Part 1: The Kullback-Leibler divergence, more properties

Recap of what we already saw on KL-divergences:

These properties heavily rely on the facts that $\varphi : x \mapsto x \ln x$ is strictly convex and $\geq -1/e$

Definition: $$\mathrm{KL}(\mathbb{P},\mathbb{Q}) = \begin{cases} +\infty & \text{if } \mathbb{P} \not\ll \mathbb{Q} \\ \displaystyle\int_\Omega \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \ln \frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{Q} & \text{if } \mathbb{P} \ll \mathbb{Q} \end{cases}$$

First properties: $\mathrm{KL}(\mathbb{P},\mathbb{Q}) \geq 0$ with $\mathrm{KL}(\mathbb{P},\mathbb{Q}) = 0 \iff \mathbb{P} = \mathbb{Q}$

Data-processing inequality (with X r.v.): $\mathrm{KL}(\mathbb{P}^X, \mathbb{Q}^X) \leq \mathrm{KL}(\mathbb{P},\mathbb{Q})$

Joint convexity of KL: $$\mathrm{KL}\big((1-\alpha)\mathbb{P}_1 + \alpha\mathbb{P}_2,\ (1-\alpha)\mathbb{Q}_1 + \alpha\mathbb{Q}_2\big) \leq (1-\alpha)\,\mathrm{KL}(\mathbb{P}_1,\mathbb{Q}_1) + \alpha\,\mathrm{KL}(\mathbb{P}_2,\mathbb{Q}_2)$$

First new result today:

KL for product measures ($\leftrightarrow$ the independent case, while the dependent case will be considered in a few pages)

Prop: 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:

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

 Lecture 5 — Winter 2022 term Lecture 5 — Winter 2022 term

<u>Proof:</u> We have $\mathbb{P}\otimes\mathbb{P}' \ll \mathbb{Q}\otimes\mathbb{Q}' \iff \big[\mathbb{P}\ll\mathbb{Q}$ and $\mathbb{P}'\ll\mathbb{Q}'\big]$

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, by Tonelli + we use that if f,g are ≥ 0, then $\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu$

we add $+2/e$

$$\mathrm{KL}(\mathbb{P}\otimes\mathbb{P}', \mathbb{Q}\otimes\mathbb{Q}') = \int_{\Omega\times\Omega'} \overbrace{\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\frac{d\mathbb{P}'}{d\mathbb{Q}'}\ln\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\frac{d\mathbb{P}'}{d\mathbb{Q}'}\right)\right)}^{\text{we add } +2/e} d(\mathbb{Q}\otimes\mathbb{Q}')$$

$$= \underbrace{\int_{\Omega'}\Bigg(\underbrace{\int_{\Omega}\overbrace{\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\ln\frac{d\mathbb{P}}{d\mathbb{Q}}\right)}^{+1/e} d\mathbb{Q}}_{=\mathrm{KL}(\mathbb{P},\mathbb{Q})}\Bigg)\underbrace{\frac{d\mathbb{P}'}{d\mathbb{Q}'}\,d\mathbb{Q}'}_{d\mathbb{P}'}}_{=\mathrm{KL}(\mathbb{P},\mathbb{Q})} + \underbrace{\text{similar term with } \ln\frac{d\mathbb{P}'}{d\mathbb{Q}'},}_{=\mathrm{KL}(\mathbb{P}',\mathbb{Q}')} \text{ also with } +1/e$$

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

We ensure here that $f,g\geq 0$ by the translations by $1/e$

<u>Consequence</u> (Garivier, Ménard, Stoltz, 2016):

Data-processing inequality with expectations of random variables

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

Let $X : (\Omega,\mathcal{F}) \to \big([0,1], \mathcal{B}([0,1])\big)$ 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:

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

<u>Proof:</u> We denote by $\mathfrak{m}$ the Lebesgue measure over $[0,1]$ and augment the underlying measurable space into $\big(\Omega\times[0,1], \mathcal{F}\otimes\mathcal{B}([0,1])\big)$, 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:

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

$$\overset{\text{cf. product distributions}}{=} \mathrm{KL}(\mathbb{P},\mathbb{Q}) + \mathrm{KL}(m,m) = \mathrm{KL}(\mathbb{P},\mathbb{Q})$$

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

$$\mathbb{P}\otimes m(E) = \mathbb{E}_{\mathbb{P}}[X] \quad \text{and} \quad \mathbb{Q}\otimes m(E) = \mathbb{E}_{\mathbb{Q}}[X]$$

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

By Tonelli's theorem:

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

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 $\forall B \in \mathcal{F}'$ $\quad \omega \mapsto K(\omega, B)$ 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?)

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

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

An extension of Fubini(–Tonelli) theorem

immediate extension to φ 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 ≥ 0 or KP-integrable.

Then $\omega \mapsto \int_{\Omega'} \varphi(\omega, \omega')\, K(\omega, d\omega')$ 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) The result is true for $\varphi = \mathbb{1}_{A \times B}$ by definition of KP including measurability of $\omega \mapsto \int \varphi(\omega, \cdot)\, K(\omega, d\cdot)$ by regularity of K

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 $\begin{cases} \varphi \geq 0 \\ \varphi \in \mathbb{L}^1 \end{cases}$ by monotone convergence

Question: Does anyone have a simpler argument?

