Théorie des probabilités et processus aléatoires

Ecole Centrale de Lyon

18 janvier 2018

Table des matières

Pour compléter le cours, on pourra consulter :

Valérie Girardin et Nikolaos Limnios : Probabilités en vue des applications, tomes I et II, Vuibert 2008.

(ouvrage disponible à la bibliothèque en plusieurs exemplaires ainsi que sur scholarvox par un lien à partir du catalogue de la bibliothèque).

Chapitre 1

Evénements et probabilités.

Sommaire

1.1 Espaces mesurés.

1.1.1 Tribus.

Définition 1.1 (*Algèbre de Boole et tribu*)

Soit Ω un ensemble non vide.

(i) Une famille $\mathcal{A}$ de parties de Ω est appelée *algèbre de Boole* si

. $\Omega \in \mathcal{A}$

. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

. $A,\ B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.

(ii) Une famille $\mathcal{F}$ de parties de Ω est appelée *tribu sur* Ω si

. $\Omega \in \mathcal{F}$

. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

. $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

Quand $\mathcal{F}$ est une tribu, $(\Omega, \mathcal{F})$ s'appelle un *espace mesurable ou probabilisable*. Les éléments de $\mathcal{F}$ s'appellent des *ensembles mesurables ou événements*.

Remarque : $\varnothing \in \mathcal{A}$ et $\varnothing \in \mathcal{F}$.

Exemple : Les unions finies d'intervalles de $\mathbb{R}$ et les unions finies de pavés de $\mathbb{R}^n$ sont des algèbres de Boole.

Exemple :

. L'ensemble $\mathcal{P}(\Omega)$ des parties de Ω est une tribu sur Ω, dite la plus fine. Elle contient toutes les tribus sur Ω.

. $\{\varnothing, \Omega\}$ est la plus petite tribu sur Ω, appelée la tribu triviale. C'est la plus grossière et elle est contenue dans toutes les tribus sur Ω.

Exemple : Si $A \subset \Omega$ est strict (i.e. $A \neq \varnothing$ et $A \neq \Omega$), alors $\mathcal{F} = \{\varnothing, A, A^c, \Omega\}$ est une tribu. C'est la plus petite tribu contenant A.

Question : Si $\mathcal{F}$ tribu, $B \in \mathcal{F}$ et $A \subset B$ alors $A \in \mathcal{F}$?

Question : Si $A \in \mathcal{A}$ et $B \in \mathcal{A}$ alors $A \cap B \in \mathcal{A}$?

Proposition 1.1 (*Premières propriétés*)

(i) Toute algèbre de Boole est stable par intersections et unions finies. Toute tribu est stable par intersections et unions dénombrables.

(ii) L'intersection quelconque de tribus est une tribu.

Démonstration :

(i) On remarque que pour tout $(A_n)_{n\geq 1}$,

$$\bigcap_{n\geq 1} A_n = \left(\bigcup_{n\geq 1} A_n^c\right)^c.$$

(ii) Soit $(\mathcal{F}_i)_{i\in I}$ une famille quelconque de tribus. Alors

- $\Omega \in \bigcap_{i\in I} \mathcal{F}_i$;
- Si $A \in \bigcap_{i\in I} \mathcal{F}_i$ alors pour tout $i \in I$, $A^c \in \mathcal{F}_i$ et donc $A^c \in \bigcap_{i\in I} \mathcal{F}_i$.

Idem pour la stabilité par union dénombrable. □

Définition 1.2 (*Tribu engendrée*)

Soit $\mathcal{C}$ un ensemble non vide de parties de Ω. On appelle *tribu engendrée par* $\mathcal{C}$ l'intersection de toutes les tribus contenant $\mathcal{C}$.

Remarque : L'ensemble des tribus contenant $\mathcal{C}$ est non vide car contient $\mathcal{P}(\Omega)$. De plus, l'intersection des tribus contenant $\mathcal{C}$ est bien une tribu par la proposition précédente.

Proposition 1.2

La tribu engendrée par $\mathcal{C}$ est la plus petite tribu contenant $\mathcal{C}$.

Démonstration : On a vu que c'est bien une tribu. Elle contient $\mathcal{C}$ puisque si $C \in \mathcal{C}$, C est dans toutes les tribus contenant $\mathcal{C}$ et donc dans leur intersection. Enfin c'est la plus petite par construction puisqu'elle est incluse dans toutes les autres. □

Notation : Dans la suite, on notera $\sigma(\mathcal{C})$ la tribu engendrée par $\mathcal{C}$.

Proposition 1.3 (*Tribu engendrée par une partition dénombrable*)

Soit Ω un ensemble non vide et soit $\mathcal{A} = \{A_i,\ i \in \mathbb{N}\}$ une partition dénombrable de Ω. Alors la plus petite tribu contenant $\mathcal{A}$ est

$$\sigma(\mathcal{A}) = \left\{\bigcup_{i\in I} A_i /\ I \subset \mathbb{N}\right\}.$$

Exemple : On l'a vu pour $\mathcal{A} = \{A, A^c\}$, A strict. Et si $\mathcal{A} = \{A_1, A_2, A_3\}$ partition, alors

$$\sigma(\mathcal{A}) = \{\emptyset, A_1, A_2, A_3, A_1 \cup A_2, A_1 \cup A_3, A_2 \cup A_3, \Omega\}.$$

Démonstration : On pose $\mathcal{F} = \left\{ \bigcup_{i \in I} A_i / \ I \subset \mathbb{N} \right\}$ et on montre que $\mathcal{F} = \sigma(\mathcal{A})$, i.e. que $\mathcal{F}$ est bien une tribu contenant $\mathcal{A}$ puis que c'est la plus petite.

- C'est bien une tribu car
  . $\Omega = \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$.
  . Si $A \in \mathcal{F}$, $A = \bigcup_{i \in I} A_i$ et $A^c = \bigcup_{i \in I^c} A_i \in \mathcal{F}$;
  . Si $(A^k)_{k \geq 1}$ est une suite de $\mathcal{F}$, alors pour tout $k \geq 1$, $A^k = \bigcup_{i \in I^k} A_i$ et donc

$$\bigcup_{k \geq 1} A^k = \bigcup_{i \in \cup_{k \geq 1} I_k} A_i \in \mathcal{F}.$$

- $\mathcal{F}$ contient $\mathcal{A}$ car tous les A_i sont dedans (on choisit $I = \{i\}$).
- C'est bien la plus petite car toute tribu contenant $\mathcal{A}$ doit au moins contenir $\mathcal{F}$. □

Exemple : On essaye toujours de se ramener à une partition. Par exemple,

$$\sigma - \{A, B\} = \sigma - \{A/B, B/A, A \cap B, (A \cup B)^c\}.$$

Exemple : [Tribu borélienne ou tribu de Borel]

On appelle *tribu borélienne* sur $\mathbb{R}$ (resp. sur $\mathbb{R}^n$) la tribu engendrée par les intervalles de $\mathbb{R}$ (resp. les pavés de $\mathbb{R}^n$). On peut montrer qu'elle est aussi la tribu engendrée par chacune des familles suivantes :

- les unions finies d'intervalles de $\mathbb{R}$ (resp. de pavés de $\mathbb{R}^n$) ;
- les ouverts de $\mathbb{R}$ (resp. de $\mathbb{R}^n$) (on montre que tout ouvert de $\mathbb{R}$ est union dénombrable d'intervalles ouverts disjoints) ;
- les fermés de $\mathbb{R}$ (resp. de $\mathbb{R}^n$) ;
- les intervalles (resp. pavés) à bornes dans $\mathbb{Q}$ (resp. $\mathbb{Q}^n$), qui forment une famille dénombrable.

Dans le cas particulier de $\mathbb{R}$, on peut aussi montrer que la tribu borélienne est engendrée par les intervalles de la forme $]-\infty, x]$ pour $x \in \mathbb{R}$ (ou $x \in \mathbb{Q}$) (cf. exercice).

Les ensembles de la tribu borélienne s'appellent les ensembles boréliens ou les boréliens.

1.1.2 Mesures.

Définition 1.3

Soit $(\Omega, \mathcal{F})$ un espace mesurable. Une application $\mu : \mathcal{F} \to \overline{\mathbb{R}}_+$ est une *mesure positive* si

(i) $\mu(\emptyset) = 0$

(ii) $\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n)$ pour toute suite $(A_n)_n$ de $\mathcal{F}$ famille d'éléments disjoints (cette propriété s'appelle la σ-additivité).

Remarque :

1. $(\Omega, \mathcal{F}, \mu)$ s'appelle un *espace mesuré*. Si μ est une probabilité, i.e. si sa masse totale $\mu(\Omega) = 1$, alors $(\Omega, \mathcal{F}, \mu)$ s'appelle un *espace probabilisé*.
2. Une application μ définie sur une algèbre de Boole et à valeurs dans $\overline{\mathbb{R}}_+$ est appelée mesure si elle satisfait (i) et si elle satisfait (ii) pour les suites $(A_n)_n$ telles que $\cup_n A_n$ appartient à l'algèbre de Boole.

Exemple : Soit Ω un ensemble non vide.

- *Mesure de Dirac sur* Ω. Soit $x_0 \in \Omega$. La mesure de Dirac en x_0 est définie par : $\forall A \in \mathcal{P}(\Omega)$,

$$\delta_{x_0}(A) = \begin{cases} 1 \text{ si } x_0 \in A \\ 0 \text{ sinon.} \end{cases}$$

C'est une probabilité sur $\mathcal{P}(\Omega)$. En effet, $\delta_{x_0}(\varnothing) = 0$, $\delta_{x_0}(\Omega) = 1$ et si $(A_n)_{n\in\mathbb{N}}$ est une suite d'ensembles disjoints alors $x_0 \in \cup_n A_n$ si et seulement si x_0 est dans l'un des A_n et comme ils sont disjoints, x_0 ne peut appartenir qu'à un seul A_n. Autrement dit $\delta_{x_0}(\cup_n A_n) = 1$ si $\sum_n \delta x_0(A_n) = 1$, et l'autre valeur possible est 0, ce qui conclut la preuve.

- *Mesures discrètes sur* Ω. Soit $(x_k)_{k\geq 1}$ une suite d'éléments de Ω et $(a_k)_{k\geq 1}$ une suite de réels positifs. Alors la fonction, définie sur $\mathcal{P}(\Omega)$ par

$$\mu = \sum_{k\geq 1} a_k \delta_{x_k}$$

est une mesure sur $\mathcal{P}(\Omega)$. Si $\sum_k a_k = 1$, c'est une mesure de probabilité.

Proposition 1.4

Soit μ une application sur une tribu $\mathcal{F}$ à valeurs dans $\overline{\mathbb{R}}_+$ et telle que $\mu(\varnothing) = 0$. Alors μ est σ-additive si et seulement si

(i) μ est additive, i.e. si $\mu(A \cup B) = \mu(A) + \mu(B)$, $\forall\ A,\ B \in \mathcal{F}$ disjoints.

(ii) μ est σ-sous-additive, i.e. pour toute famille $(A_n)_n \subset \mathcal{F}$,

$$\mu\left(\bigcup_n A_n\right) \leq \sum_n \mu(A_n).$$

Proposition 1.5 (*Propriétés d'une mesure*)

Soient A, $B \in \mathcal{F}$.

1. *$A \subset B \Rightarrow \mu(A) \leq \mu(B)$ et $\mu(B/A) = \mu(B) - \mu(A)$ si $\mu(A) < +\infty$.*
2. *Si $\mu(A \cap B) < +\infty$, alors $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.*

Définition 1.4 (*Ensembles négligeables, propositions presque sûres*)

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré.

(i) Si $N \in \mathcal{F}$ est tel que $\mu(N) = 0$, on dit que N est *μ-négligeable ou négligeable*.

(ii) Si $B \in \mathcal{F}$ est négligeable et si $A \subset \Omega$ est tel que $A \subset B$, on dira aussi que A est *négligeable*.

(iii) Un *espace mesuré* est *complet* si N négligeable $\Rightarrow N \in \mathcal{F}$ (on parle alors d'espace mesuré complet ou de *mesure complète*).

(iv) Une propriété dépendant de $\omega \in \Omega$ est dite vraie *(μ-)presque sûrement* si elle est vraie en dehors d'un négligeable.

Proposition 1.6

Toute union dénombrable de négligeables est négligeable (démonstration par σ-sous-additivité).

Définition 1.5

Une mesure μ sur un espace mesurable $(\Omega, \mathcal{F})$ est dite *σ-finie* si Ω admet une partition dénombrable $(A_n)_n$ telle que $\mu(A_n) < +\infty$ pour tout n.

Théorème 1.1 (*Prolongement de Carathéodory*)

Théorème admis, de preuve constructive.

Soit μ une mesure sur une algèbre de Boole $\mathcal{A}$ sur Ω. Alors μ admet un prolongement qui est une mesure sur $\sigma(\mathcal{A})$. Si de plus μ est σ-finie alors le prolongement est unique.

Exemple : [Mesures boréliennes]

- *Mesure de Lebesgue*

  Soit I un intervalle de $\mathbb{R}$. On note $\lambda(I)$ la longueur de I, éventuellement infinie. On note $\mathcal{A}$ l'ensemble des unions finies d'intervalles de $\mathbb{R}$. Alors $\mathcal{A}$ est une algèbre de Boole. De plus, on peut prolonger λ à $\mathcal{A}$ en conservant la σ-additivité et donc en une mesure. En remarquant que $\mathbb{R}$ est l'union des intervalles de longueur 1 $[n, n+1[$, pour $n \in \mathbb{Z}$, on montre que λ est σ-finie et donc la mesure se prolonge de façon unique à $\mathcal{B}(\mathbb{R})$.

  On peut faire la même chose sur les pavés de $\mathbb{R}^n$ en remplaçant la notion de longueur d'intervalle par celle d'aire dans $\mathbb{R}^2$, de volume dans $\mathbb{R}^3$, etc. On définit ainsi les mesures de tous les boréliens de $\mathbb{R}^n$ et on note λ_n cette mesure.

  La tribu borélienne, complétée par les ensembles négligeables s'appelle la *tribu de Lebesgue*, notée $\mathcal{L}(\mathbb{R})$ (resp. $\mathcal{L}(\mathbb{R}^n)$). La version complète de la mesure λ (resp. λ_n) se note encore λ (resp. λ_n) et s'appelle la *mesure de Lebesgue*.

- *Mesures boréliennes.* On peut construire d'autres mesures sur la tribu de Borel en définissant une mesure sur le clan des unions finies d'intervalles ou de pavés. Par exemple, sur $\mathbb{R}$, étant donné une fonction F croissante et continue à droite, si on définit $\mu(]a, b]) = F(b) - F(a)$ pour tous $a < b$ alors μ se prolonge en une unique mesure borélienne. La mesure de Lebesgue est le cas particulier où $F = Id$.

1.1.3 Espaces et mesures produit.

Définition 1.6 (*Tribu produit*)

Soient $(\Omega_1, \mathcal{F}_1)$,..., $(\Omega_n, \mathcal{F}_n)$ n espaces mesurables. On appelle tribu produit sur $\Omega_1 \times ... \times \Omega_n$ et on note $\mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n$ la tribu engendrée par les pavés mesurables i.e. les ensembles $A_1 \times ... \times A_n$ tels que $A_1 \in \mathcal{F}_1$, ..., $A_n \in \mathcal{F}_n$.

Exemple : [Propriété du produit des tribus boréliennes] (cf. exercice)

La tribu engendrée par les pavés mesurables $B_1 \times ... \times B_d$, où $B_i \in \mathcal{B}(\mathbb{R})$ pour tout i, est la tribu $\mathcal{B}(\mathbb{R}^d)$ des boréliens sur $\mathbb{R}^d$. Cela s'écrit

$$\mathcal{B}(\mathbb{R})^{\otimes d} \stackrel{\text{def}}{=} \underbrace{\mathcal{B}(\mathbb{R}) \otimes ... \otimes \mathcal{B}(\mathbb{R})}_{d \text{ fois}} = \mathcal{B}(\mathbb{R}^d).$$

Cela signifie que la tribu engendrée par les pavés mesurables coïncide avec celle engendrée par les pavés.

Proposition 1.7

Soient μ_1,...,μ_n des mesures σ-finies sur $(\Omega, \mathcal{F}_1)$, ..., $(\Omega, \mathcal{F}_n)$ respectivement. On définit μ sur les pavés mesurables par

$$\mu(A_1 \times ... \times A_n) = \mu_1(A_1)...\mu_n(A_n) \ \forall \ (A_1, ..., A_n) \in \mathcal{F}_1 \times ... \times \mathcal{F}_n. \tag{1.1}$$

Soit $\mathcal{A}$ l'algèbre de Boole des unions finies des pavés mesurables. Alors on peut prolonger μ de façon unique à $\mathcal{A}$ de sorte que μ soit une mesure σ-finie sur $\mathcal{A}$.

Proposition 1.8

Il existe une unique mesure μ sur $(\Omega_1 \times ... \times \Omega_n, \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n)$ satisfaisant (1.1). C'est une conséquence directe du théorème de Carathéodory.

Définition 1.7 (*mesure produit*)

La mesure μ s'appelle la mesure produit (ou probabilité produit), sur l'espace mesurable $(\Omega_1 \times ... \times \Omega_n, \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n)$, des mesures μ_1,..., μ_n. On note ce produit $\mu = \mu_1 \otimes ... \otimes \mu_n$.

Exemple : [Proposition sur le produit des mesures de Lebesgue] (Proposition admise)
La mesure de Lebesgue sur $\mathbb{R}^d$ est le produit des mesures de Lebesgue sur $\mathbb{R}$:

$$\lambda_d = \underbrace{\lambda \otimes ... \otimes \lambda}_{d \text{ fois}}.$$

1.2 Espaces probabilisés.

1.2.1 Généralités.

Définition 1.8

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Un événement $A \in \mathcal{F}$ est dit $\mathbb{P}$-presque sûr ou $\mathbb{P}$-presque certain si $\mathbb{P}(A) = 1$.

Remarque : Les calculs de probabilités vus dans des cours d'introduction aux probabilités ne seront pas rappelés ici. Notons qu'ils peuvent dans certain cas être généralisés aux mesures bornées, voire aux mesures σ-finies ou quelconques.

Exemple : [probabilité uniforme] (A faire en exercice)

- ☐ Si Ω fini et $\mathcal{F} = \mathcal{P}(\Omega)$, la fonction définie sur $(\Omega, \mathcal{F})$ par

$$\mathbb{P}_{ud}(A) = \frac{\#A}{\#\Omega}$$

(le symbole # désignant le cardinal de l'ensemble qui le suit), est une probabilité sur $(\Omega, \mathcal{F})$.

- ☐ Si B_0 est un borélien de $\mathbb{R}^d$ tel que $0 < \lambda_d(B_0) < +\infty$ alors

$$\mathbb{P}_{uc}(B) = \frac{\lambda_d(B \cap B_0)}{\lambda_d(B_0)} \text{ pour } B \in \mathcal{B}(\mathbb{R}^d)$$

est une probabilité appelée la probabilité uniforme sur B_0.

1.2.2 Probabilités conditionnelles.

Définition 1.9

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et soit A un événement non négligeable. L'application $\mathbb{P}[\cdot|A] : \mathcal{F} \to [0,1]$ qui à B associe $\mathbb{P}[B|A] = \mathbb{P}[B \cap A]/\mathbb{P}[A]$ est une probabilité sur $(\Omega, \mathcal{F})$ appelée la probabilité conditionnelle de B sachant A.

Démonstration : La preuve est la même que celle de l'exemple précédent avec la mesure borélienne. Il suffit de monter pour conclure que $\mathbb{P}[\Omega|A] = 1$.

1.2.3 Indépendance de familles de tribus.

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

Définition 1.10

Soient $\mathcal{F}_1$, ..., $\mathcal{F}_n$ des sous-tribus de $\mathcal{F}$ (i.e. pour tout i, $A \in \mathcal{F}_i \Rightarrow A \in \mathcal{F}$) sur Ω. Elles sont dites indépendantes si $\mathbb{P}[A_1 \cap ... \cap A_n] = \mathbb{P}[A_1] ... \mathbb{P}[A_n]$ pour tous $A_1 \in \mathcal{F}_1, ..., A_n \in \mathcal{F}_n$.

Remarque : L'indépendance des événements dépend de la mesure choisie.

Définition 1.11

Soient A_1, ..., $A_n \in \mathcal{F}$. On dit que A_1, ..., A_n sont indépendants si $\sigma - \{A_1\}$, ..., $\sigma - \{A_n\}$ le sont.

Exemple :

- ☐ Ω et $\varnothing$ sont indépendants de tous les événements. De même que les événements de probabilité 0 ou 1.
- ☐ Les seuls événements indépendants d'eux-même sont les événements de probabilité 0 ou 1.
- ☐ Si A_1, ..., A_n sont des événements tels que $\mathbb{P}[A_1 \cap ... \cap A_n] = \mathbb{P}[A_1] ... \mathbb{P}[A_n]$ alors A_1, ..., A_n sont indépendants.

Définition 1.12 (*Familles indépendantes dénombrables*)

- ☐ Les sous-tribus de $\mathcal{F}$ formant la famille $\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ sont dites indépendantes si toute sous-famille finie de $\mathbb{F}$ forme une famille de tribus indépendantes.
- ☐ Les événements de la suite $(A_n)_{n \in \mathbb{N}}$ sont dits indépendants si les tribus de la suite $(\sigma - \{A_n\})_{n \in \mathbb{N}}$ sont indépendantes.

Proposition 1.9 (*Propriété*)

Soit $\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ une famille dénombrable de tribus indépendantes. Soient $N \geq 1$ un entier et $(I_k)_{k=1,..,N}$ une partition de $\mathbb{N}$. Pour tout $k \in \{1, .., N\}$, on définit

$$\mathcal{G}_k = \sigma\left(\bigcup_{i \in I_k} \mathcal{F}_i\right).$$

Alors les tribus $(\mathcal{G}_k)_{k=1..N}$ sont indépendantes.

1.3 Exercices

Dans tous les exercices, Ω est un ensemble non vide.

Exercice 1.1.
Soient A, B deux sous-ensembles stricts de Ω. Donner une condition nécessaire et suffisante pour que $\sigma\{A\} \cup \sigma\{B\}$ soit une tribu.

Exercice 1.2.
Soient $\mathcal{A}$, $\mathcal{B} \subset \mathcal{P}(\Omega)$ et non vides. Montrer que

$$\sigma(\mathcal{A} \cup \mathcal{B}) = \sigma\{A \cup B,\ A \in \mathcal{A}, B \in \mathcal{B}\} \text{ si } \varnothing \in \mathcal{A} \cap \mathcal{B}$$

et

$$\sigma(\mathcal{A} \cup \mathcal{B}) = \sigma\{A \cap B,\ A \in \mathcal{A}, B \in \mathcal{B}\} \text{ si } \Omega \in \mathcal{A} \cap \mathcal{B}.$$

Exercice 1.3.
Soit (Ω) un ensemble non vide et on note $\mathcal{P}$ une partition de Ω. On s'intéresse au cardinal de $\sigma(\mathcal{P})$.

