AKTAR YALÇIN

PROBABILITÉS AVANCÉES EISTI ING1-GM 2018-2019

MERCI DE ME SIGNALER TOUTE COQUILLE, YAR@EISTI.EU

EISTI

Table des matières

Préliminaires

On s'est basé sur le livre de Jacod et Protter [1].

1. J. Jacod and P. Protter. *L'essentiel en théorie des probabilités*. Collection Enseignement des mathématiques. Cassini, 2003. ISBN 9782842250508. URL https://books.google.fr/books?id=GHolAAAACAAJ

0.1 *tribu (ou σ-algèbre),mesure, mesure σ-finie, fonction mesurable, fonction borélienne*

Définition 1 *Soit X un ensemble. On appelle tribu (ou σ-algèbre) sur X, un ensemble $\mathcal{F}$ de parties de X qui vérifie*

1. $\mathcal{F} \neq \emptyset$
2. *$\mathcal{F}$ est stable par complémentaire :*
$$\forall B \in \mathcal{F},\, B^c \in \mathcal{F}.$$
3. *$\mathcal{F}$ est stable par union dénombrable :*
$$(\forall n \in \mathbb{N},\, B_n \in \mathcal{F}) \Rightarrow \bigcup_{n\in\mathbb{N}} B_n \in \mathcal{F}.$$

L'union est dite dénombrable parce que l'ensemble des indices l'est.

Quelques exemples

- La tribu dite grossière : $\mathcal{F} = \{\emptyset, X\}$.
- La tribu dite discrète : $\mathcal{F} = \mathcal{P}(X)$ où $\mathcal{P}(X)$ représente l'ensemble de toutes les parties de X.
- Si $X = \{a, b, c, d\}$, alors $\mathcal{F} = \{\emptyset, \{a\}, \{b, c, d\}, X\}$ est une tribu sur X. C'est la plus petite tribu contenant l'ensemble $\{a\}$.
- Pour tout X, $\{A \in \mathcal{P}(X) \,|\, A$ ou A^c fini ou dénombrable$\}$ est une tribu sur X.
- $(\forall n \in I,\, B_n \in \mathcal{F}) \Rightarrow \bigcap_{n\in I} B_n$, $\liminf_{n\to+\infty} A_n$, $\limsup_{n\to+\infty} A_n \in \mathcal{F}$.
- $A \in \mathcal{P}(X) \Rightarrow \bigcap_{A\in\mathcal{F}} \mathcal{F}$ est une tribu, appelée la tribu engendrée par A, notée par $\sigma(A)$.

Espace mesurable

Soit X un ensemble et $\mathcal{F}$ une tribu sur X. Le couple $(X, \mathcal{F})$ est appelé espace mesurable ou espace probabilisable en fonction du contexte. Sur les espaces mesurables on définit des mesures ; sur les espaces probabilisables on s'intéresse spécifiquement aux probabilités.

Les parties de X qui appartiennent à la tribu $\mathcal{F}$ sont appelés ensembles mesurables. Dans un contexte probabiliste, on les appelle événements et X est appelé l'univers.

Mesure

Une application μ définie sur $\mathcal{F}$ à valeurs dans $[0, +\infty]$, est appelée mesure lorsque les deux propriétés suivantes sont satisfaites :

$\mu(\emptyset) = 0$,

μ est σ-additive : si $E_1, E_2, \ldots$ est une famille dénombrable de parties de X appartenant à $\mathcal{F}$, et si ces parties sont deux à deux disjointes, alors la mesure $\mu(E)$ de leur réunion E est égale à la somme des mesures des parties :

$$\mu\big(\cup_{k\in\mathbb{N}^*} E_k\big) = \sum_{k\in\mathbb{N}^*} \mu(E_k).$$

Lorsqu'on dispose d'une mesure μ sur $(X, \mathcal{F})$, on dit que le triplet $(X, \mathcal{F}, \mu)$ est un espace mesuré.
Tout élément de $\mathcal{F}$ est appelé ensemble mesurable, la valeur $\mu(S)$ est appelée la mesure de S.
Lorsque $\mu(X)$ est fini, on parle de mesure finie ou mesure bornée.
Lorsque $\mu(X) = 1$, on parle de mesure de probabilité ; le triplet $(X, \mathcal{F}, \mu)$ est alors appelé un espace probabilisé ou de probabilité.
Un sous-ensemble S de X est dit négligeable lorsqu'il est inclus dans un T appartenant à la tribu $\mathcal{F}$ et de mesure nulle.
La mesure μ est dite complète lorsque tout ensemble négligeable appartient à la tribu $\mathcal{F}$.

Définition 2 (Mesure σ-finie) *Soit $(X, \mathcal{F}, \mu)$ un espace mesuré. On dit que la mesure μ est σ-finie lorsqu'il existe un recouvrement dénombrable de X par des sous-ensembles de mesure finie, c'est-à-dire lorsqu'il existe une suite $(E_n)_{n\in\mathbb{N}}$ d'éléments de la tribu $\mathcal{F}$, tous de mesure finie, avec*

$$X = \cup_{n\in\mathbb{N}} E_n.$$

Définition 3 (Mesure absolument continue) *Soient μ_1 et μ_2 deux mesures complexes sur un espace mesurable $(X, \mathcal{F})$. On dit que μ_1 est absolument continue par rapport à μ_2 si et seulement si pour tout ensemble mesurable A :*

$$\mu_2(A) = 0 \Rightarrow \mu_1(A) = 0,$$

ce que l'ont note $\mu_1 \ll \mu_2$.

Fonctions mesurables

Définition 4 *Soit E et F des espaces mesurables munis de leurs tribus respectives $\mathcal{E}$ et $\mathcal{F}$. Une fonction $f : E \to F$ est dite $(\mathcal{E}, \mathcal{F})$-mesurable si la tribu image réciproque par f de la tribu $\mathcal{F}$ est incluse dans $\mathcal{E}$, c'est-à-dire si :*

$$\forall B \in \mathcal{F},\ f^{-1}(B) \in \mathcal{E}.$$

Si F est l'ensemble des réels et si $\mathcal{F}$ est sa tribu borélienne, on dira simplement que f est une fonction mesurable sur $(E, \mathcal{E})$, et on utilisera la notation $f \in \mathcal{F}$.

Soient E un espace mesurable et $(f_n)_n$ une suite de fonctions mesurables de E dans $\mathbb{R}$ (ou même dans $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$). Alors la fonction f définie par $f = \inf_n f_n$ (à valeurs dans $\bar{\mathbb{R}}$) est mesurable. En effet, l'image réciproque par f de $]a, +\infty]$ peut s'écrire

$$\bigcup_{n \in \mathbb{N}} \{x \in E \,|\, f_n(x) > a\}$$

et cet ensemble est une réunion dénombrable d'éléments de $\mathcal{E}$, donc un ensemble mesurable.

Par passage aux opposés, on en déduit que, si les fonctions f_n de E dans $\bar{\mathbb{R}}$ sont toutes mesurables, alors la fonction $\sup_n f_n$ l'est également.

On peut alors montrer que les fonctions limites inférieure et supérieure $\liminf_{n \to +\infty} f_n$ et $\limsup_{n \to +\infty} f_n$ sont, elles aussi, mesurables.

Définition 5 *Toute fonction $f : (R, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ qui est mesurable est appelée Borel mesurable, ou borélienne.*

Quelques exemples :

- Toute fonction continue de $\mathbb{R}^n$ dans $\mathbb{R}$ est borélienne puisque l'image réciproque de tout ouvert de $\mathbb{R}$ est un ouvert de $\mathbb{R}^n$ (c'est une des définitions de la continuité des fonctions).
- Les fonctions monotones sur $\mathbb{R}$ sont boréliennes car l'image réciproque d'un intervalle est un intervalle, et donc l'image réciproque d'un ouvert est borélienne. En particulier, toute fonction de répartition est borélienne.
- Toute fonction dérivée est borélienne, car si f est dérivable sur $\mathbb{R}$, sa dérivée f' est la limite simple de la suite des fonctions continues

$$(n(f \circ (\mathrm{Id} + \tfrac{1}{n}) - f))_{n \in \mathbb{N}^*}.$$

- La somme de deux fonctions boréliennes f et g est borélienne. Par exemple,

$$\{x : f(x) + g(x) < a\} = \bigcup_{r \in \mathbb{Q}} \big(\{x : f(x) < r\} \cap \{x : g(x) < a - r\}\big).$$

Définition 6 *Soit* $(\Omega, \mathcal{F})$ *un espace mesurable. Soient* $I \subset \mathbb{N}^*$, $(A_i)_{i\in I} \in \mathcal{F}^I$ *tel que* $i \neq j \Rightarrow A_i \cap A_j = \emptyset$ *et* $\Omega = \cup_{i\in I} A_i$.

1. *Alors on appelle* $\mathcal{P} = (A_i)_{i\in I}$ *une partition de* Ω *(ou* $\mathcal{F}$*-mesurable de* Ω*).*
2. *Une partition* $\mathcal{P}' = (B_j)_{j\in J}$ *de* Ω *est dite plus fine que* $\mathcal{P}$ *(ou un raffinement de* $\mathcal{P}$*) si pour tout i il existe* $J_i \subset J$ *tel que* $A_i = \cup_{j\in J_i} B_j$ *ou, ce qui est équivalent à, ainsi qu'on le montre facilement, si pour tout j il existe i tel que* $B_j \subset A_i \in \mathcal{P}$ *et on note cela par* $\mathcal{P}' < \mathcal{P}$ *ou* $\mathcal{P} > \mathcal{P}'$.

Propriété 1

1. $\mathcal{P}' < \mathcal{P} \Leftrightarrow \sigma(\mathcal{P}) \subset \sigma(\mathcal{P}')$.
2. *Si* $\mathcal{P}_1$ *et* $\mathcal{P}_2$ *sont deux partitions différentes de* Ω*, il existe toujours une partition* $\mathcal{P}$ *de* Ω *à la fois plus fine que* $\mathcal{P}_1$ *et* $\mathcal{P}_2$

En effet : si $\mathcal{P}_1 = (A_i)_{i\in I}$ et $\mathcal{P}_2 = (B_j)_{j\in J}$ il suffira de prendre pour $\mathcal{P}$ la famille des $A_i \cap B_j$ non vides ($i \in I, j \in J$) (montrer-le). On note cela par $\mathcal{P} = \mathcal{P}_1 \vee \mathcal{P}_2$.

0.2 *Espaces* L^p

On ne considère que des fonctions à valeurs réelles (ou dans $[0, \infty]$). Le cas des fonctions à valeurs complexes s'en déduit en séparant parties réelles et imaginaires. Dans la première partie on travaille dans un espace mesuré quelconque puis, dans la deuxième partie on spécifiera à la mesure de Lebesgue sur $\mathbb{R}^k$.

0.2.1 *Approximation par des fonctions étagées*

Soit $(X, \mathcal{A}, \mu)$ un espace mesuré.

Définition 7 (Fonction étagée) *Une fonction f sur X est dite étagée lorsqu'il existe*

1. $I \subset \mathbb{N}$, $\mathrm{card}(I) < \infty$.
2. $(\alpha_i)_{i\in I} \in \mathbb{R}^I$.
3. $(A_i)_{i\in I} \in \mathcal{A}^I$.

tels que

$$f = \sum_{i\in I} \alpha_i \mathbf{1}_{A_i}.$$

Le théorème suivant permet d'approcher toute fonction mesurable positive par une suite de fonctions étagées :

Éspérances Conditionnelles

0.3 *Probabilité Conditionnelle et Indépendance*

Définition 8 (Probabilité Conditionnelle) *Pour tous événements $A, B \in \mathcal{F}$ t.q. $\mathbb{P}(B)) \neq 0$, la* ***probabilité conditionnelle*** *de A sachant B est définie par*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{1}$$

Exemple 1 (Formule de probabilités totales) *Pour tout événement $A \in \mathcal{F}$ et toute partition $(B_i)_{i\in\mathbb{N}^*} \in \mathcal{F}^{\mathbb{N}^*}$ de Ω t.q. $\mathbb{P}(B_i) > 0$ pour tout $i \in \mathbb{N}^*$,*

$$\mathbb{P}(A) = \sum_{i\in\mathbb{N}*} \mathbb{P}(A \cap B_i) = \sum_{i\in\mathbb{N}*} \mathbb{P}(A|B_i)\mathbb{P}(B_i),$$

où la première égalité et la deuxième égalité viennent respectivement de σ-additivité dénombrable de $\mathbb{P}$ et de (1).

Définition 9 (Indépendance des événements) *Soient $n \geq 2$ et $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Deux événements A et B de $\mathcal{F}$ sont dits* ***indépendants*** *(par rapport à $\mathbb{P}$) si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*

Une famille quelconque d'événements $A_i \in \mathcal{F}$, $i \in I$, est mutuellement indépendante si pour tout $J \subset I$ fini

$$\mathbb{P}\big(\cap_{j\in J} A_j\big) = \prod_{j\in J} \mathbb{P}(A_j),$$

est une famille d'événements deux à deux indépendants si pour tout $J \subset I$ fini et pour tous $i, j \in J$,

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

Exemple 2 *On lance un dé à 6 et on note A = "obtenir un nombre pair" et B = "obtenir un multiple de 3". On a : $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{1}{3}$ et $\mathbb{P}(A \cap B) = \frac{1}{6}$. Les événements A et B sont indépendants.*

On lance deux dés rouges et verts, et on note A = "La somme des numéros fait 6" et B= "Sur le dé rouge, on obtient un nombre pair". Un petit dénombrement de tous les cas possibles montre que $\mathbb{P}(A) = \frac{5}{36}$, $\mathbb{P}(B) = \frac{18}{36}$ et $\mathbb{P}(A \cap B) = \frac{2}{36}$. Les événements A et B ne sont pas indépendants.

Définition 10 (Indépendance de tribus) *Soient $n \geq 2$ et $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Une famille quelconque de sous-tribus $\mathcal{G}_i \subset \mathcal{F}$, $i \in I$, est mutuellement indépendante si toute famille d'événements $G_i \in \mathcal{G}_i$, $i \in I$, est mutuellement indépendante.*

Exemple 3 *Bien sûr, des événements mutuellement indépendants sont indépendants deux à deux, la réciproque est fausse, en effet en lançant 2 fois une pièce de monnaie et en posant A = "on obtient pile au premier lancer", B = "on obtient face au deuxième lancer" et C = "on obtient la même chose aux deux lancers", il vient très facilement $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$, $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}$, $\mathbb{P}(A \cap B \cap C) = 0$; les événements A, B et C sont donc deux à deux indépendants, mais ne sont pas mutuellement indépendants.*

Définition 11 (Indépendance de variables aléatoires) *Soient $n \geq 2$ et $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Une famille quelconque de variables* *aléatoires réelles X_i, $i \in I$, sur $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans $(E, \mathcal{B})$ est (mutuellement) indépendante si la famille des tribus engendrées par les X_i est (mutuellement) indépendante, i.e. pour tout $J \subset I$ fini, et tous les ensembles mesurables $B_j \in \mathcal{B}$, $j \in J$,*

$$\mathbb{P}\big(\cap_{j\in J}\{X_j \in B_j\}\big) = \prod_{j\in J}\mathbb{P}(X_j \in B_j).$$

Exemple 4 *Considérons l'exemple du dé, on a $(\Omega, \mathcal{F}, \mathbb{P})$ où $\Omega = \{1,2,3,4,5,6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et $\mathbb{P}(\{\omega\}) = \frac{1}{6}$ pour tout $\omega \in \Omega$. Soient*

$$A_1 = \{1,2,3\},\ A_2 = \{2,5\},\ X_i = \mathbf{1}_{A_i},\ \mathcal{F}_i = \sigma(\{A_i\}),\ i = 1,2.$$

Indépendance des événements :$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2) = \frac{1}{6}$, donc A_1 et A_2 sont indépendants.
Indépendance des variables aléatoires : remarquons que $\sigma(X_1) = \{\emptyset, \Omega, A_1, A_1^c\}$ et $\sigma(X_2) = \{\emptyset, \Omega, A_2, A_2^c\}$. Par calculs élémentaires $\mathbb{P}(B_1)\mathbb{P}(B_2) = \mathbb{P}(B_1)\mathbb{P}(B_2)$ pour tous $B_1 \in \sigma(X_1), B_2 \in \sigma(X_2)$, d'où X_1 et X_2 sont indépendants.

