t = (U_0, X_{a_1}(1), U_1, \dots, X_{a_t}(t), U_t)}_{\text{history of observations + randomization for first } t \text{ rounds}} \longmapsto \underbrace{\pi_{t+1}(H_t) = a_{t+1}}_{\text{arm picked at } t+1}$$

Defn - a strategy is consistent w.r.t a model $\mathcal{D}$ if:

for all bandit instances $\nu \in \mathcal{D}^K$, $\forall \alpha \in (0,1]$, $\forall k$ s.t. $\Delta_k > 0$,

$$\mathbb{E}[N_k(T)] = o(T^\alpha).$$

for well behaved models, there exist consistent strategies

e.g. UCB with $\mathcal{D} = \mathcal{P}([0,1])$.

- Typical "(asymptotic) bounds for good strategies

$$\forall \nu \in \mathcal{D}^K, \ \forall k \text{ s.t. } \Delta_k > 0, \quad \limsup_{T \to \infty} \frac{\mathbb{E}[N_k(T)]}{\ln T} \leq \underbrace{C_k(\nu)}$$

problem dependent term

- optimal such term: $$c_k(\nu) = \frac{1}{K_{inf}(\nu_k, \mu^*; \mathcal{D})}$$

where $$K_{inf}(\nu_k, \mu^*; \mathcal{D}) = \inf \left\{ KL(\nu_k, \nu_k') \;\middle|\; \begin{array}{l} \nu_k' \in \mathcal{D} \\ \mathbb{E}(\nu_k') > \mu^* \end{array} \right\}$$

we will now prove one part of this optimality: a lower bound on $c_k(\nu)$.

Theorem (Lai and Robbins, 1985, Burnetas and Katehakis, 1996)

For all bandit models $\mathcal{D} \in \mathcal{P}_1(\mathbb{R})$,
for any consistent strategy wrt $\mathcal{D}$,
for any bandit instance $\nu \in \mathcal{D}^K$,
for all suboptimal arms k (ie $\Delta_k > 0$), $$\liminf_{T \to \infty} \frac{\mathbb{E}[N_k(T)]}{\ln T} \geq \frac{1}{K_{inf}(\nu_k, \mu^*; \mathcal{D})}$$

Corollary

For all bandit models $\mathcal{D}$, any consistent strategy wrt $\mathcal{D}$, all bandit instances $\nu \in \mathcal{D}^K$:

$$\liminf_{T \to +\infty} \frac{R_T}{\ln T} \geq \sum_{a,\ \Delta_a > 0} \frac{\Delta_a}{K_{\inf}(\nu_a, \mu^*, \mathcal{D})}$$

To prove this theorem (and other lower bounds), we need the following fundamental inequality.

Notation: for a strategy π, we note $H_t = (U_0, X_{a_1}(1), U_1, X_{a_2}(2), \dots, X_{a_t}(t), U_t)$

Recall that a_{t+1} is $\sigma(H_t)$-measurable.

depends on π

Lemma: (fundamental inequality for stochastic bandits)

For all bandit problems $\nu = (\nu_k)_{k \in [K]}$ and $\nu' = (\nu'_k)_{k \in [K]}$ in $\mathcal{D}^K$ with $\nu_k \ll \nu'_k$ for all k,

for all strategies and random variables Z taking values in $[0,1]$ that are $\sigma(H_T)$-measurable,

law of H_T under ν (and π)

$$\sum_{k=1}^{K} \mathbb{E}_\nu[N_k(T)]\, KL(\nu_k, \nu'_k) = KL(\mathbb{P}_\nu^{H_T}, \mathbb{P}_{\nu'}^{H_T})$$

$$\geq KL(\mathrm{Ber}(\mathbb{E}_\nu[Z]), \mathrm{Ber}(\mathbb{E}_{\nu'}[Z]))$$

dependence in strategy π hidden everywhere here.

Note: this lemma is our key to perform an implicit change of measures in the proof of the theorem.

