Let's first complete the proof of the Lemma: ["Hoeffding-Azuma inequality with a random number of summands"]

Setting: Probability distributions $\nu_1, \dots \nu_K$ over $[0,1]$ with respective expectations $\mu_1, \dots \mu_K$

At each round, $I_t \in \{1, \dots K\}$ is picked in a $\sigma(Y_1, \dots Y_{t-1})$-measurable way

then Y_t is drawn independently at random according to ν_{I_t}, given I_t

ie: $Y_t \mid I_t \sim \nu_{I_t}$

We denote $N_a(t) = \sum_{s=1}^{t} \mathbb{1}_{\{I_s = a\}}$ and assume that each arm a was pulled once in the first K rounds, so that:

$N_a(t) \geqslant 1 \quad \forall t \geqslant K$

Then, for $t \geqslant K$: $\hat{\mu}_{a,t} = \frac{1}{N_a(t)} \sum_{s=1}^{t} Y_s \mathbb{1}_{\{I_s = a\}}$

Lemma: $\forall \delta \in (0,1)$, $\mathbb{P}\left\{ \mu_a > \hat{\mu}_{a,t} - \sqrt{\frac{\ln(1/\delta)}{2 N_a(t)}} \right\} \geqslant 1 - t\delta$

(and a similar symmetric statement with $\mu_a < \hat{\mu}_{a,t} + \sqrt{\dots}$)

The proof will be based on the fact that $(Z_t)_{t \geqslant 0}$, where

$$Z_t = \sum_{s=1}^{t} (Y_s - \mu_a) \mathbb{1}_{\{I_s = a\}}$$

is a martingale wrt. $(\mathcal{F}_t) = (\sigma(Y_1, \dots Y_t))_{t \geqslant 0}$, which we already proved last time:

$$\mathbb{E}\left[(Y_t - \mu_a) \mathbb{1}_{\{I_t = a\}} \mid Y_1 \dots Y_{t-1} \right] = \mathbb{E}\left[(Y_t - \mu_a) \mathbb{1}_{\{I_t = a\}} \mid I_t, Y_1 \dots Y_{t-1} \right]$$

$$= \left(\mathbb{E}[Y_t \mid I_t, Y_1 \dots Y_{t-1}] - \mu_a \right) \mathbb{1}_{\{I_t = a\}}$$

where we used the bandit model

$$= (\mu_{I_t} - \mu_a) \mathbb{1}_{\{I_t = a\}} = 0 \quad \text{a.s.}$$

Remark: How does this bound compare to what the classical version of the Hoeffding-Azuma says?

Martingale increment $(Y_s - \mu_a) \mathbb{1}_{\{I_s = a\}}$ bounded between

$a_t = -\mu_a$ and $b_t = 1 - \mu_a$

so that

(actually in the version I stated, I can have $\leqslant$ or $<$) $(b_t - a_t)^2 = 1$

 Lecture 5 — March 7, 2018 Lecture 5 — March 7, 2018

$$1-t\delta \leqslant \mathbb{P}\left\{ Z_t < \sqrt{\frac{t}{2}\ln\frac{1}{t\delta}} \right\} = \mathbb{P}\left\{ N_a(t)\left(\hat{\mu}_{a,t}-\mu_a\right) < \sqrt{\frac{t}{2}\ln\frac{1}{t\delta}} \right\}$$

$$= \mathbb{P}\left\{ \hat{\mu}_{a,t} - \sqrt{\frac{t}{N_a(t)}}\sqrt{\frac{\ln(1/t\delta)}{2\,N_a(t)}} < \mu_a \right\}$$

versus the bound of our lemma: $1-t\delta \leqslant \mathbb{P}\left\{ \hat{\mu}_{a,t} - \sqrt{\frac{\ln(1/\delta)}{2N_a(t)}} < \mu_a \right\}$

The proposed deviations essentially differ from a $\sqrt{t/N_a(t)}$ factor, and it is so nice to get rid of it!

