Lecture #5: some properties of the KL

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

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

Is it possible to do better?

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

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

ie $\forall k \in [K], \; \nu_k \in \mathcal{D}$

unknown known

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

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

where

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

Kullback-Leibler divergence

We will only prove the lower bound part

(and the upper bound in exercise session/homework for two specific cases).

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

then

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

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

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

$\hookrightarrow$ optimal up to constant factor

can be made optimal with finer version

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

then

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

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

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

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

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

Basic Facts

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

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

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

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

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

A useful rewriting:

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

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

see proof in exercise session 3.

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

Lemma (data processing inequality)

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

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

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

Proof: we can assume $P \ll Q$, since otherwise $KL(P,Q) = +\infty$ and it holds.

We show that we then have $P^X \ll Q^X$, with $\frac{dP^X}{dQ^X} = \underbrace{\mathbb{E}_Q\left[\frac{dP}{dQ} \middle| X = \cdot\right]}_{= \gamma}$

i.e. $\gamma(X) = \mathbb{E}_Q\left[\frac{dP}{dQ} \middle| X\right]$

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

$$\mathbb{P}^X(B) = \mathbb{P}(X \in B) = \int_\Omega \mathbb{1}_B(X) \frac{dP}{dQ} dQ \overset{\text{tower rule}}{=} \int_\Omega \mathbb{1}_B(X) \mathbb{E}_Q\left[\frac{dP}{dQ} \middle| X\right] dQ$$

$$= \int_\Omega \mathbb{1}_B(X) \gamma(X) \, dQ \underset{\text{by def of } Q^X}{=} \int_{\Omega'} \mathbb{1}_B \gamma \, dQ^X$$

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

$$= \int_\Omega \left(\mathbb{E}_Q\left[\frac{dP}{dQ} \middle| X\right] \ln \left(\mathbb{E}_Q\left[\frac{dP}{dQ} \middle| X\right] \right) \right) dQ$$

ψ is convex, conditional Jensen inequality

$$\leq \int_\Omega \mathbb{E}_Q\left[\frac{dP}{dQ} \ln \frac{dP}{dQ} \middle| X\right] dQ$$

$$\overset{\text{tower rule}}{=} \int_\Omega \frac{dP}{dQ} \ln \frac{dP}{dQ} dQ = KL(\mathbb{P}, Q)$$

$\square$

References

- The proof above is due to Ali and Silvey ('66), but it's far from being well-known.

- Typical proofs in the more recent literature:
  - either focus on the discrete case (Cover and Thomas, 2006)
  - or use the duality/variational formula for the KL (Massart 2007; Boucheron, Lugosi, Massart 2013)

- The joint convexity of KL, given below, is typically proved in a tedious way, relying on the joint convexity of $(x,y) \in \mathbb{R}_+^2 \longmapsto \left(\frac{x}{y} \ln \frac{x}{y}\right) y$. We may see it instead as a consequence of the data processing inequality.

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 $\lambda \in (0,1)$:

$$\mathrm{KL}\Big((1-\lambda)\mathbb{P}_1 + \lambda \mathbb{P}_2\,,\ (1-\lambda)\mathbb{Q}_1 + \lambda \mathbb{Q}_2\Big) \leqslant (1-\lambda)\,\mathrm{KL}(\mathbb{P}_1, \mathbb{Q}_1) + \lambda\, \mathrm{KL}(\mathbb{P}_2, \mathbb{Q}_2).$$

Proof: we augment $(\Omega, \mathcal{F})$ into $(\Omega', \mathcal{F}')$ where

$$\Omega' = \Omega \times \{1,2\}$$

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

we define the random pair (X,J) by the projections $X : \begin{array}{l} \Omega\times\{1,2\} \to \Omega \\ (\omega,j) \mapsto \omega \end{array}$

and $J : \begin{array}{l} \Omega\times\{1,2\} \to \{1,2\} \\ (\omega,j) \mapsto j \end{array}$

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

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

(and a similar def for $\mathbb{Q}$, with $\mathbb{Q}_1, \mathbb{Q}_2$)

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

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

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

and as we prove below $KL(\mathbb{P},\mathbb{Q}) = (1-\lambda)\,KL(\mathbb{P}_1,\mathbb{Q}_1) + \lambda\,KL(\mathbb{P}_2,\mathbb{Q}_2)$ so that the result follows from the data processing inequality.

