Part 1: The Kullback-Leibler divergence: the data-processing inequality

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

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

Basic facts: • Existence of the defining integral when $\mathbb{P} \ll \mathbb{Q}$: because $\varphi : x \mapsto x \ln x$ is bounded from below on $[0,+\infty[$

• $\mathrm{KL}(\mathbb{P},\mathbb{Q}) \geq 0$ and $\mathrm{KL}(\mathbb{P},\mathbb{Q}) = 0$ if and only if $\mathbb{P} = \mathbb{Q}$:
because φ is strictly convex, Jensen's inequality indicates that

$$\mathrm{KL}(\mathbb{P},\mathbb{Q}) = \int_\Omega \varphi\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{Q} \geq \varphi\Bigg(\underbrace{\int_\Omega \frac{d\mathbb{P}}{d\mathbb{Q}}\, d\mathbb{Q}}_{=1}\Bigg) = 0,$$

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

Lemma (contraction of entropy; also known as 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:

$$\mathrm{KL}(\mathbb{P}^X, \mathbb{Q}^X) \leq \mathrm{KL}(\mathbb{P},\mathbb{Q}).$$

Proof: We may assume that $\mathbb{P} \ll \mathbb{Q}$, otherwise, $\mathrm{KL}(\mathbb{P},\mathbb{Q}) = +\infty$ and the inequality is true. Then, we also have $\mathbb{P}^X \ll \mathbb{Q}^X$, with

$$\frac{d\mathbb{P}^X}{d\mathbb{Q}^X} = \mathbb{E}_\mathbb{Q}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X = \cdot\,\right] \overset{\text{not.}}{=} \gamma,$$

ie, $\mathbb{E}_\mathbb{Q}\left[\frac{d\mathbb{P}}{d\mathbb{Q}}\right] = \gamma(X)$.

Indeed, for all $B \in \mathcal{F}'$:

$$\mathbb{P}^X(B) = \mathbb{P}\{X \in B\} = \int_\Omega \mathbb{1}_B(X) \frac{d\mathbb{P}}{d\mathbb{Q}}\, d\mathbb{Q} \underset{\text{tower rule}}{=} \int_\Omega \mathbb{1}_B(X)\, \mathbb{E}_\mathbb{Q}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X\right] d\mathbb{Q}$$

$$= \int_\Omega \mathbb{1}_B(X)\, \gamma(X)\, d\mathbb{Q} \underset{\text{definition of } \mathbb{Q}^X}{=} \int_{\Omega'} \mathbb{1}_B\, \gamma\, d\mathbb{Q}^X.$$

Therefore,

$$KL(\mathbb{P}^X, \mathbb{Q}^X) = \int_{\Omega'} (\gamma \ln \gamma)\, d\mathbb{Q}^X = \int_{\Omega} \gamma(X) \ln \gamma(X)\, d\mathbb{Q}$$

def. of γ

$$= \int_{\Omega} \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X\right] \ln \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X\right] d\mathbb{Q}$$

conditional version of Jensen's inequality

$$\leqslant \int_{\Omega} \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \ln \frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X\right] d\mathbb{Q} \underset{\text{tower rule}}{=} KL(\mathbb{P}, \mathbb{Q})$$

References:
- The proof above is a combination of arguments used by Gray (2011) and Pardo (2006), to be published by Gerchinovitz, Ménard, Stoltz (2017?).
- Typical proofs in the literature
  - either focus on the discrete case (Cover and Thomas, 2006),
  - or use the duality formula for the KL (Massart 2007; Boucheron, Lugosi, Massart, 2013).