<u>Proof:</u> (1) We prove that $\forall x\in\mathbb{R}$, $\mathbb{E}\left[e^{x Z_t - x^2/8\, N_a(t)} \right] \leqslant 1$

We do so by showing that $M_t = \exp\left(x Z_t - \frac{x^2}{8} N_a(t) \right)$ is a supermartingale, so that $\mathbb{E}[M_t] \leqslant \mathbb{E}[M_0] = 1$.

<u>Indeed</u>, by the conditional version of Hoeffding's lemma,

$$\mathbb{E}\left[e^{x(Y_t-\mu_a)\mathbb{1}_{\{I_t=a\}}} \,\middle|\, \mathcal{F}_{t-1} \right] \leqslant e^{x^2/8} \quad \text{a.s.}$$

but we can do better!

Since I_t and thus also $\mathbb{1}_{\{I_t=a\}}$ are $\mathcal{F}_{t-1}$-measurable, we get:

$$\mathbb{E}\left[e^{x(Y_t-\mu_a)\mathbb{1}_{\{I_t=a\}}} \,\middle|\, \mathcal{F}_{t-1} \right] = \mathbb{E}\left[e^{x(Y_t-\mu_a)\mathbb{1}_{\{I_t=a\}}} \left(\mathbb{1}_{\{I_t=a\}} + \mathbb{1}_{\{I_t\neq a\}} \right) \,\middle|\, \mathcal{F}_{t-1} \right]$$

$$= \mathbb{E}\left[e^{x(Y_t-\mu_a)\mathbb{1}_{\{I_t=a\}}} \,\middle|\, \mathcal{F}_{t-1} \right] \mathbb{1}_{\{I_t=a\}} + e^0\, \mathbb{1}_{\{I_t\neq a\}}$$

given what we had before

$$\leqslant e^{x^2/8}\, \mathbb{1}_{\{I_t=a\}} + \mathbb{1}_{\{I_t\neq a\}} = \exp\left(\frac{x^2}{8} \mathbb{1}_{\{I_t=a\}} \right)$$

Put differently, $\mathbb{E}\left[e^{x(Y_t-\mu_a)\mathbb{1}_{\{I_t=a\}} - \frac{x^2}{8}\mathbb{1}_{\{I_t=a\}}} \,\middle|\, \mathcal{F}_{t-1} \right] \leqslant 1$

which entails that $\exp\left(x\sum_{s=1}^{t} (Y_s-\mu_a)\mathbb{1}_{\{I_s=a\}} - \frac{x^2}{8}\sum_{s=1}^{t}\mathbb{1}_{\{I_s=a\}} \right)$

$$= \exp\left(x Z_t - \frac{x^2}{8} N_a(t) \right) = M_t$$

is a supermartingale wrt $\mathcal{F}_t = \sigma(Y_1 \dots Y_t)$.

(2) We prove that $\forall \varepsilon > 0$, $\forall \ell \geq 1$, $\mathbb{P}\{Z_t \geq \varepsilon \text{ and } N_a(t) = \ell\}$
$\leq \exp(-2\varepsilon^2/\ell)$

Indeed, by a Markov-Chernov bounding,

$\forall x > 0$,

$$\begin{aligned}
\mathbb{P}\{Z_t \geq \varepsilon \text{ and } N_a(t) = \ell\} &\leq e^{-x\varepsilon}\, \mathbb{E}\left[e^{xZ_t}\, \mathbb{1}_{\{N_a(t)=\ell\}}\right] \\
&= e^{-x\varepsilon + \frac{x^2\ell}{8}}\, \mathbb{E}\left[e^{xZ_t - \frac{x^2}{8}N_a(t)}\, \mathbb{1}_{\{N_a(t)=\ell\}}\right] \\
&\leq e^{-x\varepsilon + x^2\ell/8}\, \underbrace{\mathbb{E}\left[e^{xZ_t - \frac{x^2}{8}N_a(t)}\right]}_{\leq 1 \text{ by } (1)}
\end{aligned}$$

Optimizing over $x > 0$
(take $x = 4\varepsilon/\ell$) yields the claimed bound.

