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

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

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

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

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

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

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

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

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

  which is equivalent to minimizing (controlling from above)

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

- Rewriting by tower rule:

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

  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

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

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

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

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

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

E.g.: at least $\mathcal{D} = \mathcal{M}_1([0,1])$; $\quad \mathcal{D}$ = a one-dimensional regular exponential family

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

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

where is a problem-dependent constant.

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

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

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

<u>THEOREM:</u> For all bandit models $\mathcal{D} \subset \mathcal{M}_1(\mathbb{R})$,

For all strategies $\Psi$ consistent wrt $\mathcal{D}$,

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

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

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

To do so (and to prove other lower bounds), we will need the following fundamental inequality:

LEMMA: Denote by $\mathbb{P}_\nu^{H_T}$ the law of $H_T$ in the bandit problem $\nu$ given a strategy $\psi$.

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,

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

References: The equality is very standard and follows from the chain rule; already present in Auer, Cesa-Bianchi, Freund & Schapire, 2002, for Bernoulli distributions.

The inequality had already been stated for events: $Z = \mathbb{1}_A$ where $A$ is a $\sigma(H_T)$-measurable event; see Combes and Proutière, 2014, Kaufmann, Cappé and Garivier, 2016.

Proof: The inequality is a direct application of the data-processing inequality for random variables, see previous lecture for its statement.

For the equality: it is convenient to take $\Omega = [0,1] \times (\mathbb{R} \times [0,1])^T$ as the underlying probability space,

$U_0, Y_1, U_1, \ldots Y_T, U_T$ are the projections,

$\mathbb{P}_\nu$ is given by

$\forall$ Borel sets $B \subset [0,1]$, $\mathbb{P}_\nu(U_0 \in B) = \mathfrak{m}(B)$ where $\mathfrak{m}$ is the Lebesgue measure

$\forall$ Borel sets $B' \subset \mathbb{R}$ and $B \subset [0,1]$, $\forall t \geq 0$, $\mathbb{P}_\nu(Y_{t+1} \in B' \text{ and } U_{t+1} \in B \mid H_t) = \nu_{\psi_t(H_t)}(B')\,\mathfrak{m}(B)$

That is, $\mathbb{P}_\nu^{H_{t+1}} = K_t \, \mathbb{P}_\nu^{H_t}$ where $K_t(h,\cdot) = \nu_{\psi_t(h)} \otimes m$ is a stochastic kernel

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

Let us check the assumptions of our chain rule:

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

with $(h,(y,u)) \longmapsto \dfrac{d K_t(h,\cdot)}{d K'_t(h,\cdot)}(y,u) = \sum_{a=1}^{K} \mathbb{1}_{\{\psi_t(h)=a\}} \dfrac{d\nu_a}{d\nu'_a}(y)$

being measurable

and (*) $h \longmapsto KL(K_t(h,\cdot), K'_t(h,\cdot)) = \sum_{a=1}^{K} \mathbb{1}_{\{\psi_t(h)=a\}} KL(\nu_a, \nu'_a)$

being measurable.

Therefore,

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

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

see which values $\psi_t(H_t) = I_{t+1}$ can take

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

cf. $I_{t+1} = \psi_t(H_t)$

Then, an induction: the hypothesis being $KL(\mathbb{P}_\nu^{H_t}, \mathbb{P}_{\nu'}^{H_t}) = \sum_{a=1}^{K} \mathbb{E}_\nu[N_a(t)] \, KL(\nu_a, \nu'_a)$

For $t=0$: $H_0 = U_0$ and $\mathbb{P}_\nu^{H_0} = m = \mathbb{P}_{\nu'}^{H_0}$, as well as $N_a(0) = 0$

From $t \geq 0$ to $t+1$: see the equality we proved above.

