Lecture 6 — Winter 2019 term
Lecture 6 — Winter 2019 term

<u>KL for product measures.</u> ( $\leftrightarrow$ The independent case)

<u>Proposition:</u> Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces,
let $\mathbb{P}, \mathbb{Q}$ be two probability measures over $(\Omega, \mathcal{F})$
$\mathbb{P}', \mathbb{Q}'$ over $(\Omega', \mathcal{F}')$
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:

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

Proof: We have $\mathbb{P}\otimes\mathbb{P}' \ll \mathbb{Q}\otimes\mathbb{Q}' \iff [\mathbb{P}\ll\mathbb{Q}$ and $\mathbb{P}'\ll\mathbb{Q}']$

So we can assume that all $\ll$ statements hold, and then

$$\frac{d(\mathbb{P}\otimes\mathbb{P}')}{d(\mathbb{Q}\otimes\mathbb{Q}')} = \frac{d\mathbb{P}}{d\mathbb{Q}}\,\frac{d\mathbb{P}'}{d\mathbb{Q}'}$$

(this is a fundamental result in measure theory and one of the best characterizations of independence!).

Therefore,

$$\mathrm{KL}(\mathbb{P}\otimes\mathbb{P}', \mathbb{Q}\otimes\mathbb{Q}') = \int_{\Omega\times\Omega'} \left(\frac{d\mathbb{P}}{d\mathbb{Q}}\,\frac{d\mathbb{P}'}{d\mathbb{Q}'} \ln\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\,\frac{d\mathbb{P}'}{d\mathbb{Q}'}\right)\right) d(\mathbb{Q}\otimes\mathbb{Q}')$$

We use that if $f,g$ are $\geqslant$ a constant, then $\int(f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu$

$$= \underbrace{\int_{\Omega'}\Bigg(\underbrace{\int_{\Omega}\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\ln\frac{d\mathbb{P}}{d\mathbb{Q}}\right)d\mathbb{Q}}_{=\mathrm{KL}(\mathbb{P},\mathbb{Q})}\Bigg)\underbrace{\frac{d\mathbb{P}'}{d\mathbb{Q}'}\,d\mathbb{Q}'}_{d\mathbb{P}'}}_{=\mathrm{KL}(\mathbb{P},\mathbb{Q})} + \underbrace{\text{similar term with } \ln\frac{d\mathbb{P}'}{d\mathbb{Q}'}}_{=\mathrm{KL}(\mathbb{P}',\mathbb{Q}')}$$

here we apply Tonelli's theorem (again because $x\mapsto x\ln x$ is lower bounded)

Consequence (Garivier, Ménard, Stoltz, 2016):

Data-processing inequality with expectations of random variables

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

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:

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

Proof: We denote by $\mathfrak{m}$ the Lebesgue measure over $[0,1]$ and augment the underlying measurable space into $(\Omega\times[0,1], \mathcal{F}\otimes\mathcal{B}([0,1]))$, over which we consider the product-distributions $\mathbb{P}\otimes\mathfrak{m}$ and $\mathbb{Q}\otimes\mathfrak{m}$.

For any event $E\in\mathcal{F}\otimes\mathcal{B}([0,1])$, we have, by the data-processing inequality:

 Lecture 6 — Winter 2019 term Lecture 6 — Winter 2019 term

$$\mathrm{KL}\Big( \underbrace{(\mathbb{P}\otimes m)^{\mathbb{1}_E}}_{\mathrm{Ber}(\mathbb{P}\otimes m(E))}, \underbrace{(\mathbb{Q}\otimes m)^{\mathbb{1}_E}}_{\mathrm{Ber}(\mathbb{Q}\otimes m(E))} \Big) \leqslant \mathrm{KL}(\mathbb{P}\otimes m, \mathbb{Q}\otimes m)$$

$$= \mathrm{KL}(\mathbb{P},\mathbb{Q}) + \mathrm{KL}(m,m)$$

cf. product distributions

$$= \mathrm{KL}(\mathbb{P},\mathbb{Q})$$

Thus: $\mathrm{KL}\big( \mathrm{Ber}(\mathbb{P}\otimes m(E)), \mathrm{Ber}(\mathbb{Q}\otimes m(E))\big) \leqslant \mathrm{KL}(\mathbb{P},\mathbb{Q})$

The proof is concluded by picking $E \in \mathcal{F}\otimes\mathcal{B}([0,1])$ such that
$\mathbb{P}\otimes m(E) = \mathbb{E}_{\mathbb{P}}[X]$ and $\mathbb{Q}\otimes m(E) = \mathbb{E}_{\mathbb{Q}}[X]$

Namely, $E = \big\{ (\omega,x) \in \Omega\times[0,1] : x \leqslant X(\omega) \big\}$

By Tonelli's theorem:

$$\mathbb{P}\otimes m(E) = \int_\Omega \left( \int_{[0,1]} \mathbb{1}_{\{x \leqslant X(\omega)\}} \, dm(x) \right) d\mathbb{P}(\omega)$$

$$= \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \mathbb{E}_{\mathbb{P}}[X]$$

and a similar equality for $\mathbb{Q}\otimes m(E)$.