(3) Conclusion: we prove that $\mathbb{P}\left\{\mu_a \leq \hat{\mu}_{at} - \sqrt{\frac{\ln(1/\delta)}{2N_a(t)}}\right\} \leq t\delta$

Indeed, by distinguishing according to the values taken by $N_a(t)$:

$$\begin{aligned}
&\mathbb{P}\left\{\mu_a \leq \hat{\mu}_{at} - \sqrt{\frac{\ln(1/\delta)}{2N_a(t)}}\right\} \\
&= \sum_{\ell=1}^{t} \mathbb{P}\left\{N_a(t) = \ell \text{ and } \mu_a \leq \hat{\mu}_{at} - \sqrt{\ln(1/\delta)/2\ell}\right\} \\
&= \sum_{\ell=1}^{t} \mathbb{P}\left\{N_a(t) = \ell \text{ and } \frac{Z_t}{N_a(t)} \geq \sqrt{\ln(1/\delta)/2\ell}\right\} \\
&= \sum_{\ell=1}^{t} \mathbb{P}\left\{N_a(t) = \ell \text{ and } Z_t \geq \sqrt{\ell \ln(1/\delta)/2}\right\} \\
&\leq \sum_{\ell=1}^{t} \exp\Big(-2\left(\ell \ln(1/\delta)/2\right)/\ell\Big) = t\delta.
\end{aligned}$$

by (2)

($\sum_{\ell=1}^{t-K+1}$ would be enough)

Stochastic bandits : What about arms indexed by a continuum?

Setting 1 : Arms indexed by $x \in A$, where A is some possibly large set. With each arm $x \in A$ is associated a probability distribution ν_x over $\mathbb{R}$ s.t. $E(\nu_x)$ exists.
At each round, the decision-maker picks $I_t \in A$, gets a reward Y_t drawn at random according to ν_{I_t} (given I_t); and this is the only feedback she gets.

Definition : $f : x \in A \longmapsto E(\nu_x)$ is the mean-payoff function

Regret : $\overline{R}_T = T \sup_{x \in A} f(x) - E\left[\sum_{t=1}^{T} Y_t \right]$

Setting 2 : [special case] $\longrightarrow$ Noisy optimization of a function.
We fix $f : A \rightarrow \mathbb{R}$
The noise is given by a sequence of iid random variables $\varepsilon_1, \varepsilon_2, \ldots$
When $I_t \in A$ is picked, $Y_t = f(I_t) + \varepsilon_t$

↳ Special case of setting #1 where ν_x is the distribution of $f(x) + \varepsilon_1$ (all these distributions have the same shape given by the common distribution of the ε_j)

We of course need conditions for the regret to be minimized.

Definition : Let $\mathcal{F}$ be a set of possible bandit problems $\nu = (\nu_x)_{x \in A}$

The regret can be controlled (in a non-uniform way) against $\mathcal{F}$ if:

we also say that $(A, \mathcal{F})$ is tractable

there exists a strategy s.t. $\forall \nu \in \mathcal{F}$, $\overline{R}_T = o(T)$.

 Lecture 5 — March 7, 2018 Lecture 5 — March 7, 2018

Ex: $A = \{1, \dots K\}$ and $\underbrace{\mathcal{F} = \left(\mathcal{P}([0,1])\right)^K}_{\text{the case of finitely many arms with bounded distributions}}$, the set of all K-tuples of probability distributions over $[0,1]$

$\rightarrow$ UCB does the job.

Counter-example: $A = [0,1]$ and $\underbrace{\mathcal{F} = \left(\mathcal{P}([0,1])\right)^{[0,1]}}$

illustrating that continuity is a minimal requirement.

all bandit problems $(\nu_x)_{x \in [0,1]}$ with distributions ν_x having support $[0,1]$.