1. Si $card(\mathcal{P}) = 1$, quel est le cardinal de $\sigma(\mathcal{P})$?
2. Si $card(\mathcal{P}) = 2$, quel est le cardinal de $\sigma(\mathcal{P})$?
3. Si $card(\mathcal{P}) = n, n \geq 3$, quel est le cardinal de $\sigma(\mathcal{P})$?
4. Qu'obtient-on pour une tribu engendrée par une partition dénombrable ? (fini, dénombrable ou indénombrable ?)

Indication : On rappelle que si $\mathcal{P} = \{A_i\}_{i\in\mathbb{N}}$, alors tout $A \in \sigma(\mathcal{P})$ s'écrit

$$A = \bigcup_{i\in I} A_i \text{ pour un } I \subset \mathbb{N}.$$

Pour tout $A \in \sigma(\mathcal{P})$, on note $N(A)$ l'ensemble des indices i qui interviennent dans la définition de A et on définit l'applicaton

$$\begin{aligned} c_A : \mathcal{P} &\to \{0,1\} \\ A_i &\mapsto \begin{cases} 1 & si\ i \in N(A) \\ 0 & sinon \end{cases} \end{aligned}$$

et on remarque que $card(\sigma(\mathcal{P}))$ est le nombre de ces applications.

Exercice 1.4.
Montrer que $\mathcal{B}(\mathbb{R}) = \sigma\{]-\infty, x],\ x \in \mathbb{R}\}$.

Exercice 1.5.
Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré, montrer les résultats suivants

1. Pour toute suite croissante $(A_n)_{n\geq 1}$ d'ensembles mesurables, on a

$$\mu\left(\bigcup_{n\geq 1} A_n\right) = \lim_{n\to+\infty} \mu(A_n) = \sup_{n\geq 1} \mu(A_n);$$

2. Pour toute suite décroissante $(A_n)_{n\geq 1}$ d'ensembles mesurables tel qu'il existe $n_0 > 0$ tel que $\mu(A_{n_0}) < +\infty$ on a

$$\mu\left(\bigcap_{n\geq 1} A_n\right) = \lim_{n\to+\infty} \mu(A_n) = \inf_{n\geq 1} \mu(A_n).$$

Exercice 1.6. [*Mesure de Lebesgue-Stieltjes*]
Soit F une fonction de $\mathbb{R} \to \mathbb{R}$ croissante et continue à droite. On définit, pour tous a, $b \in \mathbb{R}$,

$$\lambda_F(]a,b]) = F(b)-F(a), \quad \lambda_F(]-\infty,a]) = F(a)-\lim_{t\to-\infty} F(t) \text{ et } \lambda_F(]b,+\infty[) = \lim_{t\to+\infty} F(t)-F(b).$$

On peut alors démontrer que λ_F st une mesure borélienne appelée mesure de Lebesgue-Stieltjes.

Calculer la mesure de Lebesgue-Stieltjes associée à la fonction F croissante continue à droite qui vaut 0 dans les négatifs et 1 dans les positifs.

Exercice 1.7. [*Combinaison convexe de probabilités*]

1. Soit $(\mathbb{P}_n)_{n\geq 1}$ une suite de probabilités sur un espace mesurable $(\Omega, \mathcal{F})$ et $(a_n)_{n\geq 1}$ une suite de réels positifs de somme égale à 1. Montrer que l'application

$$\mathbb{P}[A] = \sum_{i\geq 1} a_i \mathbb{P}_i[A], \ A \in \mathcal{F},$$

est une probabilité.

2. Soit Ω un ensemble dénombrable. Pour tout $A \in \Omega$, on note $|A| = card(A)$.
   a. Montrer que $A \to |A|$ est une mesure sur $(\Omega, \mathcal{P}(\Omega))$.
   b. Soit $\{B_i : i \in \mathbb{N}\}$ une partition de Ω. On suppose que pour tout $i \in \mathbb{N}, 0 < |B_i| < +\infty$. Soit alors $\mathbb{P}$ l'application définie sur $\mathcal{P}(\Omega)$ par

$$\mathbb{P}[A] = \sum_{i\in\mathbb{N}} a_i \frac{|B_i \bigcap A|}{|B_i|}.$$

   Montrer que sous les conditions du 1., $\mathbb{P}$ est une probabilité.

3. Soit $(x_k)_{k\geq 1}$ une suite de $\mathbb{R}^n$. Pour tout borélien A de $\mathbb{R}^n$, on note

$$\mathbb{P}[A] = \sum_{k\geq 1} a_k \mathbf{1}_A(x_k) = \sum_{k\geq 1} a_k \delta_{x_k}(A).$$

Montrer que sous les conditions du 1., $\mathbb{P}$ est une probabilité.

Exercice 1.8. [*Restriction des tribus et de probabilités*]
Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $A \subset \Omega$.

1. Montrer que $\mathcal{F}_A = \{A \cap B : B \in \mathcal{F}\}$ est une tribu sur A.
2. Montrer que si $A \in \mathcal{F}$, $\mathcal{F}_A$ est incluse dans $\mathcal{F}$. Donner une condition pour que la restriction de $\mathbb{P}$ à $(A, \mathcal{F}_A)$ soit une probabilité.

Exercice 1.9.[*Tribu et mesure produit*]

1. Montrer que $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

2. Montrer que la mesure de Lebesgue sur $\mathbb{R}^2$ coïncide avec la mesure produit $\lambda \otimes \lambda$ sur $\mathcal{B}(\mathbb{R}^2)$.

Exercice 1.10. [*Evénements négligeables et indépendance*]
Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $A \in \mathcal{F}$ et $B \in \mathcal{F}$. Etudier l'équivalence deux à deux entre les propositions suivantes :

1. "A est de probabilité un" et "A est indépendant de lui-même" ;
2. "A est négligeable" et "A est indépendant de tout événement" ;
3. "A et B sont indépendants et disjoints" et "A ou B est négligeable".

Exercice 1.11.
Les assertions suivantes sont elles vraies ?

1. A et B disjoints $\Rightarrow$ A et B indépendants ?
2. A et B indépendants $\Rightarrow$ A et B disjoints ?

Exercice 1.12.
Trouver deux intervalles indépendants pour la mesure de Lebesgue sur $[0, 1]$.

Chapitre 2

Fonctions mesurables, variables aléatoires.

Sommaire

2.1 Autonomie : Image directe, image réciproque

2.1.1 Images directes et réciproques d'ensembles

Pour commencer un rappel des définitions :

Définition 2.1

Soit $f : E \to F$ une application d'un ensemble E dans un ensemble F.

(i) Si $A \subset E$, on appelle image directe de A par f le sous-ensemble $f(A)$ de F défini par

$$f(A) = \{f(x), x \in A\}.$$

(ii) Si $B \subset F$, on appelle image réciproque de B par f le sous-ensemble $f^{-1}(B)$ de E défini par

$$f^{-1}(B) = \{x \in E, f(x) \in B\}.$$

Le but est, avant de passer à des choses plus compliquées, de faire quelques manipulations ensemblistes.

Inclusions On suppose $A_1 \subset A_2 \subset E$ et on veut savoir si on peut en déduire que $f(A_1) \subset f(A_2)$. On choisit donc un point de $f(A_1)$, soit $f(x)$ pour un $x \in A_1$. Mais si $x \in A_1$ alors $x \in A_2$ puisque $A_1 \subset A_2$ et donc $f(x) \in f(A_2)$, d'où $f(A_1) \subset f(A_2)$.

Pouvez-vous montrer maintenant que si $B_1 \subset B_2 \subset F$ alors $f^{-1}(B_1) \subset f^{-1}(B_2)$?

Stabilité par union ou intersection finie ou dénombrable Une autre question naturelle est "l'image directe ou réciproque d'une union (respectivement une intersection) finie ou dénombrable d'ensembles est elle égale à l'union (respectivement l'intersection) finie ou dénombrable des images directes et réciproques ?".

Commençons par les images réciproques qui sont très stables :

Soient B_1, $B_2 \subset F$ et soit $x \in f^{-1}(B_1 \cup B_2)$, c'est-à-dire $f(x) \in B_1 \cup B_2$. Alors $f(x) \in B_1$ ou $f(x) \in B_2$, c'est-à-dire soit $x \in f^{-1}(B_1)$ soit $x \in f^{-1}(B_2)$, ce qui s'écrit $x \in f^{-1}(B_1) \cup f^{-1}(B_2)$, et donc $f^{-1}(B_1 \cup B_2) \subset f^{-1}(B_1) \cup f^{-1}(B_2)$. Réciproquement, un élément de $f^{-1}(B_1) \cup f^{-1}(B_2)$ est soit un élément de $f^{-1}(B_1)$ soit un élément de $f^{-1}(B_2)$, c'est-à-dire soit $f(x) \in B_1$ soit $f(x) \in B_2$ ou encore $f(x) \in B_1 \cup B_2$, c'est-à-dire $x \in f^{-1}(B_1 \cup B_2)$. On a donc $f^{-1}(B_1) \cup f^{-1}(B_2) \subset f^{-1}(B_1 \cup B_2)$ et finalement $f^{-1}(B_1) \cup f^{-1}(B_2) = f^{-1}(B_1 \cup B_2)$.

A ce stade, vous devriez voir aisément que si on change union par intersection, la même preuve reste valable. De même le caractère fini de l'union n'intervient pas dans la démonstration et donc le résultat se prolonge naturellement au cas dénombrable.

Passons au cas des images directes (qui sont toujours plus délicates à manipuler). Pour l'union, finie ou dénombrable, il n'y a à nouveau aucun problème. En revanche, pour l'intersection, la preuve est mise en défaut, même dans le cas fini.

Question :

1. Retrouvez à quel endroit exactement la preuve est mise en défaut et voyez quelle inclusion est fausse en toute généralité. Donner un exemple.
2. Cette inclusion peut être obtenue sous l'une des conditions suivantes :
   (a) f injective (un seul antécédent) : $f(x) = f(y) \Rightarrow x = y$.
   (b) f surjective (un antécédent au moins) : $f(E) = F$.
   Laquelle ?

Ensembles complémentaires Les images directes et indirectes sont-elles stables par passage au complémentaire ? Sinon, sous quelle condition supplémentaire obtient-on les inclusions ?

Applications composées Soient E, F et G trois ensembles et soient $f : E \to F$ et $g : F \to G$ deux applications. Pour un $A \subset E$, $g \circ f(A)$ est l'ensemble des $g \circ f(x)$ pour $x \in A$, soit des $g(f(x))$ pour $x \in \mathcal{A}$. Mais l'ensemble des $f(x)$ pour $x \in A$ est $f(A)$ et donc $g \circ f(A) = g(f(A))$.

Soit $C \subset F$. Montrer de même que $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$.

Stabilité des ensembles par $f \circ f^{-1}$ et $f^{-1} \circ f$. Quand f n'est pas bijective, les ensembles ne sont pas stables par ces applications a priori, mais certaines inclusions restent vraies. En particulier, si $x \in A$, $f(x) \in f(A)$ et donc $x \in f^{-1}(f(A))$, et donc $A \subset f^{-1}(f(A))$. L'inclusion inverse est fausse en général. En effet, si $x \in f^{-1}(f(A))$ alors $f(x) \in f(A)$ mais si f n'est pas injective, cela n'implique pas $x \in A$ mais seulement $f(x) = f(y)$ pour un $y \in A$.

De même, trouver une inclusion entre les ensembles B et $f(f^{-1}(B))$ dans F et une condition sur f pour avoir égalité.

Proposition 2.1

Soient E, F et G trois ensembles et soient $f : E \to F$ et $g : F \to G$ deux applications.

1. *Pour tout $A \subset E$, $g \circ f(A) = g(f(A))$.*
2. *Pour tout $C \subset G$, $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$.*

Proposition 2.2

Soit f une application d'un ensemble E dans un ensemble F.

1. *Soit $A \subset E$*
   (a) $A \subset f^{-1}(f(A))$.
   (b) *Si f est injective,* $f^{-1}(f(A)) \subset A$.
2. *Soit $B \subset F$.*
   (a) $f(f^{-1}B)) \subset B$.
   (b) *Si f est surjective* $B \subset f(f^{-1}(B))$.

Proposition 2.3

Soit f une application d'un ensemble E dans un ensemble F.

1. *Soit $A \subset E$.*
   (a) *Si f est injective,* $f(A^c) \subset f(A)^c$.
   (b) *Si f est surjective,* $f(A)^c \subset f(A^c)$.
2. *Soit $B \subset F$. Alors* $f^{-1}(B^c) = f^{-1}(B)^c$.
3. *Soient A_1 et A_2 deux sous-ensembles de E.*
   (a) *Si $A_1 \subset A_2$ alors* $f(A_1) \subset f(A_2)$.
   (b) $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$
   (c) $f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$ *et on a égalité si f injective.*
4. *Soient B_1 et B_2 deux sous-ensembles de F.*
   (a) *Si $B_1 \subset B_2$ alors* $f^{-1}(B_1) \subset f^{-1}(B_2)$.
   (b) $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$.
   (c) $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.
5. *Les propriétés ci-dessus de stabilité par union et intersection sont encore vraies pour des familles dénombrables d'ensembles.*

2.1.2 Images directes et réciproques de familles d'ensembles

Dans la suite, si Ω est un ensemble, $\mathcal{P}(\Omega)$ désigne l'ensemble des parties de Ω. Pour $\mathcal{A}_1$ et $\mathcal{A}_2$ deux parties de $\mathcal{P}(\Omega)$ l'inclusion $\mathcal{A}_1 \subset \mathcal{A}_2$ signifie que tous les ensembles A de $\mathcal{A}_1$ sont dans $\mathcal{A}_2$.

Définition 2.2

Soit $f : \Omega \to \Omega'$ une application d'un ensemble Ω dans un ensemble Ω'.

(i) Pour tout $\mathcal{A} \subset \mathcal{P}(\Omega)$, on appelle image directe de $\mathcal{A}$ par f la famille $f(\mathcal{A})$ de parties de Ω' définie par
$$f(\mathcal{A}) = \{B \subset \Omega', f^{-1}(B) \in \mathcal{A}\}.$$

(ii) Pour tout $\mathcal{B} \subset \mathcal{P}(\Omega')$, on appelle image réciproque de $\mathcal{B}$ par f la famille $f^{-1}(\mathcal{B})$ de parties de Ω définie par
$$f^{-1}(\mathcal{B}) = \{f^{-1}(B),\ B \in \mathcal{B}\}.$$

Question : A-t-on $f(\mathcal{A}) = \{f(A),\ A \in \mathcal{A}\}$ en général ? Sous quelle condition est-ce vrai ?

On se pose alors les mêmes questions que dans la section précédente.

Inclusions. Soient $\mathcal{A}_1$ et $\mathcal{A}_2$ deux parties de $\mathcal{P}(\Omega)$. On suppose que $\mathcal{A}_1 \subset \mathcal{A}_2$. Soit $B \in \mathcal{A}_1$. Alors $f^{-1}(B) \subset \mathcal{A}_1 \subset \mathcal{A}_2$ et donc $B \in f(\mathcal{A}_2$ et $f(\mathcal{A}_1)...f(\mathcal{A}_2)$.

Sans difficulté, on peut montrer le même type de stabilité par f^{-1} pour l'inclusion.

Proposition 2.4

1. *Si* $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \mathcal{P}(\Omega)$, $f(\mathcal{A}_1) \subset f(\mathcal{A}_2)$.
2. *Si* $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \mathcal{P}(\Omega')$, $f^{-1}(\mathcal{B}_1) \subset f^{-1}(\mathcal{B}_2)$.

Le point important est l'analogue de la proposition 2.2.

Proposition 2.5

Soit f une application d'un ensemble Ω dans un ensemble Ω'.

1. *Soit* $\mathcal{A} \subset \mathcal{P}(\Omega)$
   (a) $f^{-1}(f(\mathcal{A})) \subset \mathcal{A}$.
   (b) *Si f est injective,* $\mathcal{A} \subset f^{-1}(f(\mathcal{A}))$.
2. *Soit* $\mathcal{B} \subset \mathcal{P}(\Omega')$.
   (a) $\mathcal{B} \subset f(f^{-1}\mathcal{B}))$.
   (b) *Si f est surjective* $f(f^{-1}(\mathcal{B})) \subset \mathcal{B}$.

Remarquer que l'on a les inclusions inverses de celles de la proposition 2.2.

Démonstration :

1. Soit $A \in f^{-1}(f(\mathcal{A}))$. Cela signifie $A = f^{-1}(B)$ pour un B de $f(\mathcal{A})$. Mais $B \in f(\mathcal{A})$ si et seulement si $f^{-1}(B) \in \mathcal{A}$ donc $A \in \mathcal{A}$ et $f^{-1}(f(\mathcal{A})) \subset \mathcal{A}$.
   Réciproquement, si f injective, $A = f^{-1}(f(A))$ pour tout $A \in \mathcal{P}(\Omega)$. Soit donc $A \in \mathcal{A}$. Comme $A = f^{-1}(f(A))$, on en déduit $f^{-1}(f(A)) \in \mathcal{A}$ i.e. $f(A) \in f(\mathcal{A})$. Donc finalement, $A = f^{-1}(B)$ pour un $B \in f(\mathcal{A})$, ce qui signifie $A \in f^{-1}(f(\mathcal{A}))$ et $\mathcal{A} \subset fm(f(\mathcal{A}))$.
2. Si $B \in \mathcal{B}$, alors $f^{-1}(B) \in f^{-1}(\mathcal{B})$, i.e. $B \in f(f^{-1}(\mathcal{B}))$. Réciproquement, soit $B \in f(f^{-1}(\mathcal{B}))$, ou autrement dit $f^{-1}(B) \subset f^{-1}(\mathcal{B})$. On ne déduit que $f^{-1}(B) = f^{-1}(B')$ pour un B' de $\mathcal{B}$. Il s'ensuit $f(f^{-1}(B)) = f(f^{-1}(B'))$. Maintenant, si f est surjective, on ne déduit d'après la proposition 2.2 que $B = B'$ soit $B \in \mathcal{B}$, d'où $f(f^{-1}(\mathcal{B})) \subset \mathcal{B}$.

2.2 Applications mesurables.

2.2.1 Définition et exemples de base

Définition 2.3 (*Application mesurable*)

Soient $(\Omega, \mathcal{F})$ et $(\Omega, \mathcal{F}')$ deux espaces mesurables, et soit $f : \Omega \to \Omega'$ une application. On dit que f est *mesurable* si $\forall A' \in \mathcal{F}'$, $f^{-1}(A') \in \mathcal{F}$, autrement dit si $f^{-1}(\mathcal{F}') \subset \mathcal{F}$.

Remarque : Si $\mathcal{F}$ et $\mathcal{F}'$ sont des tribus boréliennes, on dit que f est borélienne.

Exemple : *A faire à titre d'exercice.*

(i) Si $\mathcal{F} = \mathcal{P}(\Omega)$, toutes les applications sont mesurables.

(ii) On suppose que $\mathcal{F}'$ contient les singletons de Ω'. Alors si $\mathcal{F} = \{\varnothing, \Omega\}$, f est mesurable si et seulement si f est constante. Plus généralement, lorsque $\mathcal{F}$ est engendrée par une partition dénombrable, les applications mesurables sont celles qui sont constantes sur chaque élément de la partition.

(iii) On suppose $\Omega' = \mathbb{R}$ et $\mathcal{F}' = \mathcal{B}(\mathbb{R})$. Alors, si $A \subset \Omega$, $\mathbf{1}_A$ mesurable si et seulement si $A \in \mathcal{F}$.

2.2.2 Tribus engendrées par une fonction

Dans cette section, Ω et Ω' sont deux ensembles non vides et $f : \Omega \to \Omega'$ est une application.

La proposition suivante est à faire à titre d'exercice.

Proposition 2.6 (*Images réciproques et directes de tribus*)

(i) Soit $\mathcal{F}'$ une tribu sur Ω'. Alors son image réciproque par f est une tribu sur Ω et c'est la plus petite tribu qui rend f mesurable de Ω dans $(\Omega', \mathcal{F}')$.

(ii) Soit $\mathcal{F}$ une tribu sur Ω. Son image directe par f est une tribu sur Ω', c'est la plus grosse tribu rendant f mesurable.

Définition 2.4 (*Tribu engendrée par une fonction*)

Soient Ω et Ω' deux ensembles non vides et soit $f : \Omega \to \Omega'$ une application. Soit $\mathcal{F}'$ une tribu sur Ω'. La *tribu engendrée* par f lorsque Ω' est muni de la tribu $\mathcal{F}'$ est la plus petite tribu $\mathcal{F}$ rendant f mesurable de $(\Omega, \mathcal{F})$ dans $(\Omega', \mathcal{F}')$. C'est la tribu

$$\sigma(f) = f^{-1}(\mathcal{F}') = \{f^{-1}(A') /\ A' \in \mathcal{F}'\}.$$

Proposition 2.7

Soit $f : \Omega \to \Omega'$ une application, soit $\mathcal{F}$ une tribu sur Ω et soit $\mathcal{A}'$ une famille de parties de Ω'. On suppose que $f^{-1}(\mathcal{A}') \subset \mathcal{F}$. Alors f est mesurable de $(\Omega, \mathcal{F})$ dans $(\Omega', \sigma(\mathcal{A}'))$.

Exemple : Soit $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une application. Alors f est mesurable si et seulement si $f^{-1}(]-\infty, t]) \in \mathcal{F}$ pour tout $t \in \mathbb{R}$.

Démonstration : On montre en fait que

$$f^{-1}(\sigma(\mathcal{A}')) = \sigma(f^{-1}(\mathcal{A}')).$$

Pour montrer la proposition, on n'a besoin que de l'inclusion

$$f^{-1}(\sigma(\mathcal{A}')) \subset \sigma(f^{-1}(\mathcal{A}')) \tag{2.1}$$

car alors $f^{-1}(\mathcal{A}') \subset \mathcal{F}$ implique $\sigma(f^{-1}(\mathcal{A}')) \subset \mathcal{F}$ et donc $f^{-1}(\sigma(\mathcal{A}')) \subset \mathcal{F}$, ce qu'il fallait démontrer. Montrons maintenant l'inclusion (2.1). On a : $\mathcal{A}' \subset f(f^{-1}(\mathcal{A}')) \subset f(\sigma(f^{-1}(\mathcal{A}')))$, cette dernière famille étant une tribu. Comme elle contient $\mathcal{A}'$, elle contient $\sigma(A')$. On en déduit

$$f^{-1}(\sigma(\mathcal{A}')) \subset f^{-1}(f(\sigma(f^{-1}(\mathcal{A}')))) \subset \sigma(f^{-1}(\mathcal{A}')).$$

Pour l'inclusion inverse, on a : $\mathcal{A}' \subset \sigma(\mathcal{A}') \Rightarrow f^{-1}(\mathcal{A}') \subset f^{-1}(\sigma(\mathcal{A}'))$ qui est une tribu donc $\sigma(f^{-1}(\mathcal{A}')) \subset f^{-1}(\sigma(\mathcal{A}'))$.