Theorem [chain rule for KL]: Assume $P \ll Q$ actually with no loss of generality

As soon as (*) $K(\omega, \cdot) \ll L(\omega, \cdot)$ for Q 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,
↳ version up to a LQ-null set

Then

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

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

Remark: see a remark stated in two pages for the (lack of) necessity of Assumptions (*) and (**).

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

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

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

we will not use this, actually

$$\int_{\Omega\times\Omega'} g \ln g \; d\,L\mathbb{Q} = \int \mathrm{KL}\big(K(\omega,\cdot), L(\omega,\cdot)\big)\, d\mathbb{Q}(\omega)$$

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

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

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

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

By Radon–Nikodym's theorem: $\dfrac{dK\mathbb{P}}{dL\mathbb{Q}} = f\,g$ $L\mathbb{Q}$-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 assumed 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, $$\mathrm{KL}(K\mathbb{P}, L\mathbb{Q}) = \int_{\Omega\times\Omega'} \Big(f(\omega)\, g(\omega,\omega') \ln f(\omega)\, g(\omega,\omega') \Big)\; d\,L\mathbb{Q}(\omega,\omega')$$

Lecture 5 — Winter 2022 term Lecture 5 — Winter 2022 term

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

what is being integrated in the Ω' integral is a sum of: function lower bounded by $-\frac{1}{e}$ and an integrable function in ω'

$$\int \big(fg \ln(fg) \big) \, dL\mathbb{Q} = \int_{\Omega} f(\omega) \left(\int_{\Omega'} \big(g(\omega,\omega') \ln g(\omega,\omega') + g(\omega,\omega') \ln f(\omega) \big) L(\omega, d\omega') \right) d\mathbb{Q}(\omega)$$

$$= \int_{\Omega} \Bigg(\underbrace{\int_{\Omega'} \big(g(\omega,\omega') \ln g(\omega,\omega') \big) L(\omega, d\omega')}_{KL(K(\omega,\cdot),\, L(\omega,\cdot))} + \big(\ln f(\omega)\big) \underbrace{\int_{\Omega'} g(\omega,\omega') L(\omega, d\omega')}_{=1} \Bigg) f(\omega) \, d\mathbb{Q}(\omega)$$

$$= \int_{\Omega} \Big(KL\big(K(\omega,\cdot), L(\omega,\cdot)\big) + \ln f(\omega) \Big) f(\omega) \, d\mathbb{Q}(\omega)$$

sum of two functions bounded from below

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

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 Ω' is a topological space with a countable base (a "second-countable space") and $\mathcal{F}'$ is the Borel σ-algebra.

  I.e., there exists some countable collection $(O_m)_{m \geq 1}$ of open sets 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 $\rightarrow$ we will consider $\Omega' = [0,1] \times \big(\mathbb{R} \times [0,1]\big)^{\mathbb{N}}$

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

Part 2: Regret lower bound for stochastic bandits

 Lecture 5 — Winter 2022 term Lecture 5 — Winter 2022 term

Lower bounds on the regret for stochastic bandits.

Here is first a summary of the setting and context of stochastic bandits:

- K arms each indexed by $a = 1, 2, \ldots K$
- 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 $\underline{\nu} = (\nu_a)_{a \in \{1,\ldots K\}}$
- Important quantities and notation:

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

  $\mu^* = \max\limits_{a=1\ldots K} \mu_a$ 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 $t = 1, 2, \ldots$

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

  1. The decision-maker picks $I_t \in \{1, \ldots K\}$ possibly at random based on an auxiliary randomization U_{t-1}
  2. She gets a reward Y_t drawn at random according to ν_{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:
$$R_T = T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} \mu_{I_t}\right] = \sum_{a=1}^{K} \Delta_a \, \mathbb{E}[N_a(T)]$$

  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 $\underline{\nu} \in \mathcal{D}^K$,

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

- 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. $\Delta_a > 0$,
$$\limsup_{T \to +\infty} \frac{\mathbb{E}[N_a(T)]}{\ln T} \leq C_a(\underline{\nu})$$
where $C_a(\underline{\nu})$ is a problem-dependent constant.