Pour montrer que deux sous-tribus $\mathcal{F}_1$ et $\mathcal{F}_2$ sont indépendantes, il faut montrer que tous les événements B_1, B_2 tels que $B_1 \in \mathcal{F}_1$ et $B_2 \in \mathcal{F}_2$ sont indépendants ; en notons que $\mathcal{F}_i$ coïncide avec $\sigma(X_i)$, pour $i \in \{1,2\}$, le précédent calcul permet donc de dire que $\mathcal{F}_1$ et $\mathcal{F}_2$ sont indépendants.

De ce qui précède on voit que si deux tribus $\mathcal{F}_1$ et $\mathcal{F}_2$ sont indépendantes, alors toute variable aléatoire $\mathcal{F}_1$-mesurable est indépendante de toute autre variable aléatoire $\mathcal{F}_2$-mesurable.

0.4 *Espérance Conditionnelle* [2]

2. M. Capinski and P.E. Kopp. *Measure, Integral and Probability*. Springer Undergraduate Mathematics Series. Springer London, 2013. ISBN 9781447106456. URL `https://books.google.fr/books?id=5d6PBAAAQBAJ`

0.4.1 *Conditionnement par rapport à un événement*

Définition 12 (Espérance conditionnelle par rapport à un événement) *Pour tout variable aléatoire $X \in L^1$ sur $(\Omega, \mathcal{F}, \mathbb{P})$ et pour tout événement $B \in \mathcal{F}$ tel que $\mathbb{P}(B) > 0$, l'espérance conditionnelle de X sachant B, que l'on note par $\mathbb{E}[X|B]$, est défini par*

$$\mathbb{E}[X|B] = \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}.$$

0.4.2 *Conditionnement par rapport à une variable aléatoire discrète*

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Soient X une variable aléatoire $L^1(\Omega)$ (donc l'espérance existe) et $Z : \Omega \to \{z_i\}_{i\in I} \subset \mathbb{R}^I$, où $I \subset \mathbb{N}^*$, une variable aléatoire discrète avec $i \neq j \Rightarrow z_i \neq z_j$. Sachant que nous avons défini l'espérance conditionnelle de X par rapport à un événement et que pour tout $B \in \mathcal{B}(\mathbb{R})$, $\{Z \in B\}$ est obtenu par les $\{Z = z_i\}$, $i \in I$, d'où la définition de l'espérance conditionnelle suivante :

Définition 13 (Espérance conditionnelle par rapport à une v.a.r. discrète) *Soient $I \subset \mathbb{N}^*$ et $X \in L^1(\Omega)$ et Z une v.a. discrète prenant valeurs dans $\{z_i\}_{i\in I} \subset \mathbb{R}^I$ avec $i \neq j \Rightarrow z_i \neq z_j$. L'espérance conditionnelle de X par rapport à Z est définie par la variable aléatoire discrète $\mathbb{E}[X|Z] : \Omega \to \mathbb{R}$ suivante :*

$$\mathbb{E}[X|Z] = \sum_{i\in I} \mathbb{E}[X|\{Z = z_i\}]\mathbf{1}_{\{Z=z_i\}}.$$

Notez que $\mathbb{E}[X|Z] \in L^1(\Omega)$, en effet $\mathbb{E}[|\mathbb{E}[X|Z]|] \leq \sum_{i\in I} \mathbb{E}[|X\mathbf{1}_{\{Z=z_i\}}|] = \mathbb{E}[|X|] < +\infty$.

Exemple 5 *Voir TD1 pour l'exercice.*

Théorème 1 *Si $X \in L^1(\Omega)$ et Z est une v.a. discrète, alors*

i) $\mathbb{E}[X|Z]$ *est* $\sigma(Z)$*-mesurable.*

ii) $\mathbb{E}[\mathbb{E}[X|Z]\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$, $\forall A \in \sigma(Z)$.

Preuve: On pose $B_i = \{Z = z_i\}$ et donc $(B_i)_{i\in I}$ est une partition de Ω, i.e. les B_i sont disjoints et $\Omega = \cup_{i\in I} B_i$. On pose $\mathcal{G} = \sigma(Z)$, et il est clair (montrer le) que $\sigma(Z) = \sigma(\{B_i, i \in I\})$, i.e. $\mathcal{G} = \sigma(\{B_i, i \in I\})$.

Montrons que $\sigma(\{B_i, i \in I\}) = \{\cup_{j\in J} B_j,\ J \subset I\} := K$. Il est facile de montrer que K est une tribu contenant les B_i, $i \in I$, donc $\mathcal{G}$ étant la plus petite tribu contenant les B_i, $i \in I$ on obtient que $\mathcal{G} \subset K$. Or $A \in K$ implique clairement que $A \in \mathcal{G}$, d'où $\mathcal{G} = K$, i.e. que $\sigma(\{B_i, i \in I\}) = \{\cup_{j\in J} B_j,\ J \subset I\}$.

On pose $\mathbb{E}[X|Z] := \sum_{i\in I} \mathbb{E}[X|B_i]\mathbf{1}_{B_i}$ et donc elle est $\sigma(Z)$-mesurable comme étant une somme de v.a. $\sigma(Z)$-mesurables. Pour le second point il est aisé de vérifier pour tout les $A = B_i \in \sigma(Z)$ que $\mathbb{E}[\mathbb{E}[X|Z]\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$, mais comme un $A \in \sigma(Z)$ s'écrit comme étant $\cup_{j\in J} B_j$ d'où par union disjoints d'événements B_j on obtient le second point vérifié. $\square$

Exemple 6 *Voir l'exercice du TD pour une application numérique du précédent théorème.*

0.5 *Conditionnement par rapport à une variable aléatoire quelconque*

Si la v.a. Z est discrète, nous pouvons définir $\mathbb{E}[X|Z]$ facilement, mais lorsque Z est une variable aléatoire quelconque (qui discrète ou continue), il n'est pas facile d'écrire une formule explicite pour $\mathbb{E}[X|Z]$ en générale.

Une voie naturelle serait d'écrire $\mathbb{E}[X|Z] = \int x \frac{f_{X,Z}(x,z)}{f_Z(z)}\, dx$, mais ce n'est pas utilisable si les densités $f_{X,Z}$ et f_Z n'existent pas.

Toutefois **Théorème** 1 nous suggère une voie pour définir l'espérance conditionnelle par rapport à une variable aléatoire quelconque :

Définition 14 (Espérance conditionnelle par rapport à une v.a. quelconque) *Soient $X \in L^1(\Omega)$ et Z une v.a. quelconque. Alors l'espérance conditionnelle de X par rapport à Z est définie comme étant une v.a. que l'on note par $\mathbb{E}[X|Z]$, telle que :*

i) $\mathbb{E}[X|Z]$ *est* $\sigma(Z)$*-mesurable,*

ii) Pour tout $A \in \sigma(A)$*,*

$$\mathbb{E}[\mathbb{E}[X|Z]\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A], \forall A \in \sigma(Z) \tag{2}$$

Comme l'espérance conditionnelle est définie implicitement par (2) au lieu d'une formule explicite, nous aurons besoin de justifier une telle définition en montrant que la v.a. $\mathbb{E}[X|Z]$ est caractérisée de façon unique ; comme il s'agit d'une v.a., cette unicité dépend donc de la mesure de probabilité $\mathbb{P}$ à travers la notion de $\mathbb{P}$-presque sûrement.

Définition 15 (Égalité $\mathbb{P}$-presque sûre)

1. *Deux événements $A, B \in \mathcal{F}$ sont égaux $\mathbb{P}$-presque sûrement ($\mathbb{P}$-p.s. ou p.s.) si $\mathbb{P}((A\backslash B) \cup (B\backslash A)) = 0$, i.e., la partie non commune de A et de B est de probabilité 0.*

2. *Deux v.a. X et Y définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ sont dites égales $\mathbb{P}$-presque sûrement si $\mathbb{P}(X = Y) = 1$, ce que l'on note par $X = Y$, $\mathbb{P}$-p.s.*

Exemple 7 *Notons que deux événements A et B sont égaux p.s. n'impliquent pas qu'ils sont les mêmes. On sait seulement que les événements $A\backslash B$ et $B\backslash A$ sont de mesures nulles. Par exemple considérons $A =]0, 0.5[$ et $B =]0, 0.5]$ sur $([0, 1], \mathcal{B}([0, 1]), \lambda_{[0,1]})$. Il est clair que $A = B$ p.s. mais $A \neq B$.*

Pour le cas de variables aléatoires, soient $\Omega = \{\omega_1, \omega_2\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}$ défini sur $\mathcal{F}$ par $\mathbb{P}(\{\omega_1\}) = 1$. Supposons que $X_i : \Omega \to \mathbb{R}$ est une v.a. avec $X_i = \mathbf{1}_{\{\omega_1\}} + (i + 1)\mathbf{1}_{\{\omega_2\}}$, $i \in \{1, 2\}$; alors $X_1 = X_2$ p.s. mais $X_1 \neq X_2$.

Théorème 2 (Unicité de l'espérance conditionnelle) $\mathbb{E}[X|Z]$ *est unique, c'est-à-dire si $X = X'$ p.s. alors $\mathbb{E}[X|Z] = \mathbb{E}[X'|Z]$.*

La preuve repose sur le lemme suivant :

Lemme 1 *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $\sigma(Z) \in \mathcal{F}$. Si Y est une v.a. $\sigma(Z)$-mesurable et $\mathbb{E}[Y\mathbf{1}_A] = 0$, pour tout $A \in \sigma(Z)$; alors $Y = 0$ p.s. C'est-à-dire, $\mathbb{P}(Y = 0) = 1$.*

Preuve: Soient $n \geq 1$ et $A = \{Y \geq \frac{1}{n}\}$. On a $A = \{Y \in [\frac{1}{n}, +\infty[\} \in \sigma(Z)$ et $0 \leq \mathbb{E}[\frac{1}{n}\mathbf{1}_A] \leq \mathbb{E}[Y\mathbf{1}_A] = 0$, d'où $\mathbb{P}(Y \geq \frac{1}{n}) = 0$; de la même façon $\mathbb{P}(Y \leq -\frac{1}{n}) = 0$ et donc $\mathbb{P}(A_n) = 1$ où $A_n = \{-\frac{1}{n} < Y < \frac{1}{n}\}$. Comme $(A_n)_{n\in\mathbb{N}^*}$ est une suite d'événements décroissante, i.e. $A_{n+1} \subset A_n$, et que $\{Y = 0\} = \cap_{n\in\mathbb{N}^*} A_n$ d'où

$$\mathbb{P}(Y = 0) = \mathbb{P}(\cap_{n=1}^{+\infty} A_n) = \lim_{m\to+\infty} \mathbb{P}(\cap_{n=1}^{m} A_n) = \lim_{m\to+\infty} \mathbb{P}(A_m) = 1. \tag{3}$$

□

Preuve: (Théorème 2) Supposons que $X = X'$ p.s., i.e. $\mathbb{P}(X = X') = 1$. On veut montrer que $\mathbb{E}[X|Z] = \mathbb{E}[X'|Z]$ p.s. On pose $Y = \mathbb{E}[X|Z] - \mathbb{E}[X'|Z]$ qui est par $\sigma(Z)$-mesurabilité des deux espérances, est $\sigma(Z)$-mesurable. Pour tout $A \in \sigma(Z)$, $\mathbb{E}[Y\mathbf{1}_A] = 0$, donc grâce au lemme précédent, $Y = 0$ p.s. i.e. $\mathbb{E}[X|Z] = \mathbb{E}[X'|Z]$ p.s.
□

Exemple 8 *Voir TD2.*

0.6 Conditionnement par rapport à une tribu

Dans les définitions précédentes nous conditionnions par rapport aux ensembles de $\sigma(Z)$, nous pouvons généraliser la définition de l'espérance conditionnelle par rapport à une tribu :

Définition 16 (Espérance conditionnelle par rapport à une tribu) *Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, $X \in L^1(\Omega)$ et soit $\mathcal{G} \subset \mathcal{F}$ une sous-tribu. Alors l'espérance conditionnelle de X par rapport à $\mathcal{G}$, que l'on note par $\mathbb{E}[X|\mathcal{G}]$, est définie comme étant une v.a. satisfaisant aux deux conditions suivantes :*

i) $\mathbb{E}[X|\mathcal{G}]$ *est une v.a. $\mathcal{G}$-mesurable*

ii) Pour tout $A \in \mathcal{G}$,

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]. \tag{4}$$

Notons que différentes variables aléatoires peuvent générer une même tribu. Par exemple, sur l'univers $\Omega = \{H, T\}$, soient $X_1 = \mathbf{1}_{\{H\}}$ et $X_2 = \mathbf{1}_{\{T\}}$. Alors X_1 et X_2 génèrent la même tribu $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$. C'est pourquoi l'espérance conditionnelle par rapport à une tribu est plus générale que l'espérance conditionnelle par rapport à une variable aléatoire.

Remarquer que l'unicité de l'espérance conditionnelle se démontre avec le même raisonnement utilisé dans la Preuve du **Théorème** 2.

Exemple 9 *Si $\mathcal{G} = \{\emptyset, \Omega\}$, alors $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ p.s.*

Pour la preuve, voir les TDs.

Exemple 10 *Pour tout $B \in \mathcal{G}$, on a $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|B] = \mathbb{E}[X|B]$.*
Pour la preuve voir les TDs.

0.7 *Dérivé de Radon-Nikodým*

Nous allons montrer l'existence de l'espérance conditionnelle, garantie par le Théorème de Radon-Nikodým.

Définition 17 (Mesure absolument continue) *Soient μ_1 et μ_2 deux mesures de probabilité sur $(\Omega, \mathcal{F})$ t.q. $\mu_1(A) = 0 \Rightarrow \mu_2(A) = 0$ pour tout $A \in \mathcal{F}$. Alors la mesure μ_2 est dite absolument continue par rapport à μ_1, et on note cela par $\mu_2 \ll \mu_1$.*

Exemple 11 *Considérons l'espace mesurable $(\Omega, \mathcal{F})$ où $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$. Pour tout $\omega \in \Omega$, soient $\mu_1(\{\omega\}) = \frac{1}{6}$, $\mu_2(\{\omega\}) = \frac{|\omega - 3.5|}{9}$, $\mu_3(\{\omega\}) = \frac{1}{3}\mathbf{1}_{\{\omega \geq 4\}}$ et $\mu_4(\{\omega\}) = \frac{1}{3}\mathbf{1}_{\omega \leq 3\}}$. Alors vérifier que $\mu_1 \ll \mu_2$, $\mu_2 \ll \mu_1$, $\mu_3 \ll \mu_i$ et $\mu_4 \ll \mu_i$ pour $i \in \{1, 2\}$.*

Exemple 12 *Sur l'espace mesurable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ on définit $\mathbb{P}_N$ et $\mathbb{P}_E$ par $\mathbb{P}_N(]a, b]) = F_N(b) - F_N(a)$ et $\mathbb{P}_E(]a, b]) = F_E(b) - F_E(a)$, où F_N et F_E sont les fonctions de répartitions respectives des lois normale standard et exponentielle standard. Alors $\mathbb{P}_E \ll \mathbb{P}_N$ est vraie alors que $\mathbb{P}_N \ll \mathbb{P}_E$ n'est pas vraie.*

Théorème 3 (Théorème de Radon-Nikodým) *Soit $(\Omega, \mathcal{F})$ un espace mesurable. Soient $\mathbb{P}$, $\mathbb{Q}$ deux mesures de probabilités sur $(\Omega, \mathcal{F})$ telles que $\mathbb{Q} \ll \mathbb{P}$. Alors il existe une unique (au sens presque-sûrement) variable aléatoire positive $Y \in L^1(\mathbb{P})$ telle que*

$$\mathbb{Q}(A) = \int_A Y \, d\mathbb{P}, \tag{5}$$

pour tout $A \in \mathcal{F}$. La variable aléatoire Y est appelée la dérivée Radon-Nikodým de $\mathbb{Q}$ par rapport à $\mathbb{P}$, que l'on note par $Y = \frac{d\mathbb{Q}}{d\mathbb{P}}$.

