Preuve du fait que les hypothèses ($\star$) et ($\star\star$) ne sont pas nécessaires pour la "chain rule"
Auteurs : Marin Ballu et Hédi Hadiji (étudiants 2017 du cours)

Proposition A.7. *On a l'implication suivante :*

$$\begin{cases} \mathbb{P} \ll \mathbb{Q} \\ K(\omega,\cdot) \ll L(\omega,\cdot) \quad \textit{pour } \mathbb{P}\textit{-presque-tout } \omega \end{cases} \quad \Rightarrow \quad K\mathbb{P} \ll L\mathbb{Q} \tag{69}$$

Si de plus, la tribu $\mathcal{F}'$ est générée par un π-système dénombrable $\Pi = (E_n)_{n\in\mathbb{N}}$, alors la réciproque est valable et on a l'équivalence :

$$\begin{cases} \mathbb{P} \ll \mathbb{Q} \\ K(\omega,\cdot) \ll L(\omega,\cdot) \quad \textit{pour } \mathbb{P}\textit{-presque-tout } \omega \end{cases} \quad \Leftrightarrow \quad K\mathbb{P} \ll L\mathbb{Q} \tag{70}$$

Remarque A.8. L'hypothèse faite sur $\mathcal{F}'$ est relativement générale. En particulier, elle est vérifiée lorsque $\mathcal{F}'$ est la tribu borélienne associée à une topologie à base dénombrable d'ouverts et par conséquent quand c'est la tribu borélienne sur un sous-ensemble de $\mathbb{R}^d$. En effet si l'on note $(O_i)_{i\in\mathbb{N}}$ la base d'ouverts, il suffit de considérer le π-système :

$$\Pi = \{O_i\} \cup \left\{ \bigcap_{k=1}^{n} O_{i_k} \right\} \tag{71}$$

c'est-à-dire les éléments de la base d'ouverts ainsi que leurs intersections finies. Alors Π est clairement un π-système dénombrable, et la tribu engendrée par Π contient tous les ouverts et par conséquent toute la tribu borélienne.

Preuve de la proposition. Nous esquissons ici une preuve de la première assertion. Soit $E \in \mathcal{F}\otimes\mathcal{F}'$ tel que $L\mathbb{Q}(E) = 0$. Supposons dans un premier temps que $E = A \times B$. Alors :

$$\int_A L(\omega, B)\,\mathrm{d}\mathbb{Q}(\omega) = 0$$

donc :

$$L(\omega, B) = 0 \quad \text{pour } \mathbb{Q}\text{-presque-tout } \omega \in A$$

donc comme $\mathbb{P} \ll \mathbb{Q}$:

$$L(\omega, B) = 0 \quad \text{pour } \mathbb{P}\text{-presque-tout } \omega \in A$$

et par conséquent grâce à la deuxième hypothèse :

$$K(\omega, B) = 0 \quad \text{pour } \mathbb{P}\text{-presque-tout } \omega \in A$$

Donc :

$$K\mathbb{P}(A\times B) = \int_A K(\omega, B)\,\mathrm{d}\mathbb{P}(\omega) = 0 \tag{72}$$

On peut alors voir que l'ensemble :

$$\big\{E \in \mathcal{F}\otimes\mathcal{F}' \,|\, L\mathbb{Q}(E) = 0 \Rightarrow K\mathbb{P}(E) = 0\big\}$$

est une classe monotone qui contient les produits de mesurables : c'est donc la tribu $\mathcal{F}\otimes\mathcal{F}'$ tout entière.

Supposons désormais que $\mathcal{F}'$ est générée par le π-système $(E_n)_{n\in\mathbb{N}}$. Supposons que $K\mathbb{P} \ll L\mathbb{Q}$. Montrons d'abord que $\mathbb{P} \ll \mathbb{Q}$. Soit $A \in \mathcal{F}$ tel que $\mathbb{Q}(A) = 0$ (l'hypothèse sur $\mathcal{F}'$ n'est pas nécessaire pour cela). Alors :

$$L\mathbb{Q}(A\times\Omega') = \mathbb{Q}(A) = 0 \tag{73}$$

Donc :

$$\mathbb{P}(A) = K\mathbb{P}(A \times \Omega') = 0 \tag{74}$$

Ce qui montre que $\mathbb{P} \ll \mathbb{Q}$.

Reste à prouver la seconde partie de l'assertion. Comme $K\mathbb{P} \ll L\mathbb{Q}$ et $\mathbb{P} \ll \mathbb{Q}$, on peut poser d'après le théorème de Radon-Nikodym :

$$f(\omega, \omega') = \frac{\mathrm{d}K\mathbb{P}}{\mathrm{d}L\mathbb{Q}}(\omega, \omega')$$

et

$$h(\omega) = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(\omega)$$

avec f (resp. h) mesurable par rapport à $\mathcal{F} \otimes \mathcal{F}'$ (resp. $\mathcal{F}$).