Finalement, si on applique le résultat de la proposition à la famille $\mathcal{A}'$ des intervalles $]-\infty, t]$ pour $t \in \mathbb{R}$ qui engendre les boréliens, on obtient bien la mesurabilité de f dès que $f^{-1}(]-\infty, t]) \in \mathcal{F}$ pour tout $t \in \mathbb{R}$.

2.2.3 Propriétés des applications mesurables

Donner brièvement les arguments de preuve pour les deux premières propositions.

Proposition 2.8

1. *Soient* $f : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ *et* $g : (\Omega', \mathcal{F}') \to (\Omega'', \mathcal{F}'')$*, où* $\mathcal{F}$*,* $\mathcal{F}'$ *et* $\mathcal{F}''$ *sont des tribus sur* Ω*,* Ω' *et* Ω'' *respectivement, deux applications mesurables. Alors* $g \circ f$ *est mesurable de* $(\Omega, \mathcal{F})$ *dans* $(\Omega'', \mathcal{F}'')$*.*
2. *Soient* $(\Omega, \mathcal{F})$ *et* $(\Omega_i, \mathcal{F}_i)$ *pour* $i = 1..n$*,* $n \geq 2$*,* $n+1$ *espaces mesurables. Soit* f *une application de* Ω *dans* $\Omega_1 \times \Omega_2 \times ... \times \Omega_n$*. Autrement dit,* $f = (f_1, ..., f_n)$ *où les* f_i *sont des applications de* Ω *dans* Ω_i *pour tout* i*.*

   L'application f *est mesurable de* $(\Omega, \mathcal{F})$ *dans* $(\Omega_1 \times ... \times \Omega_n, \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n)$ *si et seulement si les applications* f_i *sont mesurables de* $(\Omega, \mathcal{F})$ *dans* $(\Omega_i, \mathcal{F}_i)$ *pour tout* i*.*

Proposition 2.9

Soient $(E, |\cdot|_E, \mathcal{E})$ *et* $(E', |\cdot|_{E'}, \mathcal{E}')$ *deux espaces vectoriels normés sur* $\mathbb{R}$ *(ou* $\mathbb{C}$*) munis de leur tribu borélienne (engendrée par les ouverts).*

1. *Si* f *est continue de* $(E, |\cdot|_E)$ *dans* $(E', |\cdot|_{E'})$ *alors* f *est mesurable de* $(E, \mathcal{E})$ *dans* $(E', \mathcal{E}')$*.*
2. *Soient* f *et* g *deux applications mesurables de* $(\Omega, \mathcal{F})$ *dans* $(E, \mathcal{E})$*. Alors* $|f|_E$*,* $f+g$*,* αf*, pour* $\alpha \in \mathbb{R}$ *ou* $\mathbb{C}$*, sont mesurables.*

Les deux propositions suivantes sont à faire en exercice.

Proposition 2.10

1. *Soient* f *et* g *deux applications mesurables à valeurs dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*. Alors,* fg*,* $\mathbf{1}_{[g \neq 0]} f/g$*,* $f \vee g = \max(f, g)$*,* $f \wedge g = \min(f, g)$ *sont mesurables.*
2. *Soit* $(f_n)_{n \geq 1}$ *une suite de fonctions mesurables à valeurs réelles. Alors* $\sup_n f_n$*,* $\inf_n f_n$*,* $\overline{\lim}_n f_n$*,* $\underline{\lim}_n f_n$ *sont mesurables. En particulier, la limite simple de fonctions mesurables est mesurable (y compris à valeur dans un espace vectoriel normé muni de sa tribu borélienne).*

Proposition 2.11

Toute fonction croissante de $\mathbb{R}$ *dans* $\mathbb{R}$ *est mesurable.*

2.2.4 Construction de fonctions mesurables réelles

Cette section doit être lue en préliminaire à la construction de l'intégrale (Chapitre 3) et peut-être omise en première lecture du présent chapitre.

Définition 2.5

Soit $(\Omega, \mathcal{F})$ un espace mesurable. On dit que $f : \Omega \to \mathbb{R}$ est une *fonction étagée* si

$$f = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}, \quad \text{avec } n \in \mathbb{N}^* \text{ et } a_i \in \mathbb{R}, A_i \in \mathcal{F}, \ \forall i \in \{1, ..., n\}.$$

Remarque : Les fonctions étagées sont mesurables.

Théorème 2.1

1. *Toute fonction mesurable à valeurs réelles est la différence de deux fonctions mesurables positives.*
2. *Toute fonction mesurable positive est limite simple d'une suite croissante de fonctions étagées.*

Démonstration :

1. Si f mesurable, on peut écrire $f = f^+ - f^-$ avec $f^+ = f \vee 0$ et $f^- = (-f) \vee 0$. La mesurabilité de f^+ et f^- est obtenue par la Proposition 2.10.
2. Preuve constructive. Pour tout $n \geq 1$, on définit

$$f_n = n\mathbf{1}_{[f \geq n]} + \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{1}_{[\frac{k}{2^n} < f \leq \frac{k+1}{2^n}]}.$$

Alors $(f_n)_{n\geq 1}$ est une suite croissante de fonctions étagées convergeant vers f.

Théorème 2.2

Soit f une fonction mesurable d'un espace mesurable $(\Omega, \mathcal{F})$ dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ et soit g mesurable de $(\Omega, \mathcal{F})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La fonction g est $\sigma(f)$-mesurable si e seulement si il existe $\varphi : \mathbb{R}^n \to \mathbb{R}$ borélienne telle que $g = \varphi \circ f = \varphi(f)$.

Démonstration : La preuve peut-être faite en s'aidant de l'Exercice 2.4 qui détaille les différentes étapes à suivre.

2.2.5 Mesures images

La proposition suivante peut être démontrée en exercice.

Proposition 2.12

Soit h une fonction mesurable de $(\Omega, \mathcal{F})$ dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $d \in \mathbb{N}^*$ et soit μ une mesure sur $(\Omega, \mathcal{F})$. On définit μ_h sur $\mathcal{B}(\mathbb{R}^d)$ par

$$\mu_h(B) = \mu \circ h^{-1}(B) = \mu(h^{-1}(B)), \ \forall B \in \mathcal{B}(\mathbb{R}^d).$$

Alors μ_h est une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)$. De plus, si μ est une probabilité, il en est de même pour μ_h.

Définition 2.6

Soit h une fonction mesurable de $(\Omega, \mathcal{F})$ dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $d \in \mathbb{N}^*$ et soit μ une mesure sur $(\Omega, \mathcal{F})$. La mesure $\mu_h = \mu \circ h^{-1}$ est appelée la *mesure image* de μ par h (ou *probabilité image* ou *distribution image* si μ est une probabilité).

On note souvent $[h \in B]$ l'ensemble $h^{-1}(B)$ et donc $\mu[h \in B]$ la quantité $\mu(h^{-1}(B))$.

On utilise aussi parfois la notations $h\mu$ pour μ_h pour rappeler que l'on applique h à μ pour définir une mesure sur l'espace d'arrivée de h qui est l'image de h par μ.

2.3 Variables aléatoires.

Dans cette section, $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé.

2.3.1 Lois des variables et vecteurs aléatoires.

Définition 2.7

(i) Une application X mesurable de $(\Omega, \mathcal{F})$ dans un espace vectoriel normé muni de sa tribu Borélienne $(E, |\cdot|, \mathcal{E})$ s'appelle une *variable aléatoire* définie sur $(\Omega, \mathcal{F})$ et à valeurs dans $(E, \mathcal{E})$ (ou de Ω à valeurs dans E s'il n'y a pas d'ambiguïté sur les tribus utilisées).

(ii) Si X est une variable aléatoire de Ω à valeurs dans E, on appelle *loi de* X la mesure image de $\mathbb{P}$ par X, $\mathbb{P} \circ X^{-1}$, notée $\mathbb{P}_X$. Autrement dit, $\mathbb{P}_X[B] = \mathbb{P}[X \in B]$ pour tout $B \in \mathcal{E}$.

Exemple : Quand $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on parle de variable aléatoire réelle et quand $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $d \geq 2$, on parle de vecteur aléatoire.

Notations :

- Une variable aléatoire se note *v.a.* en abrégé et on utilise *v.a.r.* pour variable aléatoire réelle.
- On utilise en général des majuscules pour désigner des variables aléatoires et des minuscules pour les valeurs prises par les v.a. Par exemple

$$[X = x] = \{\omega \in \Omega / \ X(\omega) = x\}.$$

- Si X est de loi $\mathbf{P}$, on dit que X suit la loi $\mathbf{P}$ et on note $X \rightsquigarrow \mathbf{P}$ ou $\mathcal{L}(X) = \mathbf{P}$.
- Si deux variables aléatoires X et Y ont même loi, on note $\mathcal{L}(X) = \mathcal{L}(Y)$ ou $X \sim Y$.

Définition 2.8

(i) Si X est une v.a. telle que $\mathcal{L}(X) = \mathcal{L}(-X)$, on dit que X est symétrique.

(ii) Si X et Y sont deux v.a. telles que $X = Y$ p.s. alors ont dit que Y est une version de X (et vice-versa).

Remarque : Si les v.a. X et Y sont des versions l'une de l'autre alors elles ont même loi.

Exemple : [Variables aléatoires discrètes]

Soit X une v.a. à valeurs dans $(E, \mathcal{E})$. On dit que X est discrète si, elle prend ses valeurs dans un ensemble dénombrable. On peut noter alors, par exemple, $\{x_k, \ k \in \mathbb{N}\}$ l'ensemble des valeurs prises par X. Dans ce cas, $\sigma(X)$ est la tribu engendrée par la partition $\{[X = x_k], \ k \in \mathbb{N}\}$ de Ω. La loi de X est alors

$$\mathbb{P}_X = \sum_{k \in \mathbb{N}} p_k \delta_{x_k}$$

où $p_k = \mathbb{P}[X = x_k]$ pour tout $k \in \mathbb{N}$. On en déduit que

$$\mathbb{P}_X[B] = \sum_{k / \ X_k \in B} p_k \text{ pour tout } B \in \mathcal{E}.$$

Toutes ces remarques sont vraies à un ensemble de mesure nulle près, aussi X est discrète si elle prend presque sûrement un ensemble dénombrable de valeurs.

2.3.2 Lois de vecteurs aléatoires

Par définition, la loi d'un vecteur aléatoire $X = (X_1, ..., X_n)$ de $\mathbb{R}^n$ est donnée par

$$\mathbf{P}_X[B] = \mathbb{P}[X \in B] \text{ pour } B \in \mathcal{B}(\mathbb{R}^n).$$

Mais on a vu que les mesures sur les espaces produits étaient entièrement déterminées par la donnée de leur valeur sur les pavés mesurables. Autrement dit, pour spécifier la loi du vecteur aléatoire X, il suffit de connaître

$$\mathbf{P}_X[B_1 \times ... \times B_n] = \mathbb{P}[X_1 \in B_1, ..., X_n \in B_n] \text{ pour } B_1, ..., B_n \in \mathcal{B}(\mathbb{R}).$$

Comme l'espace d'arrivée est réel, on peut même se contenter de connaître ces valeurs pour des pavés de $\mathbb{R}^n$ (i.e. des produits d'intervalles).

Remarquons que cette simplification reste valable pour le produit d'espaces probabilisés quelconques à l'arrivée.

Définition 2.9 (*Loi jointe, lois marginales*)

Soit $X = (X_1, ..., X_d)$ un vecteur aléatoire.

(i) La loi de X est également appelée *loi jointe* de $(X_1, ..., X_d)$ (l'ordre compte), ou encore *vraisemblance* de $(X_1, ..., X_d)$, en particulier en statistique lorsque ce vecteur est un échantillon d'une variable aléatoire.

(ii) Les *lois marginales* de X sont les lois des r-uplets $(X_{i_1}, ..., X_{i_r})$ pour tout entier r compris entre 1 et d et pour tout ensemble de r indices distincts i_k de $\{1, .., d\}$, $k = 1$ à r.

2.3.3 Variables aléatoires indépendantes.

Définition 2.10

Soient X_1, ..., X_n des variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans des espaces mesurés $(E_i, \mathcal{E}_i)$. On dit que les v.a. X_1, ..., X_n sont indépendantes si $\sigma(X_1)$, ,..., $\sigma(X_n)$ le sont, c'est-à-dire si

$$\mathbb{P}[X_1 \in B_1, ..., X_n \in B_n] = \mathbb{P}[X_1 \in B_1] ... \mathbb{P}[X_n \in B_n] \ \forall B_1 \times ... \times B_n \in \mathcal{E}_1 \times ... \times \mathcal{E}_n.$$

Remarque : Autrement dit, $\mathbb{P}_{(X_1,...,X_n)} = \mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_n}$. Donc les variables aléatoires qui composent un vecteur aléatoire sont indépendantes si la loi du vecteur est le produit des lois des variables.

2.4 Exercices

Exercice 2.1. [*Images de tribus*]
Soient Ω, Ω' deux ensembles non vides, $\mathcal{F}$ une tribu sur Ω et $f : \Omega \to \Omega'$. On pose

$$\mathcal{F}' = \{f(A), A \in \mathcal{F}\}.$$

Donner un exemple pour montrer que $\mathcal{F}'$ n'est pas une tribu en général. Montrer que si f est injective, $\mathcal{F}'$ est une tribu sur $f(\Omega)$.

Exercice 2.2.[*Un critère de mesurabilité*]
Soient Ω, Ω' deux ensembles non vides. Soient $\mathcal{F}$ (resp. $\mathcal{F}'$) une tribu sur Ω (resp. Ω') et f une application de Ω dans Ω'. Soit $(A_n, n \in \mathbb{N})$ une partition mesurable de Ω.
Montrer que f est $\mathcal{F}$-mesurable si et seulement si pour tout $n \geq 0$ la restriction $f_{|A_n}$ de f à A_n est mesurable par rapport à la tribu induite par $\mathcal{F}$ sur A_n noté $\mathcal{F}_{|A_n} = \{A_n \cap B, B \in \mathcal{F}\}$.

Exercice 2.3. [*Fonctions mesurables*] On pourra traiter la question 3 après avoir fait l'autonomie sur les limites supérieures et inférieures.

1. Montrer que toute fonction croissante $f : \mathbb{R} \to \mathbb{R}$ est borélienne.
2. Montrer que le supremum de deux fonctions mesurables est mesurable.
3. *Soit* $(f_n)_{n\geq 1}$ *une suite de fonctions mesurables. Montrer que* $\overline{\lim}_{n\to\infty} f_n$ *est mesurable. En déduire que la limite simple de fonctions mesurables est mesurable.*

Exercice 2.4. [*Theorème 2.2*] Faire la démonstration du théorème 2.2. On commencera par la réciproque. Pour le sens direct, on commencera par $g = \mathbf{1}_A$, puis on continuera par g étagée, g mesurable positive et enfin g mesurable et on construira φ à chaque étape.

Exercice 2.5. [*Indépendance*]
Soient X et Y deux v.a. indépendantes définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, de lois respectives $\mathbb{P}_X$ et $\mathbb{P}_Y$. Donner une condition nécessaire et suffisante pour que $\mathbb{P}_{(X,Y)} = \mathbb{P}_{(Y,X)}$.

Exercice 2.6. [*Conditionnement et Indépendance*]
Soient X_1 et X_2 deux v.a. indépendantes définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, de loi géométrique sur $\mathbb{N}^*$ de paramètre q. On définit les ensembles

$$A_1 = \{\omega \in \Omega | X_1(\omega) < X_2(\omega)\},\ A_2 = \{\omega \in \Omega | X_2(\omega) < X_1(\omega)\}$$

et

$$H = \{\omega \in \Omega | X_1(\omega) \neq X_2(\omega)\}.$$

1. On considère les v.a
$$X = \min(X_1, X_2) \text{ et } J = \mathbf{1}_{A_1} + 2\mathbf{1}_{A_2}$$
On note $\mathbb{P}^H$ la probabilité conditionnelle sur l'espace $(\Omega, \mathcal{F})$ définie , pour tout $A \in \mathcal{F}$, par $\mathbb{P}^H(A) = \mathbb{P}[A|H]$.
   (a) Calculer, pour $k \in \mathbb{N}^*$, la probabilité $\mathbb{P}[H \cap (J = 1) \cap (X > k)]$.
   (b) Justifier l'égalité $\mathbb{P}(A_1) = \mathbb{P}(A_2)$ et en déduire $\mathbb{P}[H]$.
   (c) Calculer pour tout $k \in \mathbb{N}$, la valeur de la probabilité $\mathbb{P}^H[(J = 1) \cap (X > k)]$.
   (d) Que vaut alors $\mathbb{P}^H[(J = 1)]$?
2. On étudie une propriété d'indépendance des v.a X et J sous la probabilité $\mathbb{P}^H$.
   (a) Remarquer que l'on a l'égalité
$$\mathbb{P}[H \cap (J = 1) \cap (X > k)] = \mathbb{P}[H \cap (J = 2) \cap (X > k)]$$
   et en déduire de ce qui précède la valeur de la probabilité de $\mathbb{P}^H(X > k)$.
   (b) Montrer que les v.a. X et J sont indépendantes lorsqu'on les considère comme définies sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P}^H)$.

Chapitre 3

Intégrale de Lebesgue et moments de variables aléatoires.

Sommaire

3.1 Intégrale de Lebesgue.

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré.

3.1.1 Définition. Construction.

Définition 3.1 (*Intégrale des fonctions étagées*)

Soit f une fonction étagée définie par

$$f = \sum_{i=1}^{N} \alpha_i \mathbf{1}_{A_i}$$

pour $N \in \mathbb{N}^*$, $\alpha_i \in \mathbb{R}$ et $A_i \in \mathcal{F}$ pour tout $i = 1$ à N. On suppose

- soit $\mu(A_i) < +\infty$ pour tout $i = 1$ à N,
- soit $\alpha_i \geq 0$ pour tout $i = 1$ à N.

Alors l'intégrale de Lebesgue de f sur Ω par rapport à (relativement à) la mesure μ est le réel

$$\int_\Omega f d\mu = \sum_{i=1}^{N} \alpha_i \mu(A_i) \in \overline{\mathbb{R}},$$

avec la convention $0.\mu(A_i) = 0$ quand $\mu(A_i) = +\infty$.

Pour construire l'intégrale des fonctions mesurables positives, on va approcher celles-ci par des suites croissantes de fonctions étagées et définir l'intégrale de la fonction mesurable positive comme limite croissante de la suite des intégrales des fonctions étagées.

Lemme 3.1

Soient g et h deux fonctions étagées telles que $g \leq h$. Alors $\int_\Omega g d\mu \leq \int_\Omega h d\mu$.

Démonstration : On remarque que l'on peut trouver un entier $N \in \mathbb{N}^*$, $(\alpha_i)_{i=1..N}$, $(\beta_i)_{i=1..N}$ des suites réelles et $(B_i)_{i=1..N}$ une suite d'éléments disjoints de $\mathcal{F}$ tels que

$$g = \sum_{i=1}^{N} \alpha_i \mathbf{1}_{B_i} \text{ et } h = \sum_{i=1}^{N} \beta_i \mathbf{1}_{B_i}.$$

Alors $g \leq h$ implique $\alpha_i \leq \beta_i$ pour tout $i = 1..N$ et la définition de l'intégrale donne clairement le résultat escompté. □

Lemme 3.2

Soit $(h_n)_{n\in\mathbb{N}}$ une suite croissante de fonctions étagées positives et soit g une fonction étagée telles que $\lim_{n\to+\infty} h_n \geq g$. Alors

$$\lim_{n\to+\infty} \int_\Omega h_n d\mu \geq \int_\Omega g d\mu.$$

Démonstration : On suppose

$$g = \sum_{i=1}^{N} \alpha_i \mathbf{1}_{A_i}$$

pour $N \in \mathbb{N}^*$, $\alpha_i \in \mathbb{R}$ et $A_i \in \mathcal{F}$ pour tout $i = 1$ à N. Pour tout $\varepsilon > 0$ et pour tout $n \in \mathbb{N}$, on note $E_n = \{\omega \in \Omega / \; h_n(\omega) \geq (1-\varepsilon)g(\omega)\}$. On montre facilement que les E_n sont des ensembles mesurables (h_n et g mesurables), qu'ils forment une suite croissante de $\mathcal{F}$ (la suite $(h_n)_{n\in\mathbb{N}}$ est croissante) et que $\bigcup_{n\in\mathbb{N}} E_n = \Omega$. En effet, si $\omega \in \Omega$ est tel que $g(\omega) \leq 0$, $\omega \in E_n$ pour tout n et si $g(\omega) > 0$,

$$(1-\varepsilon)g(\omega) < g(\omega) \leq \lim_{n\to+\infty} h_n(\omega)$$

et donc $\omega \in E_n$ pour n assez grand.
Pour tout $n \in \mathbb{N}$ et pour tout $\varepsilon > 0$, comme l'intégrale est croissante,

$$\int_\Omega h_n d\mu \geq \int_\Omega \mathbf{1}_{E_n} h_n d\mu \geq (1-\varepsilon) \int_\Omega \mathbf{1}_{E_n} g d\mu = (1-\varepsilon) \sum_{i=1}^{N} \alpha_i \mu(A_i \cap E_n).$$

Comme $(E_n)_{n\in N}$ est croissante et d'union Ω, $(A_i \cap E_n)_{n\in\mathbb{N}}$ est croissante et d'union A_i pour tout $i = 1..N$. Comme la suite des intégrales des h_n est une suite réelle croissante, sa limite est bien définie et on déduit des inégalités précédentes

$$\lim_{n\to+\infty} \int_\Omega h_n d\mu \geq (1-\varepsilon) \int_\Omega g d\mu$$

pour tout $\varepsilon > 0$ d'où le résultat. □

Etant donnée une fonction f mesurable positive, il existe au moins une suite croissante de fonctions étagées positives qui converge vers f. Si $(h_n)_{n\in\mathbb{N}}$ et $(g_n)_{n\in\mathbb{N}}$ sont deux telles suites alors pour tout $N \in \mathbb{N}$, on a par exemple

$$\lim_{n\to+\infty} h_n \geq g_N.$$

On en déduit par le précédent lemme que

$$\lim_{n\to+\infty} \int_\Omega h_n d\mu \geq \int_\Omega g_N d\mu$$

pour tout N et donc

$$\lim_{n\to+\infty} \int_\Omega h_n d\mu \geq \lim_{N\to+\infty} \int_\Omega g_N d\mu.$$

En inversant les rôles des deux suites on conclut que ces limites sont égales et donc que la limite ne dépend pas de la suite choisie. Ceci conduit aux définitions suivantes.