<u>The chain rule</u> — 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.

<u>Definition:</u> 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 stochastic kernel (regular) $K$ is a mapping

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

such that $\forall B \in \mathcal{F}'$ $\omega \mapsto K(\omega,B)$ 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: (Carathéodory?)

$\forall A \in \mathcal{F}, \quad \forall B \in \mathcal{F}', \qquad KP(A \times B) = \displaystyle\int_\Omega \mathbb{1}_A(\omega) \underbrace{K(\omega, B)}_{\text{is indeed measurable}} \, dP(\omega)$

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

An extension of Fubini(–Tonelli) theorem

immediate extension to $\varphi$ lower bounded by $m \in \mathbb{R}$

**Lemma:** Let $\varphi : \Omega \times \Omega' \to \mathbb{R}$ be $\mathcal{F} \otimes \mathcal{F}'$-measurable and either $\geqslant 0$ or $KP$-integrable.

Then $\omega \mapsto \displaystyle\int_{\Omega'} \varphi(\omega, \omega')\, K(\omega, d\omega')$ is $\mathcal{F}$-measurable and

$$\int_{\Omega \times \Omega'} \varphi \, dKP = \int_\Omega \left( \int_{\Omega'} \varphi(\omega, \omega')\, K(\omega, d\omega') \right) dP(\omega)$$

**Proof:** (sketch)

including measurability of $\omega \mapsto \int \varphi(\omega, \cdot)\, K(\omega, d\cdot)$ by regularity of $K$

The result is true for $\varphi = \mathbb{1}_{A \times B}$ by definition of $KP$

Extension to $\mathbb{1}_E$ for any $E \in \mathcal{F} \otimes \mathcal{F}'$ by an argument of σ-algebra contained / monotone class theorem, using monotone convergence (including the $\omega \mapsto \int_{\Omega'} \ldots$ measurability)

Extension to $\varphi \geqslant 0$ by monotone convergence

Extension to $\varphi \in \mathbb{L}^1$ by dominated convergence

**Question:** Does anyone have a simpler argument?

**Theorem** [chain rule for KL]:

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

with (**) the existence of a version $g : (\omega, \omega') \mapsto \dfrac{dK(\omega, \cdot)}{dL(\omega, \cdot)}(\omega')$ being $\mathcal{F} \otimes \mathcal{F}'$-measurable,

Then

$$KL(KP, LQ) = KL(P, Q) + \int_\Omega KL\big(K(\omega, \cdot), L(\omega, \cdot)\big)\, dP(\omega)$$

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

Remark: see last page of these lecture notes for the (lack of) necessity of assumptions (*) and (**).

 Lecture 6 — Winter 2019 term Lecture 6 — Winter 2019 term

Proof: * By bi-measurability of $g \ln g$, and since $g \ln g$ is lower bounded,

(an immediate extension of) the previous lemma can be applied to get

$$\omega \longmapsto \int_{\Omega'} g(\omega,\cdot)\,\ln\big(g(\omega,\cdot)\big)\,L(\omega,d\cdot) = KL\big(K(\omega,\cdot),\,L(\omega,\cdot)\big)$$

is $\mathcal{F}$-measurable and $\geqslant 0$, with:

we will not use this, actually $\Big\{$
$$\int_{\Omega\times\Omega'} g\ln g \; d\,L\mathbb{P} = \int KL\big(K(\omega,\cdot),\,L(\omega,\cdot)\big)\,d\mathbb{P}(\omega)$$

* We assume $\mathbb{P} \ll \mathbb{Q}$, let $f = \dfrac{d\mathbb{P}}{d\mathbb{Q}}$ : what can we say about $(\omega,\omega') \mapsto f(\omega)\,g(\omega,\omega')$?