Corollary (data-processing inequality with expectations of random variables):
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 expectations of X under $\mathbb{P}$ and $\mathbb{Q}$:

$$KL\big(\mathrm{Ber}(\mathbb{E}_{\mathbb{P}}[X]),\ \mathrm{Ber}(\mathbb{E}_{\mathbb{Q}}[X])\big) \leqslant KL(\mathbb{P}, \mathbb{Q})$$

Reference: Garivier, Ménard, Stoltz (2016)

Proof: We denote by 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 $\mathbb{P} \otimes m$ and $\mathbb{Q} \otimes m$. For any event $E \in \mathcal{F} \otimes \mathcal{B}([0,1])$, we have, by the data-processing inequality:

$$KL\big((\mathbb{P} \otimes m)^{\mathbb{1}_E}, (\mathbb{Q} \otimes m)^{\mathbb{1}_E}\big) \leqslant KL(\mathbb{P}, \mathbb{Q})$$

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

The proof is concluded by picking E 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, $E = \{ (\omega, x) \in \Omega \times [0,1] \ : \ x \leq X(\omega) \}$, since by Tonelli's theorem,

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

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

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

Corollary (joint convexity of KL): for all probability distributions $\mathbb{P}_1$, $\mathbb{P}_2$ and $\mathbb{Q}_1$, $\mathbb{Q}_2$ over the same measurable space $(\Omega, \mathcal{F})$, and all $\alpha \in (0,1)$,

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

References:

- Typically proved in the following way: continuous log-sum inequality is that $(p,q) \in [0,+\infty)^2 \longmapsto \left(\frac{p}{q} \ln \frac{p}{q}\right) q$ is jointly convex. Introduce $\mu = \mathbb{P}_1 + \mathbb{P}_2 + \mathbb{Q}_1 + \mathbb{Q}_2$, consider the densities $p_j = \frac{d\mathbb{P}_j}{d\mu}$ and $q_j = \frac{d\mathbb{Q}_j}{d\mu}$; and apply pointwise the continuous log-sum inequality.
- We (Gerchinovitz, Ménard, Stoltz, 2017?) see it as a consequence of the data-processing inequality:

Proof: We augment $(\Omega, \mathcal{F})$ into $(\Omega \times \{1,2\}, \mathcal{F}')$ where

$$\mathcal{F}' = \mathcal{F} \otimes \big\{ \emptyset, \{1\}, \{2\}, \{1,2\} \big\}$$

We define the random pair (J, X) by the projections

$$X : (\omega, j) \longmapsto \omega \qquad \text{and} \qquad J : (\omega, j) \longmapsto j$$

Let $\mathbb{P}$ be a measure on $(\Omega \times \{1,2\}, \mathcal{F}')$ such that

$$\begin{cases} J \sim 1 + \mathrm{Ber}(\alpha) \\ X \mid J = j \ \sim \ \mathbb{P}_j \end{cases}$$

that is,

$$\forall j \in \{1,2\}, \quad \forall A \in \mathcal{F}, \qquad \mathbb{P}(A \times \{j\}) = \left((1-\alpha) \mathbb{1}_{\{j=1\}} + \alpha \mathbb{1}_{\{j=2\}} \right) \mathbb{P}_j(A)$$

Now,
$$\mathbb{P}^X = (1-\alpha)\mathbb{P}_1 + \alpha \mathbb{P}_2$$
$$\mathbb{Q}^X = (1-\alpha)\mathbb{Q}_1 + \alpha \mathbb{Q}_2$$

and $\mathrm{KL}(\mathbb{P}, \mathbb{Q}) = (1-\alpha)\,\mathrm{KL}(\mathbb{P}_1, \mathbb{Q}_1) + \alpha \, \mathrm{KL}(\mathbb{P}_2, \mathbb{Q}_2)$

so that the result follows from the data-processing inequality.

