Lecture #5 : Lower bound

Last lecture, we proposed algorithms with (pseudo) regrets bounded as

$$R_T \leqslant c \sum_{k, \Delta_k > 0} \frac{\ln T}{\Delta_k} \qquad \text{(instance dependent regret)}$$

Is it possible to do better ?

This lecture focuses on lower bounding the achievable regret by any algorithm

For that we consider a model where the rewards distributions belong to some known distribution set $\mathcal{D}$.

ie $\qquad \forall k \in [K], \ \underset{\text{unknown}}{\nu_k} \in \underset{\text{known}}{\mathcal{D}}$

One can show matching upper and lower bounds (with associated strategies) :

$$R_T \text{ is at best of order } \left(\sum_{k, \Delta_k > 0} \frac{\Delta_k}{K_{\inf}(\nu_k, \mu^*, \mathcal{D})} \right) \ln T$$

where

$$K_{\inf}(\nu_k, \mu^*, \mathcal{D}) = \inf \left\{ \underset{\text{Kullback-Leibler divergence}}{KL(\nu_k, \nu')} \ \middle| \ \begin{array}{l} \nu' \in \mathcal{D} \\ E[\nu'] > \mu^* \end{array} \right\}.$$

We will only prove the lower bound part

- **Case 1**: $\mathcal{D} = \{ \mathcal{N}(\mu, \sigma^2) \mid \mu \in \mathbb{R} \}$

then

$$K_{\inf}(\nu_k, \mu^*, \mathcal{D}) = \frac{\Delta_k^2}{2\sigma^2}$$

Best possible regret of order $2\sigma^2 \sum_{k, \Delta_k > 0} \frac{\ln T}{\Delta_k}$

UCB has regret $\lesssim 32\sigma^2 \sum_{k, \Delta_k > 0} \frac{\ln T}{\Delta_k}$

↳ optimal up to constant factor
can be made optimal with finer version

- **Case 2**: $\mathcal{D} = \{ \mathrm{Ber}(p) \mid p \in [0,1] \}$

then

$$K_{\inf}(\nu_k, \mu^*, \mathcal{D}) = \mu_k \ln \frac{\mu_k}{\mu^*} + (1-\mu_k) \ln \frac{1-\mu_k}{1-\mu^*}$$

But before proving the lower bound, I guess that some reminder of basic and non-basic results about KL divergences would be needed!

For sake of time, I will only give these key properties, without any proof.

Definition let $\mathbb{P}, \mathbb{Q}$ be two probability measures over $(\Omega, \mathcal{F})$

$$KL(\mathbb{P}, \mathbb{Q}) = \begin{cases} +\infty & \text{if } \mathbb{P} \text{ is not absolutely continuous wrt } \mathbb{Q} \\ \int_\Omega \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \ln\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)\right) d\mathbb{Q} = \int_\Omega \ln\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{P} & \text{if } \mathbb{P} \ll \mathbb{Q}. \end{cases}$$

$\mathbb{Q}(A) = 0 \Rightarrow \mathbb{P}(A) = 0$

Basic Facts

- existence of the defining integral when $\mathbb{P} \ll \mathbb{Q}$, because $\varphi : x \longmapsto x \ln x$ is bounded from below on $[0, +\infty)$

- $KL(\mathbb{P}, \mathbb{Q}) \geqslant 0$ and $KL(\mathbb{P}, \mathbb{Q}) = 0$ if and only if $\mathbb{P} = \mathbb{Q}$.

indeed, φ is strictly convex. Jensen's inequality indicates that

$$KL(\mathbb{P},\mathbb{Q}) = \int_\Omega \varphi\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{Q} \geqslant \varphi\left(\int_\Omega \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{Q}\right) = \varphi(1) = 0, \text{ with}$$

equality if and only if $\frac{d\mathbb{P}}{d\mathbb{Q}}$ is $\mathbb{Q}$-almost surely constant, ie $\mathbb{P}=\mathbb{Q}$.

A useful rewriting:

Assume $\mathbb{P} \ll \mathbb{Q}$ and let ν be any probability measure over $(\Omega, \mathcal{F})$ with $\mathbb{P} \ll \nu$, $\mathbb{Q} \ll \nu$. Denote $f = \frac{d\mathbb{P}}{d\nu}$, $g = \frac{d\mathbb{Q}}{d\nu}$,

Then $KL(\mathbb{P},\mathbb{Q}) = \int_\Omega \ln\left(\frac{f}{g}\right) f \, d\nu$.

useful when $\mathbb{P}$ and $\mathbb{Q}$ both admit densities over a classical reference measure (eg Lebesgue).