Posons alors :

$$\varphi(\omega, \omega') = \frac{f(\omega, \omega')}{h(\omega)} \mathbb{1}_{\{h \neq 0\}}(\omega) \tag{75}$$

φ est bien bi-mesurable.

La preuve de la proposition est conclue par un lemme que nous énonçons à part car il nous resservira dans la suite. □

Lemme A.9. *Avec les hypothèses et les notations qui précèdent, pour $\mathbb{P}$-presque-tout ω, on a :*

$$\forall B \in \mathcal{F}' \quad \int_B \varphi(\omega, \omega') L(\omega, \mathrm{d}\omega') = K(\omega, B) \tag{76}$$

Et par conséquent si ω vérifie la propriété ci-dessus :

$$K(\omega, \cdot) \ll L(\omega, \cdot) \tag{77}$$

et

$$\varphi(\omega, \omega') = \frac{\mathrm{d}K(\omega, \cdot)}{\mathrm{d}L(\omega, \cdot)}(\omega') \qquad \textit{pour } L(\omega, \cdot)\textit{-presque-tout } \omega' \tag{78}$$

Preuve du lemme. Soit $E_i \in \Pi$ générant $\mathcal{F}'$. Alors pour tout $A \in \mathcal{F}$ d'après la proposition A.6 :

$$\begin{aligned} \int_A \left(\int_{E_i} \varphi(\omega, \omega') L(\omega, \mathrm{d}\omega') \right) \mathrm{d}\mathbb{Q}(\omega) &= \int_{A \times E_i} \varphi(\omega, \omega') \, \mathrm{d}L\mathbb{Q}(\omega, \omega') \\ &= \int_{A \times E_i} \frac{f(\omega, \omega')}{h(\omega)} \mathbb{1}_{\{h \neq 0\}}(\omega) \, \mathrm{d}L\mathbb{Q}(\omega, \omega') \end{aligned}$$

Par définition de f, puis par la proposition A.6, cette intégrale vaut :

$$\begin{aligned} \int_{A \times E_i} \frac{1}{h(\omega)} \mathbb{1}_{\{h \neq 0\}}(\omega) \, \mathrm{d}K\mathbb{P}(\omega, \omega') &= \int_A \left(\int_{E_i} \frac{1}{h(\omega)} \mathbb{1}_{\{h \neq 0\}}(\omega) K(\omega, \mathrm{d}\omega') \right) \mathrm{d}\mathbb{P}(\omega) \\ &= \int_A \frac{1}{h(\omega)} \mathbb{1}_{\{h \neq 0\}}(\omega) K(\omega, E_i) \, \mathrm{d}\mathbb{P}(\omega) \\ &= \int_A \frac{1}{h(\omega)} \mathbb{1}_{\{h \neq 0\}}(\omega) K(\omega, E_i) \, h(\omega) \, \mathrm{d}\mathbb{Q}(\omega) \qquad \text{par définition de } h \\ &= \int_A \mathbb{1}_{\{h \neq 0\}}(\omega) K(\omega, E_i) \, \mathrm{d}\mathbb{Q}(\omega) \end{aligned}$$

On a donc montré que pour tout ensemble $A \in \mathcal{F}$:

$$\int_A \left(\int_{E_i} \varphi(\omega, \omega') L(\omega, \, \mathrm{d}\omega') \right) \mathrm{d}\mathbb{Q}(\omega) = \int_A \mathbb{1}_{\{h \neq 0\}}(\omega) K(\omega, E_i) \, \mathrm{d}\mathbb{Q}(\omega) \tag{79}$$

Par conséquent, pour $\mathbb{Q}$-presque-tout ω :

$$\mathbb{1}_{\{h \neq 0\}}(\omega) K(\omega, E_i) = \int_{E_i} \varphi(\omega, \omega') L(\omega, \, \mathrm{d}\omega')$$

Donc puisque $\mathbb{P} \ll \mathbb{Q}$, pour $\mathbb{P}$-presque-tout ω :

$$\mathbb{1}_{\{h \neq 0\}}(\omega) K(\omega, E_i) = \int_{E_i} \varphi(\omega, \omega') L(\omega, \, \mathrm{d}\omega') \tag{80}$$

Or, $\mathbb{P}[h = 0] = 0$ puisque :

$$\mathbb{P}[h = 0] = \int_\Omega \mathbb{1}_{\{h=0\}} \, \mathrm{d}\mathbb{P}(\omega) = \int_\Omega \mathbb{1}_{\{h=0\}} \, h(\omega) \, \mathrm{d}\mathbb{Q}(\omega) = 0 \tag{81}$$