Preuve: Cette preuve qui est constructive et elle se fait en 4 étapes :

i) Supposer $\mathbb{Q}(A) \leq \mathbb{P}(A)$ pour tout $A \in \mathcal{F}$. Pour tout $n \in \mathbb{N}^*$, construire une v.a. Y_n sur une tribu $\mathcal{F}_n \subset \mathcal{F}$, engendrée par une partition finie de Ω.

ii) Prendre la limite $(Y_n)_{n \in \mathbb{N}^*}$ pour obtenir Y.

iii) Montrer que Y trouvé satisfait bien à l'égalité (5).

iv) Montrer que $\mathbb{Q}(A) \leq \mathbb{P}(A)$ pour tout $A \in \mathcal{F}$.

i) Supposons que $\mathbb{Q}(A) \leq \mathbb{P}(A)$ pour tout $A \in \mathcal{F}$. Soit $n \in \mathbb{N}^*$. Soit $(F_j)_{j \in \{1, 2, \ldots, r_n\}}$ une partition finie de Ω, i.e. $\Omega = \cup_{j=1}^{r_n} F_j$ où $r_n \in \mathbb{N}^*$ et $F_i \in \mathcal{F}$ avec $i \neq j \Rightarrow F_i \cap F_j = \emptyset$. Soit $\mathcal{F}_n = \sigma(F_1, \ldots, F_{r_n})$. On définit $Y_n : \Omega \to \mathbb{R}_+$ par $Y_n = \sum_{j=1}^{r_n} (\mathbb{Q}/\mathbb{P})(F_j) \mathbf{1}_{F_j} \mathbf{1}_{\mathbb{P}(F_j)>0}$, qui est bien intégrable puisque

$$\mathbb{E}[|Y_n|] = \int_\Omega Y_n \, d\mathbb{P} = \sum_{i=1}^{r_n} \int_{F_i} Y_n \, d\mathbb{P} = \sum_{i=1}^{r_n} \mathbb{Q}(F_i) = \mathbb{Q}(\Omega) = 1 < \infty.$$

De plus, pour tout $1 \leq j \leq r_n$ $\mathbf{1}_{F_j}$ est $\mathcal{F}$-mesurable car $F_j \in \mathcal{F}$, et pour tout $\omega \in F_j$, $(\mathbb{Q}/\mathbb{P})(F_j)$ est constante ; d'où Y_n est $\mathcal{F}$-mesurable comme somme de variables aléatoires $\mathcal{F}$-mesurables.

Comme tout $F \in \mathcal{F}_n$ est union finie de F_j, d'où (5) est vérifié pour tout $A \in \mathcal{F}_n$.

Soit $\mathcal{F}_{n+1} = \sigma(\tilde{F}_1, \ldots, \tilde{F}_{r_{n+1}})$ où $(\tilde{F}_k)_{1 \leq k \leq r_{n+1}}$ est une partition plus fine que $(F_k)_{1 \leq k \leq r_n}$. On définit $Y_{n+1} : \Omega \to \mathbb{R}_+$ par $Y_{n+1} = \sum_{j=1}^{r_{n+1}} (\mathbb{Q}/\mathbb{P})(\tilde{F}_j) \mathbf{1}_{\tilde{F}_j} \mathbf{1}_{\mathbb{P}(\tilde{F}_j)>0}$. De façon analogue, on démontre que Y_{n+1} est $\mathbb{P}$-intégrable et $\mathcal{F}_{n+1}$-mesurable, et $\mathbb{Q}(A) = \int_A Y_{n+1} \, d\mathbb{P}$ pour tout $A \in \mathcal{F}_{n+1}$. De plus, $\int_\Omega Y_{n+1}^2 \, d\mathbb{P} \geq \int_\Omega Y_n^2 \, d\mathbb{P}$, en effet $\int_\Omega Y_n (Y_{n+1} - Y_n) \, d\mathbb{P} = \sum_{j=1}^{r_n} \int_{F_j} Y_n (Y_{n+1} - Y_n) \, d\mathbb{P} = \sum_{j=1}^{r_n} \frac{\mathbb{Q}(F_j)}{\mathbb{P}(F_j)} \left(\int_{F_j} Y_{n+1} \, d\mathbb{P} - \int_{F_j} \frac{\mathbb{Q}(F_j)}{\mathbb{P}(F_j)} \, d\mathbb{P} \right) = 0$.

ii) Donc, lorsque la partition de Ω devient de plus en plus fine (par exemple la suite des tribus $\mathcal{F}_1 \subset \ldots \subset \mathcal{F}_k \subset \ldots$), la suite des dérivées Radon-Nikodým $(Y_k)_{k \in \mathbb{N}^*}$ devient une suite infinie croissante dans le sens où $\int_\Omega Y_{k+1}^2 \, d\mathbb{P} \geq \int_\Omega Y_k^2 \, d\mathbb{P}$ pour tout $k \in \mathbb{N}^*$. De l'autre côté $\left(\int_\Omega Y_k^2 \, d\mathbb{P} \right)_k$ est borné parce que $\mathbb{P}(\Omega) = 1$ et $Y_k \leq 1$ (en effet $\mathbb{Q}(A) \leq \mathbb{P}(A)$ pour tout $A \in \mathcal{F}$ et la définition de Y_k donne cela). Donc d'après le théorème de convergence monotone, la limite $\ell := \lim_{k \to +\infty} \int_\Omega Y_k^2 \, d\mathbb{P}$ existe.

Avec les mêmes arguments de limite on montre que $\lim_{k \to +\infty} Y_k := Y$ existe ; en effet il existe $N \in \mathbb{N}^*$ tel que pour tout $n \geq N$, $\ell - 4^{-n} \leq \int_\Omega Y_n^2 \, d\mathbb{P} \leq \int_\Omega Y_{n+1}^2 \, d\mathbb{P} \leq \ell$. Donc $\int_\Omega (Y_{n+1} - Y_n)^2 \, d\mathbb{P} = \int_\Omega (Y_{n+1}^2 - Y_n^2) \, d\mathbb{P} < 4^{-n}$. En utilisant l'inégalité de Cauchy-Schwartz avec $|Y_{n+1} - Y_n|$ et 1, on obtient $\int_\Omega |Y_{n+1} - Y_n| \, d\mathbb{P} < 2^{-n}$. Comme $(\sum_{k=N}^m |Y_{n+1} - Y_n|)_{m \geq N}$ est suite croissante de fonctions mesurables à valeurs positives, d'où d'après le théorème de convergence monotone (ou en version anglaise le théorème de Beppo-Levi) on a $\sum_{k=N}^m |Y_{n+1} - Y_n|$ est mesurable et $\int_\Omega \sum_{k=N}^\infty |Y_{n+1} - Y_n| \, d\mathbb{P} = \sum_{k=N}^\infty \int_\Omega |Y_{n+1} - Y_n| \, d\mathbb{P} < \infty$. Et donc on a $\sum_{k=N}^\infty |Y_{n+1} - Y_n| < \infty$ p.s. (sinon on ne peut pas avoir $\int_\Omega \sum_{k=N}^\infty |Y_{n+1} - Y_n| \, d\mathbb{P} < \infty$), par conséquent $\sum_{k=N}^\infty (Y_{n+1} - Y_n) < \infty$ p.s., c'est-à-dire $\lim_{k \to \infty} Y_k < \infty$ p.s.

iii) Comme Y_n est $\mathcal{F}_n$-mesurable où $\mathcal{F}_n \subset \mathcal{F}$ et $Y = \lim_{n \to \infty} Y_n$, d'où Y est $\mathcal{F}$-mesurable. Pour tout $A \in \mathcal{F}$, soit $\mathcal{R}_n = (A_i)_{1 \leq i \leq r_n} \vee \{A, A^c\}$ où $\mathcal{F}_n = \sigma(A_1, \ldots, A_{r_n})$. Il est facilement vérifiable que $\mathbb{Q}(A) = \int_A Y_{\mathcal{R}_n} \, d\mathbb{P}$ où $Y_{\mathcal{R}_n}$ est définie comme a été définie Y_n mais sur $\mathcal{R}_n$ au lieu de $\mathcal{F}_n$. Par l'étape ii), on a $\ell - 4^{-n} \leq \int_\Omega Y_n^2 \, d\mathbb{P} \leq \int_\Omega Y_{\mathcal{R}_n}^2 \, d\mathbb{P} \leq \ell$ et donc $\int_\Omega (Y_{\mathcal{R}_n} - Y_n)^2 \, d\mathbb{P} = \int_\Omega (Y_{\mathcal{R}_n}^2 - Y_n^2) \, d\mathbb{P} < 4^{-n}$ et en utilisant l'inégalité de Cauchy-Schwartz avec $\mathbf{1}_A$ et $|Y_{\mathcal{R}_n} - Y_n|$: $\left|\int_A (Y_{\mathcal{R}_n} - Y_n) \, d\mathbb{P}\right| \leq \int_A |Y_{\mathcal{R}_n} - Y_n| \, d\mathbb{P} \leq 2^{-n}$. Et donc en écrivant $\mathbb{Q}(A) = \int_A Y_{\mathcal{R}_n} \, d\mathbb{P} = \int_A (Y_{\mathcal{R}_n} - Y_n) \, d\mathbb{P} + \int_A Y_n \, d\mathbb{P}$; la première intégrale tend vers 0 lorsque $n \to \infty$ alors que la seconde tend vers $\int_A Y \, d\mathbb{P}$ par le théorème de convergence dominé (puisque $0 \leq Y_n \leq 1$ et $\mathbb{P}(\Omega) = 1$), donc on a $\mathbb{Q}(A) = \int_A Y \, d\mathbb{P}$ pour tout $A \in \mathcal{F}$.

iv) Dans le cas général, soit $\mathbb{S} = \mathbb{P} + \mathbb{Q}$, alors $\mathbb{P}(A) \leq \mathbb{S}(A)$ et $\mathbb{Q}(A) \leq \mathbb{S}(A)$ pour tout $A \in \mathcal{F}$. Les précédents résultats impliquent qu'il existe deux variables aléatoires positives intégrables mesurables $Y_\mathbb{Q}$ et $Y_\mathbb{P}$ telles que $\mathbb{Q}(A) = \int_A Y_\mathbb{Q} \, d\mathbb{S}$ et $\mathbb{P}(A) = \int_A Y_\mathbb{P} \, d\mathbb{S}$ pour tout $A \in \mathcal{F}$. Posons $B = \{Y_\mathbb{P} = 0\}$ et $Y = \frac{Y_\mathbb{Q}}{Y_\mathbb{P}} \mathbf{1}_{B^c}$. Alors $\mathbb{Q}(A) = \int_A Y_\mathbb{Q} \, d\mathbb{S} = \int_{A \cap B^c} \frac{Y_\mathbb{Q}}{Y_\mathbb{P}} Y_\mathbb{P} \, d\mathbb{S} + \int_{A \cap B} Y_\mathbb{Q} \, d\mathbb{S}$. Comme $\mathbb{P}(B) = \int_B Y_\mathbb{P} \, d\mathbb{S} = 0$ et $\mathbb{Q} \ll \mathbb{P}$, d'où $\mathbb{Q}(B) = 0$ et donc $\mathbb{S}(B) = 0 = \mathbb{S}(A \cap B)$ puisque $A \cap B \subset B$.

On obtient donc $\mathbb{Q}(A) = \int_A Y Y_\mathbb{P} \, d\mathbb{S}$.

En utilisant la démarche suggérée dans l'Exercice 1 on obtient

$$\mathbb{Q}(A) = \int_A Y \, d\mathbb{P}, \; \forall A \in \mathcal{F}.$$

□

Exercice 1 *En utilisant $\int_A Y_\mathbb{P} \, d\mathbb{S} = \mathbb{P}(A) = \int_A d\mathbb{P}$ montrer que $\mathbb{E}_\mathbb{S}[Y Y_\mathbb{P}] = \mathbb{E}_\mathbb{P}[Y]$ pour toute fonction indicatrice puis pour toute fonction simple Y sur $\mathcal{F}$, puis montrer la même égalité pour Y v.a. $\mathcal{F}$-mesurable, en utilisant le théorème de convergence monotone.*

Théorème 4 (Existence de l'espérance conditionnelle) *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $\mathcal{G} \subset \mathcal{F}$ une sous-tribu. Alors pour toute v.a. $X \in L^1(\Omega)$ il existe une v.a. positive $X_\mathcal{G}$ $\mathcal{G}$-mesurable t.q. pour tout $A \in \mathcal{G}$,*

$$\mathbb{E}[X \mathbf{1}_A] = \mathbb{E}[X_\mathcal{G} \mathbf{1}_A]. \tag{6}$$

$X_\mathcal{G}$ est l'espérance conditionnelle $\mathbb{E}[X|\mathcal{G}]$.

Preuve: **1.** *via Le Théorème de Radon-Nikodým* :

On se donne une v.a. positive $X \in L^1(\Omega, \mathcal{F})$ telle que $\mathbb{E}[X] = 1$. Soit $\mathbb{Q} : \mathcal{F} \to [0, +\infty[$ défini par

$$\mathbb{Q}(A) = \int_A X \, d\mathbb{P},$$

pour tout $A \in \mathcal{F}$.

Tout d'abord $\mathbb{Q}$ est bien une mesure de probabilité sur $(\Omega, \mathcal{F})$; en effet $\mathbb{Q}(\emptyset) = 0$, $\mathbb{Q}(\Omega) = 1$ et $\mathbb{Q}$ est σ-additive (c'est-à-dire que

$\mathbb{Q}(\cup_{k=1}^{+\infty} E_k) = \sum_{k=1}^{+\infty} \mathbb{Q}(E_k)$ où $E_1, E_2, \ldots \in \mathcal{F}$ sont deux à deux disjoints). $\mathbb{Q}$ est telle que $\mathbb{Q} \ll \mathbb{P}$ sur $\mathcal{F}$, c'est-à-dire absolument continu par rapport $\mathbb{P}$; en effet, si $A \in \mathcal{G}$ vérifie $\mathbb{P}(A) = 0$, alors pour tout $M > 0$, $\mathbb{E}[X\mathbf{1}_A] \leq \mathbb{E}[X\mathbf{1}_{0 \leq X \leq M}] + \mathbb{E}[X\mathbf{1}_{X>M}\mathbf{1}_A] M\mathbb{P}(A) + \mathbb{E}[X\mathbf{1}_{X>M}]$. Le premier terme de la somme ci-dessus est nul et par le théorème de convergence dominée (on domine par $X \in L^1(\Omega, \mathbb{F}, \mathbb{P})$), le deuxième tend vers 0 lorsque $M \to +\infty$, ce qui implique que $\mathbb{Q}(A) = 0$.