Lemma (data processing inequality)

Let $\mathbb{P}, \mathbb{Q}$ be two probability measures over $(\Omega, \mathcal{F})$.
Let $X: (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ be any random variable.

Denote by $\mathbb{P}^X$ and $\mathbb{Q}^X$ the laws of X under $\mathbb{P}$ and $\mathbb{Q}$.

Then $$KL(\mathbb{P}^X, \mathbb{Q}^X) \leqslant KL(\mathbb{P},\mathbb{Q})$$

Proposition (KL for product measures, independent case)

Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces.

Let $\mathbb{P}, \mathbb{Q}$ be two probability measures over $(\Omega, \mathcal{F})$

$\mathbb{P}', \mathbb{Q}'$ $(\Omega', \mathcal{F}')$

and denote by $\mathbb{P} \otimes \mathbb{P}'$ and $\mathbb{Q} \otimes \mathbb{Q}'$ the product distributions over $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$. Then

$$KL(\mathbb{P} \otimes \mathbb{P}', \mathbb{Q} \otimes \mathbb{Q}') = KL(\mathbb{P}, \mathbb{Q}) + KL(\mathbb{P}', \mathbb{Q}').$$

Consequence (Garivier, Menard, Stoltz 2016)

Data-processing inequality with expectations of random variables.

Let $X : (\Omega, \mathcal{F}) \to ([0,1], \mathcal{B}([0,1]))$ be any $[0,1]$-valued random variable.

Then, denoting by $\mathbb{E}_{\mathbb{P}}[X]$ and $\mathbb{E}_{\mathbb{Q}}[X]$ the respective expectations of X under $\mathbb{P}$ and $\mathbb{Q}$, we have:

$$\mathbb{E}_{\mathbb{P}}[X] \ln \frac{\mathbb{E}_{\mathbb{P}}[X]}{\mathbb{E}_{\mathbb{Q}}[X]} + (1 - \mathbb{E}_{\mathbb{P}}[X]) \ln \frac{1 - \mathbb{E}_{\mathbb{P}}[X]}{1 - \mathbb{E}_{\mathbb{Q}}[X]} = KL(\mathrm{Ber}(\mathbb{E}_{\mathbb{P}}[X]), \mathrm{Ber}(\mathbb{E}_{\mathbb{Q}}[X])) \leq KL(\mathbb{P}, \mathbb{Q}).$$

Proof by upper-bounding $KL\Big(\underbrace{(\mathbb{P} \otimes \mu)^{\mathbb{1}_E}}_{\mathrm{Ber}((\mathbb{P} \otimes \mu)(E))}, \underbrace{(\mathbb{Q} \otimes \mu)^{\mathbb{1}_E}}_{\mathrm{Ber}(\mathbb{Q} \otimes \mu(E))}\Big)$

The chain rule – A generalization of the decomposition of the KL between product-distributions.

We will need it in a special case only, when the joint distributions follow from one of the marginal distributions via a stochastic kernel.

Definition Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces; we denote by $\mathcal{P}(\Omega', \mathcal{F}')$ the set of probability measures over $(\Omega', \mathcal{F}')$.

A (regular) stochastic kernel K is a mapping

$$\begin{aligned} (\Omega, \mathcal{F}) &\longrightarrow \mathcal{P}(\Omega', \mathcal{F}') \\ \omega &\longmapsto K(\omega, \cdot) \end{aligned}$$

such that $\forall B \in \mathcal{F}'$, $\omega \longmapsto K(\omega, B)$ is $\mathcal{F}$-measurable

Now consider two such kernels K and L, and two probability measures $\mathbb{P}$ and $\mathbb{Q}$ over $(\Omega, \mathcal{F})$. Then $K\mathbb{P}$ and $L\mathbb{Q}$ defined below are probability measures over $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$, by some extension theorem (Caratheodory)

$\forall A \in \mathcal{F}$, $\forall B \in \mathcal{F}'$,

$$K\mathbb{P}(A \times B) = \int_\Omega \underbrace{\mathbb{1}_A(\omega)\ K(\omega, B)}_{\text{is indeed measurable}} \, d\mathbb{P}(\omega)$$

$$LQ(A\times B) = \int_{\Omega} \mathbb{1}_A(\omega)\, L(\omega,B)\, dQ(\omega)$$

actually with no loss of generality.