Indeed: Consider $(\delta_0)_{x \in [0,1]}$ the bandit problem in which each arm x is associated with the Dirac mass on 0

Fix any strategy: it gets $Y_t = 0$ $\forall t$ and uses a sequence of (possibly) random choices I_t, $t \geq 1$

Since probability distributions can only have at most countably many atoms, $\mathcal{Y} = \{x \in [0,1] : \exists t \mid \mathbb{P}\{I_t = x\} > 0 \text{ under } (\delta_0)_{x \in [0,1]}\}$ is countable. In particular, $[0,1] \setminus \mathcal{Y}$ is non empty.

But the strategy behaves the same under the problem $(\nu'_x)_{x \in [0,1]}$ in which

$$\begin{cases} \nu'_x = \delta_0 & \forall x \neq x_0 \\ \nu'_{x_0} = \delta_1 & \text{for one fixed } x_0 \in [0,1] \setminus \mathcal{Y} \end{cases}$$

With probability 1, it thus never hits x_0.

Therefore, $Y_t = 0$ a.s. $\forall t$ and $\bar{R}_T = \underbrace{T}_{=T} - \mathbb{E}\left[\sum_{t=1}^{T} Y_t\right]$

Actually, continuity is sufficient for the regret to be controlled, as long as A is not too large.

Theorem: Let A be a metric space and let $\mathcal{F}^{\text{cont}}$ be the set of bandit problems $(\nu_x)_{x \in A}$ with:

$\rightarrow$ $\forall x$, ν_x is a distribution over $[0,1]$

$\rightarrow$ a continuous mean-payoff function $f : x \mapsto \mathbb{E}(\nu_x)$

The regret can be controlled against $\mathcal{F}^{\text{cont}}$ if and only if A is separable.

Corollary. Let A be any set. Let $\mathcal{F}^{\text{all}}$ be the family of all bandit models $(\nu_x)_{x\in A}$ with distributions ν_x over $[0,1]$. Then the regret against $\mathcal{F}^{\text{all}}$ can be controlled if and only if A is at most countable.

Before we prove these facts, consider the following more concrete example, in which, by strengthening the regularity requirement on the mean-payoff function, we can even get rates.

Exercise: (Lipschitz bandits) Let $A = [0,1]$ and let $\mathcal{F}^{\text{Lip.}}$ be the family of bandit models $(\nu_x)_{x\in[0,1]}$ with distributions ν_x over $[0,1]$ and with mean-payoff functions that are Lipschitz.

Exhibit a strategy based on UCB + a sequence of discretizations of $[0,1]$ into K bins (to be refined over time) such that:

$$\forall \nu \in \mathcal{F}^{\text{Lip}}, \qquad \overline{R}_T \leq \left(3L + 6\sqrt{8\ln T + 2}\right)T^{2/3} + 2$$

where L is the Lipschitz constant of the mean-payoff function of ν.

Hint: First, prove a performance bound by splitting $[0,1]$ into $[(i-1)/K,\ i/K]$ with $i = 1,\dots,K$ for a fixed K, where each bin $[(i-1)/K,\ i/K]$ plays the role of an i in a bandit problem with finitely many arms. Then discuss how to pick K over the time, as we do in the next proof.

Proof of the Corollary:

We endow A with the discrete topology, ie, chose the distance $d(x,y) = \mathbb{1}_{\{x\neq y\}}$. Then:

1. All applications $f : A \to \mathbb{R}$ are continuous
2. A is separable if and only if A is at most countable.

Proof of the Theorem: It relies on the possibility or impossibility of uniform exploration of the arms.

1) **If A is separable:** let $(x_n)_{n\in\mathbb{N}}$ be a collection of points in A that is dense.