Définition 3.2 (*Intégrale des fonctions mesurables positives*)

Soit f une fonction mesurable positive. Etant donnée $(f_n)_{n\in\mathbb{N}}$ une suite croissante de fonctions étagées convergeant vers f, on définit l'intégrale de Lebesgue de f comme l'élément de $\overline{\mathbb{R}}_+$

$$\int_\Omega f d\mu = \lim_{n\to+\infty} \int_\Omega f_n d\mu.$$

Définition 3.3 (*fonctions intégrables et intégrale*)

Soit f une fonction mesurable. On dit qu'elle est intégrable si

$$\int_\Omega |f| d\mu < +\infty$$

et on définit alors son intégrale de Lebesgue comme

$$\int_\Omega f d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu.$$

Remarque :

(i) Soit $E \in \mathcal{F}$. Pour tout f mesurable positive, on définit l'intégrale de f sur E par

$$\int_E f d\mu = \int_\Omega \mathbf{1}_E f d\mu.$$

Cette définition coïncide avec l'intégrale définie sur l'espace mesuré $(E, \mathcal{F}_E, \mu_E)$ où

$$\mathcal{F}_E = \{A \cap E /\ A \in \mathcal{F}\} \text{ est la tribu des événements restreints à } E,$$

et μ_E est la restriction de μ à $\mathcal{F}_E \subset \mathcal{F}$. Si f mesurable de signe quelconque est telle que

$$\int_E |f| d\mu < +\infty,$$

on définit son intégrale sur E comme

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu.$$

Dans $\mathbb{R}$, on utilise la notation

$$\int_a^b f d\mu = \int_{[a,b]} f d\mu.$$

(ii) Si $f = (f_1, ..., f_d)$ est mesurable à valeurs dans $\mathbb{R}^d$, on dit que f est intégrable si tous les f_i sont intégrables et on note

$$\int_\Omega f d\mu = \left(\int_\Omega f_1 d\mu, ..., \int_\Omega f_d d\mu \right).$$

(iii) De même on peut définir une intégrale des fonction à valeurs complexes. Si f à valeurs dans $\mathbb{C}$ est telle que $\mathrm{Re}(f)$ et $\mathrm{Im}(f)$ sont intégrables, on définit l'intégrale de f par

$$\int_\Omega f d\mu = \int_\Omega \mathrm{Re}(f) d\mu + i \int_\Omega \mathrm{Im}(f) d\mu.$$

Exemple : Avec $\Omega = \mathbb{R}^d$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^d)$ ou $\mathcal{L}(\mathbb{R}^d)$, $\mu = \lambda_d$, l'intégrale sur $\mathbb{R}^d$ coïncide avec l'extension de l'intégrale de Riemann à $\mathbb{R}^d$ quand elle est définie. On note indifféremment la mesure $d\lambda_d$ ou dx (ou dy...). Les applications intégrables pour la mesure de Lebesgue s'appellent des fonctions Lebesgue intégrables.

3.1.2 Propriétés de l'intégrale

Proposition 3.1

1. *L'ensembles des fonctions intégrables est stable par addition et multiplication par un scalaire. De plus, l'application* $f \to \int_\Omega f d\mu$ *est linéaire.*
2. *Si f est mesurable positive alors son intégrale est positive.*

On laisse au lecteur le soin de démontrer cette proposition en exercice en deuxième lecture. Pour le premier point, on montre que le résultat est vrai pour les fonctions étagées, puis pour les fonctions mesurables positives en les approchant par des suites croissantes de fonctions étagées. La preuve pour les fonctions intégrables est plus fastidieuse (beaucoup de cas à envisager) car les partie positives et négatives d'un nombre ne sont pas linéaires.

Pour le second point, on procèdera en suivant les même étapes. Il faut néanmoins remarquer que, si f est étagée, on peut toujours se ramener au cas où

$$f = \sum_{i=1}^N \alpha_i \mathbf{1}_{A_i}, \text{ avec } A_i \cap A_j = \varnothing \text{ pour } i \neq j.$$

Dans ce cas, si f est positive alors $\alpha_i \geq 0$ pour tout i.

Proposition 3.2

1. *Soit N un ensemble mesurable négligeable. Pour toute fonction f mesurable positive ou intégrable, on a :*
$$\int_N f d\mu = 0.$$
2. *Soit f une fonction mesurable nulle presque partout (ou p.p. en abrégé : cela signifie que $f = 0$ en dehors d'un ensemble négligeable). Alors f est intégrable et*
$$\int_\Omega f d\mu = 0.$$
3. *Soit f une application mesurable. Alors*
$$\int_\Omega |f| d\mu = 0 \Longrightarrow f = 0 \ \mu - \textit{presque partout}.$$
4. *Soient f et g deux applications mesurables positives ou intégrables. Alors*
$$f = g \ \mu - \text{p.p.} \ \Rightarrow \int_\Omega f d\mu = \int_\Omega g d\mu$$
$$f \leq g \ \mu - \text{p.p.} \ \Rightarrow \int_\Omega f d\mu \leq \int_\Omega g d\mu$$

On fera cette proposition en exercice à la fin de la section 3.1.

Proposition 3.3 (*Beppo-Levi*)

Soit $(f_n)_{n\geq 1}$ une suite croissante de fonctions mesurables positives. Alors $\lim_n f_n$ est mesurable positive (cf. chapitre 2) et
$$\lim_{n\to+\infty} \int_\Omega f_n d\mu = \int_\Omega \lim_{n\to+\infty} f_n d\mu.$$

Démonstration : Soit $f = \lim_n f_n$. La fonction f est mesurable positive et comme $f_n \leq f$ pour tout n, on a
$$\int_\Omega f_n d\mu \leq \int_\Omega f d\mu \text{ et donc } \lim_{n\to+\infty} \int_\Omega f_n d\mu \leq \int_\Omega f d\mu.$$
Pour l'autre inégalité, on utilise l'analogue du Lemme 3.2 pour les fonctions mesurables positives dont la preuve est similaires à celle de ce même lemme.

Lemme 3.3

Soit $(f_n)_n$ une suite croissante de fonctions mesurables positives telle que $\lim_n f_n \geq g$ où g est une fonction étagée. Alors
$$\lim_{n\to+\infty} \int_\Omega f_n d\mu \geq \int_\Omega g d\mu.$$

Soit alors $(g_n)_n$ une suite croissante de fonctions étagées convergeant vers f. D'après le lemme précédent,
$$\int_\Omega g_N d\mu \leq \lim_{n\to+\infty} \int_\Omega f_n d\mu.$$

Mais, par définition de l'intégrale,

$$\int_\Omega f d\mu = \lim_N \int_\Omega g_N d\mu$$

et on obtient ainsi la seconde inégalité.

Corollaire 3.1 (*Beppo-Levi pour les séries de fonctions mesurables positives*)

Soit $(u_n)_n$ une suite de fonction mesurables positives. Alors

$$\int_\Omega \sum_{n\geq 1} u_n d\mu = \sum_n \int_\Omega u_n d\mu.$$

Démonstration : Il suffit de poser $f_N = \sum_{n=1}^{N} u_n$ et d'appliquer Beppo-Levi à la suite $(f_N)_N$.

Proposition 3.4 (*Relation de Chasles*)

Soit $(A_n)_n$ une suite d'événements de $\mathcal{F}$ deux à deux disjoints. Alors, pour tout f mesurable positive ou intégrable

$$\int_{\bigcup_n A_n} f d\mu = \sum_n \int_{A_n} f d\mu.$$

On fera la preuve en exercice en utilisant le corollaire précédent et en montrant que

$$\mathbf{1}_{\bigcup_n A_n} = \sum_n \mathbf{1}_{A_n}.$$

Proposition 3.5 (*Exemples fondamentaux*)

1. *On suppose $\mu = \delta_{\omega_0}$ la masse de Dirac en ω_0 où ω_0 est un élément particulier de Ω. Alors toutes les fonctions mesurables à valeurs dans $\mathbb{R}$ sont intégrables et*

$$\int_\Omega f d\mu = f(\omega_0).$$

2. *On suppose $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ où*

$$\mu(B) = \int_B f(x)dx, \ \forall B \in \mathcal{B}(\mathbb{R}^d)$$

pour une certaine fonction f mesurable positive. Alors pour toute fonction ϕ mesurable telle que $|\phi| f$ est Lebesgue intégrable, on a

$$\int_{\mathbb{R}^d} \phi d\mu = \int_{\mathbb{R}^d} \phi(x) f(x) dx.$$

Démonstration : La preuve de ces résultats est constructive : on commence par définir l'intégrale des fonctions étagées, puis de toutes les fonctions mesurables positives puis des fonctions intégrables en suivant le processus de construction des fonctions intégrables et en appliquant les résultats de convergence appropriés. Il est très important de savoir faire ce type de preuves (comme la linéarité de l'intégrale vue plus haut).

1. Soit f une fonction étagée définie par
$$f = \sum_{i=1}^{N} \alpha_i \mathbf{1}_{A_i}.$$
Par définition de l'intégrale,
$$\int_\Omega f d\mu = \sum_{i=1}^{N} \alpha_i \mu(A_i).$$
Il faut donc pour commencer savoir définir $\mu(A)$ pour un $A \in \mathcal{F}$. Or
$$\mu(A) = \delta_{\omega_0}(A). \tag{3.1}$$
Ceci vaut 1 si $\omega_0 \in A$ et 0 si $\omega_0 \notin A$. On en déduit donc que
$$\delta_{\omega_0}(A) = \mathbf{1}_A(\omega_0).$$
En appliquant ce résultat à l'application f, on obtient, par la définition de l'addition des fonctions,
$$\int_\Omega f d\mu = \sum_{i=1}^{N} \alpha_i \mathbf{1}_{A_i}(\omega_0) = \left(\sum_{i=1}^{N} \alpha_i \mathbf{1}_{A_i} \right)(\omega_0) = f(\omega_0)$$
ce qui établit le résultat souhaité pour f étagée.
Pour la seconde étape, on choisit f mesurable positive et $(f_n)_n$ une suite croissante de fonctions étagées convergeant vers f. Alors d'une part, par définition de l'intégrale,
$$\int_\Omega f d\mu = \lim_n \int_\Omega f_n d\mu$$
et d'autre part, par application du résultat pour les fonctions étagées et par convergence simple de la suite $(f_n)_n$,
$$\lim_{n\to+\infty} \int_\Omega f_n d\mu = \lim_{n\to+\infty} f_n(\omega_0) = f(\omega_0) \tag{3.2}$$
ce qui conclut la preuve pour f mesurable positive. Enfin, on suppose que f est une fonction mesurable quelconque. Par définition de l'intégrale,
$$\int_\Omega f d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu.$$
On utilise alors successivement le résultat obtenu pour les fonctions mesurables positives et la définition de l'addition des fonctions pour obtenir
$$\int_\Omega f d\mu = f^+(\omega_0) - f^-(\omega_0) = (f^+ - f^-)(\omega_0) = f(\omega_0) \tag{3.3}$$
ce qui permet de conclure.
2. On procède dans le second cas exactement de la même façon (exercice). Attention, les justifications à apporter sont différentes à chaque étape : on veillera à bien trouver l'équivalent du calcul de l'étape (3.1), à justifier la convergence de l'étape (3.2) et enfin l'étape (3.3), la première étape étant fondamentale, la seconde technique et la troisième assez évidente en général.

Remarque : Soient μ_1 et μ_2 deux mesures définies sur $(\Omega, \mathcal{F})$. Si
$$\int_\Omega h d\mu_1 = \int_\Omega h d\mu_2 \text{ pour toute fonction mesurable bornée}$$
alors $\mu_1 = \mu_2$. Ceci est également vrai si on prend comme fonctions-test les fonctions mesurables positives. Pour s'en convaincre, il suffit de choisir les fonctions $h = \mathbf{1}_A$ pour $A \in \mathcal{F}$, ces fonctions étant mesurables positives et bornées.

Théorème 3.1 (*Formule de transfert*)

Soient $(\Omega', \mathcal{F}')$ un espace mesurable et h une application mesurable de $(\Omega, \mathcal{F})$ dans $(\Omega', \mathcal{F}')$. On définit $\mu_h = \mu \circ h^{-1}$ la mesure image de μ par h définie sur $(\Omega', \mathcal{F}')$. Alors pour toute fonction f mesurable positive sur $(\Omega', \mathcal{F}')$,

$$\int_{\Omega'} f d\mu_h = \int_{\Omega} f \circ h d\mu \tag{3.4}$$

De plus, si f est une application mesurable sur $(\Omega', \mathcal{F}')$, alors f est μ_h-intégrable si et seulement si $f \circ h$ est μ-intégrable sur $(\Omega, \mathcal{F})$ et, lorsque f est μ_h-intégrable, l'égalité (3.4) est satisfaite.

On laisse la preuve en exercice : il faut procéder exactement comme pour la preuve de la proposition 3.5.

Nous présentons ici un résultat connus dans le cas de réel dans un cadre plus abstrait.

Théorème 3.2 (*Théorème de Fubini*)

Soient $(\Omega, \mathcal{F}, \mu)$ et $(\Omega', \mathcal{F}', \nu)$ deux espaces mesurés et soit f une fonction réelle borélienne définie sur $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$. On suppose f positive (resp. $\mu \otimes \nu$-intégrable). Alors, l'application définie μ- presque sûrement

$$\omega \mapsto \int_{\Omega'} f(\omega, \omega') \nu(d\omega')$$

est borélienne positive (resp. μ-intégrable) sur $(\Omega, \mathcal{F})$, et l'application définie ν-presque sûrement

$$\omega' \mapsto \int_{\Omega} f(\omega, \omega') \mu(d\omega)$$

est borélienne positive (resp. ν-intégrable) sur $(\Omega', \mathcal{F}')$. De plus

$$\begin{aligned} \int_{\Omega \times \Omega'} f \, d\mu \otimes \nu &= \int_{\Omega} \left(\int_{\Omega'} f(\omega, \omega') \nu(d\omega') \right) \mu(d\omega) \\ &= \int_{\Omega'} \left(\int_{\Omega} f(\omega, \omega') \mu(d\omega) \right) \nu(d\omega'). \end{aligned}$$

Notation : Ici, afin de pouvoir précise la variable d'intégration, on utilise la notation usuelle

$$\int_{\Omega} f d\mu = \int_{\Omega} f(\omega) \mu(d\omega).$$

Enfin, nous rappelons le théorème de changement de variable que nous utilisons fréquemment en modélisation.

Théorème 3.3 (*changement de variable dans $\mathbb{R}^d$*)

Soient D et D' deux ouverts de $\mathbb{R}^d$ et ψ un C^1-difféomorphisme de D dans D' (ψ est bijective de D sur D', de classe C^1 et d'inverse C^1) . On note $D\psi$ la différentielle de ψ . Soit f une fonction borélienne réelle définie sur D'. La fonction f est intégrable sur D' si et seulement si $x \mapsto f(\psi(x))|\det D\psi(x)|$ est intégrable sur D et si f est positive ou intégrable sur D',

$$\int_{D'} f(x')dx' = \int_D f(\psi(x))|\det D\psi(x)|dx.$$

Remarque : $|\det D\psi(x)|$ s'appelle le jacobien de ψ. On a la relation

$$|\det D\psi^{-1}(y)| = \frac{1}{|\det D\psi(\psi^{-1}(y))|}.$$

Remarque : A ce stade, il faudrait se remémorer (et démontrer) tous les résultats classiques utilisés pour le calcul intégral, ce qui est hors de portée de ce cours qui doit présenter de nombreux autres aspects de la théorie des probabilités. Voici une liste non limitative de résultats que vous devriez connaître, veillez à vérifier vos connaissances :

- Lemme de Fatou (voir chapitre 4 sur les suites)
- Théorème de convergence dominée
- Intégrales dépendant d'un paramètre (convergence, continuité, dérivabilité et dérivation par rapport au paramètre)
- Etude des espaces vectoriels normés de fonctions intégrables, de carré intégrable, etc. L^1, L^2, L^p et de l'espace vectoriel normé des fonctions essentiellement bornées (i.e. bornées en dehors d'un ensemble négligeable) L^∞.
- Inclusions remarquables entre ces espaces, inégalités remarquables sur les normes, comparaison des convergence dans les espaces de mesure finie ($\mu(\Omega) < +\infty$).

3.1.3 Absolue continuité des mesures

Définition 3.4

Soient μ et ν deux mesures définies sur $(\Omega, \mathcal{F})$. La mesure μ est dite absolument continue par rapport à la mesure ν, ce que l'on note $\mu << \nu$, si

$$\forall\ A \in \mathcal{F},\ \nu(A) = 0 \Rightarrow \mu(A) = 0.$$

Théorème 3.4 (*Radon-Nikodym - admis*)

Soient μ et ν deux mesures σ-finies. Si $\mu << \nu$ alors il existe f mesurable de $(\Omega, \mathcal{F})$ dans $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ telle que

$$\mu(A) = \int_A f d\nu.$$

La fonction f s'appelle la densité de μ par rapport à ν. On la note aussi

$$f = \frac{d\mu}{d\nu}.$$

Remarque : Réciproquement, si μ admet une densité f par rapport à ν alors μ est absolument continue par rapport à ν. (Exercice).

3.2 Espérance mathématique

Sot $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

3.2.1 Espérance, matrice de covariance

Définition 3.5

Soit $X = (X_1, ..., X_d)$ une variable aléatoire définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs ans $\mathbb{R}^d$. Si X est $\mathbb{P}$-intégrable, on appelle espérance de X le vecteur

$$\mathbb{E}[X] = \left(\int_\Omega X d\mathbb{P}, ..., \int_\Omega X_d \, d\mathbb{P}\right)^t.$$

Remarque : [Application de la formule de transfert]
Soit $\varphi : \mathbb{R}^d \to \mathbb{R}$ borélienne positive. Alors

$$\mathbb{E}[\varphi(X)] = \int_\Omega \varphi(X) d\mathbb{P} = \int_{\mathbb{R}^d} \varphi \, d\mathbb{P}_X \tag{3.5}$$

où $\mathbb{P}_X$ est la loi de X. Si φ est borélienne de signe quelconque alors $\varphi(X)$ est intégrable si et seulement si φ est $\mathbb{P}_X$-intégrable et dans ce cas (3.5) est satisfaite.
On en déduit que X est intégrable si et seulement si l'aplication $x \to |x|$ est $\mathbb{P}_X$-intégrable, c'est-à-dire si

$$\int_{\mathbb{R}^d} |x| \mathbb{P}_X(dx) < +\infty$$

où l'on a utilisé la notation $\mathbb{P}_X(dx)$ pour $d\mathbb{P}_X$ pour faire apparaître le nom de la variable d'intégration.
Si X est intégrable, on a

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P} = \int_{\mathbb{R}^d} x \mathbb{P}_X(dx).$$

Proposition 3.6

Soient X et Y deux vecteurs aléatoires de $\mathbb{R}^d$, intégrables sur Ω et soit $a \in \mathbb{R}^d$.

1. *Si $X = a$ p.s. alors $\mathbb{E}[X] = a$.*
2. $\mathbb{E}[aX + Y] = a\mathbb{E}[X] + \mathbb{E}[Y]$.
3. *Si X v.a.r. est telle que $X \geq 0$ p.s. alors $\mathbb{E}[X] \geq 0$.*

Définition 3.6 (*Moments d'ordre k*)

On note $L^k(\Omega, \mathcal{F}, \mathbb{P})$ l'espace vectoriel des fonctions mesurables X à valeurs dans $\mathbb{R}$ telles que $|X|^k$ est intégrable sur Ω. Pout tout $X \in L^k(\Omega, \mathcal{F}, \mathbb{P})$, on appelle moment d'ordre k de X, pour $k \in \mathbb{N}^*$, la valeur $\mathbb{E}\left[X^k\right]$ et moment centré d'ordre k de X la quantité $\mathbb{E}\left[(X - \mathbb{E}[X])^k\right]$.

Théorème 3.5 (*Caractérisation de la loi d'un vecteur aléatoire*)

Soit X un vecteur aléatoire de $\mathbb{R}^d$. La loi de X est caractrisée par la donnée des $\mathbb{E}[\varphi(X)]$ pour toute fonction $\varphi : \mathbb{R}^d \to \mathbb{R}$ borélienne bornée.

Démonstration : On remarque que pour toute fonction borélienne bornée φ, $\varphi(X)$ est intégrable. On suppose la loi de X connue. Alors $\mathbb{E}[\varphi(X)]$ est donnée par la formule de transfert. Réciproquement, si toutes les espérances sont connues alors il suffit de choisir les $\varphi = \mathbf{1}_B$ pour B borélien de $\mathbb{R}^d$ pour calculer $\mathbb{P}_X(B)$.

Proposition 3.7 (*Inégalité de Jensen*)

Soit X une v.a.r. à valeurs dans un intervalle $I \subset \mathbb{R}$ et soit φ une fonction convexe sur I. Si X et $\varphi(X)$ sont intégrables alors

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

L'inégalité est stricte si φ est strictement convexe et X non égale à une constante presque sûrement.

Démonstration : Comme φ est convexe, on a la proposition suivante : pour tout $z \in I$, il existe $\lambda_z \in \mathbb{R}$ tel que

$$\varphi(x) \geq \varphi(z) + \lambda_z(x-z) \text{ pour tout} x \in I.$$

(Faire un dessin !). On l'applique à $z = \mathbb{E}[X]$ et à $x = X(\omega)$ pour tout $\omega \in \Omega$. Il vient

$$\varphi(X(\omega)) \geq \varphi(\mathbb{E}[X]) + \lambda_z(X(\omega) - \mathbb{E}[X]) \text{ pour tout } \omega \in \Omega.$$

On intègre alors sur Ω et on conclut en remarquant que l'intégrale sur Ω vaut 1 et qu'elle est linéaire.

Corollaire 3.2 (*Applications*)

Soit X à valeurs réelles.

. *Si X est intégrable,* $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

. *Si X est de carré intégrable,* $\mathbb{E}[X]^2 \leq \mathbb{E}\left[X^2\right]$.

. *Si X a un moment exponentiel (i.e. si $\exp(X)$ est intégrable),* $\exp(\mathbb{E}[X]) \leq \mathbb{E}[\exp(X)]$.