Indeed: we may assume with no loss of generality that $\mathbb{P}_1 \ll \mathbb{Q}_1$, $\mathbb{P}_2 \ll \mathbb{Q}_2$, so that $\mathbb{P} \ll \mathbb{Q}$ with

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(\omega, j) = \mathbb{1}_{\{j=1\}} \frac{d\mathbb{P}_1}{d\mathbb{Q}_1}(\omega) + \mathbb{1}_{\{j=2\}} \frac{d\mathbb{P}_2}{d\mathbb{Q}_2}$$

This entails that

$$\mathrm{KL}(\mathbb{P}, \mathbb{Q}) = \int_{\Omega \times \{1,2\}} \ln \frac{d\mathbb{P}}{d\mathbb{Q}}(\omega, j)\, d\mathbb{Q}(\omega, j)$$

$$= (1-\alpha) \underbrace{\int_\Omega \ln \frac{d\mathbb{P}}{d\mathbb{Q}}(\omega, 1)\, d\mathbb{Q}_1(\omega)}_{= \int_\Omega \left(\ln \frac{d\mathbb{P}_1}{d\mathbb{Q}_1}(\omega)\right) d\mathbb{Q}_1(\omega) = \mathrm{KL}(\mathbb{P}_1, \mathbb{Q}_1)} + \alpha \underbrace{\int_\Omega \ln \frac{d\mathbb{P}}{d\mathbb{Q}}(\omega, 2)\, d\mathbb{Q}_2(\omega)}_{= \mathrm{KL}(\mathbb{P}_2, \mathbb{Q}_2)}$$

<u>A final useful result</u>: <u>The chain rule.</u>

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

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 K is a mapping

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

such that $\forall B \in \mathcal{F}'$, $\omega \mapsto K(\omega, B)$ 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 $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$:

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

$$K\mathbb{P}(A \times B) = \int_\Omega \mathbb{1}_A(\omega)\, K(\omega, B)\, d\mathbb{P}(\omega)$$

$$L\mathbb{Q}(A \times B) = \int_\Omega \mathbb{1}_A(\omega)\, L(\omega, B)\, d\mathbb{Q}(\omega).$$

<u>Theorem</u>: Provided that (*) $\omega \mapsto \mathrm{KL}(K(\omega, \cdot), L(\omega, \cdot))$ is $\mathcal{F}$-measurable,

and that (**) $K(\omega, \cdot) \ll L(\omega, \cdot)$ for $\mathbb{P}$-almost all ω,

with $(\omega, \omega') \longmapsto \dfrac{dK(\omega, \cdot)}{dL(\omega, \cdot)}(\omega')$ being $\mathcal{F} \otimes \mathcal{F}'$-measurable,

then

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

Questions: Are these assumptions « real » assumptions?

I.e.: would it be that (*) is automatically satisfied?

would (**) follow from $K\mathbb{P} \ll LQ$, as $\mathbb{P} \ll Q$ does?

Proof: Given (**) we have $K\mathbb{P} \ll LQ$ if and only if $\mathbb{P} \ll Q$; we thus assume with no loss of generality that $K\mathbb{P} \ll LQ$ and $\mathbb{P} \ll Q$ (otherwise, both members of the putative equality equal $+\infty$).

We write $\frac{d\mathbb{P}}{dQ} = f$ and $\frac{dK(\omega,\cdot)}{dL(\omega,\cdot)} = g(\omega,\cdot)$

We then have $\frac{d\, K\mathbb{P}}{d\, LQ}(\omega,\omega') = f(\omega)\, g(\omega,\omega')$

Thus, $$KL(K\mathbb{P}, LQ) = \int_{\Omega\times\Omega'} \Big(\ln \underbrace{\frac{dK\mathbb{P}}{dLQ}(\omega,\omega')}_{f(\omega)\, g(\omega,\omega')}\Big)\ \underbrace{d\, K\mathbb{P}(\omega,\omega')}_{K(\omega,\, d\omega')\, d\mathbb{P}(\omega)}$$

Fubini / Fubini-Tonelli

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

Special case: independence / product-distributions. [with constant stochastic kernels]

Let $\mathbb{P}_1, Q_1$ be two probability measures over $(\Omega, \mathcal{F})$
$\mathbb{P}_2, Q_2$ $(\Omega', \mathcal{F}')$
then the product measures $\mathbb{P}_1 \otimes \mathbb{P}_2$, $Q_1 \otimes Q_2$ over $(\Omega\times\Omega', \mathcal{F}\otimes\mathcal{F}')$ are such that

$$KL(\mathbb{P}_1 \otimes \mathbb{P}_2,\ Q_1 \otimes Q_2) = KL(\mathbb{P}_1, Q_1) + KL(\mathbb{P}_2, Q_2)$$

Part 2 : Lower bounds on the regret for stochastic bandits.