Proof of the theorem (based on the lemma)

$$K_{inf}(\nu_k, \mathcal{D}, \mu^*) = \inf \{ KL(\nu_k, \nu_k') \mid \nu_k' \in \mathcal{D},\ \nu_k \ll \nu_k' \text{ and } \mathbb{E}(\nu') > \mu^* \}.$$

Convention $\inf \emptyset = +\infty$

This is why we will:

- fix $\mathcal{D}$, strategy π, ν and k s.t. $\Delta_k > 0$ (π is consistent wrt $\mathcal{D}$)

- fix an alternative model ν' with

$$\begin{cases} \nu_i' = \nu_i & \text{for all } i \neq k \\ \nu_k' \text{ s.t. } & \nu_k' \in \mathcal{D},\ \nu_k \ll \nu_k' \text{ and } \mathbb{E}(\nu_k') > \mu^* \end{cases}$$

That is ν and ν' only differ at k, the unique optimal arm in ν'.

- Take $Z = \frac{N_k(T)}{T}$ which is $[0,1]$-valued
$\sigma(H_T)$-measurable

Our fundamental inequality (lemma) yields, since ν and ν' only differ at k:

$$\mathbb{E}_\nu[N_k(T)]\, KL(\nu_k, \nu_k') \geq KL\left(Ber\left(\mathbb{E}_\nu\left[\frac{N_k(T)}{T}\right]\right), Ber\left(\mathbb{E}_{\nu'}\left[\frac{N_k(T)}{T}\right]\right)\right)$$

$$\geq -\ln(2) + \left(1 - \mathbb{E}_\nu\left[\frac{N_k(T)}{T}\right]\right) \ln\left(\frac{1}{1 - \mathbb{E}_{\nu'}\left[\frac{N_k(T)}{T}\right]}\right)$$

indeed $KL(Ber(p), Ber(q)) = p \ln\left(\frac{p}{q}\right) + (1-p)\ln\left(\frac{1-p}{1-q}\right)$

$$= p\ln\left(\frac{1}{q}\right) + (1-p)\ln\left(\frac{1}{1-q}\right) + \left(p\ln(p) + (1-p)\ln(1-p)\right)$$

$\underset{\geq 0}{\sim}$ $\qquad$ $\underbrace{\qquad\qquad}_{\geq -\ln 2}$

$\geq -\ln 2 + (1-p)\ln\left(\frac{1}{1-q}\right)$ for all $(p,q) \in [0,1)$ (and even for $p,q \in [0,1]$)

π is consistent, so:

- instance ν → k is suboptimal $\mathbb{E}_\nu\left[\frac{N_k(T)}{T}\right] \xrightarrow[T\to\infty]{} 0$

- instance ν' → all $i \neq k$ are suboptimal:

for any $\alpha \in [0,1]$, $\mathbb{E}_{\nu'}[N_i(T)] = o(T^\alpha)$

In particular: $T - \mathbb{E}_{\nu'}[N_k(T)] = \sum_{i\neq k} \mathbb{E}_{\nu'}[N_i(T)] = o(T^\alpha)$

so: $\frac{1}{1-\mathbb{E}_{\nu'}\left[\frac{N_k(T)}{T}\right]} = \frac{T}{T - \mathbb{E}_{\nu'}[N_k(T)]} = \frac{T}{o(T^\alpha)}$

$\geq T^{1-\alpha}$ for T large enough

substituting back and dividing by $\ln T$: for any $\alpha \in (0,1)$ and T large enough

$$\frac{\mathbb{E}_\nu[N_k(T)]\ KL(\nu_k,\nu_k')}{\ln T} \geq -\frac{\ln 2}{\ln T} + \left(1 - \mathbb{E}_\nu\left[\frac{N_k(T)}{T}\right]\right)\frac{\ln T^{1-\alpha}}{\ln T}$$

thus $\liminf_{T\to+\infty} \frac{\mathbb{E}_\nu[N_k(T)]}{\ln T} \geq \frac{(1-\alpha)}{KL(\nu_k,\nu_k')}$ (true whether the KL is $<+\infty$ or $=+\infty$ (it is necessarily >0))

for any $\alpha \in (0,1)$, so

$$\liminf_{T\to+\infty} \frac{\mathbb{E}_\nu[N_k(T)]}{\ln T} \geq \frac{1}{KL(\nu_k,\nu_k')}$$