$$\int \mathbb{1}_{A\times B}(\omega,\omega')\,f(\omega)\,g(\omega,\omega')\;dL\mathbb{Q}(\omega,\omega')$$

$$\underset{\text{cf. extension of Tonelli}}{=} \int_{\Omega}\Bigg(\underbrace{\int_{\Omega'} \mathbb{1}_B(\omega')\,g(\omega,\omega')\,L(\omega,d\omega')}_{\substack{=\int_{\Omega'}\mathbb{1}_B(\omega')\,K(\omega,d\omega') \\ = K(\omega,B) \\ \text{given the definition of } g}}\Bigg)\,\mathbb{1}_A(\omega)\,f(\omega)\,d\mathbb{Q}(\omega)$$

$$= \int \underbrace{\mathbb{1}_A(\omega)\,K(\omega,B)}_{\mathcal{F}\text{-measurable}}\;\underbrace{f(\omega)\,d\mathbb{Q}(\omega)}_{\substack{d\mathbb{P}(\omega) \\ \text{since } f = d\mathbb{P}/d\mathbb{Q}}} = K\mathbb{P}(A\times B) \quad \text{by def. of } K\mathbb{P}$$

By Radon–Nikodym's theorem: $\dfrac{dK\mathbb{P}}{dL\mathbb{Q}} = f\,g \quad L\mathbb{Q}\text{-a.s.}$

* It is easily seen that $K\mathbb{P} \ll L\mathbb{Q} \Rightarrow \mathbb{P} \ll \mathbb{Q}$ (in all cases, even without (*) and (**))

* Therefore, we have $K\mathbb{P} \ll L\mathbb{Q} \iff \mathbb{P} \ll \mathbb{Q}$ under (*) and (**), we thus assume with no loss of generality that $K\mathbb{P} \ll L\mathbb{Q}$ and $\mathbb{P} \ll \mathbb{Q}$ (otherwise both $= +\infty$ and the putative equality is $+\infty = +\infty$).

Then, $$KL(K\mathbb{P},\,L\mathbb{Q}) = \int_{\Omega\times\Omega'} \Big(f(\omega)\,g(\omega,\omega')\,\ln f(\omega)\,g(\omega,\omega')\Big)\;d\,L\mathbb{Q}(\omega,\omega')$$

$\varphi = fg \ln(fg)$ is lower bounded, the lemma (extension of Fubini-Tonelli) extends to it:

$$\int fg \ln(fg)\, d\mathbb{L}\mathbb{Q}$$

$$= \int_{\Omega} \Bigg( \bigg( \int_{\Omega'} \underbrace{g(\omega,\omega') \ln g(\omega,\omega')\, L(\omega, d\omega')}_{\mathrm{KL}(K(\omega,\cdot),\, L(\omega,\cdot))} \bigg) + \ln f(\omega) \Bigg) f(\omega)\, d\mathbb{Q}(\omega)$$

$$= \int_{\Omega} \Big( \underbrace{f(\omega)\, \mathrm{KL}(K(\omega,\cdot), L(\omega,\cdot))}_{\geqslant 0} + \underbrace{f(\omega) \ln f(\omega)}_{\text{lower bounded}} \Big)\, d\mathbb{Q}(\omega)$$

$\varphi$ sum of two bounded functions

$$= \int_{\Omega} \mathrm{KL}(K(\omega,\cdot), L(\omega,\cdot))\, \underbrace{f(\omega) d\mathbb{Q}(\omega)}_{d\mathbb{P}(\omega)} + \underbrace{\int f \ln f \, d\mathbb{Q}}_{\mathrm{KL}(\mathbb{P},\mathbb{Q})}$$

## REMARKS ON THE ASSUMPTIONS

- The assumptions $(*)$ and $(**)$ will be satisfied for the applications we have in mind.
- They can be relaxed: it suffices to assume that $\Omega'$ is a topological space with a countable base (a "second-countable space") and $\mathcal{F}'$ is the Borel $\sigma$-algebra.

  I.e., there exists some countable collection $(O_m)_{m \geqslant 1}$ of open sets of $\Omega'$ such that each open set $V$ of $\Omega'$ can be written

$$V = \bigcup_{i:\, O_i \subseteq V} O_i$$

  that is, as a countable union of elements of $(O_m)_{m\geqslant 1}$.