Définition 3.7 (*Covariance d'un vecteur aléatoire*)

Soit $X = (X_1, ..., X_d)$ un vecteur aléatoire. On suppose que toutes les composantes sont de carré intégrable (on dit alors que X est de carré intégrable). On définit la matrice de covariance de X par

$$\Gamma_X = \mathbf{Cov}[X] = (\mathrm{cov}(X_i, X_j))_{\substack{i=1..d \\ j=1..d}} = \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^t\right].$$

Proposition 3.8 (*Propriété de la matrice de covariance*)

Soit X un vecteur aléatoire de $\mathbb{R}^d$ de carré intégrable et $a \in \mathbb{R}^d$.

1. $\Gamma_X = \mathbb{E}[XX^t] - \mathbb{E}[X]\mathbb{E}[X]^t$.
2. $\Gamma_{X+a} = \Gamma_X$.
3. *La matrice Γ_X est symétrique, semi-définie-positive.*

Démonstration : Faire les deux premiers points en exercice. Dans le troisième, la symétrie est évidente. Pour le dernier point, soit $u \in \mathbb{R}^d$. On doit montrer que $\langle \Gamma_X u, u \rangle \geq 0$, où $\langle \cdot, \cdot \rangle$ désigne le produit scalaire dans $\mathbb{R}^d$. On pose $Y = X - \mathbb{E}[X]$. Il vient $\Gamma_X = \mathbb{E}\left[YY^t\right]$ et

$$\langle \Gamma_X u, u \rangle = \langle \mathbb{E}\left[YY^t\right] u, u \rangle = u^t \mathbb{E}\left[YY^t\right] u = \mathbb{E}\left[(Y^t u)^t Y^t u\right] = \mathbb{E}\left[|Y^t u|^2\right] \geq 0.$$

Proposition 3.9

Soit X un vecteur aléatoire de $\mathbb{R}^d$, soit A une application linéaire de $\mathbb{R}^d$ dans $\mathbb{R}^{d'}$ et soit b un vecteur de $\mathbb{R}^{d'}$. On suppose X de carré intégrable. Alors $AX+b$ est de carré intégrable et

$$\mathbb{E}\left[AX+b\right] = A\mathbb{E}\left[X\right] + b, \text{ et } \mathbf{Cov}\left[AX+b\right] = A\mathbf{Cov}\left[X\right]A^t.$$

Cette proposition est à démontrer pour s'entraîner, tout comme la remarque suivante.

Remarque : Si les composantes d'un vecteur aléatoire sont décorrélées ou indépendantes, la matrice de covariance du vecteur est diagonale.

Définition 3.8 (*Covariance de deux vecteurs*)

Soient X et Y deux vecteurs aléatoires de carré intégrable, $X \in \mathbb{R}^d$ et $Y \in \mathbb{R}^{d'}$. La matrice de covariance de X est Y est définie par

$$\mathbf{Cov}\left[X,Y\right] = \left(\mathrm{cov}(X_i, Y_j)\right)_{\substack{i=1..d \\ j=1..d'}} = \mathbb{E}\left[(X-\mathbb{E}\left[X\right])(Y-\mathbb{E}\left[Y\right])^t\right] \in \mathcal{M}_{d\times d'}(\mathbb{R}).$$

Remarque :

(i) Attention à la confusion, la notation **Cov** est utilisée indistinctement pour les vecteurs ou les couples de vecteurs. Ici, on doit lire $\Gamma_X = \mathbf{Cov}\left[X,X\right]$.

(ii) Tout matrice de taille $d\times d'$ est la matrice de deux vecteurs aléatoires de tailles d et d' respectivement. Pour qu'un matrice Γ de taille $d\times d$ soit la covariance d'un vecteur X de $\mathbb{R}^d$, il faut et il suffit que Γ soit positive et symétrique.

Proposition 3.10

Sous les hypothèses de la définition 3.8, on a :

1. $\mathbf{Cov}\left[X,Y\right] = \mathbf{Cov}\left[Y,X\right]^t$.
2. *Si*
$$Z = \begin{pmatrix} X \\ Y \end{pmatrix}$$
alors
$$\mathbf{Cov}\left[Z\right] = \begin{pmatrix} \Gamma_X & \mathbf{Cov}\left[X,Y\right] \\ \mathbf{Cov}\left[Y,X\right] & \Gamma_Y \end{pmatrix}.$$

3.2.2 Variables aléatoires discrètes

Définition 3.9 (*Variables aléatoires discrètes*)

Une variable aléatoire X à valeurs dans $(E, \mathcal{B}(E))$ est dite *discrète* si elle prend p.s. ses valeurs dans un ensemble fini ou dénombrable, c'est-à-dire s'il existe un ensemble dénombrable $F \in \mathcal{B}(E)$ tel que $\mathbb{P}\left[X \in F\right] = 1$.

Remarque : [Notations et formules usuelles]

Si X est discrète et $F = \{x_i,\ i \in \mathbb{N}\}$, on a

$$X = \sum_{i\in\mathbb{N}} x_i \mathbf{1}_{[X=x_i]} \text{ et } \mathbb{P}_X = \sum_{i\in\mathbb{N}} \mathbb{P}\left[X=x_i\right]\delta_{x_i}.$$

La condition “X est intégrable” s'écrit

$$\sum_{i\in\mathbb{N}} |x_i|\mathbb{P}\left[X = x_i\right] < +\infty$$

et si X est à valeur dans $\mathbb{R}$ ou $\mathbb{R}^d$ et intégrable,

$$\mathbb{E}\left[X\right] = \sum_{i\in\mathbb{N}} x_i\mathbb{P}\left[X = x_i\right].$$

Enfin, pour toute fonction borélienne φ à valeurs dans $\mathbb{R}$ telle que $\varphi(X)$ est intégrable,

$$\mathbb{E}\left[\varphi(X)\right] = \sum_{i\in\mathbb{N}} \varphi(x_i)\mathbb{P}\left[X = x_i\right].$$

Les formules se démontrent en utilisant la formule de transfert et l'expression de l'intégrale pour la mesure de Dirac de la proposition 3.5.

3.2.3 Variables aléatoires à densité

Définition 3.10 (*Variables à densité*)

Une v.a.r X à valeurs dans un espace $(E, \mathcal{B}(E))$ ($E = \mathbb{R}$ ou $\mathbb{R}^d$) est dite *absolument continue* ou *continue* si sa loi est absolument continue par rapport à la mesure de Lebesgue sur l'espace d'arrivée. La densité de la loi par rapport à la mesure de Lebesgue sur $(E, \mathcal{B}(E))$ est appelée densité de X.

Si on note λ la mesure de Lebesgue sur $(E, \mathcal{B}(E))$, alors pour tout $B \in \mathcal{B}(E)$,

$$\mathbb{P}\left[X \in B\right] = \mathbb{P}_X[B] = \int_B f d\lambda$$

où $f \geq 0$ λ-p.p. et

$$\int_E f d\lambda = 1.$$

Dans ce cas, pour toute fonction borélienne $\varphi : E \to \mathbb{R}$,

$$\mathbb{E}\left[\varphi(X)\right] = \int_E \varphi f d\lambda$$

dès que

$$\int_E |\varphi| f d\lambda < +\infty.$$

Cela se démontre à nouveau en utilisant la formule de transfert et la proposition 3.5.

3.3 Transformée de Fourier et fonction caractéristique

Définition 3.11 (***Transformée de Fourier***)

1. Si $f \in L^1(\mathbb{R}^d)$, sa transformée de Fourier $\hat{f}$ est une fonction définie sur $\mathbb{R}^d$ par
$$\hat{f}(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} f(x)dx,\ t \in \mathbb{R}^d,\ \text{où } \langle t,x\rangle = \sum_{i=1}^{d} t_i x_i.$$
2. Si μ est une mesure finie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, sa transformée de Fourier est la fonction définie sur $\mathbb{R}^d$ par
$$\hat{\mu}(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} \mu(dx),\ t \in \mathbb{R}^d.$$

Remarque : Si μ admet une densité par rapport à la mesure de Lebesgue alors sa transformée de Fourier est celle de sa densité.

Proposition 3.11

Si $f \in L^1(\mathbb{R}^d)$, $\hat{f}$ est uniformément continue.

Théorème 3.6 (***Formule d'inversion de Fourier***)

Si f et $\hat{f}$ sont intégrables, alors f est continue et
$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(t) e^{-i\langle t,x\rangle} dt \text{ pour } \lambda\text{-presque tout } x \in \mathbb{R}^d.$$

Proposition 3.12

Si μ est une mesure finie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ de transformée de Fourier $\hat{\mu}$, alors
$$\mu(]x_1, y_1[\times ... \times]x_d, y_d[) = \lim_{t\to+\infty} \int_{-t}^{t} ... \int_{-t}^{t} \prod_{k=1}^{d} \frac{e^{-iu_k x_k} - e^{-iu_k y_k}}{iu_k} \hat{\mu}(u_1, ..., u_d) du_1 ... du_d.$$

Définition 3.12

Soit X un vecteur aléatoire de dimension d. La fonction caractéristique de X est la transformée de Fourier φ_X de sa loi, soit
$$\varphi_X(t) = \mathbb{E}\left[e^{i\langle t,X\rangle}\right],\ t \in \mathbb{R}^d,$$
soit, par la formule de transfert,
$$\varphi_X(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} \mathbb{P}_X(dx).$$

Proposition 3.13 (*Propriétés de la fonction caractéristique*)

1. *$\varphi_X(0) = 1$ et $|\varphi_X(t)| \leq 1$ pour tout $t \in \mathbb{R}^d$.*
2. *La fonction caractéristique caractérise la loi : deux vecteurs aléatoires de même fonction caractéristique ont même loi.*
3. *Si X a une densité f_X alors $\varphi_X = \hat{f}_X$.*
4. *$\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$ pour tout t. On en déduit X symétrique $\Leftrightarrow$ φ_X paire $\Leftrightarrow$ φ_X réelle.*
5. *Soit X un vecteur aléatoire de $\mathbb{R}^d$ et soient $A \in \mathcal{M}_{d' \times d}(\mathbb{R})$ et $b \in \mathbb{R}^{d'}$. Alors pour tout $t \in \mathbb{R}^{d'}$,*
$$\varphi_{AX+b}(t) = e^{i\langle t,b\rangle}\varphi_X({}^t\!At).$$
Si $d = d' = 1$, $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$.

Proposition 3.14

Si φ_X est intégrable, alors X a une densité par rapport à la mesure de Lebesgue donnée par
$$f_X(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-i\langle t,x\rangle}\varphi_X(t)dt \ \lambda\text{-p.p}$$

Exemple : Calculer la fonction caractéristique d'une v.a. binomiale de paramètres n et p.

Exemple : Calcul de la fonction caractéristique d'une loi normale : montrer que la fonction
$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{itx}e^{-\frac{1}{2}x^2}dx, \ x \in \mathbb{R},$$
est dérivable, calculer sa dérivée et en déduire qu'elle satisfait une équation différentielle ordinaire. La résoudre pour déterminer la fonction caractéristique d'une v.a. gaussienne centrée réduite. En déduire la fonction caractéristique d'une $\mathcal{N}(m, \sigma^2)$.

Théorème 3.7

Soit X une v.a. réelle.

1. *Si X^n est intégrable alors $\varphi_X \in C^n(\mathbb{R})$ et $\varphi_X^{(m)}(t) = i^m\mathbb{E}\left[X^m e^{itX}\right]$ pour tout $m \in \{0, ..., n\}$. En particulier, $\varphi_X^{(m)}(0) = i^m\mathbb{E}\left[X^m\right]$ pour tout $m \in \{0, ..., n\}$.*
2. *Si φ_X est k fois dérivable en 0 ($k \geq 2$), alors X admet des moments jusqu'à l'ordre $2\left[\frac{k}{2}\right]$. Ils sont donnés par $\varphi_X^{(m)}(0) = i^m\mathbb{E}\left[X^m\right]$ pour tout $1 \leq m \leq 2\left[\frac{k}{2}\right]$.*

Démonstration : Quelques éléments.

1. La proposition est vraie en $m = 0$ pour tout $t \in \mathbb{R}$. Elle se démontre ensuite par récurrence en dérivant sous le signe somme.
2. On remarque que $2\left[\frac{k}{2}\right] = k$ si k est pair. Sinon, c'est l'entier précédent. On admet ce résultat.

Remarque : Si $X = (X_1, ..., X_d)$, $\mathbb{E}\left[X_iX_j\right] = -\dfrac{\partial^2\varphi_X}{\partial t_i\partial t_j}(0)$ quand cela existe.

Exemple : Calcul des moments d'une loi normale. On eut montrer que si $X \rightsquigarrow \mathcal{N}(0, \sigma^2)$ alors
$$\mathbb{E}\left[X^{2k+1}\right] = 0 \text{ pour tout } k \geq 0$$
et
$$\mathbb{E}\left[X^{2k}\right] = \frac{(2k)!}{k!}\frac{\sigma^{2k}}{2^k}.$$

3.4 Vecteurs aléatoires indépendants

3.4.1 Caractérisation des vecteurs aléatoires indépendants

Théorème 3.8

Soient X_1, ...,X_n, n vecteurs aléatoires de dimensions respectives d_1,...,d_n. Les propositions suivantes sont équivalentes.

1. *X_1, ..., X_n sont indépendantes.*
2. *$\forall\ h_i : \mathbb{R}^{d_i} \to \mathbb{R}$ boréliennes bornées pour $i = 1..n$,*

$$\mathbb{E}\left[\prod_{i=1}^{n} h_i(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}\left[h_i(X_i)\right].$$

3. *$\varphi_{(X_1,...,X_n)}(t_1, ..., t_n) = \varphi_{X_1}(t_1)...\varphi_{X_n}(t_n)$ for all $(t_1, ..., t_n) \in \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_n}$.*

Remarque : L'expression dans 2. reste vraie dès que les intégrales ont un sens et que l'on est dans les conditions d'application d'un théorème de Fubini.

Démonstration :

☐ On montre l'équivalence entre 1. et 2. pour deux variables aléatoires, le passage à n se faisant par récurrence. Soient donc X et Y deux v.a. de $\mathbb{R}^d$ et $\mathbb{R}^{d'}$ respectivement et soient φ et ψ deux fonctions boréliennes bornées définies sur $\mathbb{R}^d$ et $\mathbb{R}^{d'}$ respectivement.

$$\mathbb{E}\left[\varphi(X)\psi(Y)\right] = \int_{\mathbb{R}^d\times\mathbb{R}^{d'}} \varphi(x)\psi(y)\mathbb{P}_{(X,Y)}(dx, dy).$$

Comme les v.a. sont indépendantes, $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$, et par Fubini,

$$\int_{\mathbb{R}^d\times\mathbb{R}^{d'}} \varphi(x)\psi(y)\mathbb{P}_X \otimes \mathbb{P}_Y(dx, dy) = \int_{\mathbb{R}^d} \varphi(x)\mathbb{P}_X(dx) \int_{\mathbb{R}^{d'}} \psi(y)\mathbb{P}_Y(dy).$$

On en déduit bien

$$\mathbb{E}\left[\varphi(X)\psi(Y)\right] = \mathbb{E}\left[\varphi(X)\right]\mathbb{E}\left[\psi(Y)\right].$$

Réciproquement, on remarque que, pour $A \in \mathcal{B}(\mathbb{R}^d)$ et $B \in \mathcal{B}(\mathbb{R}^{d'})$,

$$\mathbb{P}\left[X \in A, Y \in B\right] = \mathbb{E}\left[\mathbf{1}_{[X\in A, Y\in B]}\right] = \mathbb{E}\left[\mathbf{1}_{[X\in A]}\mathbf{1}_{[Y\in B]}\right] = \mathbb{E}\left[\mathbf{1}_A(X)\mathbf{1}_B(Y)\right]$$

et en appliquant 2. à $\varphi = \mathbf{1}_A$ et $\psi = \mathbf{1}_B$, il vient

$$\mathbb{E}\left[\mathbf{1}_A(X)\mathbf{1}_B(Y)\right] = \mathbb{E}\left[\mathbf{1}_A(X)\right]\mathbb{E}\left[\mathbf{1}_B(Y)\right]$$

soit

$$\mathbb{P}\left[X \in A, Y \in B\right] = \mathbb{P}\left[X \in A\right]\mathbb{P}\left[X \in B\right].$$

☐ 2. ⇒ 3. pour $(t_1, ..., t_n)$ fixé avec avec $h_i(x) = e^{i\langle t_i, x\rangle}$ pour $x \in \mathbb{R}^{d_i}$ pour tout $i \in \{1, ..., n\}$.

☐ 3. ⇒ 1. se montre en utilisant la formule d'inversion et Fubini. C'est assez technique.

Théorème 3.9

Soient X_1,...,X_n des vecteurs aléatoires de dimensions d_1,...,d_n respectivement.

1. *Si les vecteurs X_1, ..., X_n admettent des densités f_{X_1},...,f_{X_n} et sont indépendants alors $(X_1, ..., X_n)$ admet une densité définie par*

$$f_{(X_1,...,X_n)}(x_1, ..., x_n) = f_{X_1}(x_1)...f_{X_n}(x_n),\ \forall\ (x_1, ..., x_n) \in \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_n}. \quad (3.6)$$

2. *Si $(X_1, ..., X_n)$ admet une densité alors les X_i admettent aussi une densité. Dans ce cas, les v.a. X_1,..., X_n sont indépendantes si et seulement si (3.6) est satisfaite.*

Proposition 3.15

Soient X et Y deux vecteurs aléatoires indépendants de carré intégrable. Alors ils sont décorrélés.

Question : Justifier ce résultat avec les théorèmes et propositions précédents.

3.4.2 Convolution de mesures et somme de vecteurs aléatoires indépendants

Dans ce chapitre, on voit que la loi de la somme de v.a. indépendantes est la convolution de leurs lois.

Définition 3.13

Soient μ et ν deux mesures sur $(E, \mathcal{B}(E))$ espace vectoriel normé muni de sa tribu borélienne. La convolution ou produit de convolution de μ et ν, notée $\mu * \nu$ est la mesure image de $\mu \otimes \nu$ par l'application

$$\begin{array}{rccc} T: & E \times E & \to & E \\ & (x, y) & \to & x + y. \end{array}$$

Autrement dit, $\mu * \nu = \mu \otimes \nu \circ T^{-1}$ est définie sur $\mathcal{B}(E)$ et pour tout $B \in \mathcal{B}(E)$,

$$\mu * \nu(B) = \mu \otimes \nu(\{(x, y) \in E^2 / \ x + y \in B\}.$$

Proposition 3.16

(i) Pour toute fonction h borélienne positive sur E,

$$\int_E h\ d(\mu * \nu) = \int_{E^2} h(x + y)\mu \otimes \nu(dx, dy). \quad (3.7)$$

(ii) Pour toute fonction h borélienne h est $\mu * \nu$-intégrable si et seulement si $h \circ T$ est $\mu \otimes \nu$-intégrable sur E^2, et dans ce cas (3.7) est satisfaite.

Démonstration : Il suffit d'appliquer la formule de transfert, théorème 3.1.

Proposition 3.17

On suppose que $E = \mathbb{R}^d$ et que μ et ν sont respectivement de densités f et g par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, f et g étant intégrables. Alors $\mu * \nu$ admet pour densité $f * g$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ et $f * g$ est intégrable.

Démonstration : (à compléter). On montre que

$$\int_{\mathbb{R}^d} h\, d(\mu * \nu) = \int_{\mathbb{R}^d} h(f * g) d\lambda$$

pour toute fonction h borélienne bornée. Soit donc h une telle fonction. On utilise (3.7) puis le théorème de Fubini et enfin l'absolue continuité de μ et ν par rapport à la mesure de Lebesgue. On obtient donc

$$\int_{\mathbb{R}^d} h\, d(\mu * \nu) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} h(x+y) f(x) dx \right) g(y) dy.$$

Reste alors à faire le changement de variable $x \to z = x + y$ puis à appliquer à nouveau le théorème de Fubini. On obtient donc

$$\int_{\mathbb{R}^d} h\, d(\mu * \nu) = \int_{\mathbb{R}^d} h(z) \left(\int_{\mathbb{R}^d} f(z-y) g(y) dy \right) dz$$

ce qu'il fallait démontrer.

Proposition 3.18 (*Propriétés du produit de convolution*)

(i) Si μ et ν sont des probabilités, $\mu * \nu$ également.

(ii) Si μ et ν sont discrètes alors $\mu * \nu$ également. Plus précisément, si

$$\mu = \sum_k \mu_k \delta_{x_k} \quad \text{et} \quad \nu = \sum_k \nu_k \delta_{y_k}$$

alors

$$\mu * \nu = \sum_k \sum_l \mu_k \nu_l \delta_{x_k + y_l}.$$

En particulier, si $x_k = y_k = k$ pour tout k, alors

$$\mu * \nu = \sum_k \left(\sum_{l=0}^{k} \mu_l \nu_{k-l} \right) \delta_k.$$

Démonstration : A faire en exercice.

(i) Montrer que $\mu * \nu(E) = 1$ si $\mu(E) = \nu(E) = 1$.

(ii) On admet

$$\mu * \nu = \sum_k \sum_l \mu_k \nu_l \delta_{x_k} * \delta_{y_l}$$

et on calcule seulement, pour $x, y \in E$, $\delta_x * \delta_y(B)$ pour un $B \in \mathcal{B}(E)$. Enfin, on appliquera au cas entier (i.e. lorsque $x_k = y_k = k$ pour tout k).

Théorème 3.10

La transformée de Fourier du produit de convolution de deux fonctions ou deux mesures est le produit des transformées de Fourier.

Démonstration : Soient μ et ν deux mesures boréliennes finies. Pour tout $t \in \mathbb{R}^d$,

$$\widehat{\mu * \nu}(t) = \int_{\mathbb{R}^d} e^{i\langle t, x\rangle} (\mu * \nu)(dx) = \int_{\mathbb{R}^{2d}} e^{i\langle t, x+y\rangle} \mu(dx) \nu(dy).$$

On applique alors Fubini, il vient :

$$\widehat{\mu * \nu}(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\mu(dx) \int_{\mathbb{R}^d} e^{i\langle t,y\rangle}\nu(dy) = \hat{\mu}(t)\hat{\nu}(t).$$

Théorème 3.11

Soient X_1,..., X_n n vecteurs aléatoires indépendants à valeurs dans $\mathbb{R}^d$. Alors

1. $\mathbb{P}_{X_1+...+X_n} = \mathbb{P}_{X_1} * ... * \mathbb{P}_{X_n}$.
2. *Si X_1, ..., X_n ont respectivement pour densité f_1,..., f_n alors $X_1 + ... + X_n$ a pour densité $f_1 * ... * f_n$.*
3. $\varphi_{X_1+...+X_n}(t) = \varphi_{X_1}(t)...\varphi_{X_n}(t)$ *pour tout $t \in \mathbb{R}^d$.*

Démonstration : A faire en exercice.

Question : Soient X et Y deux v.a. exponentielles indépendantes de paramètres λ et μ respectivement, $\lambda \neq \mu$. Calculer la densité de $X + Y$.