Indeed, we may assume with no loss of generality for $\lambda\in(0,1)$ that $\mathbb{P}_1 \ll \mathbb{Q}_1$, $\mathbb{P}_2 \ll \mathbb{Q}_2$, so that $\mathbb{P}\ll\mathbb{Q}$ with

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

This entails that:

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

$$= \int_{\Omega'} \left(\frac{d\mathbb{P}_1}{d\mathbb{Q}_1}(\omega) \ln \frac{d\mathbb{P}_1}{d\mathbb{Q}_1}(\omega) \right) \mathbb{1}_{\Omega \times \{1\}}(\omega, j)\, d\mathbb{Q}(\omega, j)$$

$$+ \int_{\Omega'} \left(\frac{d\mathbb{P}_2}{d\mathbb{Q}_2}(\omega) \ln \frac{d\mathbb{P}_2}{d\mathbb{Q}_2}(\omega) \right) \mathbb{1}_{\Omega \times \{2\}}(\omega, j)\, d\mathbb{Q}(\omega, j)$$

$$= \int_{\Omega} \left(\frac{d\mathbb{P}_1}{d\mathbb{Q}_1}(\omega) \ln \frac{d\mathbb{P}_1}{d\mathbb{Q}_1}(\omega) \right) (1-\lambda)\, d\mathbb{Q}_1(\omega) + \dots$$

$$= (1-\lambda)\, KL(\mathbb{P}_1, \mathbb{Q}_1) + \lambda\, KL(\mathbb{P}_2, \mathbb{Q}_2). \qquad \square$$

Proposition (KL for product measures, independent case)

Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces.
Let $\mathbb{P}, \mathbb{Q}$ be two probability measures over $(\Omega, \mathcal{F})$
$\mathbb{P}', \mathbb{Q}'$ $(\Omega', \mathcal{F}')$

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

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

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. 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 of the best characterizations of independence).

Therefore by Tonelli

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

$f = \frac{d\mathbb{P}'}{d\mathbb{Q}'}\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\ln\frac{d\mathbb{P}}{d\mathbb{Q}} + \frac{1}{e}\right)$

$g = \frac{d\mathbb{P}}{d\mathbb{Q}}\left(\frac{d\mathbb{P}'}{d\mathbb{Q}'}\ln\frac{d\mathbb{P}'}{d\mathbb{Q}'} + \frac{1}{e}\right)$ here

if $f, g \geq 0$, then

$\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu$

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

here we apply Tonelli again

$\square$

Consequence (Garivier, Menard, Stoltz 2016)

Data-processing inequality with expectations of random variables.

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

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

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

Proof: we denote by μ 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 $\mathbb{P}\otimes\mu$ and $\mathbb{Q}\otimes\mu$.

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

$$KL(\underbrace{(\mathbb{P}\otimes\mu)^{\mathbb{1}_E}}_{\mathrm{Ber}((\mathbb{P}\otimes\mu)(E))}, \underbrace{(\mathbb{Q}\otimes\mu)^{\mathbb{1}_E}}_{\mathrm{Ber}(\mathbb{Q}\otimes\mu(E))}) \leq KL(\mathbb{P}\otimes\mu, \mathbb{Q}\otimes\mu) = KL(\mathbb{P},\mathbb{Q}) + KL(\mu,\mu)$$
$$= KL(\mathbb{P},\mathbb{Q}).$$

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

Is it possible?

Yes, taking $E = \{(\omega, x)\in\Omega\times[0,1] : x\leq X(\omega)\} \in \mathcal{F}\otimes\mathcal{B}([0,1])$ as X is measurable.

By Tonelli's theorem:

$$\mathbb{P}\otimes\mu(E) = \int_\Omega\left(\int_{[0,1]}\mathbb{1}_{\{x\leq X(\omega)\}}\,d\mu(x)\right)d\mathbb{P}(\omega) = \int_\Omega X(\omega)\,d\mathbb{P}(\omega) \quad \text{and same for } \mathbb{Q}.$$

$\boxtimes$

<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.