Back to our theorem:

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

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

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

This is why we will

- Fix $\mathcal{D}$, $\psi$, $\nu$ and $a$ s.t. $\Delta_a > 0$
- Fix an alternative model $\nu'$ 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, $\nu$ and $\nu'$ only differ at $a$; $a$ is the unique optimal arm in $\nu'$

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

Our fundamental inequality yields, since $\nu$ and $\nu'$ only differ at $a$:

$$E_\nu[N_a(T)]\ KL(\nu_a, \nu'_a) \geq KL\left(Ber\left(E_\nu[N_a(T)/T]\right), Ber\left(E_{\nu'}\left[\frac{N_a(T)}{T}\right]\right)\right)$$

$$\geq -\ln 2 + \left(1 - E_\nu[N_a(T)/T]\right) \ln \frac{1}{1 - E_{\nu'}[N_a(T)/T]}$$

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

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

$$= \underbrace{p \ln \frac{1}{q}}_{\geq 0} + (1-p) \ln \frac{1}{1-q} + (\underbrace{p \ln p + (1-p)\ln(1-p)}_{\geq -\ln 2})$$

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

for all $p, q \in (0,1)$ and even for all $p, q \in [0,1]$

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

- in the problem $\nu$, $a$ is suboptimal: $E_\nu[N_a(T)/T] \longrightarrow 0$

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

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

↳ in particular, for $T$ large enough,

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

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

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

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

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

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

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

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

<u>Remark:</u> Non-asymptotic lower bounds are possible (for « super-consistent » strategies and « well behaved » models), but are heavily technical; see Garivier, Ménard, Stoltz, 2016.

Second THEOREM: initial linear regime for the $\mathbb{E}_\nu[N_a(T)]$ when $a$ is suboptimal

We observe on simulations that even for not so small $T$, we have

$\forall a$, $\mathbb{E}_\nu[N_a(T)]$ of the order of $T/K$

ie: a long preliminary almost uniform exploration phase.

Restrictions: As before we can only consider « reasonable » strategies (we should discard strategies like "Always pull arm #1" that will perform perfectly on some bandit problems just by chance... )

Our simpler result is associated with the following restriction; we have more complicated results with nicer (milder) restrictions, like only very mild symmetry requirements.

Definition: A strategy $\psi$ is smarter than the uniform strategy (on $\mathcal{D}$) if for all bandit problems $\nu \in \mathcal{D}^K$, for all optimal arms $a^*$,

$$\forall T \geq 1, \qquad \mathbb{E}_\nu[N_{a^*}(T)] \geq T/K.$$

THEOREM: For all strategies $\psi$ that are smarter than the uniform strategy, for all bandit problems $\nu$, for all arms $a$, for all $T \geq 1$,

$$\mathbb{E}_\nu[N_a(T)] \geq \frac{T}{K}\left(1 - \sqrt{2T\, \mathcal{K}_{\inf}(\nu_a, \mu^*)}\right)$$

In particular: $\forall \nu \in \mathcal{D}^K$, $\forall a$, $\forall T \leq \dfrac{1}{8\, \mathcal{K}_{\inf}(\nu_a, \mu^*)}$, $\mathbb{E}_\nu[N_a(T)] \geq \dfrac{T}{2K}$.

Proof: It suffices to consider suboptimal arms $a$ (the smarter-than-the-uniform-strategy requirement on $\psi$ takes care of optimal arms $a^*$).

As in the previous proof: we consider any $\nu'_a \in \mathcal{D}$ with $\nu_a \ll \nu'_a$ and $E(\nu'_a) > \mu^*$, we show that for this $\nu'_a$,

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

and can conclude by taking the supremum of the right-hand side over those $\nu'_a$ (the left-hand side is independent of $\nu'_a$).

We form $\nu'$ given by $\begin{cases} \nu'_k = \nu_k & k \neq a \\ \nu'_a & \text{as fixed above} \end{cases}$

ie: we pick the same alternative bandit problem as in the previous proof.