Question : Soient X et Y deux lois normales indépendantes de paramètres (m, σ^2) et (m', σ'^2) respectivement. En utilisant les fonctions caractéristiques, déterminer la loi de $X + Y$.

3.5 Vecteurs gaussiens

Définition 3.14

Un vecteur aléatoire $X = (X_1, ..., X_d)^t$ est dit gaussien si toute combinaison linéaire de ses composantes est gaussienne.

Remarque :

(i) Si X est un vecteur gaussien, alors toutes ses composantes sont gaussiennes.

(ii) Pour toute fonction $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ affine, si X vecteur gaussien, alors $f(X)$ gaussien.

(iii) Notations usuelles : l'espérance d'un vecteur gaussien est notée $M = \mathbb{E}[X] \in \mathbb{R}^d$, sa matrice de covariance $\Gamma = \mathbf{Cov}[X] \in \mathcal{M}_d(\mathbb{R})$. On note $X \rightsquigarrow \mathcal{N}_d(M, \Gamma)$.

Exemple :

(i) $X = a \in \mathbb{R}^d$, $X \rightsquigarrow \mathcal{N}_d(a, 0)$.

(ii) Si $X \in \mathbb{R}^d$ a toutes ses composantes gaussiennes et indépendantes, X est gaussien.

Théorème 3.12

La fonction caractéristique φ d'un vecteur gaussien de loi $\mathcal{N}_d(M, \Gamma)$ est définie par

$$\varphi(t) = e^{i\langle M,t\rangle} e^{-\frac{1}{2}\langle \Gamma t,t\rangle} \text{ pour tout } t \in \mathbb{R}^d.$$

Démonstration : A faire en exercice. On remarquera que

$$\varphi_X(t) = \varphi_{\langle t,X\rangle}(1)$$

et que si X est gaussien, alors pour $t \in \mathbb{R}^d$, $\langle t, X\rangle$ est une v.a. gaussienne dont on calculera l'espérance et la variance.

Proposition 3.19

Soit $X \rightsquigarrow \mathcal{N}_d(M,\Gamma)$ et soit $Y = AX + B$, avec $A \in \mathcal{M}_{d'\times d}(\mathbb{R})$ et $B \in \mathbb{R}^{d'}$. Alors $Y \rightsquigarrow \mathcal{N}_{d'}(AM + B, A\Gamma A^t)$.

Théorème 3.13

Soit $X = (X_1, ..., X_d)^t \rightsquigarrow \mathcal{N}_d(M,\Gamma)$. Les composantes X_1,..., X_d sont mutuellement indépendantes si et seulement si elles sont deux à deux décorrélées, i.e. si Γ est diagonale.

Démonstration : On a montré l'une des implications proposition 3.15. Réciproquement, si Γ est diagonale, alors pour tout $t \in \mathbb{R}^d$,

$$\varphi_X(t) = e^{i\langle M,t\rangle} e^{-\frac{1}{2}\langle \Gamma t,t\rangle} = \prod_{i=1}^{d} e^{im_i t_i} \prod_{i=1}^{d} e^{-\frac{1}{2}\mathbf{Var}[X_i]t_i^2} = \prod_{i=1}^{d} \varphi_{X_i}(t_i),$$

donc les X_i sont indépendants.

Corollaire 3.3

Si $X \rightsquigarrow \mathcal{N}_d(0, I_d)$ alors X a pour densité

$$f_X(x) = \frac{1}{\sqrt{2\pi}^d} e^{-\frac{1}{2}\sum_{i=1}^d x_i^2}, \quad \textit{pour } x = (x_1, ..., x_n) \in \mathbb{R}^d.$$

Démonstration : La densité est le produit des densité des d v.a. gausiennes centrées réduites.

Remarque : Des v.a. gausiennes ne forment pas nécessairement un vecteur gaussien, même si elles sont décorrélées. Elles ne sont alors pas indépendantes. Voici un exemple à faire à titre d'exercice.

On considère les v.a. indépendantes X_1 gausienne centrée réduite et ε de Bernoulli qui prend les valeurs -1 et 1 avec probabilité 1/2. On définit alors $X_2 = \varepsilon X_1$. On montre que X_2 est une v.a. gaussienne centrée réduite et que X_1 et X_2 sont décorrélées. Mais on montre ensuite que X_1 et X_2 ne sont pas indépendantes et que (X_1, X_2) n'est pas un vecteur gaussien.

Théorème 3.14

Soit $X \rightsquigarrow \mathcal{N}_d(M,\Gamma)$ avec $\Gamma \neq 0$. Soit r le rang de Γ. Alors il existe $B \in \mathcal{M}_{d\times r}(\mathbb{R})$ et $Y \rightsquigarrow \mathcal{N}_r(0, I_r)$ tels que $X = BY + M$ et $BB^t = \Gamma$.

Démonstration : C'est un résultat d'algèbre et une transformation affine d'un vecteur gaussien. Dans les exercices, on construit B explicitement sur un exemple. La démonstration du résultat n'est que la généralisation évidente de cette preuve.

Corollaire 3.4

Soit $X \rightsquigarrow \mathcal{N}_d(M,\Gamma)$. X admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ si et seulement si Γ est inversible. Dans ce cas, elle s'écrit

$$f(x) = \frac{1}{\sqrt{2\pi}^d \sqrt{\det\Gamma}} e^{-\frac{1}{2}\langle \Gamma^{-1}(x-M), x-M\rangle} \quad \textit{pour } x \in \mathbb{R}^d.$$

Remarque : Si X a pour densité f et si $Y = AX + B$ avec A inversible, alors Y a pour densité la fonction

$$y \mapsto \frac{1}{|\det A|} f(A^{-1}(y - B))$$

Démonstration : (du corollaire). $X = BY + M$ avec $Y \rightsquigarrow \mathcal{N}_d(0, I_d)$ et $\Gamma = BB^t$. Si Γ est inversible, le rang de Γ est plein, B est carrée et $\det B \neq 0$ donc B est inversible. Calculer alors la densité de X à partir de la densité de Y.

Les vecteurs gaussiens sont très utilisés en statistique, notamment à cause du théorème limite central. On rappelle à titre d'exemple ici le théorème de Cochran : si $(X_1, ..., X_n)$ est un échantillon de la loi $\mathcal{N}(m, \sigma^2)$ (des v.a. indépendantes et toutes distribuées selon la loi $\mathcal{N}(m, \sigma^2)$), alors

$$\overline{X} = \frac{1}{n} X_i \text{ et } \Sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

sont indépendantes et

$$\overline{X} \rightsquigarrow \mathcal{N}(m, \frac{\sigma^2}{n}) \text{ et } \frac{n\Sigma_n^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}.$$

Ces résultats sont utilisés de façon asymptotique pour les grands échantillons gaussiens.

3.6 Exercices

Exercice 3.1. Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré et $f : \Omega \to \mathbb{R}^+$ une application mesurable et positive. Soit $(E, \mathcal{B})$ un espace mesurable et $\phi : \Omega \to E$ une bijection bi-mesurable (ϕ et ϕ^{-1} sont mesurables).
On suppose que $\tilde{\mu}$ a pour densité f par rapport à μ. Démontrer que la mesure image de $\tilde{\mu}$ par ϕ admet une densité, que l'on déterminera, par rapport à la mesure image de μ par ϕ.

Exercice 3.2.
Montrer que si X est une v.a.r de loi normale $\mathcal{N}(0, 1)$, alors X^2 suit une loi de χ^2_1.

Exercice 3.3. [*Variante du calcul des moments*]
Soit X une v.a. positive

1. Montrer que pour tout entier $p \geq 1$ tel que $\mathbb{E}[X^p]$ est finie,
$$\mathbb{E}[X^p] = \int_{\mathbb{R}^+} pt^{p-1} \mathbb{P}[X > t] dt.$$
Ind : on pourra remarquer que $\mathbb{P}[X > t] = \int_t^{+\infty} d\mathbb{P}_X(s)$.
2. On suppose X intégrable. Déduire de la question 1 que
$$\mathbb{E}[X] = \sum_{n \geq 0} \int_n^{n+1} \mathbb{P}[X > t] dt.$$
En déduire alors que
   (a) Si X est entière $\mathbb{E}[X] = \sum_{n \geq 0} \mathbb{P}[X > n]$.
   (b) (Cas général) $\sum_{n \geq 0} \mathbb{P}[X > n+1] \leq \mathbb{E}[X] \leq \sum_{n \geq 0} \mathbb{P}[X > n]$.

Exercice 3.4.

1. Soient $(\Omega_1, \mathcal{T}_1, \mu_1)$, $(\Omega_2, \mathcal{T}_2, \mu_2)$ deux espaces mesurés σ-finis. On considère des applications $f : \Omega_1 \to \mathbb{R}^+$ μ_1-intégrable et positive, et $g : \Omega_2 \to \mathbb{R}^+$ μ_2-intégrable et positive.
Démontrer que la mesure produit $(fd\mu_1) \otimes (gd\mu_2)$ des mesures $fd\mu_1$ et $gd\mu_2$ est la mesure admettant $f \times g$ pour densité par rapport à $\mu_1 \otimes \mu_2$, où $f \times g$ désigne l'application sur $\Omega_1 \times \Omega_2$ définie par : $(f \times g)(x, y) = f(x)g(y)$.

2. Soit $I =]0, +\infty[$ muni de sa tribu borélienne $\mathcal{B}$ et de la mesure de Lebesgue λ. Soient f, g deux applications de I dans $\mathbb{R}^+$ boréliennes et λ-intégrables, et l'application $p : (x, y) \to xy$ définie sur $I \times I$.
   Démontrer que la mesure-image μ de $(fd\lambda) \otimes (gd\lambda)$ par p est une mesure à densité et donner l'expression de sa densité au moyen d'une intégrale.

Exercice 3.5. [*Loi de la transformée linéaire d'un vecteur*]

1. Soit X un vecteur aléatoire de dimension n, de densité f. Soit $A : \mathbb{R}^n \to \mathbb{R}^n$, une application linéaire inversible et $Y = AX$ un vecteur aléatoire. Déterminer la densité de Y.
2. Soit $X = (X_1, X_2)$ un vecteur de $\mathbb{R}^2$ de densité f. Déduire de la question précédente la densité du vecteur $(X_1 + X_2, X_1 - X_2)$. Si X_1 et X_2 sont indépendantes, quelle forme prend la densité de $X_1 + X_2$?

Exercice 3.6.
Soient $X_1, \ldots, X_n$ des v.a. i.i.d. de loi $\mathcal{E}(\lambda)$. On pose $S_k = X_1 + \cdots + X_k$, pour $1 \leq k \leq n$. Calculer la densité du vecteur $(S_1, S_2, \ldots, S_n)$.

Exercice 3.7. [*Loi du quotient de deux variables*]
Soit (X, Y) un vecteur aléatoire de dimension 2, de densité f. On pose $U = X/Y$.

1. Vérifier que U est bien définie.
2. Calculer la densité de U en fonction de f.
3. Quelle est la loi de X/Y si X et Y sont deux lois $\mathcal{N}(0, 1)$ indépendantes ?

Exercice 3.8. [*Coordonnées polaires*]

1. Soient X et Y deux v.a. de densité jointe f. Soient R et Θ les coordonnées polaires du point de coordonnées cartésiennes (X, Y). Déterminer la densité jointe de (R, Θ).
2. Soit $(X, Y) \rightsquigarrow \mathcal{N}_2(0, \sigma^2 I_2)$. Déterminer la loi de R et celle de Θ. Ces v.a. sont elles indépendantes ?

Exercice 3.9. [*Calcul de fonctions caractéristiques usuelles*] Calculer la fonction caractéristique de la v.a. Y suivant :

1. une loi de Poisson $\mathcal{P}(\lambda)$,
2. une loi exponentielle $\mathcal{E}(\lambda)$,
3. une loi uniforme sur $[-a, a]$.

Exercice 3.10. [*Calcul de fonctions caractéristiques*]
Soit la fonction définie sur $\mathbb{R}$ par $f(y) = \alpha e^{-\alpha|y|}/2$. Calculer la transformée de Fourier de f et en déduire la fonction caractéristique d'une v.a de loi de Cauchy $\mathcal{C}(\alpha)$.

Exercice 3.11. [*Propriétés de fonctions caractéristiques*]

1. Soit X une v.a.r. Montrer que $\varphi_X(u)$ est à valeurs réelles si et seulement si X a une loi symétrique, i.e $\mathbb{P}_X = \mathbb{P}_{-X}$.

2. Montrer que si X et Y sont des variables aléatoires indépendantes et de même loi, alors $Z = X - Y$ a une loi symétrique.

Exercice 3.12. [*Somme de v.a indépendantes*]
Soient X et Y deux v.a.r indépendantes.

1. Montrer que si $X \rightsquigarrow \mathcal{P}(\lambda)$ et $Y \rightsquigarrow \mathcal{P}(\mu)$, alors $X + Y \rightsquigarrow \mathcal{P}(\lambda + \mu)$.
2. Montrer que si $X \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2)$ et $Y \rightsquigarrow \mathcal{N}(m_2, \sigma_2^2)$, alors $X + Y \rightsquigarrow \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Exercice 3.13.
Soit $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ un vecteur gaussien centré de matrice de variance $\Gamma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, où ρ est un réel tel que $-1 < \rho < 1$.

1. Soit $U = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ et $V = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. On pose $Y = \begin{pmatrix} < U, X > \\ < V, X > \end{pmatrix}$. Montrer que Y est un vecteur gaussien dont on précisera la loi.
2. Les variables aléatoires $< U, X >$ et $< V, X >$ sont-elles indépendantes ?
3. De manière générale, soit $U = \begin{pmatrix} a \\ b \end{pmatrix}$ et $V = \begin{pmatrix} c \\ d \end{pmatrix}$.
   a) Quelle est la loi de $< U, X >$?
   b) Montrer que $\begin{pmatrix} < U, X > \\ < V, X > \end{pmatrix}$ est un vecteur gaussien.
   c) Quelle est la covariance de $< U, X >$ et $< V, X >$?
   d) A quelle condition nécessaire et suffisante les variables $< U, X >$ et $< V, X >$ sont-elles indépendantes ?
   e) Donner un exemple de vecteurs U et V tels que $< U, X >$ et $< V, X >$ soient des variables aléatoires indépendantes.

Exercice 3.14. [*Fonctions d'un vecteur gaussien*]
Soient $a \in \mathbb{R}^* \setminus \{\pm 1\}$ et $X = (X_0, ..., X_d)'$ un $(d+1)$-vecteur gaussien tel que $X_i \rightsquigarrow \mathcal{N}(0, 1)$ pour tout i et

$$\Gamma_X = \left\{ \begin{array}{cccccc} 1 & a & a & \dots & \dots & a \\ a & 1 & a^2 & \dots & \dots & a^2 \\ a & a^2 & 1 & a^2 & & a^2 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ a & a^2 & & a^2 & 1 & a^2 \\ a & a^2 & \dots & \dots & a^2 & 1 \end{array} \right\}$$

1. Soit $Y = (X_0, Y_1, ..., Y_d)'$ tel que
$$X_i = aX_0 + \sqrt{1 - a^2} Y_i, i = 1, \dots, d.$$
Déterminer la loi du vecteur aléatoire Y.
2. Exprimer X en fonction de Y.
3. (a) Déterminer la loi de $S_d = (X_1 + \cdots + X_d)/d$, $d \geq 1$.
   (b) On pose $Z = S_d / X_0$. Déterminer la loi de $\sqrt{d}(Z - a)/\sqrt{1 - a^2}$ et déduire celle de Z.

Exercice 3.15.

Soit X un vecteur gaussien centré de matrice de covariance $\Gamma = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$. Montrer qu'il existe une vecteur Y bi-dimensionnel centré réduit tel que $X = BY$ où B vérifie $BB' = \Gamma$.

Chapitre 4

Suites aléatoires

Sommaire

4.1 Autonomie : Limites supérieures et inférieures

4.1.1 Cas des suites réelles

Définition 4.1

Soit $(x_n)_{n\in\mathbb{N}}$ une suite réelle. On définit la limite supérieure de la suite $(x_n)_{n\in\mathbb{N}}$ par

$$\overline{\lim_{n\to+\infty}}\, x_n = \lim_{n\to+\infty} \sup_{m\geq n} x_m$$

et sa limite inférieure par

$$\underline{\lim_{n\to+\infty}}\, x_n = \lim_{n\to+\infty} \inf_{m\geq n} x_m.$$

Ces limites existent dans $\overline{\mathbb{R}}$ car les suites

$$\left(\sup_{m\geq n} x_m\right)_{n\in\mathbb{N}} \quad \text{et} \quad \left(\inf_{m\geq n} x_m\right)_{n\in\mathbb{N}}$$

sont monotones (respectivement décroissante et croissante).

Si la suite n'est ni majorée ni minorée

$$\overline{\lim_{n\to+\infty}}\, x_n = +\infty \text{ et } \underline{\lim_{n\to+\infty}}\, x_n = -\infty.$$

Si la suite est majorée mais pas minorée

$$\overline{\lim_{n\to+\infty}}\, x_n = \in \overline{\mathbb{R}} \text{ et } \underline{\lim_{n\to+\infty}}\, x_n = -\infty.$$

Si la suite est minorée mais pas majorée

$$\overline{\lim_{n\to+\infty}} x_n = +\infty \text{ et } \underline{\lim_{n\to+\infty}} x_n \in \overline{\mathbb{R}}.$$

Proposition 4.1

On a

$$\underline{\lim_{n\to+\infty}} x_n \leq \overline{\lim_{n\to+\infty}} x_n.$$

Si la suite est bornée, les limites supérieure et inférieure sont réelles et ces deux valeurs sont les valeurs d'adhérence maximale et minimale de la suite. En outre, toujours dans le cas borné, la suite converge si et seulement si les deux limites coïncident et dans ce cas

$$\lim_{n\to+\infty} x_n = \underline{\lim_{n\to+\infty}} x_n = \overline{\lim_{n\to+\infty}} x_n.$$

4.1.2 Cas des fonctions réelles

Si $(f_n)_{n\in\mathbb{N}}$ est une suite de fonctions réelles définie sur un ensemble Ω, on définit les limites supérieure et inférieure de la suite ponctuellement :

$$\left(\overline{\lim_{n\to+\infty}} f_n\right)(x) = \overline{\lim_{n\to+\infty}} f_n(x), \text{ pour tout } x \in \Omega$$

et

$$\left(\underline{\lim_{n\to+\infty}} f_n\right)(x) = \underline{\lim_{n\to+\infty}} f_n(x), \text{ pour tout } x \in \Omega.$$

L'intérêt de ces limites est qu'elles sont définies même lorsqu'il n'y a pas convergence et elles sont souvent utilisées pour montrer des étapes intermédiaires dans les théorèmes de convergence. Voici par exemple un résultat dont on se sert pour démontrer le théorème de convergence dominée de Lebesgue.

Théorème 4.1 (*Lemme de Fatou*)

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré et soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions boréliennes positives sur $(\Omega, \mathcal{F})$. Alors

$$\int_\Omega \underline{\lim_{n\to+\infty}} f_n d\mu \leq \underline{\lim_{n\to+\infty}} \int_\Omega f_n d\mu.$$

4.1.3 Cas des ensembles

Soit $(A_n)_{n\in\mathbb{N}}$ une suite de sous-ensembles d'un ensemble non vide Ω. On peut définir une notion de limite de la suite $(A_n)_{n\in\mathbb{N}}$ lorsqu'elle est monotone de la façon suivante :

- Si la suite $(A_n)_{n\in\mathbb{N}}$ est croissante i.e. si $A_n \subset A_{n+1}$ pour tout n, on définit

$$\lim_{n\to+\infty} A_n = \bigcup_{n\in\mathbb{N}} A_n;$$

- Si la suite $(A_n)_{n\in\mathbb{N}}$ est décroissante i.e. si $A_{n+1} \subset A_n$ pour tout n, on définit

$$\lim_{n\to+\infty} A_n = \bigcap_{n\in\mathbb{N}} A_n.$$

On ne cherche pas à généraliser la notion de limite à d'autres cas, mais on définit des limites supérieure et inférieure de n'importe quelle suite d'ensembles $(A_n)_{n\in\mathbb{N}}$ selon le même principe que celui utilisé pour les suites réelles.

Définition 4.2

Soit $(A_n)_{n\in\mathbb{N}}$ une suite de sous-ensembles d'un ensemble non vide Ω. On définit les limites inférieure et supérieure de la suite par

$$\limsup_{n\to+\infty} A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m \text{ et } \liminf_{n\to+\infty} A_n = \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} A_m.$$

On remarque que

$$\left(\bigcup_{m\geq n} A_m\right)_{n\in\mathbb{N}} \quad \text{est une suite décroissante et}$$

$$\left(\bigcap_{m\geq n} A_m\right)_{n\in\mathbb{N}} \quad \text{est une suite croissante}$$

ce qui justifie l'emploi du terme de limite.

En outre, on vérifie aisément que, comme pour le cas des suites réelles, les limites supérieure et inférieure coïncident avec la notion de limite dans le cas monotone, i.e. si $(A_n)_{n\in\mathbb{N}}$ est monotone

$$\limsup_{n\to+\infty} A_n = \liminf_{n\to+\infty} A_n = \lim_{n\to+\infty} A_n.$$

Proposition 4.2

On a toujours

$$\liminf_{n\to+\infty} A_n \subset \limsup_{n\to+\infty} A_n.$$

Démonstration : Soit $x \in \liminf_n A_n$. Alors il existe n_0 tel que pour tout $m \geq n_0$, $x \in A_m$, autrement dit x est dans tous les A_m à partir d'un certain rang. Pour un point y, être dans la limite supérieure de la suite signifie que pour tout n, il existe $m \geq n$ tel que $x \in A_m$. Donc cela signifie que y est dans une suite infinie de A_m. Il est clair que si x est dans tous les A_m à partir d'un certain rang, x est dans une suite infinie de A_m et donc l'inclusion est vérifiée.

4.2 Suites infinies d'événements

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

Pour commencer, nous rappelons quelques résultats sur les familles indépendantes (cf. fin chapitre 1).

Proposition 4.3 (***Principe d'associativité pour les tribus indépendantes***)

Si $(\mathcal{G}_n)_{n\in\mathbb{N}}$ est une suite de tribus indépendantes, alors pour tous I_1, $I_2 \subset \mathbb{N}$ tels que $I_1 \cap I_2 = \varnothing$, les suites $(\mathcal{G}_n)_{n\in I_1}$ et $(\mathcal{G}_n)_{n\in I_2}$ sont indépendantes, c'est-à-dire que les tribus

$$\sigma\left(\bigcup_{n\in I_1} \mathcal{G}_n\right) \text{ et } \sigma\left(\bigcup_{n\in I_2} \mathcal{G}_n\right)$$

sont indépendantes.

