 Lecture 6 — March 14, 2018 Lecture 6 — March 14, 2018

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 stochastic kernel K is a mapping
(regular)

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

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

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

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

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 $\varphi \geq 0$ 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 $(**)$.

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 ≥ 0, with:

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

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

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

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

given the definition of g

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

since $f = d\mathbb{P}/d\mathbb{Q}$

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

* Therefore, we have $K\mathbb{P} \ll L\mathbb{Q}$ as soon as $\mathbb{P} \ll \mathbb{Q}$, (and the converse is easily seen) 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, $$KL(K\mathbb{P}, L\mathbb{Q}) = \int_{\Omega\times\Omega'} \Big(f(\omega)\, g(\omega,\omega')\, \ln f(\omega)\, g(\omega,\omega') \Big)\; dL\mathbb{Q}(\omega,\omega')$$

 Lecture 6 — March 14, 2018 Lecture 6 — March 14, 2018

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

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

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

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

φ: sum of two bounded functions

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

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

  Ex: Ω' 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: Martin Bellu + Hédi Hadiji, M2 students of Spring 2017

 Lecture 6 — March 14, 2018 Lecture 6 — March 14, 2018

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

  $\mu^* = \max_{a=1\ldots K} \mu_a$ 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)

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

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

  $\forall \underline{\nu} \in \mathcal{D}^K$, $\quad \forall a$ 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 $K_{\inf}(\nu_a, \mu^*, \mathcal{D}) = K_{\inf}(\nu_a, \mu^*) = \inf \left\{ KL(\nu_a, \nu_a') : \begin{array}{l} \nu_a' \in \mathcal{D} \\ E(\nu_a') > \mu^* \end{array} \right\}$

  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 $\Delta_a > 0$),

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

<u>Corollary</u>: 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 $\nu = (\nu_a)_{a\in\{1..K\}} \in \mathcal{D}^K$,

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

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

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

<u>Lemma</u> (Fundamental inequality for stochastic bandits):
For all bandit problems $\nu = (\nu_a)_{a\in\{1..K\}}$ and $\nu' = (\nu'_a)_{a\in\{1..K\}}$ in $\mathcal{D}^K$ with $\nu_a \ll \nu'_a$ for all a,
For all random variables Z taking values in $[0,1]$ and that are $\sigma(H_T)$-measurable,

$$\begin{aligned}\sum_{a=1}^K \mathbb{E}_\nu[N_a(T)]\, \mathrm{KL}(\nu_a, \nu'_a) &= \mathrm{KL}\left(\mathbb{P}_\nu^{H_T}, \mathbb{P}_{\nu'}^{H_T}\right) \\ &\geq \mathrm{KL}\left(\mathrm{Ber}(\mathbb{E}_\nu[Z]), \mathrm{Ber}(\mathbb{E}_{\nu'}[Z])\right)\end{aligned}$$

<u>Note</u>: This lemma is our key to perform an IMPLICIT change of measures in the proof of the theorem.

<u>Proof of the theorem</u> (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 $\mathcal{D}$, ψ, ν and a s.t. $\Delta_a > 0$
- Fix an alternative model ν' 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, ν and ν' only differ at a; a is the unique optimal arm in ν'

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

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

$$\mathbb{E}_\nu[N_a(T)] \ KL(\nu_a, \nu'_a) \geqslant KL\left(Ber\left(\mathbb{E}_\nu[N_a(T)/T]\right), Ber\left(\mathbb{E}_{\nu'}\left[\frac{N_a(T)}{T}\right]\right)\right)$$

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

indeed: $KL(Ber(p), Ber(q))$

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

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

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

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

↳ 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:

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

thus

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

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

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 ν'_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 $\mathbb{P}_\nu^{H_T}$ is constructed.

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 Ω, i.e., that the $U_0, Y_1, U_1, \dots Y_T, U_T$ are the projections on each component,
- $\mathbb{P}_\nu$ is given by

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

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

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

In particular: $\mathbb{P}_\nu^{H_t}$ refers to the first $1+2t$ marginals of $\mathbb{P}_\nu^{H_T} = \mathbb{P}_\nu$.

A similar construction can be done for the bandit problem ν'.

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

details on next page

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

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 ψ_t

Similarly, $\mathbb{P}_{\nu'}^{H_{t+1}} = K'_t\, \mathbb{P}_{\nu'}^{H_t}$ for $K'_t(h,\cdot) = \nu'_{\psi_t(h)} \otimes \eta$

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

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

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

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

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

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:

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

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:

$$\forall \underline{\nu} \in \mathcal{P}([0,1])^K, \qquad \forall a \text{ s.t. } \Delta_a > 0,$$
$$\mathbb{E}_{\underline{\nu}}[N_a(T)] \leqslant \frac{8}{\Delta_a^2} \ln T + 2.$$

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

(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:

$$\|\mathbb{P} - \mathbb{Q}\|_{TV} \underset{\substack{\uparrow \\ \text{the total variation} \\ \text{distance between } \mathbb{P} \text{ and } \mathbb{Q}}}{=} \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)| \leqslant \sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}, \mathbb{Q})}$$

Even better, show the stronger form:

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

(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!