  Ex: $\Omega'$ a separable metric space $\longrightarrow$ we will consider $\Omega' = [0,1] \times (\mathbb{R} \times [0,1])^{\mathbb{N}}$

↳ See details in the additional document.
CREDITS: Marin Ballu + Hédi Hadiji, M2 students of Spring 2017

 Lecture 6 — Winter 2019 term Lecture 6 — Winter 2019 term

# Lower bounds on the regret for stochastic bandits.

Here is first a <u>summary</u> of the setting and context of stochastic bandits:

- $K$ arms each indexed by $a = 1, 2, \dots K$
- With each arm is associated a probability distribution $\nu_a \in \mathcal{D}$
- $\mathcal{D}$ is the bandit model: a subset of $\mathcal{M}_1(\mathbb{R})$, the set of probability distributions over $\mathbb{R}$ with an expectation
- A bandit problem is denoted by $\underline{\nu} = (\nu_a)_{a \in \{1,\dots K\}}$
- Important quantities and notation:

  $\mu_a = \mathbb{E}(\nu_a)$ is the expectation of $\nu_a$

  $\mu^* = \max\limits_{a=1\dots K} \mu_a$ is the largest expectation within $\underline{\nu}$

  $\Delta_a = \mu^* - \mu_a$ is the gap for arm $a$

  Arm $a$ is suboptimal if $\Delta_a > 0$

(Note: the decision-maker knows $\mathcal{D}$ but does not know the specific $\underline{\nu} \in \mathcal{D}^K$ at hand)

- Protocol: at each round $t = 1, 2, \dots$

  $U_0, U_1, U_2, \dots$ iid $\sim \mathcal{U}_{[0,1]}$

  1. The decision-maker picks $I_t \in \{1, \dots K\}$ possibly at random based on an auxiliary randomization $U_{t-1}$
  2. She gets a reward $Y_t$ drawn at random according to $\nu_{I_t}$ (given $I_t$); this is the only piece of information she gets.

- Aim / regret: maximize $\mathbb{E}\left[\sum_{t=1}^{T} Y_t\right]$

  which is equivalent to minimizing (controlling from above)

  $$R_T = T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} Y_t\right]$$

- Rewriting by tower rule:

  $$\overline{R}_T = T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} \mu_{I_t}\right] = \sum_{a=1}^{K} \Delta_a \, \mathbb{E}\left[N_a(T)\right]$$

  where $N_a(T) = \sum_{t=1}^{T} \mathbb{1}_{\{I_t = a\}}$ is the number of times arm $a$ was pulled between $1$ and $T$

! It is thus necessary and sufficient to control $\mathbb{E}[N_a(T)]$ for suboptimal arms $a$

- What is a (randomized) strategy?

  A sequence of measurable functions $(\Psi_t)_{t \geq 0}$ with

  $$\Psi_t : \underbrace{H_t = (U_0, Y_1, U_1, \dots Y_t, U_t)}_{\text{history for the first } t \text{ rounds}} \longmapsto \underbrace{\Psi_t(H_t) = I_{t+1}}_{\text{arm picked at round } t+1}$$

- Strategies that are consistent wrt a model $\mathcal{D}$:

  If for all bandit problems $\underline{\nu} \in \mathcal{D}^K$,

  $\forall \alpha \in (0,1]$, $\forall a$ s.t. $\Delta_a > 0$, $\mathbb{E}[N_a(T)] = o(T^\alpha)$.

- Result: For "well behaved" models $\mathcal{D}$, there exist consistent strategies.

  E.g.: at least $\mathcal{D} = \mathcal{M}_1([0,1])$, see the UCB strategy.

- Typical bounds for good strategies (stated in an asymptotic way, even though non-asymptotic bounds are available)

  $\forall \underline{\nu} \in \mathcal{D}^K$, $\forall a$ s.t. $\Delta_a > 0$, $\displaystyle \limsup_{T \to +\infty} \frac{\mathbb{E}[N_a(T)]}{\ln T} \leq C_a(\underline{\nu})$

  where $C_a(\underline{\nu})$ is a problem-dependent constant.

- Optimal (in some sense) such constant: $\displaystyle C_a(\underline{\nu}) = \frac{1}{K_{\inf}(\nu_a, \mu^*, \mathcal{D})} = \frac{1}{K_{\inf}(\nu_a, \mu^*)}$

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

  with the convention: $\inf \emptyset = +\infty$.

  We will only prove one part of this optimality: a lower bound on $C_a(\underline{\nu})$.