Theorem (chain rule for KL): Assume $\mathbb{P} \ll \mathbb{Q}$

As soon as (*) $K(\omega,\cdot) \ll L(\omega,\cdot)$ for $\mathbb{Q}$-almost all $\omega\in\Omega$

with (**) the existence of a function $g:(\omega,\omega') \longmapsto \frac{dK(\omega,\cdot)}{dL(\omega,\cdot)}(\omega')$ being $\mathcal{F}\otimes\mathcal{F}'$-measurable, ↑ up to a $L\mathbb{Q}$-null set

Then
$$KL(K\mathbb{P}, L\mathbb{Q}) = KL(\mathbb{P},\mathbb{Q}) + \int_{\Omega} KL(K(\omega,\cdot),L(\omega,\cdot))\, d\mathbb{P}(\omega)$$

where $\omega \longmapsto KL(K(\omega,\cdot),L(\omega,\cdot))$ is indeed $\mathcal{F}$-measurable and ≥ 0 so that the integral in the right-hand side is well defined.

Remark:

1) the assumptions (*) and (**) will be satisfied for the applications we have in mind.

2) They can be relaxed: - it suffices to assume that Ω' is a topological space with a countable base and $\mathcal{F}'$ is the Borel σ-algebra.

i.e there exists some countable collection $(O_m)_{m \geq 1}$ of open ets of Ω' such that each open set V of Ω' can be written

$V = \bigcup_{i : O_i \subseteq V} O_i$, that is, as a countable union of elements of

$(O_m)_{m \geq 1}$.

Ex: Ω' a separable metric space → we will consider

$$\Omega' = [0,1] \times (\mathbb{R} \times [0,1])^{\mathbb{N}}$$

3) A typical kernel is given by $K(\omega, B) = \mathbb{P}(Y \in B \mid X = X(\omega))$

The chain rule then rewrites: $KL(\mathbb{P}^{(X,Y)}, \mathbb{Q}^{(X,Y)}) = KL(\mathbb{P}^{X|Y}, \mathbb{Q}^{X|Y}) + KL(\mathbb{P}^{Y}, \mathbb{Q}^{Y})$

Now we have stated the useful properties of the KL, let's get back to the 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

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

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)

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

dependence in strategy π hidden everywhere here.

Proof

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

heck
that it
egular

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

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

defn of the model and strategy

$$= \mathbb{E}_\nu[\mathbb{1}_A(H_t) \cdot \nu_{\pi_{t+1}(H_t)}(B') \cdot \lambda_0(B)]$$

defn of K_{t+1}

$$= \mathbb{E}_\nu[\mathbb{1}_A(H_t)\, K_{t+1}(H_t, B' \times B)]$$

rewriting

$$= \int \mathbb{1}_A(h)\, K_{t+1}(h, B' \times B)\, d\mathbb{P}_\nu^{H_t}(h)$$

defn of $K_{t+1}\mathbb{P}_\nu^{H_t}$

$$= K_{t+1}\mathbb{P}_\nu^{H_t}(A \times B' \times B)$$

$\longrightarrow$ 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

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

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

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

$$= \mathbb{E}[\mathbb{1}_{(a_{r+1}=k)}]$$

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

by induction

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

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

We are now equipped to prove the lower bound.

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

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

problem dependent term

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

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 $\Delta_k > 0$), $\liminf_{T \to \infty} \dfrac{\mathbb{E}[N_k(T)]}{\ln T} \geq \dfrac{1}{K_{\inf}(\nu_k, \mu^*; \mathcal{D})}$

Corollary

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

$$\liminf_{T \to +\infty} \frac{\mathbb{E}[R_T]}{\ln T} \geq \sum_{k,\, \Delta_k > 0} \frac{\Delta_k}{K_{\inf}(\nu_k, \mu^*; \mathcal{D})}$$

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

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

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

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

π is consistent, so:

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

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

so:

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

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

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

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

thus $$\liminf_{T\to+\infty} \frac{\mathbb{E}_\nu[N_a(T)]}{\ln T} \geq \frac{(1-\alpha)}{KL(\nu_a, \nu_a')}$$

(True whether the KL is $<+\infty$ or $=+\infty$ (it is necessarily ≥ 0))

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

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

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:

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

Comments on the lower bound :

- see a comparison with our upper bounds in exercises sheet

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

- this is an asymptotic lower bound for $T \to +\infty$.
  what about small T? → see exercises sheet

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

Theorem (minimax lower bound)