Holds for any $\nu_a' \in \mathcal{D}$ s.t. $\nu_a \ll \nu_a'$ and $\mathbb{E}(\nu_a') > \mu^*$, so that taking the supremum of the right hand side on these ν_a' yields the lower bound:

$$\liminf_{T\to\infty} \frac{\mathbb{E}_\nu[N_a(T)]}{\ln T} \geq \frac{1}{K_{inf}(\nu_a, \mathcal{D}, \mu^*)} \quad \square.$$

Proof of the lemma

• The inequality $\geq$ is a direct application of the data processing inequality with expectations.

• For the equality: and same for ν': $\mathbb{P}_{\nu'}^{H_T} = K_T'(K_{T-1}' \cdots)$

(1) we show by induction that $\mathbb{P}_\nu^{H_T} = K_T(K_{T-1}(\cdots(K_1\lambda_0)))$

we check below that it is regular

where K_t is the transition kernel:

$$h \in [0,1] \times (\mathbb{R}\times[0,1])^{t-1} \longmapsto \underbrace{K_t(h,\cdot)}_{\text{prob. measure on } \mathbb{R}\times[0,1]} = \nu_{\pi_t(h)} \otimes \lambda_0$$

with $U_t \sim \lambda_0$ "$U[0,1]$"

$\underline{T=0}$: $H_0 = U_0 \sim \lambda_0$: $\mathbb{P}_\nu^{U_0} = \lambda_0$

$\underline{t \to t+1}$ $\forall A \in \mathcal{B}([0,1] \times (\mathbb{R}\times[0,1])^t)$, $\forall B' \in \mathcal{B}(\mathbb{R})$, $\forall B \in \mathcal{B}([0,1])$;

$$\mathbb{P}_\nu^{H_{t+1}}(A\times B'\times B) = \mathbb{P}_\nu(H_t \in A \text{ and } X_{a_{t+1}}(t+1) \in B' \text{ and } U_{t+1} \in B)$$

tower rule

$$= \mathbb{E}_\nu\left[\mathbb{1}_A(H_t)\, \mathbb{P}_\nu\left[X_{a_{t+1}}(t+1) \in B' \text{ and } U_{t+1} \in B \mid H_t\right]\right]$$

$$= \mathbb{E}_\nu \left[\mathbb{1}_A(H_r) \cdot \nu_{\pi_{r+1}(H_r)}(B') \cdot \lambda_0(B) \right]$$

↙ defn of the model and strategy

$$= \mathbb{E}_\nu \left[\mathbb{1}_A(H_r) \; K_{r+1}(H_r, B' \times B) \right]$$

↓ defn of K_{r+1}

$$= \int \mathbb{1}_A(h) \; K_{r+1}(h, B' \times B) \, d\mathbb{P}_\nu^{H_r}(h)$$

↓ rewriting

↓ defn of $K_{r+1} \mathbb{P}_\nu^{H_r}$

$$= K_{r+1} \mathbb{P}_\nu^{H_r} (A \times B' \times B) \longrightarrow \text{we've shown the induction}$$

(2) we check that the assumptions of the chain rule are satisfied:

- the K_r are regular transition kernels: $\forall E \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}([0,1])$,

$$h \longmapsto K_r(h, E) = \sum_{k=1}^{K} \mathbb{1}_{(\pi_r(h) = k)} \, (\nu_k \otimes \lambda_0)(E)$$ is measurable as

π_r is measurable (with respect to considered spaces)

- Assumption (✱): $\forall h$, $K_r(h, \cdot) \ll K'_r(h, \cdot)$ as $\forall k$, $\nu_k \ll \nu'_k$ by ass.