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

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 ψ 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 $\Delta_a > 0$),

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

 Lecture 5 — Winter 2022 term Lecture 5 — Winter 2022 term

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

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

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 ψ 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,

$$\sum_{a=1}^{K} \mathbb{E}_{\underline{\nu}}[N_a(T)]\, \mathrm{KL}(\nu_a,\nu'_a) = \mathrm{KL}\big(\mathbb{P}_{\underline{\nu}}^{H_T}, \mathbb{P}_{\underline{\nu}'}^{H_T}\big)$$
$$\geqslant \mathrm{KL}\Big(\mathrm{Ber}\big(\mathbb{E}_{\underline{\nu}}[Z]\big), \mathrm{Ber}\big(\mathbb{E}_{\underline{\nu}'}[Z]\big)\Big)$$

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

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

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

This is why we will

- Fix $\mathcal{D}$, ψ, $\underline{\nu}$ 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:

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

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

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

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

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

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

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 ψ is consistent and:

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

– in the problem ν', all arms $k \neq a$ are suboptimal:
for all $\alpha \in (0,1]$, $T - \mathbb{E}_{\nu'}[N_a(T)] = \sum_{k \neq a} \mathbb{E}_{\nu'}[N_k(T)] = o(T^\alpha)$

↳ 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:

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

thus

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

Letting $\alpha \to 0$:

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

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

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

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

 Lecture 5 — Winter 2022 term Lecture 5 — Winter 2022 term

Proof of the lemma:

- The inequality $\geqslant$ is a direct application of the data-processing inequality with expectations

and similarly $\mathbb{P}_{\underline{\nu}'}^{H_T} = K'_T\big(\cdots(K'_1 \mathfrak{m})\cdots\big)$

- For the equality:

(1) We show by induction that $\mathbb{P}_{\underline{\nu}}^{H_T} = K_T\Big(K_{T-1}\big(\cdots(K_1 \mathfrak{m})\big)\cdots\Big)$

we check below that it is regular

$$\left\{ \begin{array}{l} \text{where } K_t \text{ is the transition kernel:} \\ h \in [0,1] \times (\mathbb{R}\times[0,1])^{t-1} \longmapsto K_t(h,\cdot) = \nu_{\psi_{t-1}(h)} \otimes \mathfrak{m} \end{array} \right.$$

Indeed: $T=0$: $H_0 = U_0 \sim \mathcal{U}_{[0,1]}$: $\mathbb{P}_{\underline{\nu}}^{U_0} = \mathfrak{m}$

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

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

tower rule

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

definition of the bandit model and of the strategy

$$\mathbb{E}_{\underline{\nu}}\Big[\mathbb{1}_A(H_t)\ \nu_{\psi_t(H_t)}(B')\ \mathfrak{m}(B)\Big]$$

definition of K_{t+1}

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

mere rewriting

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

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

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

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

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

$$h \longmapsto K_t(h,E) = \sum_{a=1}^{K} \mathbb{1}_{\{\psi_{t-1}(h)=a\}}\ \nu_a\otimes\mathfrak{m}(E)$$

is measurable as ψ_{t-1} is measurable

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

* Assumption (**): $\displaystyle (h,(y,u)) \longmapsto \frac{dK_t(h,\cdot)}{dK'_t(h,\cdot)}(y,u)$

is indeed bi-measurable
(product of measurable functions)

$$= \sum_{a=1}^{K} \mathbb{1}_{\{\psi_{t-1}(h)=a\}} \frac{d\nu_a}{d\nu'_a}(y)$$

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

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

- For $t \geq 0$,

$$\mathrm{KL}(\mathbb{P}_{\nu}^{H_{t+1}}, \mathbb{P}_{\nu'}^{H_{t+1}})$$

$$= \mathrm{KL}(K_{t+1}\,\mathbb{P}_{\nu}^{H_t}, K'_{t+1}\,\mathbb{P}_{\nu'}^{H_t})$$

$$= \mathrm{KL}(\mathbb{P}_{\nu}^{H_t}, \mathbb{P}_{\nu'}^{H_t}) + \int \mathrm{KL}(K_{t+1}(h,\cdot), K'_t(h,\cdot))\, d\mathbb{P}_{\nu}^{H_t}(h)$$

$$= \mathrm{KL}(\mathbb{P}_{\nu}^{H_t}, \mathbb{P}_{\nu'}^{H_t}) + \int \mathrm{KL}(\nu_{\psi_t(h)} \cancel{\otimes \eta}, \nu'_{\psi_t(h)} \cancel{\otimes \eta})\, d\mathbb{P}_{\nu}^{H_t}(h)$$

$$= \mathrm{KL}(\mathbb{P}_{\nu}^{H_t}, \mathbb{P}_{\nu'}^{H_t}) + \sum_{a=1}^{K} \mathrm{KL}(\nu_a, \nu'_a) \underbrace{\int \mathbb{1}_{\{\psi_t(h)=a\}}\, d\mathbb{P}_{\nu}^{H_t}(h)}_{\mathbb{E}\left[\mathbb{1}_{\{\psi_t(H_t)=a\}}\right]}$$

$$= \mathbb{E}\left[\mathbb{1}_{\{I_{t+1}=a\}}\right]$$

 Lecture 5 — Winter 2022 term Lecture 5 — Winter 2022 term

Exercise: $\quad 1/\mathcal{K}_{\inf}(\nu_a, \mu^*, \mathcal{D})$ 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:

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

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

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

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

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

(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:

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

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

Even better, show the stronger form:

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

(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!

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 (↔ 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. $\Delta_j > 0$,

$$\mathbb{E}[N_j(T)] \geq \frac{1}{2} \frac{T}{K}$$

at least half of the time — uniform exploration

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