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:
$$N_a(t) \geq 1 \quad \forall t \geq K$$

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

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

The proof will be based on the fact that $(Z_t)_{t\geq 0}$, where
$$Z_t = \sum_{s=1}^{t} (Y_s - \mu_a) \mathbb{1}_{\{I_s = a\}}$$
is a martingale w.r.t. $(\mathcal{F}_t) = (\sigma(Y_1, \dots Y_t))_{t \geq 0}$, which we already proved last time:
$$\begin{aligned} \mathbb{E}\left[(Y_t - \mu_a) \mathbb{1}_{\{I_t = a\}} \mid Y_1 \dots Y_{t-1} \right] &= \mathbb{E}\left[(Y_t - \mu_a) \mathbb{1}_{\{I_t = a\}} \mid I_t, Y_1 \dots Y_{t-1} \right] \\ &= \left(\mathbb{E}[Y_t \mid I_t, Y_1 \dots Y_{t-1}] - \mu_a \right) \mathbb{1}_{\{I_t = a\}} \\ &= (\mu_{I_t} - \mu_a) \mathbb{1}_{\{I_t = a\}} = 0 \text{ a.s.} \end{aligned}$$
where we used the bandit model

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
$$a_t = -\mu_a \quad \text{and} \quad b_t = 1 - \mu_a$$
so that
(actually in the version I stated, I can take $\leq$ or $<$)
$$(b_t - a_t)^2 = 1$$

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

$$= \mathbb{P}\left\{ \hat{\mu}_{a,t} - \sqrt{\frac{t}{N_1(t)}} \sqrt{\frac{\ln(1/t\delta)}{2\, N_1(t)}} < \mu_a \right\}$$

versus the bound of our lemma: $1 - t\delta \leqslant \mathbb{P}\left\{ \hat{\mu}_{a,t} - \sqrt{\frac{\ln(1/\delta)}{2 N_1(t)}} < \mu_a \right\}$

The proposed deviations essentially differ from a $\sqrt{t/N_1(t)}$ 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 $\mathbb{E}[M_t] \leqslant \mathbb{E}[M_0] = 1$.

Indeed, 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 $\mathcal{F}_t = \sigma(Y_1 \ldots Y_t)$.

 Lecture 2 — February 8, 2017 Lecture 2 — February 8, 2017

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

Indeed, by a Markov-Chernov bounding,

$\forall x > 0$,

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

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

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

Indeed, by the union bound,

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

<u>Overview of the next steps:</u> Fix a model $\mathcal{D}$, known to the decision-maker, i.e., 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 knows $\nu_j \in \mathcal{D}$ $\forall j$.

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):

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

where

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

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.

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

 Lecture 2 — February 8, 2017 Lecture 2 — February 8, 2017

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

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

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

Exercise 1: 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:

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

$$= \int_\Omega \ln\left(\frac{f}{g}\right) f\, d\mu$$

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!

<u>Lemma</u> (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:

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

<u>Proof</u>: 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 $B \in \mathcal{F}'$:

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

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

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 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

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

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

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

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

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

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 (by Tonelli's theorem)

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

we integrate first over the j

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

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

KL for product measures. ($\leftrightarrow$ The independent case)

Proposition: 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:

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

Lecture 2 — February 8, 2017 Lecture 2 — February 8, 2017

Proof: It suffices to consider the case $\mathbb{P} \ll \mathbb{P}'$ and $\mathbb{Q} \ll \mathbb{Q}'$.

Then $\mathbb{P} \otimes \mathbb{P}' \ll \mathbb{Q} \otimes \mathbb{Q}'$ with

$$\frac{d(\mathbb{P}\otimes\mathbb{P}')}{d(\mathbb{Q}\otimes\mathbb{Q}')} = \frac{d\mathbb{P}}{d\mathbb{Q}} \; \frac{d\mathbb{P}'}{d\mathbb{Q}'}$$

(this is a fundamental result in measure theory and one of the best characterizations of independence!).

Therefore,

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

by Fubini-Tonelli (cf. $x \mapsto x\ln x$ is bounded from below)

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

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:

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

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:

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

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

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

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 $\mathbb{Q}\otimes m(E)$.

<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.