- Assumption (✱✱): $$(h, (y, u)) \longmapsto \frac{dK_r(h,\cdot)}{dK'_r(h,\cdot)}(y,u) = \sum_{k=1}^{K} \mathbb{1}_{(\pi_r(h) = k)} \frac{d\nu_k}{d\nu'_k}(y)$$

is indeed bi-measurable (product of measurable functions)

(3) We then may apply the chain rule and show by induction the desired result based on:

$$- \; KL(\mathbb{P}_\nu^{H_0}, \mathbb{P}_{\nu'}^{H_0}) = KL(\lambda_0, \lambda_0) = 0$$

- For $r \geq 0$, $$KL(\mathbb{P}_\nu^{H_{r+1}}, \mathbb{P}_{\nu'}^{H_{r+1}}) = KL(K_{r+1} \mathbb{P}_\nu^{H_r}, K'_{r+1} \mathbb{P}_{\nu'}^{H_r})$$

↓ chain rule

$$= KL(\mathbb{P}_\nu^{H_r}, \mathbb{P}_{\nu'}^{H_r}) + \int KL(K_{r+1}(h,\cdot), K'_{r+1}(h,\cdot)) \, d\mathbb{P}_\nu^{H_r}(h)$$

$$= KL\left(\mathbb{P}_\nu^{H_r}, \mathbb{P}_{\nu'}^{H_r}\right) + \int KL\left(\nu_{\pi_{r+1}(h)} \cancel{\otimes \lambda_0}, \nu'_{\pi_{r+1}(h)} \cancel{\otimes \lambda_0}\right) d\mathbb{P}_\nu^{H_r}(h)$$

$$= KL\left(\mathbb{P}_\nu^{H_r}, \mathbb{P}_{\nu'}^{H_r}\right) + \sum_{k=1}^{K} KL(\nu_k, \nu_k') \cdot \underbrace{\int \mathbb{1}_{\pi_{r+1}(h)=k} \, d\mathbb{P}_\nu^{H_r}(h)}_{\substack{\mathbb{E}\left[\mathbb{1}_{(\pi_{r+1}(H_r)=k)}\right] \\ = \mathbb{E}\left[\mathbb{1}_{(a_{r+1}=k)}\right]}}$$

$$= KL\left(\mathbb{P}_\nu^{H_r}, \mathbb{P}_{\nu'}^{H_r}\right) + \sum_{k=1}^{K} KL(\nu_k, \nu_k') \, \mathbb{E}\left[\mathbb{1}_{(a_{r+1}=k)}\right]$$

by induction

$$KL\left(\mathbb{P}_\nu^{H_T}, \mathbb{P}_{\nu'}^{H_0}\right) = \sum_{t=1}^{T} \sum_{k=1}^{K} KL(\nu_k, \nu_k') \, \mathbb{E}\left[\mathbb{1}_{a_t=k}\right]$$

$$= \sum_k KL(\nu_k, \nu_k') \, \mathbb{E}\left[N_k(T)\right]. \qquad \square$$

Comments on the lower bound :

- see a comparison with our upper bounds in exercise session #4.

- algorithms with optimal instance dependent bounds are known (e.g. KL-UCB, Thompson sampling) but require a long and technical analysis.

- this is an asymptotic lower bound for $T \to +\infty$.
  what about small T? $\to$ see exercise session #4

what if we fix T and choose arbitrarily the bandit instance ν?

Theorem (minimax lower bound)

Let $\mathcal{D} = \left\{ \mathcal{N}(\mu, 1) \mid \mu \in \mathbb{R} \right\}$, $K \geq 2$ and $T \geq K \cdot 1$. Then, there exists a universal constant $c > 0$ ~~such~~ that,

for any policy π, there ~~exists~~ $\nu \in \mathcal{D}^K$ s.t.

$$R_T(\pi, \nu) \geq c\sqrt{KT}$$

Proof in exercise session #4.

minimax $\longleftrightarrow$ $\min_{\pi} \max_{\nu \in \mathcal{D}^K} R_T(\pi, \nu) \geq c\sqrt{KT}$