Proposition 4.4

Si $\mathcal{A}$ et $\mathcal{B}$ sont deux algèbres de Boole indépendantes, alors $\sigma(\mathcal{A})$ et $\sigma(\mathcal{B})$ le sont également.

Voici maintenant un nouvel outil, la tribu des événements asymptotiques.

Définition 4.3 (***Tribu queue ou tribu asymptotique***)

1. Soit $(\mathcal{G}_n)_{n\in\mathbb{N}}$ une suite de sous-tribus de $\mathcal{F}$. La tribu
$$\mathcal{L} = \bigcap_{n\geq 0} \sigma\left(\bigcup_{m\geq n} \mathcal{G}_m\right)$$
est appelée *tribu asymptotique* de la suite $(\mathcal{G}_n)_{n\in\mathbb{N}}$.
2. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. La *tribu asymptotique ou tribu queue* de la suite est la tribu asymptotique de la suite $(\sigma(X_n))_{n\in\mathbb{N}}$.

Proposition 4.5

(i) Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements. Pour tout n, on pose $\mathcal{G}_n = \sigma(\{A_n\})$ et on note $\mathcal{L}$ la tribu asymptotique de cette suite. Alors $\limsup_n A_n$, $\liminf_n A_n \in \mathcal{L}$.

(ii) Soient $(\mathcal{G}_n)_{n\in\mathbb{N}}$ une suite de tribus et $\mathcal{L}$ sa tribu asymptotique. Pour tout n, on pose
$$\mathcal{F}_n = \sigma\left(\bigcup_{k\geq n} \mathcal{G}_k\right).$$
Soit alors $(A_n)_{n\in\mathbb{N}}$ une suite d'événements tels que $A_n \in \mathcal{F}_n$ pour tout n. On a également $\limsup_n A_n$, $\liminf_n A_n \in \mathcal{L}$.

Démonstration : On remarque que, pour tout $N \in \mathbb{N}$,
$$\limsup_n A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m = \bigcap_{n\geq N} \bigcup_{m\geq n} A_m$$
puisque la suite
$$\left(\bigcup_{m\geq n} A_m\right)_{n\in\mathbb{N}}$$
est décroissante. On en déduit que $\limsup_n A_n \in \mathcal{F}_N$ pour tout N et donc $\limsup_n A_n \in \mathcal{L}$ par définition. On procède de même pour la limite inférieure.

Théorème 4.2 (*Loi $0-1$ de Kolmogorov*)

Soit $(\mathcal{G}_n)_{n\in\mathbb{N}}$ une suite de sous-tribus de $\mathcal{F}$ indépendantes. Alors les événements de sa tribu asymptotique sont de probabilité 0 ou 1.

Démonstration : Pour tout n, on pose

$$\mathcal{F}_n = \sigma\left(\bigcup_{k\geq n} \mathcal{G}_k\right).$$

Comme les tribus de la suite sont indépendantes, pour tout n, $\mathcal{G}_n$ est indépendante de $\mathcal{F}_{n+k}$ pour tout $k \geq 1$, et donc de

$$\bigcap_{k\geq n+1} \mathcal{F}_k = \mathcal{L}.$$

On en déduit que, pour tout n, $\mathcal{F}_n$ est également indépendante de $\mathcal{L}$ et donc finalement, $\mathcal{L} = \bigcap_n \mathcal{F}_n$ est indépendante d'elle-même. Donc tous les événements de $\mathcal{L}$ sont indépendants d'eux-mêmes et on a vu que les événements indépendants d'eux-mêmes sont les événements de probabilité 0 ou 1.

Exemple :

1. Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements indépendants, $\limsup_n A_n$ et $\liminf_n A_n$ sont des événements négligeables ou presque certains. En outre
   - $\mathbb{P}[\liminf_n A_n] = 1$ signifie que, presque sûrement il existe un n tel que pour tout $k \geq n$, A_k est réalisé, c'est-à-dire, presque sûrement tous les A_n sont réalisés à partir d'un certain rang.
   - De même, on peut monter que $\mathbb{P}[\limsup_n A_n] = 1$ signifie que, presque sûrement, une infinité des A_n se réalise.
   - Que signifient $\mathbb{P}[\liminf_n A_n] = 0$ et $\mathbb{P}[\limsup_n A_n] = 0$? Quelles sont finalement les situations possibles ?

   Application : On joue à pile ou face. Montrer que, presque sûrement, on a une infinité de "pile" et une infinité de "face".
2. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. indépendantes. Alors soit $\sum_n X_n$ converge presque sûrement soit $\sum_n X_n$ diverge presque sûrement. En effet, on définit la tribu asymptotique de la suite comme dans la définition 4.3-2 et on remarque que

   $$\sum_{n\geq 0} X_n \text{ converge si et seulement si, pour tout } N, \sum_{n\geq N} X_n \text{ converge.}$$

   Or

   $$\left[\sum_{n\geq N} X_n \text{ converge}\right] \in \mathcal{F}_N$$

   donc l'événement $[\sum_n X_n \text{ converge}] \in \mathcal{F}_N$ pour tout N et donc $[\sum_n X_n \text{ converge}] \in \mathcal{L}$. On en déduit que la probabilité de cet événement est 0 ou 1, ce qu'il fallait démontrer.

Théorème 4.3 (*Lemme de Borel-Cantelli*)

Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements de $\mathcal{F}$.

1. *Si $\sum_n \mathbb{P}[A_n] < +\infty$ alors $\mathbb{P}[\limsup_n A_n] = 0$ (presque sûrement, seul un nombre fini de A_n se réalise).*
2. *Si $\sum_n \mathbb{P}[A_n] = +\infty$ et si les $(A_n)_{n\in\mathbb{N}}$ sont indépendants, alors $\mathbb{P}[\limsup_n A_n] = 1$ (presque sûrement une infinité de A_n se réalise).*

Démonstration :

1. On a
$$\mathbb{P}\left[\limsup_n A_n\right] = \lim_{n\to+\infty} \mathbb{P}\left[\bigcup_{m\geq n} A_m\right] \leq \lim_{n\to+\infty} \sum_{m\geq n} \mathbb{P}\left[A_m\right].$$
Or si la série converge, la queue de la série tend vers 0, et donc la probabilité de la limite supérieure est nulle.
2. On montre en fait que
$$\mathbb{P}\left[\left(\limsup_n A_n\right)^c\right] = \mathbb{P}\left[\liminf_n A_n^c\right] = 0.$$
En effet,
$$\mathbb{P}\left[\liminf_n A_n^c\right] = \lim_{n\to+\infty} \mathbb{P}\left[\bigcap_{m\geq n} A_m^c\right] = \lim_{n\to+\infty}\lim_{N\to+\infty}\prod_{k=n}^{N} \mathbb{P}\left[A_k^c\right]$$
car les événements sont indépendants. Or
$$\prod_{k=n}^{N}(1-\mathbb{P}\left[A_k\right]) \leq \prod_{k=n}^{N} \mathrm{e}^{-\mathbb{P}[A_k]} = \mathrm{e}^{-\sum_{k=n}^{N}\mathbb{P}[A_k]}$$
et cette dernière quantité tend vers 0 quand N tend vers l'infini parce que la série des probabilités diverge. On a donc bien montré le résultat souhaité.

4.3 Convergence de suites de variables aléatoires

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

4.3.1 Différents types de convergence

Définition 4.4

Soient $(X_n)_{n\in\mathbb{N}}$ et X une suite de v.a. et une v.a. définies sur $(\Omega, \mathcal{F})$ et à valeurs dans $(E, \mathcal{B}(E))$ un espace vectoriel normé muni de sa tribu borélienne. On dit que

- □ $(X_n)_{n\in\mathbb{N}}$ converge vers X presque sûrement si $\mathbb{P}\left[\lim_{n\to+\infty} X_n = X\right] = 1$.
- □ $(X_n)_{n\in\mathbb{N}}$ converge vers X en probabilités si pour tout $\varepsilon > 0$, $\mathbb{P}\left[|X_n - X| > \varepsilon\right] \to 0$ quand $n \to +\infty$.
- □ $(X_n)_{n\in\mathbb{N}}$ converge vers X en loi si $\mathbb{E}\left[h(X_n)\right] \to \mathbb{E}\left[h(X)\right]$ pour toute fonction h continue bornée sur $(E, \mathcal{B}(E))$ et à valeurs dans $\mathbb{R}$.
- □ $(X_n)_{n\in\mathbb{N}}$ converge vers X dans $L^p(\Omega, \mathcal{F}, \mathbb{P})$ (pour un $p \in [1, +\infty])$ si toutes les v.a. X_n et X sont $L^p(\Omega, \mathcal{F}, \mathbb{P})$ et $||X_n - X||_p \to 0$ quand $n \to +\infty$

Remarque : Si $(E, \mathcal{E}, \mu)$ est un espace mesuré et si l'on considère des fonctions mesurables à valeurs dans un espace vectoriel normé de norme $|\cdot|$ muni de sa tribu borélienne $(F, \mathcal{B}(F))$, alors

. Si $1 \leq p < +\infty$, la norme $||\cdot||_p$ est définie par
$$||f||_p = \left(\int_\Omega |f|^p d\mu\right)^{\frac{1}{p}}, \text{ pour tout } f \in L^p(E, \mathcal{E}, \mu).$$

- L'espace $L^\infty(E,\mathcal{E},\mu)$ est l'espace des fonctions mesurables presque sûrement bornées et la norme L^∞ est la plus petite des constantes qui majore presque sûrement la norme dans F, appelée aussi le supremum essentiel.
- Les espaces L^p sont des espaces de Banach.
- Pour tous $f \in L^p$ et $g \in L^q$ avec $1/p + 1/q = 1$, $fg \in L^1$ et on a l'inégalité de Hölder
$$||fg||_1 \leq ||f||_p ||g||_q.$$
Dans le cas $p = q = 2$, cette inégalité est connue sous le nom d'inégalité de Cauchy-Schwarz.
- Pour tout $f,\ g \in L^2$, on définit
$$\langle f, g\rangle_2 = \int_E f g d\mu.$$
Cette forme bilinéaire est un produit scalaire sur L^2 qui est un espace de Hilbert.
- Si $\mu(E) < +\infty$ alors $L^\infty \subset L^p \subset L^1$ pour tout $p \in]1, +\infty[$.
- Si $\mu(E) = 1$ alors $||f||_p \leq ||f||_{p'}$ pour tout $p' \geq p$ dès que $f \in L^{p'}$.
- Si $\mu(E) = 1$ et si $f \in L^p$ alors on a l'inégalité de Markov suivante :
$$\mu(|f| > \varepsilon) \leq \frac{1}{\varepsilon^p}\mathbb{E}\left[|f|^p\right] \text{ pour tout } \varepsilon > 0.$$
- Soit $f \geq 0$ presque partout, une fonction intégrable. Alors pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que pour tout $A \in \mathcal{E}$,
$$\mu(A) < \delta \Rightarrow \int_A f d\mu < \varepsilon.$$

Remarque : Dans les espaces L^p, on identifie les fonctions qui sont égales presque partout.

4.3.2 Critères de convergence

Proposition 4.6

La suite $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement si et seulement si
$$\mathbb{P}\left[\limsup_{n\to+\infty}[|X_n - X| > \varepsilon]\right] = 0 \textit{ pour tout } \varepsilon > 0.$$

Démonstration : $X_n \to X$ si et seulement si pour tout $\varepsilon > 0$, il existe $n \in \mathbb{N}$ tel que pour tout $m \geq n$, $|X_m - X| \leq \varepsilon$. Autrement dit
$$[X_n \to X \text{ quand } n \to +\infty] = \bigcap_{\varepsilon>0}\bigcup_{n\in\mathbb{N}}\bigcap_{m\geq n}[|X_m - X| \leq \varepsilon] = \bigcap_{\varepsilon>0}\liminf_{n\to+\infty}[|X_n - X| \leq \varepsilon].$$

Les ensembles $\liminf_n[|X_n - X| \leq \varepsilon]$ décroissent quand ε décroit vers 0. Donc
$$\mathbb{P}\left[X_n \to X\right] = 1 \text{ si et seulement si } \mathbb{P}\left[\liminf_n[|X_n - X| \leq \varepsilon]\right] = 1 \text{ pour tout } \varepsilon > 0,$$

d'où le résultat annoncé en passant au complémentaire.

Théorème 4.4 (*Convergence en loi*)

Soient $(X_n)_{n\in\mathbb{N}}$ et X une suite de v.a. et une v.a. définies sur $(\Omega, \mathcal{F})$ et à valeurs dans $(E, \mathcal{B}(E))$ un espace vectoriel normé muni de sa tribu borélienne. Les propositions suivantes sont équivalentes :

(i) $(X_n)_{n\in\mathbb{N}}$ *converge vers* X *en loi.*

(ii) Pour toute fonction h uniformément continue bornée de $(E, \mathcal{B}(E))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)]$ quand $n \to +\infty$.

(iii) $\overline{\lim}_{n\to+\infty} \mathbb{P}[X_n \in F] \leq \mathbb{P}[X \in F]$ *pour tout ensemble F fermé.*

(iv) $\underline{\lim}_{n\to+\infty} \mathbb{P}[X_n \in O] \geq \mathbb{P}[X \in O]$ *pour tout ensemble O ouvert.*

(v) Fonctions de répartition : dans le cas où $E = \mathbb{R}$, $F_{X_n}(x) \to F_X(x)$ en tout point de continuité x de F_X.

(vi) Fonctions caractéristiques : dans le cas où $E = \mathbb{R}^d$, $\varphi_{X_n}(t) \to \varphi_X(t)$ pour tout $t \in \mathbb{R}^d$.

Démonstration : On fait une preuve partielle du théorème.

(i)⇒(ii) Evident.

(ii)⇒(iii) Soit F un fermé. Pour tout $k \in \mathbb{N}^*$, on définit

$$F_k = \left\{x \in E/\ d(x,F) \leq \frac{1}{k}\right\} \text{ où } d(x,F) = \inf_{y\in F} |x-y| \text{ est la distance de } x \text{ à } F \text{ dans } E.$$

Soit g définie sur $\mathbb{R}^+$ par $g(t) = \mathbf{1}_{[0,1]}(t)(1-t)$ uniformément continue sur $\mathbb{R}^+$. Pour tout k, on définit alors sur E la fonction f_k par $f_k(x) = g(d(x,F)k)$. Alors f_k est uniformément continue car d est Lipschitzienne. De plus, $\mathbf{1}_F \leq f_k \leq \mathbf{1}_{F_k}$. On en déduit que, pour tout k,

$$\mathbb{P}[X_n \in F] = \mathbb{E}[\mathbf{1}_F(X_n)] \leq \mathbb{E}[f_k(X_n)] \text{ pour tout } n,$$

et donc en passant à la limite supérieure et en utilisant (ii),

$$\limsup_{n\to+\infty} \mathbb{P}[X_n \in F] \leq \mathbb{E}[f_k(X)].$$

Mais, en utilisant la seconde inégalité, $\mathbb{E}[f_k(X)] \leq \mathbb{P}[X \in F_k]$, d'où

$$\limsup_{n\to+\infty} \mathbb{P}[X_n \in F] \leq \mathbb{P}[X \in F_k] \text{ pour tout } k.$$

Comme $(F_k)_k$ est une suite décroissante d'ensembles, $[X \in F_k]$ est une suite décroissante d'événements et on obtient

$$\limsup_{n\to+\infty} \mathbb{P}[X_n \in F] \leq \mathbb{P}\left[\bigcap_k [X \in F_k]\right] = \mathbb{P}\left[X \in \bigcap_k F_k\right].$$

enfin, comme F est fermé,

$$F = \bigcap_k F_k$$

ce qui conclut la preuve.

(iii)⇔(iv) Par passage au complémentaire.

(iii),(iv)⇒(v) Soit $t \in \mathbb{R}$ fixé. Pour tout n, $F_{X_n}(t) = \mathbb{P}[X \in]-\infty, t]]$ donc en utilisant (iii) on obtient $\overline{\lim}_n F_{X_n}(t) \leq F_X(t)$. D'autre part, $\mathbb{P}[X_n \in]-\infty, t]] \geq \mathbb{P}[X_n \in]-\infty, t[\,]$ pour tout n et on en déduit $\underline{\lim}_n F_{X_n}(t) \geq \mathbb{P}[X \in]-\infty, t[\,]$ en utilisant (iv). Mais pour tout $s < t$,

$$\mathbb{P}[X \in]-\infty, t[\,] \geq \mathbb{P}[X \in]-\infty, s]] = F_X(s),$$

i.e. $\underline{\lim}_n F_{X_n}(t) \geq F_X(s)$ pour tout $s < t$. On en déduit donc que si F_X est continue en t, $\underline{\lim}_n F_{X_n}(t) \geq F_X(t)$, ce qu'il fallait démontrer pour conclure.

...⇒(i) Pour remonter de (ii), (iii) ou (iv), ou (v) vers (i), il faut utiliser des arguments de densité que l'on admet ici.

(ii)⇒(vi) Evident.

(vi)⇒(i) Théorème de Paul Lévy ci-dessous.

Théorème 4.5 (*Théorème de Paul Lévy*)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de vecteurs aléatoires. Si la suite des fonctions caractéristiques $(\varphi_{X_n})_{n\in\mathbb{N}}$ converge simplement vers une fonction g continue en 0 alors g est fonction caractéristique d'un vecteur aléatoire X et $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X.

Nous admettons la preuve de ce résultat qui repose sur la convergence étroite des mesures et sur le lien entre cette convergence et la convergence des transformées de Fourier des mesures.

Théorème 4.6

Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de v.a. convergeant respectivement en loi vers X et Y.

1. *Si $Y = c$ presque sûrement, avec c constante, alors le couple $((X_n, Y_n))_{n\in\mathbb{N}}$ converge en loi vers $(X, Y) = (X, c)$.*
2. *Si X_n et Y_n sont indépendantes pour tout n alors $((X_n, Y_n))_{n\in\mathbb{N}}$ converge en loi vers un couple de v.a. $(\overline{X}, \overline{Y})$ indépendantes et telles que $\mathcal{L}(\overline{X}) = \mathcal{L}(X)$ et $\mathcal{L}(\overline{Y}) = \mathcal{L}(Y)$.*

Démonstration :

1. A faire en exercice (section 4.5.3).
2. Pour tout n, si X_n et Y_n sont indépendantes, on a, pour tout (t, s), $\varphi_{(X_n,Y_n)}(t, s) = \varphi_{X_n}(t)\varphi_{Y_n}(s) \to \varphi_X(t)\varphi_Y(s)$, d'où le résultat.

Remarque : On voit donc que la convergence en loi des marginales n'entraîne pas la convergence en loi des vecteurs en général. En revanche, la convergence en loi des vecteurs entraîne bien la convergence en loi des coordonnées. Par exemple $\varphi_{X_n}(t) = \varphi_{(X_n,Y_n)}(t, 0)$.

Proposition 4.7

La suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X dans $\mathbb{R}^d$ si et seulement si, pour tout $a \in \mathbb{R}^d$, $(\langle a, X_n\rangle)_{n\in\mathbb{N}}$ converge en loi vers $\langle a, X\rangle$.

Démonstration : Si $(X_n)_n$ converge en loi vers X, alors comme $x \to \langle a, x\rangle$ est une application continue, $x \to h(\langle a, x\rangle)$ est une fonction continue bornée pour tout h continue bomée et $\mathbb{E}[h(\langle a, X_n\rangle)] \to \mathbb{E}[h(\langle a, X\rangle)]$ quand n tend vers $+\infty$, donc $(\langle a, X_n\rangle)_{n\in\mathbb{N}}$ converge bien en loi vers $\langle a, X\rangle$.

Réciproquement, on remarque que pour tout t et pour tout n $\varphi_{X_n}(t) = \varphi_{\langle t,X_n\rangle}(1)$ et donc la convergence en loi des produits entraîne bien la convergence en loi de X_n.

Remarque : Avec cette proposition, on voit bien que la convergence en loi des coordonnées ne suffit pas pour avoir la convergence en loi du vecteur : il faut la convergence en loi e toutes les combinaisons linéaires des coordonnées pour avoir la convergence en loi du vecteur.

4.3.3 Liens entre les convergences des v.a.

Théorème 4.7

$(X_n)_{n\in\mathbb{N}}$ et X sont respectivement une suite de v.a. et une v.a. définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans un espace vectoriel normé E. On a :

1. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X dans $L^{p'}$ alors $(X_n)_{n\in\mathbb{N}}$ converge vers X dans L^p pour $1 \leq p \leq p' \leq \infty$.*
2. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X dans un L^p pour $1 \leq p < +\infty$, alors $(X_n)_{n\in\mathbb{N}}$ converge vers X en probabilités.*
3. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X presque surement alors $(X_n)_{n\in\mathbb{N}}$ converge vers X en probabilités.*
4. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X en probabilités alors $(X_n)_{n\in\mathbb{N}}$ converge vers X en loi.*
5. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X dans L^∞ alors $(X_n)_{n\in\mathbb{N}}$ converge vers X presque surement.*
6. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X en probabilités alors il existe une sous-suite de $(X_n)_{n\in\mathbb{N}}$ qui converge vers X presque sûrement.*
7. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X dans un L^p pour $1 \leq p < +\infty$, alors il existe une sous-suite de $(X_n)_{n\in\mathbb{N}}$ qui converge vers X presque sûrement.*

Question : Faire un schéma pour représenter ces convergences.