Setting : K arms indexed by $a = 1, 2, \dots K$

With each arm a is associated a probability distribution ν_a over $\mathbb{R}$ (with an expectation)

$(\nu_a)_{a \in \{1..K\}}$ is called a bandit problem

At each round $t = 1, 2, \dots$

- The decision-maker picks $I_t \in \{1, \dots K\}$ possibly at random
- She gets a reward Y_t drawn at random $\sim \nu_{I_t}$ (given I_t)
- This is the only feedback / piece of information she gets.

Aim: We denote by $\mu_a = E(\nu_a)$ expectation of ν_a and $\mu^* = \max_{a \in \{1 \dots K\}} \mu_a$

We want to control the pseudo-regret

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

Disclaimer : Very weak notion of regret : very « expected »

Rewriting: Tower rule : $\mathbb{E}[Y_t \mid I_t] = \mu_{I_t}$

thus $\mathbb{E}[Y_t] = \sum_{a \in \{1..K\}} \mu_a \, \mathbb{P}\{I_t = a\}$

$$R_T = \sum_{t=1}^{T} \underbrace{(\mu^* - \mu_a)}_{\text{gap } \Delta_a} \mathbb{E}[N_a(T)] \qquad \text{where} \quad N_a(T) = \sum_{t=1}^{T} \mathbb{1}_{\{I_t = a\}}$$

↳ All boils down to controlling $\mathbb{E}[N_a(T)]$ for $\Delta_a > 0$

Upper bounds: We restrict the ν_a to be distributions with supports in $[0,1]$

* UCB strategy by Auer, Cesa-Bianchi and Fischer, 2002

For $t = 1, 2, \dots K$: pull arm $I_t = t$

For $t = K+1, K+2, \dots$

- Compute $$\hat{\mu}_{a,t-1} = \frac{1}{N_a(t-1)} \sum_{s=1}^{t-1} Y_s \mathbb{1}_{\{I_s = a\}}$$
- Pick $$I_t \in \operatorname*{argmax}_{a \in \{1..K\}} \left\{ \hat{\mu}_{a,t-1} + \sqrt{\frac{2 \ln t}{N_a(t-1)}} \right\}$$

Theorem: $\forall a$ s.t. $\Delta_a > 0$, $$\mathbb{E}[N_a(T)] \leq \frac{8 \ln T}{\Delta_a^2} + 2.$$

By Pinsker's inequality, as we discuss below, this bound is looser than the following one:

* KL-UCB strategy (Cappé, Garivier, Maillard, Munos, Stoltz, 2013; Garivier, Ménard, Stoltz, work in progress)

$x \in \mathbb{R}$, $\quad K_{\inf}(\nu_a, x) = \inf \{ KL(\nu_a, \nu'_a) : \nu'_a$ distribution over $[0,1]$ and $E(\nu'_a) > x \}$

(convention: empty infimum equals $+\infty$)

Parameter: $f : \mathbb{N}^* \to (0, +\infty)$

For $t = 1, 2, \ldots, K$: pull arm $I_t = t$

For $t = K+1, K+2, \ldots$

- Compute $\displaystyle \hat{\nu}_{a,t-1} = \frac{1}{N_a(t-1)} \sum_{s=1}^{t-1} \delta_{Y_s} \mathbb{1}_{\{I_s = a\}}$

- Compute $\displaystyle U_{a,t-1} = \sup \left\{ \mu \in [0,1] : K_{\inf}(\hat{\nu}_{a,t-1}, \mu) \leqslant \frac{f(t-1)}{N_a(t-1)} \right\}$