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

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

$$\mathbb{E}(R_T(\pi,\nu)) \geqslant c\sqrt{KT}$$

Proof in exercise below

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

This exercise aims at showing a minimax lower bound of the regret of the form $R_T \geq c\sqrt{KT}$. We restrict ourselves to the bandit model $\mathcal{D} = \{\mathcal{N}(\mu, 1) \mid \mu \in \mathbb{R}\}$, but similar arguments can be used for more general models (e.g. Bernoulli bandits). Fix in the following $K \geq 2$ and $T \geq \frac{K-1}{2}$. The minimax regret is defined as

$$R_T^\star = \inf_{\text{strategy } \pi} \sup_{\text{instance } \nu} \mathbb{E}\left[R_T(\pi, \nu)\right]$$

Let $\varepsilon > 0$. We consider in the following $K+1$ bandit instances $(\nu^j)_{j \in [K+1]}$, where

$$\begin{aligned} \nu_k^j &= \mathcal{N}(0,1) \quad \text{ for any } k \in [K] \text{ such that } j \neq k \\ \nu_k^k &= \mathcal{N}(\varepsilon, 1) \quad \text{ for any } k \in [K]. \end{aligned}$$

1) Justify that

$$R_T^\star \geq \inf_\pi \sup_{\varepsilon \in (0,1)} \max_{i \in [K]} \varepsilon (T - \mathbb{E}_{\nu^i}^\pi [N_i(T)]),$$

and that for any strategy π, there exists k_0 such that $\mathbb{E}_{\nu^0}[N_{k_0}(T)] \leq \frac{T}{K}$.

2) Use the fundamental inequality and Pinsker's inequality to show that

$$\mathbb{E}_{\nu^0}[N_{k_0}(T)] \frac{\varepsilon^2}{2} \geq 2 \left(\mathbb{E}_{\nu^0}[\frac{N_{k_0}(T)}{T}] - \mathbb{E}_{\nu^{k_0}}[\frac{N_{k_0}(T)}{T}] \right)^2.$$

3) Combine the above results to derive

$$R_T^\star \geq \sup_{\varepsilon \in (0,1)} \varepsilon T \left(1 - \frac{1}{K} - \varepsilon \sqrt{\frac{T}{2K}} \right)$$

and conclude that $R_T^\star \geq \frac{1}{8\sqrt{2}} \sqrt{KT}$.

Solution: 1) The first point is just taking a subset over all the possible instances and rewriting the regret as $\varepsilon(T - \mathbb{E}_{\nu^i}^\pi [N_i(T)])$. The second point is because the sum over all k is equal to T, so at least one of them is below (or equal) to the average.

2) Direct use of fundamental inequality and Pinsker's inequality, with the fact that for $i \neq 0$, $\mathrm{KL}(\nu^0, \nu^i) = \frac{\varepsilon^2}{2}$.

3) We have

$$\mathbb{E}_{\nu^0}[N_{k_0}(T)] \frac{\varepsilon^2}{2} \geq 2 \left(\mathbb{E}_{\nu^0}[\frac{N_{k_0}(T)}{T}] - \mathbb{E}_{\nu^{k_0}}[\frac{N_{k_0}(T)}{T}] \right)^2.$$

So that $\mathbb{E}_{\nu^{k_0}}[\frac{N_{k_0}(T)}{T}] \leq \mathbb{E}_{\nu^0}[\frac{N_{k_0}(T)}{T}] + \varepsilon \sqrt{\frac{\mathbb{E}_{\nu^0}[N_{k_0}(T)]}{2}}$. Since $\mathbb{E}_{\nu^0}[N_{k_0}(T)] \leq \frac{T}{K}$, this implies

$$\mathbb{E}_{\nu^{k_0}}[\frac{N_{k_0}(T)}{T}] \leq \frac{1}{K} + \varepsilon \sqrt{\frac{T}{2K}}.$$

This then yields

$$R_T(\pi, \nu^{k_0}) \geq \varepsilon T \left(1 - \frac{1}{K} - \varepsilon \sqrt{\frac{T}{2K}} \right).$$

Using question 1), this yields the first point, by noticing that the bound actually does not depend on the choice of the strategy π. We conclude by taking $\varepsilon = \sqrt{\frac{K}{2T} \frac{(K_1)^2}{K^2}}$, which is smaller than 1 since $T \geq \frac{K-1}{2}$, and noticing that for $K \geq 2$, $\frac{K-1}{K} \geq \frac{1}{2}$.