Theorem:

(see Lai and Robbins, 1985; Burnetas and Katehakis, 1996)

For all bandit models $\mathcal{D} \subset \mathcal{M}_1(\mathbb{R})$,

For all strategies $\Psi$ consistent wrt $\mathcal{D}$ (possibly randomized),

For all bandit problems $\underline{\nu} = (\nu_a)_{a \in \{1 \dots K\}} \in \mathcal{D}^K$,

For all suboptimal arms $a$ (ie, such that $\Delta_a > 0$),

$$\liminf \frac{\mathbb{E}[N_a(T)]}{\ln T} \geq 1 / K_{\inf}(\nu_a, \mu^*, \mathcal{D})$$

Corollary: For all bandit models $\mathcal{D} \subseteq \mathcal{M}_1(\mathbb{R})$,
For all (possibly randomized) strategies $\psi$ consistent wrt $\mathcal{D}$,
For all bandit problems $\underline{\nu} = (\nu_a)_{a \in \{1..K\}} \in \mathcal{D}^K$,

$$\liminf_{T \to +\infty} \frac{\overline{R}_T}{\ln T} \geqslant \sum_{a : \Delta_a > 0} \frac{\Delta_a}{K_{\inf}(\nu_a, \mu^*, \mathcal{D})}.$$

To prove this theorem (and to prove other lower bounds), we will need the following fundamental inequality. In its statement, $\mathbb{P}_{\underline{\nu}}$ and $\mathbb{E}_{\underline{\nu}}$ refer to the probability distribution and the expectation induced by the bandit problem $\underline{\nu} \in \mathcal{D}^K$.

Example: $\mathbb{P}_{\underline{\nu}}^{H_T}$ is the law of $H_T = (U_0, Y_1, U_1 \ldots Y_T, U_T)$ when the bandit problem is $\underline{\nu}$. Actually, $\mathbb{P}_{\underline{\nu}}^{H_T}$ strongly depends on the strategy $\psi$ used but we omit this dependency in the notation.

Lemma (Fundamental inequality for stochastic bandits):

For all bandit problems $\underline{\nu} = (\nu_a)_{a \in \{1..K\}}$ and $\underline{\nu}' = (\nu'_a)_{a \in \{1..K\}}$ in $\mathcal{D}^K$
with $\nu_a \ll \nu'_a$ for all $a$,
For all strategies
For all random variables $Z$ taking values in $[0,1]$ and that are $\sigma(H_T)$-measurable,

$$\sum_{a=1}^{K} \mathbb{E}_{\underline{\nu}}[N_a(T)] \, KL(\nu_a, \nu'_a) = KL\left(\mathbb{P}_{\underline{\nu}}^{H_T}, \mathbb{P}_{\underline{\nu}'}^{H_T}\right)$$

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

The dependence on the strategy is hidden in the $\mathbb{E}_{\underline{\nu}}[N_a(T)]$, $\mathbb{E}_{\underline{\nu}}[Z]$ and $\mathbb{E}_{\underline{\nu}'}[Z]$

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)

We have $K_{\inf}(\nu_a, \mu^*) = \inf\{ KL(\nu_a, \nu'_a) : \nu'_a \in \mathcal{D} \text{ and } E(\nu'_a) > \mu^* \}$

$= \inf\{ KL(\nu_a, \nu'_a) : \nu'_a \in \mathcal{D},\ \nu_a \ll \nu'_a \text{ and } E(\nu'_a) > \mu^* \}$

(cf. convention: $\inf \emptyset = +\infty$ and the fact that $KL(\nu_a, \nu'_a) = +\infty$ when $\nu_a \not\ll \nu'_a$)

This is why we will

- Fix $\mathcal{D}$, $\psi$, $\underline{\nu}$ and $a$ s.t. $\Delta_a > 0$
- Fix an alternative model $\underline{\nu}'$ of the form

$$\begin{cases} \nu'_k = \nu_k & \forall k \neq a \\ \nu'_a & \text{s.t. } \nu'_a \in \mathcal{D},\ \nu_a \ll \nu'_a \text{ and } E(\nu'_a) > \mu^* \end{cases}$$

That is, $\underline{\nu}$ and $\underline{\nu}'$ only differ at $a$; $a$ is the unique optimal arm in $\underline{\nu}'$