- Pick $\displaystyle I_t \in \operatorname*{argmax}_{a \in \{1..K\}} U_{a,t-1}$

<u>Theorem</u> (to be checked; consider it with some care): $f(t) = \ln t + \ln(\ln(t))$

$\forall T \geqslant 3$, $\forall \delta \in (0, \min(\mu^* - \mu_a, 1 - \mu^*))$, where a is a suboptimal arm:

$$E[N_a(T)] \leqslant \frac{\ln T}{K_{\inf}(\nu_a, \mu^*) - \frac{2\delta}{1-\mu^*}} + \underbrace{\left(1 + \frac{\ln(\ln T)}{K_{\inf}(\nu_a, \mu^*) - \frac{2\delta}{1-\mu^*}} + \frac{3e(\ln(\ln T) + 2)}{(1 - e^{-2\delta^2})^2} + \frac{1}{1 - e^{-\frac{\delta^2 (1-\mu^*)^{1/2}}{18}}} \right)}_{O(\ln(\ln T))}$$

Asymptotically:

$$\limsup_{T \to +\infty} \frac{E[N_a(T)]}{\ln T} \leqslant \frac{1}{K_{\inf}(\nu_a, \mu^*)}$$

<u>Two remarks:</u>

(1) Better than the UCB bound: by the data-processing inequality with random variables,

$$KL(\nu_a, \nu'_a) \geqslant KL\big(Ber(E(\nu_a)), Ber(E(\nu'_a))\big)$$

↳ namely, X = identity given that ν_a, ν'_a are over $[0,1]$

$$\geqslant 2 \big(E(\nu_a) - E(\nu'_a)\big)^2$$

thus $K_{\inf}(\nu_a, \mu^*) \geqslant 2(\mu^* - \mu_a)^2 = 2\Delta_a^2$

There are refinements of UCB (Garivier et Cappé, 2011) that get the main term arbitrarily close to $\frac{\ln T}{2\Delta_a^2}$; it is still a weaker bound than the one for KL-UCB.

(2) Instead of considering all distributions over $[0,1]$, we can set a model $\mathcal{D}$ of distributions over $\mathbb{R}$ with expectations, impose $\nu_a \in \mathcal{D}$, and define

$$\mathcal{K}_{\inf}(\nu_a, x) = \inf\left\{ KL(\nu_a, \nu'_a) : \nu'_a \in \mathcal{D} \text{ and } E(\nu'_a) > x \right\}$$

for $x \in \mathbb{R}$.

Lai and Robbins (1985) and Burnetas and Katehakis (1996) propose an algorithm with the following asymptotic guarantee:

$$\limsup_{T\to+\infty} \frac{\mathbb{E}[N_a(T)]}{\ln T} \leq \frac{1}{\mathcal{K}_{\inf}(\nu_a, \mu^*)}$$

In some cases ($\mathcal{D}$ = all distributions over $[0,1]$; or $\mathcal{D}$ = one-dimensional regular exponential families, see Garivier and Cappé, 2011) non-asymptotic bounds are available.

<u>Aim</u>: Prove the lower bound $\liminf_{T\to+\infty} \frac{\mathbb{E}[N_a(T)]}{\ln T} \geq \frac{1}{\mathcal{K}_{\inf}(\nu_a, \mu^*)}$ in a straightforward way:
the original proof by Lai and Robbins (1985) and Burnetas and Katehakis are difficult to follow.
There were partial attempts by Bubeck and Cesa-Bianchi (2012), Kaufmann, Cappé and Garivier (2016) to simplify the proof.

We (Garivier, Ménard, Stoltz, 2016) believe that we now extracted the core arguments. We proceed in an information-theoretic way without the need of an explicit change of measure.

[see next page for a formal definition]

<u>Definition</u>: A strategy is consistent w.r.t. a model $\mathcal{D}$ if for all bandit problems $(\nu_a)_{a\in\{1..K\}} \in \mathcal{D}^K$

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

↳ To state precise lower bounds, and in view of the $O(\ln T)$ performance bounds of UCB and KL-UCB, it is reasonable to impose consistency.