In particular, the probability distribution $d = \sum_{n\geq 0} \frac{1}{2^{n+1}} \delta_{x_n}$ is such that $d(V) > 0$ for all open sets $V \subset A$.

We pick elements $J_1 \ J_2 \ I_1^{(2)} \dots J_t \ I_1^{(t)} \dots I_{t^2-1}^{(t)}$ as follows:

$$
\begin{array}{llll}
J_1 & & \leftarrow & \text{regime } 1 \\
J_2 & I_1^{(2)} \dots I_3^{(2)} & \leftarrow & 2 \quad 2^2 \text{ rounds} \\
\vdots & \ddots & & \vdots \\
J_t & I_1^{(t)} \dots I_{t^2-1}^{(t)} & \leftarrow & t \quad t^2 \text{ rounds}
\end{array}
$$

Actually we could take $J_m = x_m$ in order, it would work also, this extra randomization is not needed...

where $\begin{cases} \text{the } J_s \text{ are drawn at random according to } d \\ \text{the } I_s^{(r)},\ 1 \leq s \leq r^2-1, \text{ follow from the UCB strategy with arms } J_1 \dots J_r \end{cases}$

In regime r: $\quad r^2 \max_{s\leq r} \mu_{J_s} - \mathbb{E}\left[\sum_{s=1}^{r^2} Y_{S_{r-1}+s}\right]$

$\leq 1 + c\sqrt{r^3 \ln r}$

regime r starts at time $S_{r-1}+1$ where $\left[1+2^2+\dots+(r-1)^2\right] = S_{r-1}$

we have it first for $\mathbb{E}[\ \cdot \mid J_1 \dots J_r]$ and then of course for $\mathbb{E}[\ ]$ by tower rule

for J_r

well-chosen numerical constant

distribution-free regret bound for UCB on r^2-1 steps with r arms (we saw this bound as an exercise)

Let $\varepsilon > 0$, let $\tilde{r}_\varepsilon$ the first (random) time when $\mu_{J_r} = f(J_r) \geq \sup_{x\in A} f(x) - \varepsilon$

We have $\tilde{r}_\varepsilon < +\infty$ a.s. because:

- by continuity of f, there exists an open set V_ε with $\forall x' \in V_\varepsilon,\ f(x') \geq \sup_A f - \varepsilon$;

– we have $\tilde{r}_\varepsilon \leqslant \inf\{ r \geqslant 1 : \ J_r \in V_\varepsilon \} < +\infty$ a.s.

as this random variable follows a geometric distribution with parameter $d(V_\varepsilon) > 0$.

For $r \geqslant \tilde{r}_\varepsilon$, $\displaystyle \max_{s \leq r} \mu_{J_s} + \varepsilon \geqslant \sup_A f$

So that

$$\overline{R}_T = T \sup_A f - \mathbb{E}\left[\sum_{t=1}^{T} Y_t \right]$$

$$\leqslant \sum_{r=1}^{\tilde{r}_\varepsilon - 1} r^2 + T\varepsilon + \sum_{r=\tilde{r}_\varepsilon}^{r_T - 1} \underbrace{\left(1 + c\sqrt{r^3 \ln T} \right)}_{\text{the "complete regimes"}} + r_T^2$$

lengths of regimes $r \leq \tilde{r}_\varepsilon - 1$
$< +\infty$ a.s.

regime r_T may be incomplete

where r_T is such that time T belongs to regime r_T: we have r_T^3 of order T
ie, r_T of order $T^{1/3}$

and

$$\sum_{r \leq r_T} \left(1 + c\, r^{3/2} \sqrt{\ln r} \right) \leqslant \sum_{r \leq r_T} \left(1 + c\, r^{3/2} \sqrt{\ln r_T} \right) = O\left(r_T^{5/2} \sqrt{\ln r_T} \right) = O\left(T^{5/6} \sqrt{\ln T} \right)$$