- Take $Z = N_a(T)/T$ which is indeed $[0,1]$-valued, $\sigma(H_T)$-measurable

Our fundamental inequality (the lemma) yields, since $\underline{\nu}$ and $\underline{\nu}'$ only differ at $a$:

$$\mathbb{E}_{\underline{\nu}}[N_a(T)]\ KL(\nu_a, \nu'_a) \geqslant KL\Big(\mathrm{Ber}\big(\mathbb{E}_{\underline{\nu}}[N_a(T)/T]\big), \mathrm{Ber}\big(\mathbb{E}_{\underline{\nu}'}[\tfrac{N_a(T)}{T}]\big)\Big)$$

$$\geqslant -\ln 2 + \big(1 - \mathbb{E}_{\underline{\nu}}[N_a(T)/T]\big) \ln \frac{1}{1 - \mathbb{E}_{\underline{\nu}'}[N_a(T)/T]}$$

indeed: $KL(\mathrm{Ber}(p), \mathrm{Ber}(q))$

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

$$= \underbrace{p \ln \frac{1}{q}}_{\geqslant 0} + (1-p) \ln \frac{1}{1-q} + \underbrace{\big(p \ln p + (1-p)\ln(1-p)\big)}_{\geqslant -\ln 2 \text{ by a simple function study over } [0,1]}$$

$$\geqslant -\ln 2 + (1-p) \ln \frac{1}{1-q}$$

for all $p, q \in (0,1)$ and even for all $p, q \in [0,1]$ (study the cases $q = 0$ and $q = 1$ separately)

Now, the considered strategy $\psi$ is consistent and:

- in the problem $\underline{\nu}$, $a$ is suboptimal: $\mathbb{E}_{\underline{\nu}}[N_a(T)/T] \longrightarrow 0$

– in the problem $\nu'$, all arms $k \neq a$ are suboptimal:

for all $\alpha \in (0,1]$, $\quad T - \mathbb{E}_{\nu'}[N_a(T)] = \sum_{k \neq a} \underbrace{\mathbb{E}_{\nu'}[N_k(T)]}_{=o(T^\alpha)}$

↳ in particular, for $T$ large enough,

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

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

$$\frac{\mathbb{E}_\nu[N_a(T)] \ \mathrm{KL}(\nu_a, \nu'_a)}{\ln T} \geqslant -\frac{\ln 2}{\ln T} + \Bigl(1 - \underbrace{\mathbb{E}_\nu\Bigl[\frac{N_a(T)}{T}\Bigr]}_{\to 0}\Bigr) \underbrace{\frac{\ln T^{1-\alpha}}{\ln T}}_{=1-\alpha}$$

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

Letting $\alpha \to 0$, $\quad \liminf_{T \to +\infty} \frac{\mathbb{E}_\nu[N_a(T)] \ \mathrm{KL}(\nu_a, \nu'_a)}{\ln T} \geqslant 1$

Whether $\mathrm{KL}(\nu_a, \nu'_a) < +\infty$ or $= +\infty$, we thus get

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

The left-hand side is independent of $\nu'_a \in \mathcal{D}$ s.t. $\nu'_a \gg \nu_a$ and $E(\nu'_a) > \mu^*$, so that taking the supremum of the right-hand side over these $\nu'_a$, we get the desired $1/\mathcal{K}_{\inf}(\nu_a, \mu^*)$ lower bound.

**Proof of the lemma:** • The inequality $\geqslant$ is a direct application of the data-processing inequality with expectations, see the previous lecture for its statement.

• For the equality: We will explain how $\mathbb{P}_{\underline{\nu}}^{H_T}$ is constructed.

Do it only for the construction of $\mathbb{P}_{\underline{\nu}}^{H_T}$

With no loss of generality, we can consider that

- the underlying probability space is $\Omega = [0,1] \times (\mathbb{R} \times [0,1])^T$
- $H_T$ is the identity over $\Omega$, i.e., that the $U_0, Y_1, U_1, \dots Y_T, U_T$ are the projections on each component,

in particular, $\mathbb{P}_{\underline{\nu}}^{H_t}$ refers to the first $1+2t$ marginals of $\mathbb{P}_{\underline{\nu}}$

Then, $\mathbb{P}_{\underline{\nu}}$ is characterized* by:

$$\forall B \in \mathcal{B}([0,1]), \qquad \mathbb{P}_{\underline{\nu}}(U_0 \in B) = \eta(B)$$

$$\forall t \in \{0, \dots T-1\}, \quad \forall B' \in \mathcal{B}(\mathbb{R}), \ \forall B \in \mathcal{B}([0,1]), \qquad \mathbb{P}_{\underline{\nu}}\big(Y_{t+1} \in B' \text{ and } U_{t+1} \in B \,\big|\, H_t\big) = \nu_{\psi_t(H_t)}(B')\,\eta(B)$$

where $\mathcal{B}(S)$ is the Borel-$\sigma$-algebra of a set $S \subseteq \mathbb{R}$
$\eta$ is the Lebesgue measure over $[0,1]$

* Why characterized?
↳ because these formulas give us the probability of any set of the form $B_0 \times B'_1 \times B_1 \times \dots \times B'_T \times B_T$