<u>Theorem</u>: For all models $\mathcal{D}$ with integrable distributions, for all strategies consistent wrt $\mathcal{D}$, for all bandit problems $(\nu_a)_{a\in\{1..K\}} \in \mathcal{D}^K$,

$$\forall a \text{ s.t. } \Delta_a > 0, \qquad \liminf_{T\to+\infty} \frac{\mathbb{E}[N_a(T)]}{\ln T} \geq \frac{1}{\mathcal{K}_{\inf}(\nu_a, \mu^*)}$$

Proof : (1) Modeling

A randomized strategy ψ associates with the information gained in the past an arm to pick, possibly thanks to an auxiliary randomization $U_1, U_2, \dots$ (iid $\sim \mathcal{U}_{[0,1]}$).

That is, $\psi = (\psi_t)_{t \geq 0}$ where

$$\psi_t : H_t = (U_0, Y_1, U_1, \dots Y_t, U_t) \longmapsto \psi_t(H_t) = I_{t+1}$$

is measurable.

↳ We may take $\Omega = [0,1] \times (\mathbb{R} \times [0,1])^{\mathbb{N}}$ as the underlying probability space, equipped with the σ-algebra of cylinder subsets of Ω; Kolmogorov's extension theorem ensures that there exists a probability measure $\mathbb{P}_\nu$, only depending on the bandit problem $\nu = (\nu_a)_{a \in \{1..K\}}$ such that

$\forall t \geq 0$, $\forall$ Borel sets B, B',

$$\mathbb{P}_\nu\left(Y_{t+1} \in B \text{ and } U_{t+1} \in B' \mid H_t \right) = \underbrace{\nu_{\psi_t(H_t)}(B) \; \mathfrak{m}(B')}_{\text{our stochastic kernel can be read here!}}$$

where $\mathfrak{m}$ is the Lebesgue measure.

(2) Fundamental (in)equality

We denote by $\mathbb{P}_\nu^{H_T}$ the law of H_T under $\mathbb{P}_\nu$; it is a distribution over $[0,1] \times (\mathbb{R} \times [0,1])^T$.

Then, $\forall\ \nu = (\nu_a)_{a \in \{1..K\}}$ and $\nu' = (\nu'_a)_{a \in \{1..K\}}$ in $\mathcal{D}^K$, with $\nu_a \ll \nu'_a$ for all a,
$\forall$ random variable Z taking values in $[0,1]$ and being $\sigma(H_T)$-measurable,

$$\sum_{a=1}^{K} \mathbb{E}_\nu[N_a(T)] \; KL(\nu_a, \nu'_a) \underset{\substack{\uparrow \\ \text{chain rule} \\ \text{(details below)}}}{=} KL\left(\mathbb{P}_\nu^{H_T}, \mathbb{P}_{\nu'}^{H_T} \right)$$

$$\underset{\substack{\uparrow \\ \text{data-processing inequality}}}{\geq} KL\Big(\mathrm{Ber}(\mathbb{E}_\nu[Z]),\ \mathrm{Ber}(\mathbb{E}_{\nu'}[Z]) \Big)$$

↳ Proof by induction:

- OK for $T = 0$ (we have $H_0 = U_0$ and $\mathbb{P}_\nu^{H_0} = \mathfrak{m} = \mathbb{P}_{\nu'}^{H_0}$, as well as $N_a(0) = 0$)

- For $T \to T+1$, we decompose $H_{T+1} = (H_T, (Y_{T+1}, U_{T+1}))$
We apply the chain rule as

$$\mathbb{P}_\nu^{H_{T+1}} = K_T \, \mathbb{P}_\nu^{H_T} \quad \text{where} \quad K_T(h,\cdot) = \nu_{\psi_T(h)} \otimes m$$
$$\mathbb{P}_{\nu'}^{H_{T+1}} = K'_T \, \mathbb{P}_{\nu'}^{H_T} \quad \text{where} \quad K'_T(h,\cdot) = \nu'_{\psi_T(h)} \otimes m$$