Soit $\mathbb{Q}^{\mathcal{G}}$ la restriction de $\mathbb{Q}$ à $\mathcal{G}$; elle est trivialement absolument continue par rapport à $\mathbb{P}^{\mathcal{G}}$, la restriction de $\mathbb{P}$ à $\mathcal{G}$. Le Théorème 3 de Radon-Nikodýmappliqué avec $\mathbb{P}^{\mathcal{G}}$, $\mathbb{Q}^{\mathcal{G}}$ définies sur $(\Omega, \mathcal{G})$ et $\mathbb{Q}^{\mathcal{G}} \ll \mathbb{P}^{\mathcal{G}}$ garantie l'existence d'une variable aléatoire $X_{\mathcal{G}} = \frac{d\mathbb{Q}^{\mathcal{G}}}{d\mathbb{P}^{\mathcal{G}}} \in L^1_+(\Omega, \mathcal{G}, \mathbb{P}^{\mathcal{G}})$ et $\mathcal{G}$-mesurable. Pour tout $A \in \mathcal{G}$, on a

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{Q}[A] = \mathbb{Q}^{\mathcal{G}}(A) = \mathbb{E}_{\mathbb{P}^{\mathcal{G}}}[X_{\mathcal{G}}\mathbf{1}_A] = \mathbb{E}[X_{\mathcal{G}}\mathbf{1}_A]$$

où la dernière égalité vient du fait que $X_{\mathcal{G}}\mathbf{1}_A$ est $\mathcal{G}$-mesurable. Par conséquent, $X_{\mathcal{G}}$ est l'espérance conditionnelle $\mathbb{E}[X|\mathcal{G}]$.

Lorsque $\mathbb{E}[X] \neq 1$, il suffit de travailler avec $\tilde{X} = \frac{1}{\mathbb{E}[X]} X$, en effet $\mathbb{E}[\tilde{X}] = 1$. On obtient donc $\tilde{X}_{\mathcal{G}} \in L^1_+(\Omega, \mathcal{G}, \mathbb{P})$ et donc $X_{\mathcal{G}} = \mathbb{E}[X]\tilde{X}_{\mathcal{G}}$. Mais pour faire toute cette démarche, il faut que $\mathbb{E}[X] > 0$; ce n'est pas grave car on peut s'en passer de ce cas, en effet pour $X \geq 0$ p.s. avec $\mathbb{E}[X] = 0$ cela donne $X = 0$ p.s. et donc $X_{\mathcal{G}} = 0$ p.s.

Jusqu'à maintenant on avait supposé que $X \geq 0$ p.s. Lorsque X n'est pas une v.a. positive, on écrit $X = X^+ - X^-$ où $X^+ = X\mathbf{1}_{X>0}$ et $X^- = -X\mathbf{1}_{X\leq 0}$ qui sont des v.a. positives intégrables ; en appliquant le résultat précédent à X^+ et X^- on prouve l'existence de $\mathbb{E}[X|\mathcal{G}]$. $\square$

0.8 *Propriétés Générales des Espérances Conditionnelles*

Propriété 2 *Soient $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $(X_n)_{n\in\mathbb{N}^*} \in (L^1(\Omega, \mathcal{F}, \mathbb{P}))^{\mathbb{N}^*}$ et $\mathcal{H}, \mathcal{G}$ deux sous-tribus de $\mathcal{F}$, alors*

1. $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$, *p.s.* *Linéarité.*
2. $\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$ *si X est $\mathcal{G}$-mesurable et $XY \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Sortir ce qui est connu.*
3. $\mathbb{E}[X|\mathcal{G}] = X$ *si X est $\mathcal{G}$-mesurable.*
4. $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ *si X est indépendant de $\mathcal{G}$.*
5. $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}]$ *si $\mathcal{H} \subset \mathcal{G}$* *Propriété de tour.*
6. $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.
7. $\mathbb{E}[X|\mathcal{G}] \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. *Intégrabilité.*
8. $X \geq 0$ *p.s.* $\Rightarrow \mathbb{E}[X|\mathcal{G}] \geq 0$ *p.s.* *Positivité.*
9. $X \leq Y$, *p.s.*$\Rightarrow \mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$, *p.s.* *Monotonicité.*

10. *ϕ convexe, $\phi(X) \in L^1(\Omega) \Rightarrow \phi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\phi(X)|\mathcal{G}]$, Inégalité de Jensen conditionnelle.*
11. *Si $X \in L^p(\Omega)$ pour $p \in [1,\infty[$, alors $\mathbb{E}[X|\mathcal{G}] \in L^p(\Omega)$ et*
$$\|\mathbb{E}[X|\mathcal{G}]\|_{L^p(\Omega)} \leq \|X\|_{L^p(\Omega)}$$
où $\|\cdot\|_{L^p(\Omega)} = (\mathbb{E}[|X|^p])^{1/p}$.
12. *Si $0 \leq X_n \leq X_{n+1}$, p.s. pour tout $n \in \mathbb{N}^*$ et $X_n \to X$, p.s., alors $\mathbb{E}[X_n|\mathcal{G}] \nearrow \mathbb{E}[X|\mathcal{G}]$, p.s.. Théorème de Convergence Monotone conditionnelle.*
13. *Si $0 \leq X_n$, p.s., pour tout $n \in \mathbb{N}^*$, et $\liminf_n X_n \in L^1$, alors $\mathbb{E}[\liminf_n X_n|\mathcal{G}] \leq \liminf_n \mathbb{E}[X_n|\mathcal{G}]$, p.s. Lemme de Fatou conditionnel.*
14. *Si $|X_n| \leq Z$, pour tout $n \in \mathbb{N}^*$ et $Z \in L^1(\Omega)$, et si $X_n \to X$, p.s., alors $L^1(\Omega) \ni \mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$, p.s. Théorème de Convergence Dominée conditionnelle.*

Preuve:

1. $a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$ est $\mathcal{G}$-mesurable car combinaison linéaire de variables aléatoires qui sont $\mathcal{G}$-mesurables, grâce à la définition de l'espérance conditionnelle. Pour tout $B \in \mathcal{G}$, $\mathbb{E}[(a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}])\mathbf{1}_B] = a\mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_B] + b\mathbb{E}[\mathbb{E}[Y|\mathcal{G}]\mathbf{1}_B] = a\mathbb{E}[X\mathbf{1}_B] + b\mathbb{E}[Y\mathbf{1}_B] = \mathbb{E}[(aX+bY)\mathbf{1}_B]$.
2. Soit Y une v.a. $\mathcal{G}$-mesurable et $XY \in L^1(\Omega)$, on doit montrer que pour tout $A \in \mathcal{G}$, $\mathbb{E}[XY\mathbf{1}_A] = \mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]\mathbf{1}_A]$. Pour cela, montrons que $\mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]]$, pour toute v.a. Z $\mathcal{G}$-mesurable t.q. $ZX \in L^1(\Omega)$; en effet après cela il suffit de prendre $Z = Y\mathbf{1}_A$.

   Pour Z une v.a. étagée, i.e. $Z = \sum_{k=1} \alpha_k \mathbf{1}_{A_k}$, ceci est évident par la définition de l'espérance conditionnelle et la linéarité.

   Supposons que Z et X sont positives p.s. et $ZX \in L^1(\Omega)$. Dans ce cas, il existe toujours une suite croissante de v.a. étagées $(Z_n)_{n\in\mathbb{N}^*}$ t.q. $Z_n \nearrow Z$. Alors $Z_nX \in L^1(\Omega)$ pour tout $n \in \mathbb{N}^*$ et le Théorème de Convergence Monotone implique que $\mathbb{E}[ZX] = \lim_{n\to\infty} \mathbb{E}[Z_nX] = \lim_{n\to\infty} \mathbb{E}[Z_n\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]]$. Maintenant que nous nous sommes occupés du cas $X \in L^1_+(\Omega)$, il faut s'occuper du cas $X \in L^1(\Omega)$. Dans ce cas, en utilisons $p = 1$ dans la propriété 10. (ou la propriété 9. avec $-|X| \leq X \leq |X|$), on obtient $|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}]$, p.s. et donc $|Z_n\mathbb{E}[X|\mathcal{G}]| \leq Z_n\mathbb{E}[|X||\mathcal{G}] \leq Z\mathbb{E}[|X||\mathcal{G}]$.

   Par le cas précédent, on sait que $\mathbb{E}[Z\mathbb{E}[|X||\mathcal{G}]] = \mathbb{E}[Z|X|]$, et donc $Z\mathbb{E}[|X||\mathcal{G}] \in L^1(\omega)$.

   On peut maintenant utiliser le théorème de convergence dominée pour conclure que $\mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]] = \lim_{n\to\infty} \mathbb{E}[Z_n\mathbb{E}[X|\mathcal{G}]] = \lim_{n\to\infty} \mathbb{E}[Z_nX] = \mathbb{E}[ZX]$.

Finalement, le cas général pour Z s'obtient par la linéarité avec $Z = Z^+ - Z^-$.

3. Il suffit de prendre $Y = 1$ dans 2. ou vérifier trivialement que X satisfait aux deux propriétés de la définition de l'espérance conditionnelle.

4. $\mathbb{E}[X]$ est constante donc $\mathcal{G}$-mesurable et pour tout $B \in \mathcal{G}$, $\mathbb{E}[X\mathbf{1}_B] = \mathbb{E}[X\mathbf{1}_B]$.

5. La seconde égalité est triviale car $\mathbb{E}[X|\mathcal{H}]$ est $\mathcal{G}$-mesurable. Pour la première égalité, remarquons que $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$ est $\mathcal{H}$-mesurable avec $\int_A \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]\, d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} = \int_A X\, d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{H}]\, d\mathbb{P}$ pour tout $A \in \mathcal{H}$.

6. Dans 5. prendre $\mathcal{H} = \{\emptyset, \Omega\}$ en utilisant $\mathbb{E}[X|\mathcal{H}] = X$.

7. Soient $Y = \mathbb{E}[X|\mathcal{G}]$, $A^+ = \{Y > 0\}$ et $A^- = \{Y < 0\}$. Puisque $A^+ = Y^{-1}(]0,\infty[)$ avec $]0,\infty[\in \mathcal{B}(\mathbb{R})$ et Y est $\mathcal{G}$-mesurable, d'où $A^+ \in \mathcal{G}$ et donc $0 \leq \mathbb{E}[Y\mathbf{1}_{A^+}] = \mathbb{E}[X\mathbf{1}_{A^+}] \leq \mathbb{E}[|X|\mathbf{1}_{A^+}]$. De même $A^- \in \mathcal{G}$ donc $0 \leq \mathbb{E}[-Y\mathbf{1}_{A^-}] = \mathbb{E}[-X\mathbf{1}_{A^-}] \leq \mathbb{E}[|X|\mathbf{1}_{A^-}]$. Finalement $\mathbb{E}[|Y|] = \mathbb{E}[Y\mathbf{1}_{A^+} - Y\mathbf{1}_{A^-}] \leq \mathbb{E}[|X|] < +\infty$.

8. Soit $X \geq 0$ p.s. On a $\mathbb{E}[X\mathbf{1}_{A^-}] = \mathbb{E}[Y\mathbf{1}_{A^-}]$ car $A^- \in \mathcal{G}$. Or $X\mathbf{1}_{A^-} \geq 0$ p.s. par hypothèse, tandis que $Y\mathbf{1}_{A^-} \leq 0$ p.s. par définition de A^-. On en déduit que $Y\mathbf{1}_{A^-} = 0$ p.s., c'est-à-dire que $Y \geq 0$ p.s.

9. Il suffit de considérer $Y - X \geq 0$ dans la propriété précédente et d'utiliser la linéarité.

10. <u>$1^{\text{ère}}$ preuve</u> : Tout d'abord montrons qu'une fonction convexe ϕ peut s'écrire comme $\phi(x) = \sup_{n\in\mathbb{N}^*}(a_n + b_n x)$, pour tout $x \in \mathbb{R}$ avec $(a_n)_{n\in\mathbb{N}^*}, (b_n)_{n\in\mathbb{N}^*} \in \mathbb{R}^{\mathbb{N}^*}$ bien choisies.

    ϕ étant convexe, on a $\phi(x) \geq \phi(y) + (x-y)\phi'_d(y)$ pour tous $x, y \in \mathbb{R}$, où ϕ'_d désigne la dérivée à droite de ϕ. On obtient facilement que $\phi(x) = \sup_{y\in\mathbb{R}}(\phi(y) + (x-y)\phi'_d(y)) := \sup_{y\in\mathbb{R}}(A_y + B_y x)$ pour tout $x \in \mathbb{R}$. Montrons que $\phi(x) = \sup_{y\in\mathbb{Q}}(A_y + B_y x)$ pour tout $x \in \mathbb{R}$:

    Le sens $\geq$ est évident. Soit $x \in \mathbb{R}$ et soit $(y_n)_{n\in\mathbb{N}^*} \in \mathbb{Q}^{\mathbb{N}^*}$ tel que $\lim_{n\to\infty} y_n = x$. Alors $A_{y_n} + B_{y_n}x = f(y_n) + (x - y_n)\phi'_d(y_n) \to f(x)$ lorsque $n \to \infty$, en effet ϕ est continue (car convexe) et $(\phi'_d(y_n))_{n\in\mathbb{N}^*}$ est clairement bornée. D'où le sens $\leq$ est vrai. Finalement presque-sûrement et pour tout $n \in \mathbb{N}^*$, on a $a_n + b_n\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[a_n + b_n X|\mathcal{G}] \leq \mathbb{E}[\sup_{k\in\mathbb{N}^*}(a_k + b_k X)|\mathcal{G}] = \mathbb{E}[\phi(X)|\mathcal{G}]$. Cela veut dire que l'inégalité ci-dessus est vrai sur $\Omega\backslash B_n$ avec $\mathbb{P}(B_n) = 0$. En prenant le supremum sur n, on obtient que $\phi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\phi(X)|\mathcal{G}]$ est vrai sur $\Omega\backslash B$ où $B = \cup_n B_n$ avec $\mathbb{P}(B) = 0$; ce qui est équivalent à

    $$\phi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\phi(X)|\mathcal{G}], \text{ p.s.}$$

<u>$2^{\text{ème}}$ preuve</u> : ϕ étant convexe, on a $\phi'_d(y)(x-y)+\phi(y) \le \phi(x)$, pour tous $x, y \in \mathbb{R}$. En prenant $x = \mathbb{E}[X|\mathcal{G}]$ et $y = X$ on obtient $\phi'_d(\mathbb{E}[Y|\mathcal{G}])(X - \mathbb{E}[X|\mathcal{G}]) \le \phi(X) - \phi(\mathbb{E}[X|\mathcal{G}])$. Comme dans la preuve de l'inégalité de Jensen avec l'espérance, on a envie de prendre des deux côtés l'espérance conditionnelle par rapport à $\mathcal{G}$; sauf qu'ici $\phi'_d(\mathbb{E}[X|\mathcal{G}])$ n'est pas constant mais c'est une variable aléatoire et il n'est pas certain que $\phi'_d(\mathbb{E}[X|\mathcal{G}])(X - \mathbb{E}[X|\mathcal{G}])$ soit intégrable. L'astuce est donc remplacer X par $\mathbf{1}_E X$, où $E = \{\omega : \mathbb{E}[X|\mathcal{G}](\omega) \le M\}$ pour $M < \infty$. $\mathbb{E}[X|\mathcal{G}]$ est $\mathcal{G}$-mesurable, donc $E \in \mathcal{G}$, donc $\mathbf{1}_E$ est $\mathcal{G}$-mesurable, d'où $\mathbb{E}[\mathbf{1}_E X|\mathcal{G}] = \mathbf{1}_E \mathbb{E}[X|\mathcal{G}]$, et donc $\mathbb{E}[\phi(\mathbf{1}_E X)|\mathcal{G}] = \mathbb{E}[\phi(X)\mathbf{1}_E + \phi(0)\mathbf{1}_{E^c}|\mathcal{G}] = \mathbf{1}_E \mathbb{E}[\phi(X)|\mathcal{G}] + (1 - \mathbf{1}_E)\phi(0)$.