Actually, the equality $\mathbb{P}_{\underline{\nu}}\big(Y_{t+1} \in B' \text{ and } U_{t+1} \in B \,|\, H_t\big) = \nu_{\psi_t(H_t)}(B')\,\eta(B)$ exactly indicates that $\mathbb{P}_{\underline{\nu}}^{H_{t+1}} = K_t\, \mathbb{P}_{\underline{\nu}}^{H_T}$ for the regular transition kernel

$$K_t(h,\cdot) = \nu_{\psi_t(h)} \otimes \eta$$

see details on next page; it follows from the above characterization of $\mathbb{P}_{\underline{\nu}}$

regularity is: for $E \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}([0,1])$, $h \mapsto K_t(h,E)$ is measurable; it follows from the measurability of $\psi_t$

Let us check the assumptions of our chain rule:

(*) $\forall h, \quad K_t(h,\cdot) \ll K'_t(h,\cdot)$ as $\forall a, \ \nu_a \ll \nu'_a$ by assumption

(**) $\displaystyle (h,(y,u)) \longmapsto \frac{dK_t(h,\cdot)}{dK'_t(h,\cdot)}(y,u) = \sum_{a=1}^{K} \mathbb{1}_{\{\psi_t(h)=a\}} \frac{d\nu_a}{d\nu'_a}(y)$
is indeed bi-measurable.

Therefore, for $t \in \{0, \dots T-1\}$,

$$\mathrm{KL}\big(\mathbb{P}_{\underline{\nu}}^{H_{t+1}}, \mathbb{P}_{\underline{\nu}'}^{H_{t+1}}\big) = \mathrm{KL}\big(\mathbb{P}_{\underline{\nu}}^{H_t}, \mathbb{P}_{\underline{\nu}'}^{H_t}\big) + \int \mathrm{KL}\big(\nu_{\psi_t(h)} \otimes \cancel{\eta}, \nu'_{\psi_t(h)} \otimes \cancel{\eta}\big)\, d\mathbb{P}_{\underline{\nu}}^{H_t}(h)$$

$$= \mathrm{KL}\big(\mathbb{P}_{\underline{\nu}}^{H_t}, \mathbb{P}_{\underline{\nu}'}^{H_t}\big) + \sum_{a=1}^{K} \mathrm{KL}(\nu_a, \nu'_a)\ \mathbb{P}_{\underline{\nu}}^{H_t}\{\psi_t(h)=a\}$$

Now, $I_{t+1} = \Psi_t(H_t)$, so that

$$\mathbb{P}_{\nu}^{H_t}\{\Psi_t(h)=a\} = \mathbb{P}_{\nu}\{\Psi_t(H_t)=a\}$$
$$= \mathbb{P}_{\nu}\{I_{t+1}=a\} = \mathbb{E}_{\nu}\left[\mathbb{1}_{\{I_{t+1}=a\}}\right]$$

Summing up:

- $\mathrm{KL}\left(\mathbb{P}_{\nu}^{H_0}, \mathbb{P}_{\nu'}^{H_0}\right) = \mathrm{KL}\left(\mathbb{P}_{\nu}^{U_0}, \mathbb{P}_{\nu'}^{U_0}\right) = \mathrm{KL}(\eta,\eta) = 0$
- $\forall t \in \{0,\ldots,T-1\}$, $\mathrm{KL}\left(\mathbb{P}_{\nu}^{H_{t+1}}, \mathbb{P}_{\nu'}^{H_{t+1}}\right) = \mathrm{KL}\left(\mathbb{P}_{\nu}^{H_t}, \mathbb{P}_{\nu'}^{H_t}\right) + \displaystyle\sum_{a=1}^{K} \mathrm{KL}(\nu_a,\nu'_a)\, \mathbb{E}_{\nu}\left[\mathbb{1}_{\{I_{t+1}=a\}}\right]$

so that the stated result follows by induction.