Thus,

$$\limsup_{T \to +\infty} \frac{\overline{R}_T}{T} \leqslant \varepsilon \quad \text{a.s.}$$

but since $\overline{R}_T$ is a deterministic quantity and this is true $\forall \varepsilon > 0$, we have

$$\lim_{T \to +\infty} \frac{\overline{R}_T}{T} = 0 \quad \text{as requested.}$$

2) If $\mathcal{A}$ is not separable:

* We use the following characterization of separability (which relies on Zorn's lemma):

  A metric space $\mathcal{X}$ is separable if and only if it contains no uncountable subset $\mathcal{D}$ s.t.
  $$\rho = \inf\{ d(x,y) : x,y \in \mathcal{D} \} > 0.$$

In particular, if $\mathcal{A}$ is not separable, there exist an uncountable subset $\mathcal{D} \subset \mathcal{A}$ and $\rho > 0$ such that the balls $B(a, \rho/2)$, with $a \in \mathcal{D}$, are all disjoint.

↳ No probability distribution over $\mathcal{A}$ can give a positive mass to all these balls.

* We consider the bandit models $\nu^{(a)}$ inducing mean-payoff functions
$$f^{(a)} : x \in \mathcal{A} \longmapsto \left(1 - \frac{d(x,a)}{\rho/2}\right)^+ ;$$ in particular, $\nu^{(a)}_x = \delta_0$ for $x \notin B(a, \rho/2)$.

$f^{(a)}$ is indeed continuous.

We proceed as in the example showing the necessity of continuity when $\mathcal{A} = [0,1]$ and consider the bandit model $(\delta_0)_{x \in \mathcal{A}}$, as well as any strategy and the laws induced by the I_t under this model: let α_t be the law of I_t under $(\delta_0)_{x \in \mathcal{A}}$ and let $\alpha = \sum_{t \geq 1} \frac{1}{2^t} \alpha_t$.

There exists $a \in \mathcal{A}$ s.t. $\alpha\big(B(a, \rho/2)\big) = 0$, that is, s.t.,
$$\forall t \geq 1, \quad \mathbb{P}\big(I_t \in B(a, \rho/2) \text{ under } (\delta_0)_{x \in \mathcal{A}} \big) = 0.$$

The considered strategy is therefore such that the I_t have the same distribution under $(\delta_0)_{x \in \mathcal{A}}$ and $\nu^{(a)}$. In particular,
$\mathbb{E}\left[\sum_{t=1}^T Y_t \right] = 0$ in both cases, but in the latter case, $\sup f^{(a)} = 1$, so that $\overline{R}_T = T$ against $\nu^{(a)}$. The regret is <u>not</u> controlled against $\nu^{(a)} \in \mathcal{F}^{\text{cont}}$.

Overview of the next steps: Fix a model $\mathcal{D}$, known to the decision-maker, ie, a collection of probability distributions over $\mathbb{R}$ with an expectation.

Assume that $\nu_1, \dots \nu_K$ are unknown but that the decision-maker knowns $\nu_j \in \mathcal{D}$ $\forall j$.

What are the best bounds on $\overline{R}_T = T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} Y_t\right]$?

We will show matching upper and lower bounds (with associated strategies):

$\overline{R}_T$ is at best of the order of $\left(\sum_{a : \Delta_a > 0} \frac{\Delta_a}{K_{\inf}(\nu_a, \mu^*, \mathcal{D})}\right) \ln T$

where

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

Kullback-Leibler divergence

expectation of ν_a'

We will do so by
- proving a universal lower bound
- exhibiting a strategy, called KL-UCB, to achieve the bound. ← if time permits (I'm not sure we will have time to do so...)

* But * before we do that, I guess that some reminder of basic and non-basic results about KL divergences would be needed!

 Lecture 5 — March 7, 2018 Lecture 5 — March 7, 2018

The Kullback-Leibler divergence: definition and basic properties.