On a l'inégalité $\phi'_d(\mathbb{E}[\mathbf{1}_E X|\mathcal{G}])(\mathbf{1}_E X - \mathbf{1}_E \mathbb{E}[X|\mathcal{G}]) \le \phi(\mathbf{1}_E X) - \phi(\mathbf{1}_E \mathbb{E}[X|\mathcal{G}])$ où le membre de droite est intégrable grâce à $\phi(X) \in L^1(\Omega)$, donc le membre de gauche est intégrable aussi. D'où en prenant l'espérance conditionnelle par rapport à $\mathcal{G}$ et en utilisant le fait que $\phi'_d(\mathbb{E}[\mathbf{1}_E X|\mathcal{G}])$, $\phi(\mathbf{1}_E \mathbb{E}[X|\mathcal{G}])$ (ϕ'_d est croissante et ϕ est continue, donc toutes les deux sont mesurables) sont $\mathcal{G}$-mesurables, on obtient $\phi(\mathbf{1}_E \mathbb{E}[X|\mathcal{G}]) \le \mathbf{1}_E \mathbb{E}[\phi(X)|\mathcal{G}] + (1 - \mathbf{1}_E)\phi(0)$, p.s. Finalement, en prenant $M \to \infty$, on a $\mathbf{1}_E \to 1$ p.s. et la preuve devient complète.

Fonctions Convexes :
Pour tous $x, y \in \mathbb{R}$, pour tout $t \in [0,1]$,

$$\phi(tx + (1-t)y) \le t\phi(x) + (1-t)\phi(y).$$

$$x_1 < x_2 < x_3 \Rightarrow k(x_2, x_1) \le k(x_3, x_1) \le k(x_3, x_2)$$

où $k(u,v) = \frac{f(u)-f(v)}{u-v}$

$$\phi'_d(a) = \lim_{t \to 0^+} k(a+t, a).$$

ϕ est dérivable à gauche et à droite.

11. Pour $p \in [1, \infty[$, inégalité de Jensen conditionnelle appliquée avec $\phi(x) = |x|^p$ donne l'inégalité.
12. Par la monotonicité, on a $\mathbb{E}[X_n|\mathcal{G}] \nearrow \xi \in L^0_+(\Omega)$, p.s. Le théorème de convergence monotone implique que, pour tout $A \in \mathcal{G}$,
$$\mathbb{E}[\xi \mathbf{1}_A] = \lim_n \mathbb{E}[\mathbf{1}_A \mathbb{E}[X_n|\mathcal{G}]] = \lim_n \mathbb{E}[\mathbf{1}_A X_n] = \mathbb{E}[\mathbf{1}_A X].$$
13. Soit $Y_n = \inf_{k \ge n} X_k$, alors $Y_n \nearrow Y = \liminf_k X_k$. Par la monotonicité,
$$\mathbb{E}[Y_n|\mathcal{G}] \le \inf_{k \ge n} \mathbb{E}[X_k|\mathcal{G}] \text{ p.s.},$$
en effet pour tout $k \ge n$, $Y_n \le X_k$ et donc $\mathbb{E}[Y_n|\mathcal{G}] \le \mathbb{E}[X_k|\mathcal{G}]$ et l'union dénombrable d'événements négligeables est négligeable ; finalement le théorème de convergence monotone conditionnel implique que
$$\begin{aligned}\mathbb{E}[\liminf_n X_n|\mathcal{G}] = \mathbb{E}[Y|\mathcal{G}] = \mathbb{E}[\lim_n Y_n|\mathcal{G}] = \lim_n \mathbb{E}[Y_n|\mathcal{G}] \\ \le \lim_n \inf_{k \ge n} \mathbb{E}[X_k|\mathcal{G}] = \liminf_n \mathbb{E}[X_n|\mathcal{G}].\end{aligned}$$
14. Comme $X_n + Z$ et $Z - X_n$ sont des v.a. positives pour tout $n \in \mathbb{N}^*$, le lemme de Fatou conditionnel appliqué donne $\mathbb{E}[X + Z|\mathcal{G}] = \mathbb{E}[\liminf_n (X_n + Z)|\mathcal{G}] \le \liminf_n \mathbb{E}[X_n + Z|\mathcal{G}]$ et $\mathbb{E}[Z - X_n|\mathcal{G}] = \mathbb{E}[\liminf_n (Z - X_n)|\mathcal{G}] \le \liminf_n \mathbb{E}[Z - X_n|\mathcal{G}]$. Alors on obtient $\liminf_n \mathbb{E}[X_n|\mathcal{G}] \ge \mathbb{E}[X|\mathcal{G}]$ et $\limsup_n \mathbb{E}[X_n|\mathcal{G}] \le \mathbb{E}[X|\mathcal{G}]$.

□

Propriété 3 *Soient X et Y deux variables aléatoires réelles telles que (X,Y) admettent une densité jointe f. Soit $g : \mathbb{R} \to \mathbb{R}$ mesurable (i.e. borélienne) tel que $g(X) \in L^1(\Omega)$. Alors*

$$\mathbb{E}[g(X)|Y] = h(Y) \text{ où } h(y) = \frac{\int_{\mathbb{R}} g(x)f(x,y)\,dx}{f_Y(y)} \mathbf{1}_{\{y\in\mathbb{R}\,:\,f_Y(y)>0\}}.$$

c'est-à-dire que $f : \mathbb{R}^2 \to \mathbb{R}_+$ est mesurable et

$f \in L^1(\mathbb{R}^2)$, $\forall B \in \mathcal{B}(R^2)$, $\mathbb{P}((X,Y) \in B) = \int_B f(z)\,dz$.

Rappelons que

$$(x,y) \mapsto \frac{f(x,y)}{f_Y(y)} \mathbf{1}_{\{y\in\mathbb{R}\,:\,f_Y(y)>0\}}$$

est appelée la fonction de densité conditionnelle de X par rapport à Y

Preuve: Supposons que $g \geq 0$ p.p. Comme Y est $\sigma(Y)$-mesurable et h est positive et borélienne (en effet $y \mapsto \frac{1}{f_Y(y)} \mathbf{1}_{\{y\in\mathbb{R}\,:\,f_Y(y)>0\}}$ est mesurable et $g(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ implique que d'après le théorème de Fubini on a $y \mapsto \int_{\mathbb{R}} g(x)f(x,y)\,dx$ mesurable), d'où $h(Y)$ est $\sigma(Y)$-mesurable, il nous reste donc à montrer que $\mathbb{E}[h(Y)\mathbf{1}_A] = \mathbb{E}[g(X)\mathbf{1}_A]$ pour tout $A \in \sigma(Y)$.

Soit $A \in \sigma(Y)$. Alors il existe $B \in \mathcal{B}(\mathbb{R})$ tel que $A = \{Y \in B\}$. Soit $C = \{y \in \mathbb{R} : f_Y(y) > 0\}$.

$(x,y) \mapsto g(x)\mathbf{1}_B(y)f(x,y)$ est mesurable et $g(X) \in L^1(\Omega) \Rightarrow g(X)\mathbf{1}_A = g(X)\mathbf{1}_B(Y) \in L^1(\Omega)$, d'où d'après le théorème de Fubini $\mathbb{E}[g(X)\mathbf{1}_A] = \int_B (\int_{\mathbb{R}} g(x)f(x,y)\,dx)\,dy = \int_{B\cap C}(\int_{\mathbb{R}} g(x)f(x,y)\,dx)\,dy + \int_{B\cap C^c}(\int_{\mathbb{R}} g(x)f(x,y)\,dx)\,dy$. Comme pour tout $y \in C$, $h(y)f_Y(y) = \int_{\mathbb{R}} g(x)f(x,y)\,dx$ et pour tout $y \in C^c$, $f(\cdot,y) = 0$ λ-p.p., λ étant la mesure de Lebesgue ; ce qui implique que pour tout $y \in C^c$, $\int_{\mathbb{R}} g(x)f(x,y)\,dx = 0$. Par conséquent, $\mathbb{E}[g(X)\mathbf{1}_A] = \int_{B\cap C} h(y)f_Y(y)\,dy = \int_{\mathbb{R}}\int_{\mathbb{R}} h(y)\mathbf{1}_A(y)f(x,y)\,dx\,dy$, comme $y \mapsto h(y)\mathbf{1}_A(y)f(x,y)$ est une fonction positive et mesurable d'où d'après le Théorème de Tonelli, la dernière double intégrale vaut $\int_{\mathbb{R}\times\mathbb{R}} h(y)\mathbf{1}_A(y)f(x,y)\,d\lambda_2(x,y)$ et elle est finie et donc $\mathbb{E}[g(x)\mathbf{1}_A] = \mathbb{E}[h(Y)\mathbf{1}_A]$. Finalement $\mathbb{E}[g(X)|Y] = h(Y)$.

Pour le cas où g n'est pas forcément positive p.p., on écrit $g = g^+ - g^-$ où $x^+ = x\mathbf{1}_{x>0}$ et $x^- = -x\mathbf{1}_{x\geq 0}$ et on pose

$$h(y) = \frac{\int_{\mathbb{R}} g(x)f(x,y)\,dx}{f_Y(y)} \mathbf{1}_{\{y\in\mathbb{R}\,:\,f_Y(y)>0\}}$$

qui est bien mesurable d'après le théorème de Fubini. Comme $g(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ on a l'existence de $\mathbb{E}[g(x)|Y]$. En utilisant les résultats précédents et la linéarité de l'espérance conditionnelle, on obtient donc

$$\begin{aligned}\mathbb{E}[g(X)|Y] &= \mathbb{E}[g^+(X) - g^-(X)|Y] \\ &= \mathbb{E}[g^+(X)|Y] - \mathbb{E}[g^-(X)|Y] \\ &= h^{g^+}(Y) - h^{g^-}(Y) = h(Y)\end{aligned}$$

où $h^{g^\pm} = \left(\int_{\mathbb{R}} g^\pm(x)f(x,y)\,dx\right)(f_Y(y))^{-1}\mathbf{1}_{\{y\in\mathbb{R}\,:\,f_Y(y)>0\}}$. □

Propriété 4 *Soient X et Y deux variables aléatoires réelles indépendantes sur $(\Omega, \mathcal{F}, \mathbb{P})$. Soit $\mathcal{G} \subset \mathcal{F}$ une sous-tribu. Alors, pour toute fonction $h : \mathbb{R}^2 \to \mathbb{R}$ borélienne, positive ou telle que $h(X,Y) \in L^1(\Omega)$ (par exemple h bornée donne cela) on a*

$$\mathbb{E}[h(X,Y)|X] = g(X), \text{ où } g(x) = \mathbb{E}[h(x,Y)].$$

Preuve: On doit prouver que $g(X)$ est une version de l'espérance conditionnelle de $h(X,Y)$. On a $g(x) = \int_{\mathbb{R}} h(x,y)\, d\mathbb{P}_Y(y)$ et par le théorème de Fubini, g est borélienne et donc $g(Y)$ est $\sigma(Y)$-mesurable.

Soit $A \in \sigma(X)$, donc il existe $B \in \mathcal{B}(\mathbb{R})$ tel que $A = \{X \in B\}$. On a $\int_A h(X,Y)\, d\mathbb{P} = \int_{B\times\mathbb{R}} h(x,y)\, d\mathbb{P}_{(X,Y)}(x,y)$ et donc par le théorème de Fubini et du fait que $\mathbb{P}_{(X,Y)} = \mu_X \otimes \mathbb{P}_Y$, on obtient

$$\begin{aligned}\int_{B\times\mathbb{R}} h(x,y)\, d\mathbb{P}_{(X,Y)}(x,y) &= \int_B \int_{\mathbb{R}} h(x,y)\, d\mathbb{P}_Y(y)\, d\mu_X(x) \\ &= \int_B g(x)\, d\mu_X(x) = \mathbb{E}[g(x)\mathbf{1}_A].\end{aligned}$$

Ce qui donne

$$\int_A h(X,Y)\, d\mathbb{P} = \int_A g(X)\, d\mathbb{P}.$$

□

0.9 L'espace $L^2(\Omega, \mathcal{F}, \mathbb{P})$

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.

On définit l'espace $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ par $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}) = \{X : \Omega \to \mathbb{R}, \mathcal{F} - \text{mesurable} : \mathbb{E}_{\mathbb{P}}[X^2] < \infty\}$ que l'on note par $\mathcal{L}^2(\mathcal{F})$. On utilise l'application $\|\cdot\|_2$ définie de $\mathcal{L}^2(\mathcal{F})$ dans $\mathbb{R}_+$ par $\|\cdot\|_2 : X \to \mathbb{E}[X^2]^{1/2}$ pour définir une norme. Nous devrions identifier les variables aléatoires $\mathbb{P}$-presque-sûrement égales, comme $\|X - Y\|_2 = 0 \Leftrightarrow X = Y\ \mathbb{P} - \text{p.s.}$

Pour cela on définit $X \sim Y$ si et seulement si $X = Y$ $\mathbb{P}$-p.s. ; cela est une relation d'équivalence sur $\mathcal{L}^2(\mathcal{F})$ et donc on peut définir l'espace quotient $\mathcal{L}^2(\mathcal{F})/\sim$ que l'on note par $L^2(\mathcal{F})$.

Remarquons $L^2(\mathcal{F})$ est un espace préhilbertien, c'est-à-dire que cet espace vectoriel (voir le paragraphe ci-dessous) de dimension infinie est muni d'un produit scalaire. Le produit scalaire dans cet espace n'est autre que $\langle X, Y\rangle := \mathbb{E}[XY]$

cette application $\langle\cdot,\cdot\rangle$ est bien un produit scalaire, en effet elle est
- symétrique : $\langle X, Y\rangle = \langle Y, X\rangle$,
- linéaire par rapport à l'une des deux variables (donc avec la symétrie elle devient bilinéaire),
- définie : $\langle X, X\rangle = 0 \Rightarrow X = 0$,
- positive : $\langle X, X\rangle \geq 0$.

On a que $(L^2(\mathcal{F}), \|\cdot\|_2)$ est un espace vectoriel normé, en effet si X et Y sont des variables aléatoires, alors $X + Y$ et cX le sont également, pour tout réel c ; comme $(cX)^2 = c^2X^2$ et $(X+Y)^2 \leq 2(X^2 + Y^2)$, il s'ensuit que $L^2(\mathcal{F})$ est un espace vectoriel et $\|cX\|_2 = |c|\|X\|_2$. Maintenant $\|X\|_2 = 0$ si et seulement si $X = 0$ (comme un élément de $L^2(\mathcal{F})$). Donc tout ce qui reste est de prouver qu'on a l'inégalité triangulaire pour cette application $\|\cdot\|_2$.

Note : on aurait pu tout simplement définir directement $\mathcal{L}^p(\mathcal{F})$ et donc $L^p(\mathcal{F})$ pour tout $p \geq 1$.

Premièrement, on a besoin d'une conséquence importante de l'inégalité de Jensen :

Propriété 5 (Inégalité de Hölder) *Soit $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$. Si $X \in L^p(\mathcal{F})$ et $Y \in L^q(\mathcal{F})$, alors $XY \in L^1(\mathcal{F})$ et*

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

Preuve: Le résultat est immédiat si p ou q est infini ou si XY est nulle presque sûrement, supposons que $1/r = 1/p + 1/q$ avec $0 < p, q < +\infty$ et que les deux réels $\|X\|_p$ et $\|Y\|_q$ sont non nuls et même, sans perte de généralité, que $\|X\|_p = \|Y\|_q = 1$ (par homogénéité). À partir de là :

En appliquant l'inégalité de Young $ab \leq \frac{a^{p'}}{p'} + \frac{b^{q'}}{q'}$ à $a = |X|^r$, $b = |Y|^r$, $p' = \frac{p}{r}$, $q' = \frac{q}{r}$ (on a bien $\frac{1}{p'} + \frac{1}{q'} = 1$), on obtient, pour presque tout $\omega \in \Omega$, $|X(\omega)Y(\omega)|^r \leq \frac{1}{p'}|X(\omega)|^p + \frac{1}{q'}|Y(\omega)|^q$ (avec égalité si et seulement si $|X|^p = |Y|^q$ presque sûrement) et, après intégration,

$$\|XY\|_r^r \leq \frac{1}{p'}\|X\|_p^p + \frac{1}{q'}\|Y\|_q^q = \frac{1}{p'} + \frac{1}{q'} = 1.$$

On a donc bien

$$\|XY\|_r \leq 1 = \|X\|_p \|Y\|_q,$$

avec égalité si et seulement si $|X|^p = |Y|^q$ presque sûrement.