* Explanation of why $\mathbb{P}_{\nu}^{H_{t+1}} = K_t\, \mathbb{P}_{\nu}^{H_t}$:

$\forall A \in \mathcal{B}\left([0,1]\times(\mathbb{R}\times[0,1])^t\right)$, $\forall B' \in \mathcal{B}(\mathbb{R})$, $B \in \mathcal{B}([0,1])$,

$$K_t\, \mathbb{P}_{\nu}^{H_t}\left(A \times (B'\times B)\right)$$

def of product

$$= \int_{[0,1]\times(\mathbb{R}\times[0,1])^t} \mathbb{1}_A(h)\; K_t(h, B'\times B)\; d\mathbb{P}_{\nu}^{H_t}(h)$$

def of image distribution

$$= \mathbb{E}_{\nu}\left[\mathbb{1}_A(H_t)\; K_t(H_t, B'\times B)\right]$$

$K_t$ defined as some conditional probability

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

tower rule

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

**Exercise:** $1/\mathcal{K}_{\inf}(\nu_a, \mu^*, \mathcal{D})$ vs. $8/\Delta_a^2$ for UCB

Recall that in the model $\mathcal{D} = \mathcal{P}([0,1])$, the UCB algorithm enjoys the following performance bound:

$$\forall \underline{\nu} \in \mathcal{P}([0,1])^K, \qquad \forall a \text{ s.t. } \Delta_a > 0,$$

$$\mathbb{E}_{\underline{\nu}}\left[N_a(T)\right] \leqslant \frac{8}{\Delta_a^2} \ln T + 2.$$

Actually, there are refinements of UCB that get the distribution-dependent constant $8/\Delta_a^2$ arbitrarily close to $2/\Delta_a^2$.

But how do those $8/\Delta_a^2$ and $2/\Delta_a^2$ constants compare to $1/\mathcal{K}_{\inf}(\nu_a, \mu^*, \mathcal{P}([0,1]))$ ?

(1) For $p, q \in [0,1]$, we denote

$$\mathrm{kl}(p,q) = \mathrm{KL}(\mathrm{Ber}(p), \mathrm{Ber}(q))$$

Show that $\forall (p,q) \in [0,1]^2, \qquad \mathrm{kl}(p,q) \geqslant 2(p-q)^2.$

(2) Show Pinsker's inequality: let $(\Omega, \mathcal{F})$ be a measurable space, let $\mathbb{P}, \mathbb{Q}$ be two distributions over $(\Omega, \mathcal{F})$, then:

$$\| \mathbb{P} - \mathbb{Q} \|_{TV} = \sup_{A \in \mathcal{F}} \left| \mathbb{P}(A) - \mathbb{Q}(A) \right| \leqslant \sqrt{\frac{1}{2} \mathrm{KL}(\mathbb{P}, \mathbb{Q})}$$

the total variation distance between $\mathbb{P}$ and $\mathbb{Q}$

Even better, show the stronger form:

$$\sup_{\substack{Z \text{ } \mathcal{F}\text{-measurable} \\ \text{taking values in } [0,1]}} \left| \mathbb{E}_{\mathbb{P}}[Z] - \mathbb{E}_{\mathbb{Q}}[Z] \right| \leqslant \sqrt{\frac{1}{2} \mathrm{KL}(\mathbb{P}, \mathbb{Q})}$$

(3) Exhibit a lower bound on $\mathcal{K}_{\inf}(\nu_a, \mu^*, \mathcal{P}([0,1]))$ and conclude that some work is needed to get an upper bound matching our lower bound!

 Lecture 6 — Winter 2019 term Lecture 6 — Winter 2019 term

Exercice : Finite-time lower bound for small values of $T$

"All algorithms explore much!" ↳ We want to model that all algorithms must first explore uniformly all arms ($\leftrightarrow$ exploration) at least half of the time, before being able to perform exploitation more often.

(1) Establish the following local version of Pinsker's inequality:

- $\forall\, 0 \leq p < q \leq 1$,

$$\mathrm{kl}(p,q) \geq \frac{1}{2 \max_{x \in [p,q]} x(1-x)} (p-q)^2$$

$$\geq \frac{1}{2q} (p-q)^2$$

- Why is it stronger than the global version of Pinsker's inequality?

(2) Show that all strategies smarter than the uniform strategy [i.e., such that for all bandit problems, $\forall a$ s.t. $\mu_a = \mu^*$, $\mathbb{E}[N_a(T)] \geq T/K$ ], we have:

$$\forall\, T \leq \frac{1}{8\, \mathrm{KL}^*}, \qquad \text{where } \mathrm{KL}^* = \max_{j : \Delta_j > 0} \mathcal{K}_{\inf}(\nu_j, \mu_j^*)$$

$\forall j$ s.t. $\Delta_j > 0$,

$$\mathbb{E}[N_j(T)] \geq \underbrace{\frac{1}{2}}_{\text{at least half of the time}} \; \underbrace{\frac{T}{K}}_{\text{uniform exploration}}$$

Hint: consider the same alternative bandit problems $\nu'$, as in the theorem giving the asymptotic lower bound.