Donc l'égalité (80) devient, pour $\mathbb{P}$-presque-tout ω :

$$K(\omega, E_i) = \int_{E_i} \varphi(\omega, \omega') L(\omega, \, \mathrm{d}\omega')$$

Comme les E_i sont en nombre dénombrable, on peut permuter $\mathbb{P}$-presque-sûrement et pour tout $i \in \mathbb{N}$. Donc, pour $\mathbb{P}$-presque-tout $\omega \in \Omega$:

$$\forall i \in \mathbb{N} \quad K(\omega, E_i) = \int_{E_i} \varphi(\omega, \omega') L(\omega, \, \mathrm{d}\omega') \tag{82}$$

On en déduit par le lemme de Dynkin (pour les π-systèmes), à ω fixé, que pour $\mathbb{P}$-presque-tout $\omega \in \Omega$:

$$\forall B \in \mathcal{F}' \quad K(\omega, B) = \int_B \varphi(\omega, \omega') L(\omega, \, \mathrm{d}\omega') \tag{83}$$

Le lemme est donc bien prouvé.

□

La proposition précédente va nous permettre d'établir la "chain rule" pour la divergence de Kullback-Leibler.

Proposition A.10 (Chain rule pour la KL)**.** *Soient $(\Omega, \mathcal{F})$ et $(\Omega', \mathcal{F}')$ des espaces mesurables. Soient $\mathbb{P}$ et $\mathbb{Q}$ des mesures de probabilité sur Ω, et $K, L : \Omega \times \mathcal{F}' \to \mathbb{R}$ des noyaux de transition.*

On suppose que la tribu $\mathcal{F}'$ est engendrée par un π-système $\Pi = (E_n)_{n \in \mathbb{N}}$ dénombrable.

Alors :

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

Remarque A.11. La formule est naturelle si l'on considère que l'égalité :

$$\frac{\mathrm{d}K\mathbb{P}}{\mathrm{d}L\mathbb{Q}}(\omega, \omega') = \frac{\mathrm{d}K(\omega, .)}{\mathrm{d}L(\omega, .)}(\omega') \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(\omega) \tag{84}$$

doit être vraie. Malheureusement, cette égalité n'a pas de sens telle quelle puisque les fonctions $\dfrac{\mathrm{d}K(\omega, \cdot)}{\mathrm{d}L(\omega, \cdot)}(\omega')$ sont définies à des ensembles $L(\omega, \cdot)$-négligeables près. Grâce aux hypothèses sous lesquelles nous nous plaçons, nous allons pouvoir surmonter ce problème via le lemme A.9.

Preuve. A priori, il nous faudrait en premier lieu vérifier que l'application $\omega \mapsto \mathrm{KL}\Big(K(\omega,\,\cdot\,), L(\omega,\,\cdot\,)\Big)$ est mesurable pour que l'intégrale dans le terme de droite ait un sens. En fait, sous nos hypothèses, nous pouvons commencer par évacuer certains cas où l'on peut fixer par convention que les deux termes valent $+\infty$.

On rappelle que sous nos hypothèses, l'équivalence dans la proposition A.7 est valable. Plusieurs cas de figure sont possibles :

Si $K\mathbb{P}$ n'est pas absolument continue par rapport à $L\mathbb{Q}$. Alors :

- soit $\mathbb{P}$ n'est pas absolument continue par rapport à $\mathbb{Q}$ et dans ce cas les deux termes dans l'égalité de la chain rule valent $+\infty$, indépendamment du sens que l'on donne à l'intégrale.
- soit il existe $A \in \mathcal{F}$ avec $\mathbb{P}(A) > 0$ et pour $\omega \in A$ on a $\mathrm{KL}\big(K(\omega,\cdot), L(\omega,\cdot)\big) = +\infty$. Dans ce cas, même si nous ne sommes pas sûrs que l'intégrande soit mesurable, on peut décréter que la valeur de l'intégrale vaut $+\infty$, et alors les deux termes de l'égalité sont infinis.

Supposons donc désormais que $K\mathbb{P} \ll L\mathbb{Q}$. Alors on peut poser comme dans le lemme A.9 :

$$f(\omega,\omega') = \frac{\mathrm{d}K\mathbb{P}}{\mathrm{d}L\mathbb{Q}}(\omega,\omega')$$

et

$$h(\omega) = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(\omega)$$

ainsi que :

$$\varphi(\omega,\omega') = \frac{f(\omega,\omega')}{h(\omega)}\mathbb{1}_{\{h\neq 0\}}(\omega)$$

Alors d'après le lemme A.9, pour $\mathbb{P}$-presque-tout ω :

$$\varphi(\omega,\omega') = \frac{\mathrm{d}K(\omega,\cdot)}{\mathrm{d}L(\omega,\cdot)}(\omega') \quad \text{pour } L(\omega,\cdot)\text{-presque-tout } \omega' \tag{85}$$