Sachez que l'inégalité de Hölder se démontre avec $1/p + 1/q = 1/r$ également (à la place de $\|\cdot\|_1$ on aura $\|\cdot\|_r$ dans le membre gauche de l'inégalité) et se généralise immédiatement à n fonctions, par récurrence :

Soient $0 < r, p_1, \ldots, p_n \leq +\infty$ tels que $\sum_{k=1}^n \frac{1}{p_k} = \frac{1}{r}$ et n fonctions $f_k \in L^{p_k}(\mathcal{F})$. Alors, le produit des f_k appartient à $L^r(\mathcal{F})$ et

$$\left\| \prod_{k=1}^n f_k \right\|_r \leq \prod_{k=1}^n \|f_k\|_{p_k}.$$

De plus, lorsque tous les p_k sont finis, il y a égalité si et seulement si les $|f_k|^{p_k}$ sont colinéaires presque sûrement, c'est-à-dire s'il existe $a_1, \ldots, a_k$ non simultanément nuls tels que $a_1|f_1|^{p_1} = \ldots = a_k|f_k|^{p_k}$ presque sûrement. $\square$

On obtient un cas spécial et familier lorsque $p = q = 2$:

Corollaire 1 (Inégalité de Schwarz) *Soient $X, Y \in L^2(\mathcal{F})$, alors $XY \in L^1(\mathcal{F})$ et*

$$\|XY\|_1 \leq \|X\|_2 \|Y\|_2.$$

On peut donc prouver l'inégalité triangulaire pour $X \to \|X\|_p$ sur $L^p(\mathcal{F})$ pour tout $p \geq 1$.

Corollaire 2 (Inégalité de Minkowski) *Si $p \geq 1$ et $X, Y \in L^p(\mathcal{F})$, alors*

$$\|X+Y\|_p \leq \|X\|_p + \|Y\|p.$$

Preuve: Le cas $p = 1$ est trivial, donc supposons $p > 1$. On a

$$|X+Y|^p \leq |X||X+Y|^{p-1} + |Y||X+Y|^{p-1}$$

et avec $\frac{1}{p} + \frac{1}{q} = 1$ on obtient $|X+Y|^{(p-1)q} = |X+Y|^p$, et donc $|X+Y|^{p-1} \in L^q(\Omega)$. En appliquant l'inégalité de Hölder à chacun des produits on obtient

$$\int_\Omega |X||X+Y|^{p-1}\, d\mathbb{P} \leq C\|X\|_p,$$

où $C = \big(\int_\Omega (|X+Y|^{p-1})^q\, d\mathbb{P}\big)^{1/q} = \|X+Y\|_p^{p/q}$, et la même inégalité s'obtient en intervertissant X et Y. D'où

$$\|X+Y\|_p^p \leq C(\|X\|_p + \|Y\|_p).$$

En divisant les deux membres de cette dernière inégalité par C (se rappeler de $p - p/q = 1$), on a bien l'inégalité de Minkowski. □

Propriété 6 *La norme $L^p(\mathcal{F})$ précédente est croissante en p si $\mathbb{P}(\Omega)$ est fini (c'est le cas ici car $\mathbb{P}$ est une probabilité !), c'est-à-dire que*

$$1 \leq p \leq q < \infty \Rightarrow L^q(\mathcal{F}) \subset L^p(\mathcal{F}).$$

Preuve: On utilise l'inégalité de Hölder avec $r > 1$ et $\frac{1}{r} + \frac{1}{s} = 1$ pour avoir $\mathbb{E}[|YZ|] \leq (\mathbb{E}[|Y|^r])^{1/r}(\mathbb{E}[|Z|^s])^{1/s}$. En particulier, pour $Z = 1$ (constant, donc intégrable puisque $\mathbb{P}(\Omega) = 1 < \infty$) on obtient $\mathbb{E}[|Y|] \leq (\mathbb{E}[|Y|^r])^{1/r}$. D'où avec $1 \leq p \leq q < \infty$ et $r = \frac{q}{p}$, $Y = |X|^p$ on obtient $\mathbb{E}[|X|^p] \leq (\mathbb{E}[|X|^{pr}])^{1/r} = (\mathbb{E}[|X|^q])^{p/q}$; et donc prenant la racine p-ième il s'ensuit l'inégalité $\|X\|_p \leq \|X\|_q$. □

Remarque importante : Une fois que vous avez une v.a. telle que $\mathbb{E}[|X|^n] < \infty$ alors pour tout entier $k < n$ on a $\mathbb{E}[|X|^k] < \infty$.

Théorème 5 *$L^2(\mathcal{F})$ est un espace vectoriel complet.*

Preuve: La vérification sur le fait qu'il est bien un espace vectoriel est facile à faire et presque déjà faite auparavant.

Montrons qu'il est complet pour la topologie induite par la norme. Soit $(X_n)_{n\in\mathbb{N}^*} \in (L^2(\mathcal{F}))^{\mathbb{N}^*}$ une suite de Cauchy, montrons donc qu'elle converge.

On a $\sup_{m,n>k} \|X_m - X_n\|_2 \to 0$ pour $k \to \infty$. On veut montrer qu'il existe $X \in L^2(\mathcal{F})$ (unique au sens de p.s.) tel que $\lim_{n\to\infty} \|X_n - X\|_2 = 0$. Soit $(k_n)_{n\in\mathbb{N}^*}$ une suite croissante telle que $\lim_{n\to\infty} k_n = +\infty$ et $\|X_s - X_m\|_2 \leq 2^{-n}$ pour tous $s, m \geq k_n$.

Alors par le théorème de convergence monotone sur la suite croissante de v.a. mesurables et positives puis par le théorème de Jensen ($\mathbb{E}[|X|]^2 \leq \mathbb{E}[|X|^2]$) on obtient $\left(\sum_{n=1}^{m} |X_{k_{n+1}} - X_{k_n}|\right)_{m\in\mathbb{N}^*}$, on obtient

$$\mathbb{E}\left[\sum_{n\in\mathbb{N}^*} |X_{k_{n+1}} - X_{k_n}|\right] \leq \sum_{n\in\mathbb{N}^*} \|X_{k_{n+1}} - X_{k_n}\|_2 \leq \sum_{n\in\mathbb{N}^*} 2^{-n} < \infty$$

et donc pour presque tout $\omega \in \Omega$ la série $S(\omega) = \sum_{n\in\mathbb{N}^*}(X_{k_{n+1}} - X_{k_n})$ est absolument convergente et donc $\lim_{n\to\infty} X_{k_n}(\omega) < \infty$ p.s.

Soit donc $X(\omega) = \limsup_n X_{k_n}(\omega)$, on a que X est $\mathcal{F}$-mesurable et que $\lim_{n\to\infty} X_{k_n} = X$ p.s. (en effet lorsque qu'une suite $(a_n)_n$ est convergente on a $\lim_n a_n = \limsup_n a_n = \liminf_n a_n$). Maintenant on observe que si $\ell \geq n$, $\mathbb{E}[|X_r - X_\ell|^2] \leq 2^{-2n}$ pour tout $r \geq k_n$, donc une application du Lemme de Fatou donne

$$\mathbb{E}[|X_r - X|^2] = \mathbb{E}[\lim_{n\to\infty} |X_r - X_{k_\ell}|^2] \leq \lim_{\ell\to+\infty} \mathbb{E}[|X_r - X_{k_\ell}|^2] \leq 2^{-2n}$$

pour tout $r \geq k_n$, qui montre que $X \in L^2(\mathcal{F})$ et que $X_n \underset{n\to\infty}{\longrightarrow} X$ dans $L^2(\mathcal{F})$. $\square$

Corollaire 3 (Projection Orthogonale) *Si $\mathcal{B} \subset \mathcal{F}$ est une sous-tribu de $\mathcal{F}$, alors $L^2(\mathcal{B})$ est un sous-espace vectoriel fermé de $L^2(\mathcal{F})$ et pour tout $X \in L^2(\mathcal{F})$ il existe une v.a. $Y \in L^2(\mathcal{B})$ (unique au sens p.s.) qui satisfait à l'une des deux propriétés suivantes, qui sont équivalentes :*

a) $\|X - Y\|_2^2 = \inf_{Z\in L^2(\mathcal{B})} \|X - Z\|_2^2$

b) $X - Y \perp L^2(\mathcal{B})$, *i.e.* $\mathbb{E}[(X - Y)Z] = 0$ *pour tout* $Z \in L^2(\mathcal{B})$.

Preuve: Par le Théorème 5, l'ensemble $L^2(\mathcal{B})$ est complet par la norme L^2 et donc fermé de $L^2(\mathcal{F})$; en effet $(L^2(\mathcal{F}), \|\cdot\|_2)$ est un espace vectoriel normé donc un espace métrique (on peut définir par exemple la distance d définie par $d(X,Y) = \|X - Y\|_2$), et $L^2(\mathcal{B}) \subset L^2(\mathcal{F})$, d'où le sous-espace métrique $(L^2(\mathcal{B}), d)$ est complet, et donc $L^2(\mathcal{B})$ est un fermé de $L^2(\mathcal{F})$.

Soit $\delta = \inf_{Z\in L^2(\mathcal{B})} \|X - Z\|_2^2 < \infty$ et $(Y_n)_{n\in\mathbb{N}^*} \in (L^2(\mathcal{B}))^{\mathbb{N}^*}$ une suite minimisante : $\|X - Y_n\|_2^2 \to \delta$ lorsque $n \to \infty$. On a donc

$$\mathbb{E}[|X - Y_n|^2] + \mathbb{E}[|X - Y_m|^2] = 2\mathbb{E}[|X - (Y_n + Y_m)/2|^2] + \mathbb{E}[|Y_n - Y_m|^2]/2,$$

en effet $\mathbb{E}[|A + B|^2] + \mathbb{E}[|A - B|^2] = 2\mathbb{E}[|A|^2] + 2\mathbb{E}[|B|^2]$ avec $A = X - (Y_n + Y_m)/2$ et $B = (Y_n - Y_m)/2$ donne cela. Mais $(Y_n + Y_m)/2 \in L^2(\mathcal{B})$ ce qui donne que

$$\mathbb{E}[|Y_n - Y_m|^2]/2 \leq \mathbb{E}[|X - Y_n|^2] + \mathbb{E}[|X - Y_m|^2] - 2\sqrt{\delta} \underset{n,m\to\infty}{\longrightarrow} 0$$

cette suite minimisante $(Y_n)_{n\in\mathbb{N}^*}$ existe toujours par construction qui vient de la définition de inf ; en effet par définition on a $\delta \leq \|X - Z\|_2^2$ pour tout $Z \in L^2(\mathcal{B})$, donc pour tout $n \in \mathbb{N}^*$, il existe $Y_n \in L^2(\mathcal{B})$ tel que $\delta \leq \|X - Y_n\|_2^2 \leq \delta + \frac{1}{n}$, car sinon on aurait $\delta + \frac{1}{n} \leq \|X - Z\|_2^2$ pour tout $Z \in L^2(\mathcal{B})$, ce qui contredit $\inf_{Z\in L^2(\mathcal{B})} \|X - Z\|_2^2$.

et donc $(Y_n)_{n\in\mathbb{N}^*}$ est une suite de Cauchy. D'où avec $L^2(\mathcal{B})$ complet et fermé on a $Y := \lim_{n\to\infty} Y_n \in L^2(\mathcal{B})$. On a que $\|X - Y\|_2 \le \|X - Y_n\|_2 + \|Y_n - Y\|_2$ (inégalité triangulaire) et donc que

$$\|X - Y\|_2 \le \sqrt{\delta} = \inf_{Z\in L^2(\mathcal{B})} \|X - Z\|_2 \le \|X - Y\|_2$$

(on a utilisé le fait que $\|Y_n - Y\|_2 \to 0$ lorsque $n \to \infty$), d'où $\|X - Y\|_2 = \sqrt{\delta}$.

Pour tous $t \in \mathbb{R}$ et $Z \in L^2(\mathcal{B})$ on a que $Y + tZ \in L^2(\mathcal{B})$ et

$$0 \le \mathbb{E}[|X - Y - tZ|^2] - \mathbb{E}[|X - Y|^2] = -2t\mathbb{E}[(X - Y)Z] + t^2\mathbb{E}[Z^2].$$

Le polynôme $P(t) = at^2 + bt$ satisfait à $P(t) \ge 0$ pour tout $t \ge 0$, où $a = \mathbb{E}[Z^2] \ge 0$ et $b = \mathbb{E}[(X - Y)Z]$, donc on doit avoir $b = 0$, c'est-à-dire $\mathbb{E}[(X - Y)Z] = 0$. L'implication réciproque est facile à établir également. Pour montrer l'unicité presque sûre de Y, on suppose que Y' est un autre projection orthogonale. On a $\mathbb{E}[(Y - Y')Z] = 0$ pour tout $Z \in L^2(\mathcal{B})$ et donc aussi pour $Z = Y - Y' \in L^2(\mathcal{B})$, mais alors $\mathbb{E}[(Y - Y')^2] = 0$, i.e. $Y - Y' = 0$ p.s. □

Exemple 13 (Théorème de Pythagore) *Soient $\mathcal{G} \subset \mathcal{F}$, $X \in L^2(\mathcal{F})$, $Z \in L^2(\mathcal{G})$ et $Y = \mathbb{E}[X|\mathcal{G}]$ $\big(Y$ existe car $L^1(\mathcal{F}) \subset L^2(\mathcal{F})\big)$. Montrer que $\mathbb{E}[|X - Z|^2] = \mathbb{E}[|X - Y|^2] + \mathbb{E}[|Y - Z|^2]$ et en déduire que $\mathbb{E}[|X - Y|^2] = \inf_{Z\in L^2(\mathcal{G})} \mathbb{E}[|X - Z|^2]$.*

Cet exercice montre que l'espérance conditionnelle $\mathbb{E}[X|\mathcal{G}]$ dans $L^2(\mathcal{F})$ est le meilleur estimateur $\mathcal{G}$-mesurable de X selon le risque quadratique, à savoir :

$$\mathbb{E}[|X - \mathbb{E}[X|\mathcal{G}]|^2] \le \mathbb{E}[|X - Z|^2],\ \forall Z \in L^2(\mathcal{G}).$$

En effet cette interprétation géométrique est à la base d'une stratégie pour montrer l'existence de l'espérance conditionnelle dans $L^2(\mathcal{G})$:

Soit $X \in L^2(\mathcal{F})$ et $\mathcal{B}$ une sous-tribu de $\mathcal{F}$. Alors la projection orthogonale Y de X sur $L^2(\mathcal{B})$ satisfait à $\mathbb{E}[(X - Y)Z] = 0$ pour tout $Z \in L^2(\mathcal{B})$. On a donc, en prenant $A \in \mathcal{B}$, $Z = \mathbf{1}_A \in L^2(\mathcal{B})$ et $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Y\mathbf{1}_A]$. Comme Y est $\mathcal{G}$-mesurable d'où $Y = \mathbb{E}[X|\mathcal{B}]$ ce qui montre l'existence de l'espérance conditionnelle pour $X \in L^2(\mathcal{F})$. Finalement lorsque $X \in L^2(\mathcal{F})$, la projection orthogonale de X sur $L^2(\mathcal{B})$ est exactement l'espérance conditionnelle de X par rapport à la tribu $\mathcal{B}$. Voici donc une autre version de l'existence de l'espérance conditionnelle d'une variable aléatoire intégrable, en utilisant le résultat précédent :

Théorème 6 *Pour tout $X \in L^1(\mathcal{F})$ l'espérance conditionnelle $\mathbb{E}[X|\mathcal{B}]$ existe et appartient à $L^1(\mathcal{B})$.*

Preuve: Pour étendre l'existence précédente (lorsque $X \in L^2(\mathcal{F})$) à toute variable aléatoire $X \in L^1(\mathcal{F})$ on procède par approximation. Soit $X \geq 0$ p.s. et dans $L^1(\mathcal{F})$.