The fundamental lemma (with $Z = N_a(T)/T$) ensures again:

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

where we used the short-hand notation:

$$kl(p,q) = KL(\mathrm{Ber}(p), \mathrm{Ber}(q)) = p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q}.$$

Since $a$ is optimal under $\nu'$: $\quad \mathbb{E}_{\nu'}[N_a(T)] \geq T/K$

We may assume with no loss of generality: $\quad \mathbb{E}_{\nu}[N_a(T)] \leq T/K$ $\quad$ otherwise, we have the result!

Finally, $\quad q \longmapsto kl(p,q)$ is increasing on $[p,1]$

Therefore:

$$kl\left(\mathbb{E}_{\nu}\left[\frac{N_a(T)}{T}\right], \mathbb{E}_{\nu'}\left[\frac{N_a(T)}{T}\right]\right) \geq kl\left(\mathbb{E}_{\nu}\left[\frac{N_a(T)}{T}\right], 1/K\right)$$

We will prove below: $\quad \forall\, 0 \leq p < q \leq 1, \quad kl(p,q) \geq \frac{1}{2q}(p-q)^2 \quad (*)$

We get:

$$\frac{T}{K} KL(\nu_a, \nu'_a) \geq \mathbb{E}_{\nu}[N_a(T)]\, KL(\nu_a, \nu'_a) \geq kl\left(\mathbb{E}_{\nu}\left[\frac{N_a(T)}{T}\right], 1/K\right)$$

$$\geq \frac{K}{2}\Big(\underbrace{\mathbb{E}_{\nu}\left[N_a(T)/T\right]}_{\leq 1/K} - 1/K\Big)^2$$

Thus, $\quad \frac{1}{K} - \mathbb{E}_{\nu}\left[N_a(T)/T\right] \leq \sqrt{\frac{2T}{K^2} KL(\nu_a, \nu'_a)}.$

Re-arranging concludes the proof.

It remains to prove $(*)$, which we will do together with many other lower bounds on kl.

List of inequalities. The last one will be used in the next session!

Pinsker's inequality: $\forall (p,q) \in [0,1]^2 \qquad kl(p,q) \geq 2(p-q)^2$

Local refinement: $\forall\, 0 \leq p < q \leq 1, \qquad kl(p,q) \geq \dfrac{1}{2 \max_{x \in [p,q]} x(1-x)} (p-q)^2$

(Credits? Is it Cappé et al., 2013? In any case, it is Aurélien Garivier who mentioned this inequality to us/me)

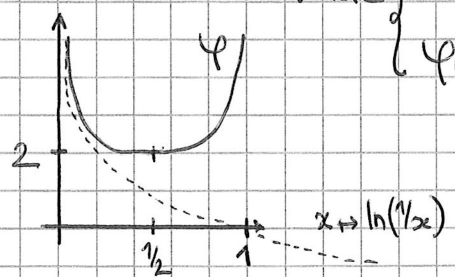
$$\geq \frac{1}{2q}(p-q)^2$$

Global optimal refinement: $\forall (p,q) \in [0,1]^2, \qquad kl(p,q) \geq \varphi(q)\,(p-q)^2$

(Weinberger, 2005)

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

where $\begin{cases} \varphi(q) = \dfrac{\ln((1-q)/q)}{1-2q} & \text{for } q \in (0,1) \setminus \{1/2\} \\ \varphi(1/2) = 2 \end{cases}$




Put differently: $|p-q| \leq \sqrt{\dfrac{kl(p,q)}{\varphi(q)}}$

Note: of course, by the data-processing inequality, all these bounds for KL of Bernoulli distributions entail bounds for any pair of probability distributions:

Eg (for the global optimal refinement): $\mathbb{P}, \mathbb{Q}$ two probability measures over $(\Omega, \mathcal{F})$, then

$$\| \mathbb{P} - \mathbb{Q} \|_{TV} = \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)| \leq \sup_{A \in \mathcal{F}} \sqrt{\frac{KL(\mathbb{P},\mathbb{Q})}{\varphi(\mathbb{Q}(A))}}$$

$$= \sqrt{\frac{KL(\mathbb{P},\mathbb{Q})}{\inf_{A \in \mathcal{F}} \varphi(\mathbb{Q}(A))}}$$