We indeed have (**) $K_T(h,\cdot) \ll K'_T(h,\cdot)$ as $\nu_a \ll \nu'_a \ \forall a$ by assumption

with $(h,(y,u)) = \dfrac{d K_T(h,\cdot)}{d K'_T(h,\cdot)}(y,u) = \sum_{a=1}^{K} \mathbb{1}_{\{\psi_T(h)=a\}} \dfrac{d\nu_a}{d\nu'_a}$ being measurable,

and (*) $h \mapsto KL(K_T(h,\cdot), K'_T(h,\cdot)) = \sum_{a=1}^{K} \mathbb{1}_{\{\psi_T(h)=a\}} KL(\nu_a, \nu'_a)$ being measurable,

Therefore,

$$KL(\mathbb{P}_\nu^{H_{T+1}}, \mathbb{P}_{\nu'}^{H_{T+1}}) = KL(\mathbb{P}_\nu^{H_T}, \mathbb{P}_{\nu'}^{H_T}) + \int KL(\nu_{\psi_T(h)} \cancel{\otimes m}, \nu'_{\psi_T(h)} \cancel{\otimes m}) \, d\mathbb{P}_\nu^{H_T}(h)$$

$$= \underbrace{\sum_{a=1}^{K} \mathbb{E}_\nu[N_a(T)] \, KL(\nu_a, \nu'_a)}_{\text{by induction}} + \underbrace{\int KL(\nu_{\psi_T(h)}, \nu'_{\psi_T(h)})}_{\text{cf. product-distributions}}$$

$$= \sum_{a=1}^{K} \mathbb{E}_\nu[N_a(T)] \, KL(\nu_a, \nu'_a) + \sum_{a=1}^{K} \underbrace{\mathbb{P}_\nu^{H_T}\{\psi_T(h)=a\}}_{\substack{= \mathbb{P}_\nu\{\psi_T(H_T)=a\} \\ = \mathbb{E}_\nu[\mathbb{1}_{\{I_{T+1}=a\}}]}} KL(\nu_a, \nu'_a)$$

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

(3) Let's pick a convenient Z: $\quad Z = N_a(T)/T$ of course!

And consider any modified bandit problem ν' of the form:

$$\begin{cases} \nu'_k = \nu_k & \text{for all } k \neq a \\ \nu'_a \text{ is a distribution in } \mathcal{D} \text{ such that } E(\nu'_a) > \mu^* \text{ and } \underbrace{\nu_a \ll \nu'_a}_{\text{was used for technical reasons...}} \end{cases}$$

Using that
$$KL(Ber(p), Ber(q)) = p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q}$$
$$= \underbrace{p \ln \frac{1}{q}}_{\geq 0} + (1-p) \ln \frac{1}{1-q} + (\underbrace{p \ln p + (1-p)\ln(1-p)}_{\geq -\ln 2})$$
$$\geq (1-p) \ln \frac{1}{1-q} - \ln 2 \quad \text{for all } p,q \in [0,1]$$

We get via the fundamental inequality, as ν and ν' only differ at a:

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

Now, the strategy ψ is consistent and:

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

- 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:

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

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

Since this holds for all $\alpha \in (0,1]$, we get

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

We may take the sup in the right-hand side (over ν'_a s.t. $E(\nu'_a) > \mu^*$ and with or without $\nu_a \ll \nu'_a$, it doesn't matter, as $KL(\nu_a, \nu'_a) = +\infty$ anyway when $\nu_a \not\ll \nu'_a$): we get

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

Remark: Non-asymptotic such lower bounds are possible, but are heavily technical! See Garivier, Ménard, Stoltz, 2016.

<u>NEXT SESSION</u>: Other such lower bounds in stochastic bandits:

- linear regime when T is small
- for the improved bounds when μ^* or the gaps Δ_a are known (see Bubeck, Perchet and Rigollet, 2013)

All will be based on our fundamental inequality!