On pose $X_n(\omega) = \min(X(\omega), n)$, d'où $X_n \in L^2(\mathcal{B})$ et Y_n la projection orthogonale correspondante sur $L^2(\mathcal{B})$ i.e. $Y_n = \mathbb{E}[X_n|\mathcal{B}]$. Alors pour tout $n \geq m$ on a que $0 \leq \mathbb{E}[\mathbf{1}_A(X_n - X_m)] = \mathbb{E}[\mathbf{1}_A(Y_n - Y_m)]$ pour tout $A \in \mathcal{B}$, ce qu'implique que $Y_n \geq Y_m$ p.s. (vérifier) et qu'il existe un ensemble de mesure nulle $N \in \mathcal{B}$ en dehors duquel la suite $(Y_n(\omega))_n$ est croissante pour tout $\omega \in N^c$.

Soit $Y = \lim_n Y_n$. On a que $\mathbb{E}[\mathbf{1}_A Y] = \lim_n \mathbb{E}[\mathbf{1}_A Y_n] = \lim_n \mathbb{E}[\mathbf{1}_A X_n] = \mathbb{E}[\mathbf{1}_A X]$ par convergence monotone et donc que $Y \in L^1(\mathcal{B})$ et que $Y = \mathbb{E}[X|\mathcal{B}]$.

• La première égalité vient du théorème de la convergence monotone (en effet $(\mathbf{1}_A Y_n)_n$ est une suite croissante p.s. de variables aléatoires positives.)

• La deuxième égalité vient de l'espérance conditionnelle (rappelons que $Y_n = \mathbb{E}[X_n|\mathcal{G}]$).

• La première égalité vient du théorème de la convergence monotone (en effet $(\mathbf{1}_A X_n)_n$ est une suite croissante p.s. de variables aléatoires positives et $\lim_n X_n = \lim_n \min(X, n) = X$ p.s.)

Pour une v.a. X tel que $X \in L^1$: soit $X = X_+ - X_-$ avec $X_+, X_- \geq 0$ et dans L^1. On pose $Y_\pm = \mathbb{E}[X_\pm|\mathcal{B}]$ et $Y = Y_+ - Y_-$. On obtient que $Y \in L^1(\mathcal{B})$ et que $\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Y]$ pour tout $A \in \mathcal{B}$.

Rappel : L'unicité au sens presque sûre de l'espérance conditionnelle a été démontrée en cours, voir la preuve qui est juste après le Lemme 1. □

0.10 *Rappels sur les Variables aléatoires vectorielles*

Notations :

- On note, pour $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $\|x\| = (x_1^2 + \ldots + x_d^2)^{1/2}$.
- On note M' la matrice transposée de la matrice M. On peut représenter $x \in \mathbb{R}^d$ par un vecteur colonne i.e. une matrice $d \times 1$ et on écrira indifféremment $x = (x_1, \ldots, x_d)$ ou $x = (x_1, \ldots, x_d)'$. Pour $x = (x_1, \ldots, x_d)'$ et $y = (y_1, \ldots, y_d)'$, on a $x'y = x_1y_1 + \ldots + x_dy_d = \langle x, y\rangle$ et xy' est la matrice de terme générale x_iy_j.
- On note $L_d^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X = (X_1, \ldots, X_d) \,:\, \mathbb{E}[|X|^p] < +\infty\}$.
- Si $X \in L_d^1$, on note $\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$.

Par définition $X = (X_1, \ldots, X_d)'$ est un vecteur aléatoire si et seulement si, pour tout $k \in \{1, \ldots, d\}$, X_k est une v.a.r. Ceci entraîne :

Propriété 7 (Doob-Dynkin) *Soient $X_1, \ldots, X_d$ d v.a.r. et $\mathcal{F}_d = \sigma(X_1, \ldots, X_d)$. Une v.a.r. Y est $\mathcal{F}_d$-mesurable si et seulement si $Y = f(X_1, \ldots, X_d)$ avec f fonction borélienne sur $\mathbb{R}^d$.*

Preuve: Admise. □

Vous aviez vu la définition de la densité d'une mesure de probabilité P définie sur $(\mathbb{R}^n, \mathcal{B}^n)$ avec $n = 1$. Voici la définition lorsque $n \geq 1$:

Définition 18 *On dit qu'une mesure de probabilité $\mathbb{P}$ sur $(\mathbb{R}^n, \mathcal{B}^n)$ a une densité de probabilité f si f est une fonction positive Borel mesurable sur $\mathbb{R}^n$ vérifiant*

$$\begin{aligned} P(A) &= \int_A f(x)\,dx = \int f(x)\mathbf{1}_A(x)\,dx & (7)\\ &= \int f(x_1,\ldots,x_n)\mathbf{1}_A(x_1,\ldots,x_n)\,dx_1,\ldots dx_n & (8)\end{aligned}$$

pour tout $A \in \mathcal{B}^n$.

En dimension 1 vous aviez vu la définition de l'espérance mathématique d'une variable aléatoire et via le théorème de transfert une autre formule permettant d'obtenir l'espérance : On considère un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, et une variable aléatoire réelle X définie sur $(\Omega, \mathcal{F}, \mathbb{P})$, de loi μ_X.

Théorème 7 (Théorème de transfert) *Soit ϕ une fonction borélienne définie sur $\mathbb{R}$. $\phi(X)$ est $\mathbb{P}$-intégrable si et seulement si ϕ est μ_x-intégrable, et l'on a :*

$$\int_\Omega \phi(X(\omega))\,d\mathbb{P}(\omega) = \int_\mathbb{R} \phi(x)\,d\mu_X(x).$$

Définition 19 (Espérance mathématique) *Si X est à valeurs positives ou si X est $\mathbb{P}$-intégrable, on appelle espérance mathématique de X, notée $\mathbb{E}[X]$, l'intégrale de X par rapport à $\mathbb{P}$:*

$$\mathbb{E}[X] = \int_\Omega X(\omega)\,d\mathbb{P}(\omega) \overset{Th.\,de.tr.}{=} \int_\mathbb{R} x\,d\mu_X(x).$$

Et voici le théorème de transfert en multidimensionnelle :

On considère un vecteur aléatoire X de dimension n défini sur $(\Omega, \mathcal{F}, \mathbb{P})$ de loi conjointe μ_X.

Théorème 8 (Théorème de transfert bis) *Soit $\phi : \mathbb{R}^n \to \mathbb{R}$ une fonction borélienne. Alors $\phi(X)$ est une variable réelle $\mathbb{P}$-intégrable si et seulement si ϕ est μ_X-intégrable, et l'on a :*

$$\mathbb{E}[\phi(X)] = \int_\Omega \phi(X(\omega))\,d\mathbb{P}(\omega) = \int\cdots\int_{\mathbb{R}^n} \phi(x_1,\ldots,x_n)\,d\mu_X(x_1,\ldots,x_n).$$

Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire. Les lois $\mu_{X_1}, \ldots, \mu_{X_d}$ s'appellent les lois marginales de X. On sait que les composants $X_1, \ldots, X_d$ sont indépendantes si et seulement si $\mu_X = \mu_{X_1} \otimes \cdots \otimes \mu_{X_d}$. Si X a une densité h, on a immédiatement :

Propriété 8 *Soit X un vecteur aléatoire de densité h. Les composants $X_1, \dots, X_d$ sont indépendants si et seulement si $h(x_1, \dots, x_d) = \prod_{j=1}^{d} g_j(x_j)$ p.p. et alors X_k a pour densité $h_{X_k}(u) = g_k(u) / \int_{\mathbb{R}} g_k(v)\, dv$.*

Rappelons la formule de changement de variables dans $\mathbb{R}^d$ (pour une preuve voir le paragraphe IV.3.4 de [p684-p690 Ramis-Warusfel, L2]). Si ϕ est un C^1-difféomorphisme de l'ouvert U sur l'ouvert V, on a, pour toute fonction $f \in \mathcal{B}(\mathbb{R}^d)$ intégrable,

Si U et V sont deux ouverts de $\mathbb{R}^n$, et ϕ une fonction de U dans V, on dit que ϕ est un C^k-difféomorphisme si ϕ est bijective et si ϕ et ϕ^{-1} sont de classe C^k.

$$\int_V f(v)\, dv = \int_U f(\phi(u)) |\det(J(\phi)(u))|\, du$$

où $\det(J(\phi)(u))$ est le déterminant de la matrice Jacobienne de ϕ au point u, i.e. $J(\phi)(u) = \frac{\partial \phi_i}{\partial u_j}$.

Propriété 9 *Soit X un vecteur aléatoire de densité f_X. On suppose que $X \in D$ p.s., D ouvert de $\mathbb{R}^d$. Soient ψ un C^1- difféomorphisme de D sur un ouvert Δ et $Y = \psi(X)$, alors*

$$f_Y(y) = f_X(\psi^{-1}(y)) |\det(J(\psi^{-1})(y))| \mathbf{1}_\Delta(y).$$

Preuve: Soit g une fonction borélienne bornée. Alors $\mathbb{E}[g(Y)] < +\infty$. Or $g(Y) = g(\psi(X)) = (g \circ \psi)(X)$ avec X admettant une densité de probabilité, donc $(g \circ \psi) f_X$ est intégrable, d'où

$$\begin{aligned} \mathbb{E}[g(Y)] = \mathbb{E}[(g \circ \psi)(X)] &= \int_D g(\psi(x)) f_X(x)\, dx \\ &= \int_\Delta g(y) f_X(\psi^{-1}(y)) |\det(J(\psi^{-1})(y))|\, dy. \end{aligned}$$

□

Pour les exemples d'applications, voir les exercices des TDs.

0.10.1 Covariance

Soient X et Y deux v.a.r. de carré intégrable. On appelle covariance de X et Y la quantité

$$\mathbb{C}\text{ov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \quad (9)$$

$(X, Y) \mapsto \mathbb{C}\text{ov}(X, Y)$ est une forme bilinéaire et $\mathbb{C}\text{ov}(X, X) = \mathbb{V}\text{ar}(X)$. Si X et Y sont indépendantes, $\mathbb{C}\text{ov}(X, Y) = 0$. On appelle coefficient de corrélation de X et Y la quantité

$$\rho_{X,Y} = \frac{\mathbb{C}\text{ov}(X, Y)}{\sigma_X \sigma_Y}.$$

On a $|\rho_{X,Y}| \le 1$ et $|\rho_{X,Y}| = 1$ si et seulement si $aX + bY + c = 0$ p.s.

Plus généralement on pose, pour $X \in L^2_d(\Omega, \mathcal{F}, \mathbb{P})$

$$\mathbb{C}\text{ov}(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])'] = \mathbb{E}[XX'] - \mathbb{E}[X]\mathbb{E}[X]'.$$

$\mathbb{C}\text{ov}(X)$ s'appelle la matrice de covariances de X. On a $\mathbb{C}\text{ov}(X)_{i,j} = \mathbb{C}\text{ov}(X_i, X_j)$. Notons que, si les composantes $X_1, \ldots, X_d$ sont indépendantes, $\mathbb{C}\text{ov}(X)$ est diagonale.

Propriété 10 *Soit $X \in L^2_d(\Omega, \mathcal{F}, \mathbb{P})$. On a*

i) 1. *Pour tout $a \in \mathbb{R}$, $\mathbb{C}\text{ov}(aX) = a^2\mathbb{C}\text{ov}(X)$*

2. *Pour tout $b \in \mathbb{R}^d$, $\mathbb{C}\text{ov}(X + b) = \mathbb{C}\text{ov}(X)$.*

3. *$\mathbb{C}\text{ov}(X)' = \mathbb{C}\text{ov}(X)$.*

ii) Pour tout $\lambda \in \mathbb{R}^d$, $\lambda'\mathbb{C}\text{ov}(X)\lambda \geq 0$.

iii) Soit M une matrice déterministe $r \times d$, on a $\mathbb{C}\text{ov}(MX) = M\mathbb{C}\text{ov}(X)M'$.

Preuve:

i) Cela résulte de la formule (9) et de la définition de $\mathbb{C}\text{ov}(X)$.

ii) Vu i), on peut supposer $\mathbb{E}[X] = 0 \in \mathbb{R}^d$. Alors

$$\lambda'\mathbb{C}\text{ov}(X)\lambda = \lambda'\mathbb{E}[XX']\lambda = \mathbb{E}[\lambda' XX'\lambda] = \mathbb{E}[(\lambda' X)^2] \geq 0.$$

iii) Vu i), on peut supposer que $\mathbb{E}[X] = 0$. Alors

$$\mathbb{C}\text{ov}(MX) = \mathbb{E}[(MX)(MX)'] = M\mathbb{E}[XX']M' = M\mathbb{C}\text{ov}(X)M'.$$

Les points i) et ii) montrent que $\mathbb{C}\text{ov}(X)$ est symétrique semi-définie positive. □

Théorème 9 *Soient $X, Y \in L^2_d(\Omega, \mathcal{F}, \mathbb{P})$ deux vecteurs aléatoires indépendants, on a $\mathbb{C}\text{ov}(X + Y) = \mathbb{C}\text{ov}(X) + \mathbb{C}\text{ov}(Y)$. En particulier, si $d = 1$, $\mathbb{V}\text{ar}(X + Y) = \mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y)$ si les v.a.r. X et Y sont indépendantes.*

Preuve: On peut supposer que $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. Alors $\mathbb{C}\text{ov}(X+Y) = \mathbb{E}[(X+Y)(X+Y)'] = \mathbb{E}[XX'] + \mathbb{E}[YY']$ puisque, vu l'indépendance, $\mathbb{E}[XY'] = \mathbb{E}[X]\mathbb{E}[Y'] = 0$ et de même $\mathbb{E}[YX'] = 0$. □

Propriété 11 *Soit $X \in L^2_d(\Omega, \mathcal{F}, \mathbb{P})$. On a $X - \mathbb{E}[X] \in \text{Im}(\mathbb{C}\text{ov}(X))$ p.s.*

Rappel : pour une matrice M, $\text{Im}(M) = \{Mx, x \in \mathbb{R}^d\}$, i.e. l'ensemble des combinaisons linéaires des colonnes de la matrice M.

Théorème du Rang pour les matrices :

Si A est une matrice $m \times n$ sur un corps K, alors $\text{rg}(A) + \dim(\ker(A)) = d$ où $\text{rg}(A) = \dim(\text{Im}(A))$.

Preuve: Soit $M = \mathbb{C}\text{ov}(X)$. Comme toujours on peut supposer que $\mathbb{E}[X] = 0$. Soit $V = \text{Im}(M)$. Si $\dim(V) = d$, il n'y a rien à démontrer, en effet $\dim(\text{Ker}(M)) = 0 \Leftrightarrow M$ est inversible et donc $MY = X$ avec $Y = M^{-1}X$ d'où $X \in \text{Im}(M)$.