Definition (intrinsic): 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 $[0,+\infty)$

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

It suffices to consider the case $\mathbb{P} \ll \mathbb{Q}$: because φ is strictly convex, Jensen's inequality indicates that

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

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

Exercise: A useful rewriting. Prove the following result:

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

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

Beware: with the usual measure theoretic conventions, if $x \neq 0$ and $y = 0$, then $x \neq y \frac{x}{y}$

↪ you therefore need to proceed with care!

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:

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

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

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

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

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

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

$$\overset{\text{not.}}{=} \int_\Omega \mathbb{1}_B(X) \, \gamma(X) \, d\mathbb{Q} \underset{\text{by definition of } \mathbb{Q}^X}{=} \int_{\Omega'} \mathbb{1}_B \, \gamma \, d\mathbb{Q}^X.$$

Therefore,

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

$$\underset{\text{definition of } \gamma}{=} \int_\Omega \left(\mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X \right] \ln \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X \right] \right) d\mathbb{Q}$$

$$\underset{\text{conditional version of Jensen's inequality}}{\leqslant} \int_\Omega \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \ln \frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X \right] d\mathbb{Q}$$

$$\underset{\text{tower rule}}{=} \int_\Omega \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \ln \frac{d\mathbb{P}}{d\mathbb{Q}} \right) d\mathbb{Q} = \mathrm{KL}(\mathbb{P}, \mathbb{Q})$$

References :

- The proof above is due to Ali and Silvey (1966), 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, which we discuss below, is typically proved in a tedious way, relying on the rewriting of Exercise 1 and the joint convexity of $(x,y) \in [0,+\infty)^2 \longmapsto \left(\frac{x}{y} \ln \frac{x}{y}\right) y$

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 $\alpha \in (0,1)$,

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

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 (X, J) by the projections

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

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

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

(and a similar definition for $\mathbb{Q}$ based on $\mathbb{Q}_1, \mathbb{Q}_2$)

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

 Lecture 5 — March 7, 2018 Lecture 5 — March 7, 2018

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

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

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

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

Indeed: we may assume with no loss of generality, given $\alpha \in (0,1)$, that $\mathbb{P}_1 \ll \mathbb{Q}_1$ and $\mathbb{P}_2 \ll \mathbb{Q}_2$, 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

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

we just use that for $f \geq$ a constant, $\int f\,d\mu = \int f\,\mathbb{1}_A\,d\mu + \int f\,\mathbb{1}_{A^c}\,d\mu$ whether f is integrable or not

$$= \int_{\Omega\times\{1,2\}} \left(\frac{d\mathbb{P}}{d\mathbb{Q}}(\omega,1)\,\ln\frac{d\mathbb{P}}{d\mathbb{Q}}(\omega,1)\right) \overbrace{\mathbb{1}_{\Omega\times\{1\}}(\omega,j)\,d\mathbb{Q}(\omega,j)}^{\text{on } \Omega\times\{1\},\ d\mathbb{Q} \text{ is } \alpha\, d\mathbb{Q}_1}$$

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

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

<u>KL for product measures.</u> $\quad$ ($\leftrightarrow$ The independent case)

<u>Proposition:</u> 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:

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

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, 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,

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

We use that if f, g are $\geqslant$ a constant, then $\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu$

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

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

Consequence (Garivier, Ménard, Stoltz, 2016): Data-processing inequality with expectations of random variables

Corollary: 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 respective expectations of X under $\mathbb{P}$ and $\mathbb{Q}$, we have:

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

Proof: We denote by $\mathfrak{m}$ 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 $\mathbb{P}\otimes\mathfrak{m}$ 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:

 Lecture 5 — March 7, 2018 Lecture 5 — March 7, 2018

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

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

Thus: $\quad KL\big(\mathrm{Ber}(\mathbb{P}\otimes m(E)), \mathrm{Ber}(\mathbb{Q}\otimes m(E)) \big) \leqslant KL(\mathbb{P},\mathbb{Q})$

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

By Tonelli's theorem:

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

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

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