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

A (regular) stochastic kernel K is a mapping
$$(\Omega, \mathcal{F}) \longrightarrow \mathcal{P}(\Omega', \mathcal{F}')$$
$$\omega \longmapsto K(\omega, \cdot)$$
such that $\forall B \in \mathcal{F}'$, $\omega \longmapsto K(\omega, B)$ is $\mathcal{F}$-measurable

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

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

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

An extension of Fubini(-Tonelli) theorem

Lemma

Let $\varphi : \Omega \times \Omega' \to \mathbb{R}$ be either $\mathcal{F} \otimes \mathcal{F}'$ measurable and ≥ 0

or $K\mathbb{P}$-integrable

Then $\omega \longmapsto \int_{\Omega'} \varphi(\omega,\omega')K(\omega,d\omega')$ is $\mathcal{F}$-measurable and $\int_{\Omega\times\Omega'} \varphi \, dK\mathbb{P} = \int_{\Omega}\left(\int_{\Omega'} \varphi(\omega,\omega')K(\omega,d\omega')\right)d\mathbb{P}(\omega)$

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

Proof (sketch): The result is true for $\varphi = \mathbb{1}_{A\times B}$ by definition of $K\mathbb{P}$.

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 $\omega \mapsto \int_{\Omega'} \cdots$ measurability)

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

Theorem (chain rule for KL): Assume $\mathbb{P} \ll \mathbb{Q}$ — actually with no loss of generality.

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

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

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

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

Remark:

1) The assumptions (*) and (**) will be satisfied for the

applications we have in mind.

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

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

Proof * by bi-measurability of $g \ln g$, and since $g \ln g$ is lowerbounded,

or immediate extension of

the previous lemma can be applied to get $w \longmapsto \int_{\Omega'} g(w,\cdot) \ln(g(w,\cdot)) \, L(w,d\cdot)$
$= KL(K(w,\cdot), L(w,\cdot))$

is $\mathcal{F}$-measurable and ≥ 0

* We assume $\mathbb{P} \ll \mathbb{Q}$, let $f = \frac{d\mathbb{P}}{d\mathbb{Q}}$. What can we say about $(w,w') \longmapsto f(w) g(w,w')$?

$$\int \mathbb{1}_{A\times B}(w,w') f(w) g(w,w') \, dL\mathbb{Q}(w,w') = \int_{\Omega} \left(\int_{\Omega'} \mathbb{1}_B(w') g(w,w') \, L(w,dw') \right) \mathbb{1}_A(w) f(w) \, d\mathbb{Q}(w)$$

extension of Tonelli

$$= \int_{\Omega'} \mathbb{1}_B(\omega') \, K(\omega, d\omega') = K(\omega, B)$$

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

By Radon-Nikodym's theorem: $\dfrac{dK\mathbb{P}}{dL\mathbb{Q}} = fg \quad L\mathbb{Q}\text{-as}$

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

indeed $L\mathbb{Q}(A \times \Omega') = \mathbb{Q}(A)$

$K\mathbb{P}(A \times \Omega') = \mathbb{P}(A)$.

- Therefore under (*), (**), we have $K\mathbb{P} \ll L\mathbb{Q} \Leftrightarrow \mathbb{P} \ll \mathbb{Q}$.

Then $KL(K\mathbb{P}, L\mathbb{Q}) = \displaystyle\int_{\Omega \times \Omega'} \big(f(\omega) g(\omega, \omega')\big) \ln\big(f(\omega) g(\omega, \omega')\big) \, dL\mathbb{Q}(\omega, \omega')$.

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

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

(again we can use the translation by $+\frac{1}{e}$ to justify this equality

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

$$= \int_\Omega \Big(KL\big(K(\omega,\cdot), L(\omega,\cdot)\big) + \ln(\rho(\omega))\Big)\rho(\omega)\, dQ(\omega)$$

again, sum of two functions bounded from below

$$= \int_\Omega KL\big(K(\omega,\cdot), L(\omega,\cdot)\big) \underbrace{\rho(\omega)\, dQ(\omega)}_{dP(\omega)} + \underbrace{\int_\Omega \rho(\omega) \ln\rho(\omega)\, dQ(\omega)}_{KL(P,Q)}$$

$\square$