Supposons que $\dim(V) = \ell < d$; d'où $1 \leq d - l = \dim(\ker(M))$. Comme pour toute matrice A de taille $n \times m$ on a $\text{Im}(A) = (\ker(A'))^\perp$ et que $M = M'$ (car symétrique) d'où $(\ker(M))^\perp = \text{Im}(M)$. De plus, pour tout $u \in \ker(M)$, vu la Proposition 10,

$$\mathbb{E}[(u'X)^2] = \mathbb{V}\text{ar}(u'X) = \mathbb{C}\text{ov}(u'X) = u'Mu = 0$$

d'où $u'X = 0$ p.s. et $X \in \ker(M)^\perp$ p.s., i.e. $X \in \text{Im}(M)$. □

0.11 Vecteurs Gaussiens

Rappels : On dit qu'une probabilité μ sur $\mathbb{R}$ est gaussienne si elle a pour densité $f_{m,\sigma}$ où $f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-m)^2}$, i.e. $\mu(B) = \int_B f_{m,\sigma^2}(x) d\lambda(x)$ avec λ la mesure de Lebesgue sur $\mathbb{R}$ et $B \in \mathcal{B}(\mathbb{R})$; ou si $\mu = \delta_m$ (la mesure de Dirac, i.e. $\delta_m(B) = \mathbf{1}_B(m)$ pour tout $B \in \mathcal{B}(\mathbb{R})$).

Il est normal d'adjoindre les mesures de Dirac aux lois gaussiennes car la mesure $\mathcal{N}(m, \sigma^2)$ converge étroitement vers δ_m.

Rappel : on dit que la suite de mesures de probabilités $(\mu_n)_n$ converge étroitement vers μ si pour tout $f \in C_b^0(\mathbb{R}^d; \mathbb{R})$ $\lim_n \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f\, d\mu$; par exemple on dit (par définition) que $(X_n)_n$ converge en loi vers X ssi $(\mu_{X_n})_n$ converge étroitement vers μ_X. Une v.a.r. est dite gaussienne si sa loi est gaussienne.

Définition 20

1. *Un vecteur aléatoire $X = (X_1, \dots, X_d)$, à valeurs dans $\mathbb{R}^d$ défini sur $(\Omega, \mathcal{F}, \mathbb{P})$, est dit gaussien si, pour tout $a \in \mathbb{R}^d$,*

$$a'X = a_1 X_1 + \dots + a_d X_d \text{ est une v.a.r. gaussienne.}$$

2. *On appelle la loi gaussienne sur $\mathbb{R}^d$ toute loi d'un vecteur gaussien.*

Remarques et Exemples :

Pour tout $k \in \{1, \dots, d\}$, X_k est une v.a.r. gaussienne mais cela ne suffit pas à assurer que le vecteur X est gaussien ; en effet $X_k = a'X$ avec $a = (0, 0, \dots, 1, 0, 0, \dots, 0) \in \mathbb{R}^d$ où k-ème élément de a vaut 1. Par exemple $Y = 2X_3 - \pi X_5$ est aussi une v.a.r. gaussienne, en effet $Y = a'X$ avec $a = (0, 0, 2, 0, -\pi, 0, 0, \dots, 0) \in \mathbb{R}^d$.

Le vecteur aléatoire constant de $\mathbb{R}^d$, $X = 0$ est un vecteur gaussien ; en effet d'après la remarque sur la convergence étroite 0 est une v.a.r. gaussienne.

Propriété 12 *Soient $d \in \mathbb{N}^*$ et $X_1, X_2, \dots, X_d$ une suite de v.a.r. i.i.d. de loi $\mathcal{N}(0,1)$ et $X = (X_1, X_2, \dots, X_d)$. Alors $a'X \sim \mathcal{N}(0, a'a)$, et donc c'est un vecteur gaussien.*

Preuve: Comme la loi d'une variable aléatoire se caractérise par la détermination de sa fonction caractéristique (ou sa fonction de répartition ou sa fonction de densité) d'où montrons que pour tout $u \in \mathbb{R}$, $\phi_{a'X}(u) \stackrel{\text{déf}}{=} \mathbb{E}[e^{iu(a'X)}]$ est de la forme $e^{imu - \frac{1}{2}u^2\sigma^2}$ (i.e. la variable aléatoire est gaussienne d'espérance m et de variance σ^2).

Soit $u \in \mathbb{R}$. On a

$$\begin{aligned}
\phi_{a'X}(u) = \mathbb{E}[e^{iu(a'X)}] &= \mathbb{E}\left[\prod_{j=1}^{d} e^{iua_j X_j}\right] \\
&= \prod_{j=1}^{d} \mathbb{E}[e^{iua_j X_j}] \text{ (car les } X_j \text{ sont indépendants)} \\
&= \prod_{j=1}^{d} \phi_{X_j}(ua_j) = \prod_{j=1}^{d} e^{-\frac{1}{2}(ua_j)^2} = e^{-\frac{1}{2}u^2 a'a}.
\end{aligned}$$

Finalement $a'X$ est une variable aléatoire réelle gaussienne d'espérance 0 et de variance $a'a$; et donc X est un vecteur aléatoire *gaussien* à valeurs dans $\mathbb{R}^d$. $\square$

La notion de vecteurs gaussiens est invariante par transformation linéaire, plus précisément :

Propriété 13 *Soit X un vecteur gaussien à valeurs dans $\mathbb{R}^d$ de moyenne $m \in \mathbb{R}^d$ et de matrice de covariances K. Pour tous $b \in \mathbb{R}^r$ et M matrice $r \times d$, $Y = b + MX$ est un vecteur gaussien à valeurs dans $\mathbb{R}^r$ de moyenne $b + Mm$ et de matrice de covariance MKM'.*

Preuve: Soit $a \in \mathbb{R}^r$. On sait que pour tout $\tilde{a} \in \mathbb{R}^d$, $\tilde{a}'X$ est une v.a.r. gaussienne, donc en prenant $\tilde{a} = M'a \in \mathbb{R}^d$ on obtient que $\tilde{a}'X = (a'M)X$ est une v.a.r. gaussienne. Comme $a'b$ est une constante d'où $a'Y = a'b + (a'M)X$ est une v.a.r. gaussienne. Finalement Y est un vecteur gaussien à valeurs dans $\mathbb{R}^r$, de moyenne $\mathbb{E}[Y] = b + M\mathbb{E}[X] = b + Mm$ et de matrice de covariances $\mathbb{C}\text{ov}(Y) = \mathbb{C}\text{ov}(b + MX) = \mathbb{C}\text{ov}(MX) = M\mathbb{C}\text{ov}(X)M' = MKM'$. $\square$

Théorème 10 *Soit X un vecteur aléatoire de moyenne m et de matrice de covariances K. Le vecteur X est gaussien ssi sa fonction caractéristique est donnée par*

$$\phi_X(t) = e^{it'm - \frac{1}{2}t'Kt}, \forall t. \tag{10}$$

Preuve: Supposons que X est un vecteur gaussien. Alors d'après la Proposition 13 on a $t'X \sim \mathcal{N}(t'm, t'Kt)$ et $\phi_{t'X}(1) = \mathbb{E}[e^{it'X}] = e^{it'm - \frac{1}{2}t'Kt}$, d'où (10).

Supposons (10). Alors $\phi_{a'X}(u) = \mathbb{E}[e^{iua'X}] = e^{iua'm - \frac{1}{2}u^2 a'Ka}$ donc $a'X$ est une v.a.r. gaussienne et X est un vecteur gaussien. $\square$

Toute loi gaussienne sur $\mathbb{R}^d$ est donc déterminée par sa moyenne m et sa matrice de covariances K. On note $\mathcal{N}_d(m, K)$ une telle loi. On a vu dans la Propriété que $\mathcal{N}_d(0, I_d)$ existe, en effet on a vu un vecteur aléatoire X tel que $\phi_X(a) = e^{ia'0 - \frac{1}{2}a'I_d a}$, mais on n'a pas établi l'existence dans le cas général. On a,

Propriété 14 *Soit K une matrice $d \times d$ symétrique semi-définie positive. Il existe une matrice $d \times d$ symétrique semi-définie positive A telle que $K = AA'$.*

Preuve: Soient $\lambda_1, \ldots, \lambda_d$ les valeurs propres de K qui sont positives. Il existe une matrice orthogonale C (i.e. $CC' = C'C = I_d$, $C^{-1} = C'$) telle que $C'KC = D := \text{diag}(\lambda_1, \ldots, \lambda_d)$ où $\text{diag}(\lambda_1, \ldots, \lambda_d)$ désigne la matrice diagonale ayant $\lambda_1, \ldots, \lambda_d$ sur la diagonale. On a alors $CDC' = K$. Soit $\Delta = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d})$. On pose $A = C\Delta C'$. On a

$$AA' = C\Delta C'(C\Delta C')' = C\Delta CC'\Delta C' = CDC' = K.$$

□

Appliquant la Propriété 13, on a que, si $X \sim \mathcal{N}_d(0, I_d)$, $Y = m + AX \sim \mathcal{N}_d(m, K)$. On a montré

Théorème 11 *Étant donnés $m \in \mathbb{R}^d$ et une matrice $d \times d$ symétrique semi-définie positive K, il existe une et une seule loi gaussienne sur $\mathbb{R}^d$ de moyenne m et de matrice de covariance K.*

0.11.1 *Vecteurs gaussiens et Indépendance*

Théorème 12 *Soient $X = (X_1, \ldots, X_d)$ un vecteur gaussien. Alors on a,*

i) les v.a.r. $X_1, \ldots, X_d$ sont indépendantes si et seulement si la matrice de covariances $\mathbb{C}\mathrm{ov}(X)$ est diagonale.

ii) en posant

$$Y_1 = (X_1, \ldots, X_{d_1})', \; Y_2 = (X_{d_1+1}, \ldots, X_{d_2})', \ldots, Y_r = (X_{d_{r-1}}, \ldots, X_d)'$$

les vecteurs aléatoires $Y_1, Y_2, \ldots, Y_r$ sont indépendants si et seulement si $\mathbb{C}\mathrm{ov}(X)_{i,j} = 0$ pour tous i, j n'appartenant pas au même intervalle $[1, d_1], [d_1 + 1, d_2], \ldots, [d_{r-1} + 1, d]$.

Preuve: Seule la suffisance demande une preuve.

i) Supposons que $\mathbb{C}\mathrm{ov}(X)$ est diagonale. On a $\mathbb{C}\mathrm{ov}(X) = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2)$ où $\sigma_k^2 = \mathrm{Var}(X_k)$. Alors, notant $m = \mathbb{E}[X]$,

$$\phi_X(t) = e^{i\sum_{k=1}^d m_k t_k - \frac{1}{2}\sum_{k=1}^d \sigma_k^2 t_k^2} = \prod_{k=1}^d e^{i m_k t_k - \frac{1}{2}\sigma_k^2 t_k^2} = \phi X_1(t_1) \cdots \phi_{X_d}(t_d)$$

et donc les X_k sont indépendants.

ii) Supposons la condition sur les covariances réalisées. Elle implique, pour tous $u_1 \in \mathbb{R}^{d_1}$, $u_2 \in \mathbb{R}^{d_2 - d_1}$,... et $p \neq q$, $\mathbb{C}\mathrm{ov}(u_p' Y_p, u_q' Y_q) = 0$. Donc, d'après *i*), les v.a.r. $u_1' Y_1, \ldots, u_r' Y_r$ sont indépendantes. On a alors

$$\phi_{Y_1,\ldots,Y_r}(u_1, \ldots, u_r) = \mathbb{E}[e^{i(u_1' Y_1 + \ldots + u_r' Y_r)}] = \mathbb{E}[e^{i u_1' Y_1}] \cdots \mathbb{E}[e^{i u_r' Y_r}],$$

i.e. $\phi_{Y_1,\ldots,Y_r}(u_1, \ldots, u_r) = \prod_{k=1}^r \phi_{Y_k}(u_k)$ et donc les vecteurs aléatoires $Y_1, \ldots, Y_r$ sont indépendants. □

Remarque : Attention à l'utilisation du Théorème 12. On peut avoir X et Y v.a.r. gaussiennes, $\mathbb{C}\mathrm{ov}(X, Y) = 0$ sans que les v.a. X et Y soient indépendantes, par exemple prendre $Y = \varepsilon X$ où ε suit une loi de Bernoulli indépendante de X telle que $\mathbb{P}(\varepsilon = \pm 1) = 2^{-1}$; en effet si (X, Y) était un vecteur gaussien on aurait, avec $\alpha = 1$ et $\beta = 1$, $\alpha X + \beta Y = X + \varepsilon Y$ variable aléatoire gaussienne et donc par continuité de la variable aléatoire gaussienne $\mathbb{P}(X + \varepsilon Y = 0) = 0$, alors que $\mathbb{P}(X + \varepsilon Y) = \mathbb{P}((1 + \varepsilon)X = 0) = \mathbb{P}(1 + \varepsilon = 0) + \mathbb{P}(X = 0) - \mathbb{P}(1 + \varepsilon = 0, X = 0) = \frac{1}{2} + 0 - \mathbb{P}(1 + \varepsilon = 0)\mathbb{P}(X = 0) = \frac{1}{2} + 0 - \frac{1}{2} \times 0 = \frac{1}{2} \neq 0$.

Rappelons qu'une variable aléatoire X est continue de densité f_X si pour tout $A \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}(X \in A) = \int_A f_X(x)\,dx$.

0.11.2 *Le cas non dégénéré*

On dit que la loi $\mathcal{N}_d(m,K)$ est non dégénérée si $\det(K) \neq 0$. Dans ce cas :

Théorème 13 *Si $X \sim \mathcal{N}_d(m,K)$ et si $\det(K) \neq 0$, X admet la densité*

$$f_X(x) = (2\pi)^{-\frac{d}{2}}(\det(K))^{-\frac{1}{2}} e^{-\frac{1}{2}(x-m)'K^{-1}(x-m)}.$$

Preuve: Soit A tel que $K = AA'$, on a $\det(A) = (\det(K))^{\frac{1}{2}}$ et A est inversible. Soit $Y \sim \mathcal{N}_d(0, I_d)$ un vecteur gaussien de densité $(2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}\|y\|^2}$. On a, d'après la Propriété 13, $X = m + AY \sim \mathcal{N}_d(m,K)$ et, pour toute fonction $f \in \mathcal{B}^+(\mathbb{R}^d)$,

$$\mathbb{E}[f(X)] = \mathbb{E}[f(m+AY)] = (2\pi)^{-\frac{d}{2}} \int f(m+Ay) e^{-\frac{1}{2}\|y\|^2}\, dy.$$

On effectue le changement de variable $y = h^{-1}(x)$ où $x = h(y) := m + Ay$ qui est bien un C^1-difféomorphisme, et donc $y = A^{-1}(x-m)$, ce qui permet d'obtenir pour tous $i,j \in \{1,\dots,d\}, K_{i,j} := \frac{\partial (h^{-1})^{(i)}}{\partial x_j}(x) = \frac{\partial}{\partial x_j}\sum_{k=1}^d (L_i(A^{-1}))^k x_k = (L_i(A^{-1}))^j = (A^{-1})_{i,j}$, et donc $(K_{i,j})_{1\le i,j\le d} = A^{-1}$, ce qui permet d'écrire $\left|\det(K_{i,j})_{1\le i,j\le d}\right| = \det(A^{-1})$. D'où

$$\mathbb{E}[f(X)] = (2\pi)^{-\frac{d}{2}} \det(A^{-1}) \int_{\mathbb{R}^d} f(x) e^{-\frac{1}{2}(x-m)'(A^{-1})'A^{-1}(x-m)}\, dx.$$

Comme $K^{-1} = (AA')^{-1} = (A^{-1})'A^{-1}$ et donc $\det(A^{-1}) = (\det(K))^{-\frac{1}{2}}$, d'où la formule de la densité ci-dessus. $\square$

Bibliographie

M. Capinski and P.E. Kopp. *Measure, Integral and Probability*. Springer Undergraduate Mathematics Series. Springer London, 2013. ISBN 9781447106456. URL https://books.google.fr/books?id=5d6PBAAAQBAJ.

J. Jacod and P. Protter. *L'essentiel en théorie des probabilités*. Collection Enseignement des mathématiques. Cassini, 2003. ISBN 9782842250508. URL https://books.google.fr/books?id=GHolAAAACAAJ.

Index