<u>Exercice</u> : Finite - time lower bound for small values of T

« <u>All algorithms explore much!</u> » ↳ We want to model that <u>all</u> algorithms must first explore uniformly all arms ($\longleftrightarrow$ 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:

– $\forall\, 0 \leq p < q \leq 1$,

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

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

– 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. $\mu_a = \mu^*$, $\mathbb{E}[N_a(T)] \geqslant T/K$], we have:

$\forall\, T \leqslant \dfrac{1}{8\, \mathrm{KL}^*}$, where $\mathrm{KL}^* = \max\limits_{j : \Delta_j > 0} \mathcal{K}_{\inf}(\nu_j, \mu_j^*)$

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

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

<u>Hint</u>: consider the same alternative bandit problems ν' as in the theorem giving the asymptotic lower bound.

 Lecture 6 — March 14, 2018 Lecture 6 — March 14, 2018

Distribution-free (i.e. uniform) lower bounds.

We prove: For all $K \geq 2$ and $T \geq K/5$

$$\inf_{\text{strategies } \psi} \ \sup_{\substack{\nu_1, \ldots, \nu_K \\ \text{in } \mathcal{P}([0,1])}} R_T \geq \frac{1}{20}\sqrt{TK}.$$

and even: sup over $\nu_1 \ldots \nu_K$ being Bernoulli distributions

(In class I did not write the correct $\underline{\nu}^{(0)}$ and $\underline{\nu}^{(j)}$, it's with $\varepsilon/2$ in)

Proof: • We consider the bandit problem $\underline{\nu}^{(0)} = \big(\mathrm{Ber}(\tfrac12 - \tfrac{\varepsilon}{2}), \ldots \mathrm{Ber}(\tfrac12 - \tfrac{\varepsilon}{2})\big)$

versus the bandit problems $\underline{\nu}^{(i)} = \big(\mathrm{Ber}(\tfrac12 - \tfrac{\varepsilon}{2}), \ldots \mathrm{Ber}(\tfrac12 - \tfrac{\varepsilon}{2}), \underbrace{\mathrm{Ber}(\tfrac12 + \tfrac{\varepsilon}{2})}_{i\text{-th position}}, \mathrm{Ber}(\tfrac12 - \tfrac{\varepsilon}{2}), \ldots \mathrm{Ber}(\tfrac12 - \tfrac{\varepsilon}{2})\big)$
for $i \in \{1..K\}$ and $\varepsilon \in (0, \tfrac12)$

For all strategies,
$$\sup_{\underline{\nu} \text{ in } \mathcal{P}([0,1])} \overline{R}_T \geq \sup_{\varepsilon \in (0,1/2)} \max_{i \in \{1..K\}} \overline{R}_T$$
$$= \sup_{\varepsilon \in (0,1/2)} \max_{i \in \{1..K\}} \sum_{k \neq i} \varepsilon \, \mathbb{E}_{\underline{\nu}^{(i)}}[N_k(T)]$$

ε: gap of arm $k \neq i$ in $\underline{\nu}^{(i)}$

$\mathbb{E}_{\underline{\nu}^{(i)}}[N_k(T)]$: expected number of times k is pulled under $\underline{\nu}^{(i)}$

And $\sum_{k \neq i} N_k(T) = T - N_i(T)$

Thus,
$$\sup_{\underline{\nu} \text{ in } \mathcal{P}([0,1])} R_T \geq \sup_{\varepsilon \in (0,1/2)} \max_{i \in \{1..K\}} \varepsilon\big(T - \mathbb{E}_{\underline{\nu}^{(i)}}[N_i(T)]\big)$$

• Now, we use the fundamental inequality to upper bound one of the $\mathbb{E}_{\underline{\nu}^{(i)}}[N_i(T)]$

There exists $k \in \{1..K\}$ such that $\mathbb{E}_{\underline{\nu}^{(0)}}[N_k(T)] \leq T/K$. For this k:

$$\sum_{a=1}^{K} \mathbb{E}_{\underline{\nu}^{(0)}}[N_a(T)] \, \mathrm{KL}(\nu_a^{(0)}, \nu_a^{(k)}) \geq \mathrm{kl}\big(\mathbb{E}_{\underline{\nu}^{(0)}}[N_k(T)/T], \mathbb{E}_{\underline{\nu}^{(k)}}[N_k(T)/T]\big)$$

since $\nu^{(0)}$ and $\nu^{(k)}$ only differ at arm k:

$$= \mathbb{E}_{\underline{\nu}^{(0)}}[N_k(T)] \, \mathrm{KL}\big(\mathrm{Ber}(\tfrac12 - \tfrac{\varepsilon}{2}), \mathrm{Ber}(\tfrac12 + \tfrac{\varepsilon}{2})\big)$$
$$\leq \underbrace{\frac{T}{K}}_{\text{by definition of } k} \mathrm{kl}\big(\tfrac12 - \tfrac{\varepsilon}{2}, \tfrac12 + \tfrac{\varepsilon}{2}\big)$$

kl = KL between Bernoulli distributions

We lower bound this side by Pinsker's inequality for Bernoulli distributions:

$$\geq 2\left(\mathbb{E}_{\underline{\nu}^{(0)}}\left[\frac{N_k(T)}{T}\right] - \mathbb{E}_{\underline{\nu}^{(k)}}\left[\frac{N_k(T)}{T}\right]\right)^2$$

We proved so far:
$$2\left(\mathbb{E}_{\underline{\nu}^{(0)}}\left[\frac{N_k(T)}{T}\right] - \mathbb{E}_{\underline{\nu}^{(k)}}\left[\frac{N_k(T)}{T}\right]\right)^2 \leq \frac{T}{K}\,\mathrm{kl}\big(\tfrac12 - \tfrac{\varepsilon}{2}, \tfrac12 + \varepsilon\big)$$

This entails in particular:

$$\mathbb{E}_{\underline{\nu}(i)}\left[\frac{N_k(T)}{T}\right] \leqslant \underbrace{\mathbb{E}_{\underline{\nu}(0)}\left[\frac{N_k(T)}{T}\right]}_{\leqslant 1/K \text{ by definition of } k} + \sqrt{\frac{T}{2K}\, \mathrm{kl}\left(\tfrac{1}{2}-\tfrac{\varepsilon}{2}, \tfrac{1}{2}+\tfrac{\varepsilon}{2}\right)}$$

- Substituting in the regret lower bound, we get (with $i=k$):

$$\sup_{\underline{\nu} \text{ in } \mathcal{D}([0,1])} \bar{R}_T \geqslant \sup_{\varepsilon\in(0,1/2)} \varepsilon T\left(1 - \mathbb{E}_{\underline{\nu}(i)}\left[\frac{N_k(T)}{T}\right]\right)$$

$$\geqslant \sup_{\varepsilon\in(0,1/2)} \varepsilon T\left(\underbrace{1 - 1/K}_{\geqslant 1/2} - \sqrt{\frac{T}{2K}\, \mathrm{kl}\left(\tfrac{1}{2}-\tfrac{\varepsilon}{2}, \tfrac{1}{2}+\tfrac{\varepsilon}{2}\right)}\right)$$

$$\geqslant \sup_{\varepsilon\in(0,1/2)} \varepsilon T\left(1/2 - \sqrt{\frac{T}{2K}\, \mathrm{kl}\left(\tfrac{1}{2}-\tfrac{\varepsilon}{2}, \tfrac{1}{2}+\tfrac{\varepsilon}{2}\right)}\right)$$

Now,

$$\mathrm{kl}\left(\tfrac{1}{2}-\tfrac{\varepsilon}{2}, \tfrac{1}{2}+\tfrac{\varepsilon}{2}\right) = \left(\cancel{\tfrac{1}{2}} - \tfrac{\varepsilon}{2}\right)\ln\frac{1-\varepsilon}{1+\varepsilon} + \left(\cancel{\tfrac{1}{2}} + \tfrac{\varepsilon}{2}\right)\ln\frac{1+\varepsilon}{1-\varepsilon}$$

$$= \varepsilon \ln \frac{1+\varepsilon}{1-\varepsilon}$$

$$\leqslant 2.5\, \varepsilon^2 \text{ on } (0, 1/2)$$

(study $\varepsilon \mapsto \varepsilon \ln\left(\frac{1+\varepsilon}{1-\varepsilon}\right)/\varepsilon^2$)

So that the regret lower bound becomes

$$\sup_{\underline{\nu} \text{ in } \mathcal{D}([0,1])} \bar{R}_T \geqslant \sup_{\varepsilon\in(0,1/2)} \frac{T}{2}\left(\varepsilon - \varepsilon^2\sqrt{5T/K}\right)$$

we pick ε s.t. $1 - 2\varepsilon\sqrt{5T/K} = 0$,

that is, $\varepsilon = \sqrt{K/T}\big/2\sqrt{5}$

which is indeed $< 1/2$ as soon as $T > K/5$.

We get a $\sqrt{TK}\left(\frac{1}{2}\left(\frac{1}{2\sqrt{5}} - \frac{1}{20}\sqrt{5}\right)\right) \geqslant \frac{1}{20}\sqrt{KT}$ bound.

<u>Note:</u> The original proof used Ber(1/2) and Ber(1/2 + ε) but the approach with Ber(1/2 − ε) and Ber(1/2 + ε) leads to slightly simpler calculation.