<u>Proof</u> [of the local refinement]:

We may assume $p>0$ and $q<1$

( $q=1 \geqslant p$ : $kl(p,1)=+\infty$, the inequality is trivial; $p>0$ to $p=0$ : by continuity )

$$\frac{\partial}{\partial p} kl(p,q) = \cancel{1}+\ln p + \ln\frac{1}{q} \cancel{-1} - \ln(1-p) - \ln\frac{1}{1-q}$$

$$\frac{\partial^2}{\partial^2 p} kl(p,q) = \frac{1}{p}+\frac{1}{1-p} = \frac{1}{p(1-p)}$$

Taylor's equality: $\exists r\in(p,q) \mid$

$$kl(p,q) = \underbrace{kl(q,q)}_{=0} + (p-q)\underbrace{\frac{\partial}{\partial p}kl(q,q)}_{=0} + \frac{(p-q)^2}{2}\underbrace{\frac{\partial^2}{\partial^2 p}kl(r,q)}_{=1/r(1-r)}$$

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

The consequence $kl(p,q)\geqslant \frac{1}{2q}(p-q)^2$ is because $\max\limits_{x\in[p,q]} x(1-x) \leq \max\limits_{x\in[p,q]} x \leq q.$

<u>Proof</u> [of the global refinement]: $q\in\{0,1\}$ treated separately ($kl(p,q)=+\infty$ in these cases unless $p=q$)

The case $q=1/2$ is addressed in a second time.

We fix $q\in(0,1)\setminus\{1/2\}$ and study $f: p\mapsto \frac{kl(p,q)}{(p-q)^2}$ for $p\neq q$ and an extension by continuity at $p=q$

We exactly show that attains its minimum over $[0,1]$ at $p=1-q$, from which the result and its <u>optimality</u> follows by noting that

$$f(1-q) = \frac{kl(1-q,q)}{(1-2q)^2} = \frac{\ln((1-q)/q)}{1-2q} \overset{\text{def}}{=} \varphi(q)$$

We replace Taylor's equality by the Taylor formula with integral remainder:

$$kl(p,q) = \cancel{kl(q,q)} + (p-q)\cancel{\frac{\partial}{\partial p}kl(q,q)} + \int_q^p (p-t)\underbrace{\frac{\partial^2}{\partial^2 p}kl(t,q)}_{1/t(1-t)}\,dt$$

After a linear change of variable: $t = q + u(p-q)$ $u \in [0,1]$

$$f(p) = \frac{kl(p,q)}{(p-q)^2} = \int_0^1 (1-u) \frac{1}{(q+u(p-q))(1-q-u(p-q))} du$$

$g : x \mapsto \frac{1}{x(1-x)}$ strictly convex $= \int_0^1 (1-u)\, g(q+u(p-q))\, du$

thus $f$ is strictly convex as well, by integration.

↳ Unique global minimum, at $x$ s.t. $f'(x) = 0$

Now, $$f'(1-q) = \int_0^1 \underbrace{g'(q+u(1-2q))\; u(1-u)}_{\text{antisymmetric in } u \text{ around } u = 1/2} du = 0$$

thus $\int_0^1 \dots du = 0$

because $u \mapsto u(1-u)$ symmetric around $u = 1/2$

because $q + (1-u)(1-2q) = 1 - (q + u(1-2q))$ is symmetric around $u = 1/2$

and $g'(x) = -\frac{1}{x^2} + \frac{1}{(1-x)^2}$ is anti-symmetric around $x = 1/2$

So, we caught the global minimum!

Which concludes the proof for the case $q \neq 1/2$.

For $q = 1/2$: by continuity (of $\varphi$, of $q \mapsto kl(p,q)$)

<u>Credits</u> for this proof (which a much simplified version of the original proof by Weinberger, 2005, and which yields the optimality in passing):

Gerchinovitz, Ménard, Stoltz, 2017?