<u>Definition:</u> 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

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

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 $(\Omega\times\Omega', \mathcal{F}\otimes\mathcal{F}')$:

$\forall A \in \mathcal{F}, \quad \forall B \in \mathcal{F}', \qquad KP(A \times B) = \int_\Omega \mathbb{1}_A(\omega)\, K(\omega, B)\, dP(\omega)$

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

Theorem (chain rule for KL):

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

with (**) $g : (\omega, \omega') \longmapsto \dfrac{dK(\omega,\cdot)}{dL(\omega,\cdot)}(\omega')$ 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 $\omega \mapsto KL\big(K(\omega,\cdot), L(\omega,\cdot)\big)$ is indeed $\mathcal{F}$-measurable, so that the integral in the right-hand side is well defined.

Question / Remark: Not sure how needed my assumptions are!

↳ See Exercise 2 on the next page.

Proof: • Since $g \ln g$ is lower bounded on $[0, +\infty)$, and in view of the measurability assumption (**), Tonelli's theorem wrt LQ ensures that

$$\omega \longmapsto \int_{\Omega'} \big(g(\omega,\omega') \ln g(\omega,\omega')\big)\, L(\omega, d\omega') = KL\big(K(\omega,\cdot), L(\omega,\cdot)\big)$$

is $\mathcal{F}$-measurable.

• Given (*), we have $KP \ll LQ$ if and only if $P \ll Q$; we thus assume with no loss of generality that $KP \ll LQ$ and $P \ll Q$ (otherwise, both sides of the putative equality equal $+\infty$).

• We write $\dfrac{dP}{dQ} = f$, we then have

$$\frac{d\, KP}{d\, LQ}(\omega,\omega') = f(\omega)\, g(\omega,\omega')$$

as can be seen by going back to the definition of LQ

And
$$KL(KP, LQ) = \int_{\Omega\times\Omega'} \Big(f(\omega) g(\omega,\omega') \ln\big(f(\omega) g(\omega,\omega')\big)\Big)\, \underbrace{d\, LQ(\omega,\omega')}_{L(\omega, d\omega')\, dQ(\omega)}$$

by Tonelli's theorem

$$= \int_{\Omega} f(\omega) \ln f(\omega) \Big(\underbrace{\int_{\Omega'} g(\omega,\omega')\, L(\omega, d\omega')}_{= K(\omega,\Omega') = 1} \Big)\, dQ(\omega)$$

$$+ \int_{\Omega} \Big(\underbrace{\int_{\Omega'} \big(g(\omega,\omega') \ln g(\omega,\omega')\big)\, L(\omega, d\omega')}_{KL(K(\omega,\cdot),\, L(\omega,\cdot))} \Big)\, \underbrace{f(\omega)\, dQ(\omega)}_{dP(\omega)}$$

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

as announced.

Exercise 2.

Try to weaken the assumptions of the chain-rule theorem.

→ Ideally, show that (*) and (**) can be relaxed into simply assuming (**) $\omega \mapsto KL\big(K(\omega,\cdot), L(\omega,\cdot)\big)$ is $\mathcal{F}$-measurable

→ At least, I feel that one could prove

$$KP \ll LQ \iff \begin{cases} P \ll Q \\ K(\omega,\cdot) \ll L(\omega,\cdot) \quad \text{for } P\text{-almost all } \omega \end{cases}$$

In which case, we could assume with no loss of generality that (*) holds and the bi-measurability assumption (**) would be the only assumption.

Note: The only non-immediate implication is $KP \ll LQ \Rightarrow K(\omega,\cdot) \ll L(\omega,\cdot)$ for P-almost all ω.

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:

  $\mu_a = \mathbb{E}(\nu_a)$ is the expectation of ν_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 $\Delta_a > 0$

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

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

  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 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 first prove one part of this optimality: a lower bound on $C_a(\underline{\nu})$.

↳ The upper bound will come later.

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 $\Delta_a > 0$),

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

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

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

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 $\nu \in \mathcal{D}^K$.

Example: $\mathbb{P}_\nu^{H_T}$ is the law of $H_T = (U_0, Y_1, U_1, \dots, Y_T, U_T)$ 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.

Lemma (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,

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

Exercise 3: Prove the theorem based on this lemma.
To that end: fix $\mathcal{D}$, ψ, $\nu \in \mathcal{D}^K$, and a s.t. $\Delta_a > 0$
Then build $\nu' \in \mathcal{D}^K$ as
$$\begin{cases} \nu'_j = \nu_j & \forall j \neq a \\ \nu'_a \in \mathcal{D} & \text{with } E(\nu'_a) > \mu^* \end{cases}$$
ν' is called an alternative problem (as opposed to ν: original bandit problem).
Note that $\begin{cases} a \text{ is suboptimal for } \nu \\ a \text{ is the only optimal arm for } \nu' \end{cases}$
and apply the lemma (fundamental inequality) with $Z = N_a(T)/T$.