Démonstration :

1. Vu dans une section précédente. Repose sur les inégalités entre les normes obtenues grâce aux inégalités de Hölder.
2. On suppose que $(X_n)_n$ converge vers X dans L^p et on utilise les inégalités de Markov de la remarque à la suite de la definition 4.4 : pour tout $\varepsilon > 0$ et pour tout n,
$$\mathbb{P}\left[|X_n - X| > \varepsilon\right] \leq \frac{1}{\varepsilon^p}\mathbb{E}\left[|X_n - X|^p\right].$$
3. On suppose $X_n \to X$ presque sûrement. Soit $\varepsilon > 0$ fixé. On a
$$\mathbb{P}\left[|X_n - X| > \varepsilon\right] = \int_\Omega \mathbf{1}_{[|X_n - X| > \varepsilon]} d\mathbb{P}.$$
Comme $X_n \to X$ presque sûrement, la fonction indicatrice est nulle pour n assez grand en presque tout point de Ω (donc converge p.s. vers 0). En outre, la fonction indicatrice est majorée par la constante 1 qui est intégrable et donc, par le théorème de convergence dominée,
$$\mathbb{P}\left[|X_n - X| > \varepsilon\right] \to 0 \text{ quand } n \to +\infty.$$
4. On suppose que $X_n \to X$ en probabilité. Soit h une fonction uniformément continue bornée. Alors, pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tous $(x, y) \in E^2$ tels que $|x - y| < \eta$, $|h(x) - h(y)| < \varepsilon$. Par ailleurs, pour tout $\eta > 0$,
$$\begin{aligned} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| &\leq \int_\Omega |h(X_n) - h(X)| d\mathbb{P} \\ &= \int_{[|X_n - X| < \eta]} |h(X_n) - h(X)| d\mathbb{P} + \int_{[|X_n - X| \geq \eta]} |h(X_n) - h(X)| d\mathbb{P}. \end{aligned}$$
Donc pour le η défini plus haut, on obtient
$$|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \leq \varepsilon \mathbb{P}\left[|X_n - X| < \eta\right] + 2\|h\|_\infty \mathbb{P}\left[|X_n - X| \geq \eta\right].$$

On majore $\mathbb{P}\left[|X_n - X| < \eta\right]$ par 1 et pour ε et η fixés, on définit n_0 tel que, pour tout, $n \geq n_0$,

$$\mathbb{P}\left[|X_n - X| \geq \eta\right] \leq \frac{\varepsilon}{2\|h\|_\infty}.$$

On a donc montré, pour tout $\varepsilon > 0$, il existe n_0 tel que pour tout $n \geq n_0$,

$$|\mathbb{E}\left[h(X_n)\right] - \mathbb{E}\left[h(X)\right]| \leq 2\varepsilon$$

ce qui conclut la preuve.

5. Evident.
6. On suppose $X_n \to X$ en probabilité et on fixe $\varepsilon > 0$. Pour α quelconque, il existe n_0 tel que pour tout $n \geq n_0$

$$\mathbb{P}\left[|X_n - X| > \varepsilon\right] \leq \alpha.$$

Alors, on peut construire une suite *croissante* d'entiers $(n_k)_{k\in\mathbb{N}^*}$ (de sorte que $(X_{n_k})_k$ soit une suite extraite de $(X_n)_{n\in\mathbb{N}}$) telle que pour tout k

$$\mathbb{P}\left[|X_{n_k} - X| > \varepsilon\right] \leq \frac{1}{k^2}.$$

On a alors en particulier

$$\sum_{k\geq 1} \mathbb{P}\left[|X_{n_k} - X| > \varepsilon\right] \leq \sum_{k\geq 1} \frac{1}{k^2} < +\infty$$

et donc, d'après le lemme de Borel-Cantelli Théorème 4.3,

$$\mathbb{P}\left[\limsup_{k\to+\infty}[|X_{n_k} - X| > \varepsilon]\right] = 0$$

ce qui, d'après le critère de convergence Proposition 4.6, équivaut à $X_{n_k} \to X$ presque sûrement. On a donc bien une sous-suite qui converge presque sûrement.

7. Il suffit d'appliquer les propositions 2. et 6. ci-dessus.

Proposition 4.8

Si $(X_n)_n$ converge en loi vers une constante c alors $(X_n)_n$ converge en probabilité vers c.

Démonstration : Soit $\varepsilon > 0$. $|X_n - c| \geq \varepsilon$ si et seulement si $[X_n \in F]$ où $F =]-\infty, c-\varepsilon] \cup [c+\varepsilon, +\infty[$ est fermé. Donc, comme $(X_n)_n$ converge vers c en loi,

$$\overline{\lim_{n\to+\infty}}\, \mathbb{P}\left[|X_n - c| \geq \varepsilon\right] \leq \mathbb{P}\left[c \in F\right] = 0,$$

et $(X_n)_n$ converge vers X en probabilité.

Théorème 4.8 (*Convergence dominée pour la convergence en probabilité*)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. qui converge en probabilité vers une v.a. X pour un $p \geq 1$. S'il existe $Y \in L^p$ telle que $|X_n| \leq |Y|$ presque sûrement pour tout n alors $(X_n)_n$ converge vers X dans L^p.

Démonstration : Comme $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers X, il existe une sous-suite qui converge presque sûrement vers X et on en déduit $|X| \leq |Y|$ presque sûrement et donc $X \in L^p$. On peut donc remplacer X_n par $X_n - X$ et X par 0 et comme la convergence L^p est la convergence L^1 de la pième puissance, on est ramené à démontrer le résultat suivant : si $(X_n)_n$ converge vers 0 en probabilité et si la suite est dominée par $|Y|$ pour $Y \in L^1$, alors $(X_n)_n$ vers 0 dans L^1, i.e. $\mathbb{E}[|X_n|] \to 0$ quand $n \to 0$. Pour tout n et pour tout $\varepsilon > 0$, on a

$$\begin{aligned}\int_\Omega |X_n| d\mathbb{P} &= \int_{[|X_n|>\varepsilon]} |X_n| d\mathbb{P} + \int_{[|X_n|\leq\varepsilon]} |X_n| d\mathbb{P} \\ &\leq \int_{[|X_n|>\varepsilon]} |Y| d\mathbb{P} + \varepsilon \mathbb{P}[|X_n| \leq \varepsilon].\end{aligned}$$

On majore le second terme par ε. Pour le premier, on utilise la dernière propriété de la remarque qui fait suite à la définition 4.4 en vérifiant qu'elle implique que si $(A_n)_n$ est une suite d'événements tels que $\mathbb{P}[A_n] \to 0$ quand $n \to +\infty$, alors

$$\int_{A_n} Z d\mathbb{P} \to 0 \text{ quand } n \to +\infty$$

pour toute fonction Z presque sûrement positive intégrable. On l' applique à $A_n = [|X_n| > \varepsilon]$ dont la probabilité converge bien vers 0 quand n tend vers l'infini et $Z = |Y|$, on en déduit que le premier terme tend vers 0 quand n tend vers $+\infty$, puis que $\mathbb{E}[|X_n|] \leq 2\varepsilon$ pour n assez grand et donc $\lim_{n\to+\infty} \mathbb{E}[|X_n|] = 0$.

4.4 Théorèmes limites

4.4.1 Lois des grands nombres

Théorème 4.9 (*Loi faible des grands nombres*)

Soit $(X_n)_n$ une suite de v.a. i.i.d. intégrables. Alors

$$\frac{X_1 + ... + X_n}{n} \to \mathbb{E}[X_1] \text{ en probabilité et dans } L^1.$$

Remarque : En 1A, on a montré une version L^2 pour des v.a. décorrélées à variances bornées dans laquelle on obtient la convergence en probabilité et dans L^2.

Théorème 4.10 (*Loi forte des grands nombres*)

Soit $(X_n)_n$ une suite de v.a. i.i.d. intégrables. Alors

$$\frac{X_1 + ... + X_n}{n} \to \mathbb{E}[X_1] \text{ presque sûrement.}$$

En outre, si la suite est dans L^2, la convergence a lieu dans L^2 également.

4.4.2 Théorème limite central

Théorème 4.11 (*Théorème limite central*)

Si $(X_n)_n$ est une suite de v.a. i.i.d. de $\mathbb{R}^d$ de carré intégrable alors

$$\sqrt{n}\left(\frac{X_1 + ... + X_n}{n} - \mathbb{E}[X_1]\right) \to \mathcal{N}_d(0, \mathbf{Cov}[X_1]).$$

Démonstration : On montre la convergence des fonctions caractéristiques.

- ☐ dans $\mathbb{R}$. On pose $\overline{X}_n = \dfrac{X_1 + ... + X_n}{n}$, on suppose $\mathbb{E}[X_1] = 0$ et on pose $\mathbf{Var}[X_1] = \sigma^2$. Alors, pour tout t,

$$\varphi_{\sqrt{n}\overline{X}_n}(t) = \varphi_{X_1+...+X_n}\left(\frac{t}{\sqrt{n}}\right) = \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 + 0 - \frac{1}{2}\frac{t^2}{n}\sigma^2 + o\Big(\frac{1}{n}\Big)\right)^n.$$

Cette quantité converge [1] vers $\exp\left(-\frac{1}{2}\sigma^2 t^2\right)$ qui est bien la valeur en t de la fonction caractéristique d'une v.a. $\mathbb{N}(0, \sigma^2)$ donc le résultat est démontré.

- ☐ dans $\mathbb{R}^d$. On utilise l'égalité $\varphi_{\sqrt{n}\overline{X}_n}(t) = \varphi_{\langle t, \sqrt{n}\overline{X}_n\rangle}(1)$ pour tout $t \in \mathbb{R}^d$ et on conclut en utilisant le résultat déjà démontré dans $\mathbb{R}$ (à faire en exercice).

4.5 Exercices

4.5.1 Limites supérieures et limites inférieures

Exercice 4.1. Soit $(x_n)_{n\in\mathbb{N}}$ une suite bornée de $\mathbb{R}$.

1. On suppose que la sous-suite $(x_{\varphi(n)})_{n\in\mathbb{N}}$ converge. Montrer que

$$\underline{\lim}_n\, x_n \leq \lim_n x_{\varphi(n)} \leq \overline{\lim}_n\, x_n.$$

2. En déduire que si $\underline{\lim}_n\, x_n = \overline{\lim}_n\, x_n$ alors la suite $(x_n)_{n\in\mathbb{N}}$ converge.

Exercice 4.2. Démontrer le lemme de Fatou.

Exercice 4.3. On suppose que $(A_n)_{n\in\mathbb{N}}$ est une suite de sous-ensembles deux à deux disjoints d'un ensemble non vide Ω. Calculer $\limsup_n A_n$ et $\liminf_n A_n$. On pourra commencer par l'exemple

$$A_n = [n, n+1[, \text{ pour tout } n \in \mathbb{N}.$$

Exercice 4.4. Soit $(x_n)_{n\in\mathbb{N}}$ une suite de réels strictement positifs. Pour tout $n \in \mathbb{N}$, on définit $A_n = [0, x_n[$. Calculer les limites inférieures et supérieures de la suite $(A_n)_{n\in\mathbb{N}}$ dans les cas suivants :

1. La suite $(x_n)_{n\in\mathbb{N}}$ converge, plus précisément on étudiera uniquement les exemples :

$$(a)\ x_n = 1 + \frac{1}{n} \quad (b)\ x_n = 1 - \frac{1}{n} \quad (c)\ x_{2n} = 1 + \frac{1}{2n} \text{ et } x_{2n+1} = 1 - \frac{1}{2n+1}.$$

2. La suite $(x_n)_{n\in\mathbb{N}}$ est quelconque (on ne cherchera pas à préciser si les intervalles limites sont ouverts ou fermés).

1. Ce résultat asymptotique est aussi valable pour des nombres complexes.

4.5.2 Calculs asymptotiques

Exercice 4.5.
Soit (X_n) une suite aléatoire i.i.d. On pose $S_n = X_1 + \cdots + X_n$. Montrer que

1. $\sup_n X_n = +\infty$ p.s ou $\sup_n X_n < +\infty$ p.s.
2. $\dfrac{S_n}{n}$ diverge ou converge p.s.

Exercice 4.6. [*Une marche aléatoire simple*]
Un joueur lance une pièce une infinité de fois. Soit $p \in]0,1[$, $p \neq 1/2$, la probabilité d'obtenir "face" à un lancer. Soit S_n la v.a. égale au nombre de face obtenu diminué du nombre de "pile", en n lancers. On pose $S_0 = 0$.

1. Calculer $\mathbb{P}[S_n = 0]$ selon la parité de n.
2. Soit $N = Card\{n \in \mathbb{N}^* : S_n = 0\}$ le nombre d'instants où la trajectoire coupe l'axe des temps. Montrer que l'espérance de N est finie.
3. Calculer $\mathbb{P}[\limsup(S_n = 0)]$.

Exercice 4.7. [*Apparition d'un mot dans un schéma de Bernoulli*]
On cherche la probabilité d'apparition d'une chaîne de piles et faces fixée dans une suite infinie de lancers indépendants d'une pièce équilibrée ou non. On fixe donc un mot de longueur k avec l piles et $k - l$ faces. On note X_1, X_2, ... les résultats des lancers et pour tout $n \geq 1$ et A_n l'événement "le mot apparaît dans la séquence X_{n+1},...,X_{n+k}".

1. Calculer la probabilité de A_n.
2. Montrer que presque sûrement le mot fixé apparaît une infinité de fois.

4.5.3 Convergence des variables aléatoires

Exercice 4.8. [*Théorème de Slutsky*]
On suppose que la suite de v.a. $(X_n)_n$ converge en loi vers une v.a. X et que la suite $(Y_n)_n$ converge en loi vers C, avec C constante presque sûrement.

1. Montrer que le couple $((X_n, Y_n))_n$ converge en loi vers (X, C).
2. En déduire que

$$X_n + Y_n \xrightarrow{\mathcal{L}} X + C, \quad X_n Y_n \xrightarrow{\mathcal{L}} XC, \quad X_n / Y_n \xrightarrow{\mathcal{L}} X/C, \ (si\ C \neq 0).$$

4.5.4 Les théorèmes de convergence

Exercice 4.9.
Soit $(X_n)_n$ une suite aléatoire i.i.d. de loi P de fonction caractéristique ϕ. Posons $Y_n = (X_n + Y_{n-1})/2$ pour tout $n \geq 1$, avec $Y_0 = X_0/2$.

1. (a) Exprimer Y_n en fonction de X_i pour $0 \leq i \leq n$.
   (b) Calculer la fonction caractéristique ϕ_n de Y_n en fonction de ϕ et de n.
   (c) Etudier la convergence en loi de la suite $(Y_n)_n$ lorsque P est la loi de Cauchy $\mathcal{C}s(1)$.
2. Etudier la convergence de $(\sum_{i=1}^{n} X_i/n)_n$ et $(\sum_{i=1}^{n} X_i/\sqrt{n})_n$ pour la loi P ci-dessus.

Chapitre 5

Conditionnement et martingales

Sommaire

5.1 Espérance conditionnelle

On suppose connu partiellement le résultat d'une expérience aléatoire et on cherche à estimer ce résultat au mieux en fonction de cette observation partielle.

5.1.1 Exemple introductif

On suppose connu le nombre de "pile" obtenus lorsque l'on lance deux fois une pièce de monnaie équilibrée et on cherche à estimer au mieux le résultat du premier lancer.

Modélisation : On note X_i le résultat du i-ème lancer avec la convention $X_i = 1$ si le lancer donne "pile" et $X_i = 0$ sinon. On suppose que les lancers sont indépendants. On cherche à estimer X_1 connaissant $X_1 + X_2$.

Question : Donner les valeurs de X_1 dans les cas suivants : $X_1 + X_2 = 0$, $X_1 + X_2 = 2$ et $X_1 + X_2 = 1$. Dans le denier cas on donnera la loi de X_1 sachant $X_1 + X_2 = 1$.

Sur l'événement $[X_1 + X_2 = 1]$, on n'a aucune information supplémentaire et on peut au mieux calculer la moyenne α de X_1.

Question : Combien vaut-elle ?

En choisissant

$$X_1 = \begin{cases} 0 \text{ si } X_1 + X_2 = 0 \\ \alpha \text{ si } X_1 + X_2 = 1 \\ 1 \text{ si } X_1 + X_2 = 2 \end{cases}$$

on a choisi de donner pour chaque valeur prise par $X_1 + X_2$, l'espérance de X_1 sur cet ensemble.

Conclusion : On remarque alors plusieurs faits concernant la nouvelle variable aléatoire que l'on vient de créer :

- elle est constante là où $X_1 + X_2$ est constante (elle est $\sigma(X_1 + X_2)$-mesurable) ;
- elle a la même moyenne que X_1 sur les ensembles où $X_1 + X_2$ est constante ;
- autrement dit, sur ces ensembles, elle est la meilleure approximation constante de X_1 au sens des moindres carrés (i.e. dans L^2).

Ces remarques constituent en fait les caractérisations de l'approximation que l'on cherche à définir. Cette approximation, qui est une variable aléatoire, s'appelle l'espérance conditionnelle de X_1 sachant $X_1 + X_2$ et se note $\mathbb{E}[X_1 | X_1 + X_2]$.

5.1.2 Définition

Cette notion ne concerne que des v.a. intégrables et l'espérance conditionnelle est elle-même une v.a. intégrable.

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

Proposition 5.1

Soit X une v.a. définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. Si X est intégrable, alors il existe une unique v.a. Z intégrable $\mathcal{G}$-mesurable telle que :

$$\text{Pour tout } A \in \mathcal{G}, \int_A Z d\mathbb{P} = \int_A X d\mathbb{P}. \tag{5.1}$$

Remarque : On retrouve les deux caractérisations discutées dans l'exemple précédent : l'espérance conditionnelle de X est mesurable par rapport à l'information qui décrit le conditionnement et la moyenne de X est préservée sur les événements constituant cette information.

Définition 5.1 (*Espérance conditionnelle*)

La v.a. Z de la proposition précédente s'appelle l'espérance conditionnelle de X sachant $\mathcal{G}$. On la note $\mathbb{E}[X | \mathcal{G}]$.

Remarque : Si X est $\mathcal{G}$-mesurable, on a alors $\mathbb{E}[X | G] = X$ presque sûrement. Ce résultat, bien que trivial est extrêmement important.

Démonstration : On commence par supposer $X \geq 0$ $\mathbb{P}$-presque sûrement et on définit l'application

$$\mu(A) = \int_A X d\mathbb{P} \text{ pour tout } A \in \mathcal{G}.$$

Comme X est intégrable, μ est une mesure bornée sur $\mathcal{G}$, et pour tout $A \in \mathcal{G}$, $\mathbb{P}(A) = 0 \Rightarrow \mu(A) = 0$. D'après le théorème de Radon-Nikodym, il existe une unique fonction Z $\mathcal{G}$-mesurable telle que

$$\mu(A) = \int_A Z d\mathbb{P} \text{ pour tout } A \in \mathcal{G}.$$

Dans ce résultat, Z est intégrable positive et $Z = \dfrac{d\mu}{d\mathbb{P}}$ est unique à un ensemble négligeable près.

Si X n'est pas positive presque sûrement, on pose

$$Z = \frac{d\mu^+}{d\mathbb{P}} - \frac{d\mu^-}{d\mathbb{P}}$$

avec

$$\mu^+(A) = \int_A X^+ d\mathbb{P} \text{ et } \mu^-(A) = \int_A X^- d\mathbb{P} \text{ pour tout } A \in \mathcal{G}$$

et le résultat reste vrai.

Définition 5.2 (*Conditionnement par rapport à une variable*)

Soit X une v.a.r. intégrable et soit T une v.a.r. On pose $\mathcal{G} = \sigma(T)$. L'espérance conditionnelle de X sachant T est $\mathbb{E}[X | T] = \mathbb{E}[X | \sigma(T)]$.

Remarque : D'après un résultat du chapitre 2, $\mathbb{E}[X | T] = g(T)$ pour une certaine fonction borélienne g. Dans l'exemple introductif, on remarque

$$\mathbb{E}[X_1 | X_1 + X_2] = 0 \cdot \mathbf{1}_{[X_1+X_2=0]} + \frac{1}{2} \cdot \mathbf{1}_{[X_1+X_2=1]} + 1 \cdot \mathbf{1}_{[X_1+X_2=2]} = \frac{X_1 + X_2}{2}.$$

5.1.3 Caractérisations et exemples

Voici une autre caractérisation possible de $\mathbb{E}[X | \mathcal{G}]$.

Proposition 5.2

Sous les hypothèses de la proposition 5.1, $Z = \mathbb{E}[X | \mathcal{G}]$ si et seulement si Z est $\mathcal{G}$-mesurable, intégrable et si

$$\text{Pour tout } Y\ \mathcal{G}\text{-mesurable et t.q. } YX \text{ intégrable, } \int_\Omega YZ d\mathbb{P} = \int_\Omega YX d\mathbb{P}. \qquad (5.2)$$

On pourra montrer ce résultat en deuxième lecture, il reprend les arguments constructifs du chapitre 3 et les résultats de convergence du chapitre 4.

On a une caractérisation dans le cas L^2 qui reprend l'idée que la meilleure approximation constante d'une fonction L^2 est sa moyenne vue dans le cas discret à la section 5.1.1.

Proposition 5.3

Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et si $\mathcal{G}$ est une sous-tribu de $\mathcal{F}$, alors $\mathbb{E}[X | \mathcal{G}]$ est la projection orthogonale de X sur $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Démonstration : L'espace $L^2(\Omega, \mathcal{G}, \mathbb{P})$ est un sous-espace vectoriel fermé de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Soit Z la projection orthogonale de X sur $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Alors Z est $\mathcal{G}$-mesurable et de carré intégrable par construction. De plus, pour tout $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$,

$$\langle Y, X - Z \rangle_2 = 0,$$

c'es-à-dire

$$\int_\Omega YZ d\mathbb{P} = \int_\Omega YX d\mathbb{P}.$$

Donc Z satisfait bien les conditions de la définition de l'espérance conditionnelle en prenant $Y = \mathbf{1}_A$ pour $A \in \mathcal{G}$.

Proposition 5.4 (*Cas discret*)

Soit X une v.a. intégrable et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$.

1. *Si $\mathcal{G} = \{\varnothing, \Omega\}$, alors $\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[X]$ presque sûrement.*
2. *Si $\mathcal{G}$ est engendrée par la partition $\mathcal{P} = \{A_i,\ i \in \mathbb{N}\}$ telle que $\mathbb{P}(A_i) \neq 0$ pour tout $i \in \mathbb{N}$, alors*

$$\mathbb{E}[X | \mathcal{G}] = \sum_{i \in \mathbb{N}} \left(\frac{1}{\mathbb{P}[A_i]} \int_{A_i} X d\mathbb{P} \right) \mathbf{1}_{A_i}.$$

Démonstration :

1. $\mathbb{E}[X]$ est $\mathcal{G}$ mesurable car constante et
$$(i)\ \int_{\varnothing} \mathbb{E}[X]d\mathbb{P} = \int_{\varnothing} Xd\mathbb{P} \text{ et } (ii)\ \int_{\Omega} \mathbb{E}[X]d\mathbb{P} = \int_{\Omega} Xd\mathbb{P}$$
puisque $\mathbb{P}(\Omega) = 1$.
2. On pose
$$Z = \sum_{i \in \mathbb{N}} \left(\frac{1}{\mathbb{P}[A_i]} \int_{A_i} Xd\mathbb{P} \right) \mathbf{1}_{A_i}.$$
Alors Z est $\mathcal{G}$-mesurable car constante sur tous les éléments de la partition. En outre, tout élément de $\mathcal{G}$ étant l'union d'un ensemble dénombrable de A_i l'égalité (5.1) aura lieu en tout A de $\mathcal{G}$ si elle a lieu en tout A_i de la partition. Soit donc $i \in \mathbb{N}$. Sur A_i, Z vaut
$$\frac{1}{\mathbb{P}[A_i]} \int_{A_i} Xd\mathbb{P}.$$
Donc
$$\int_{A_i} Zd\mathbb{P} = \mathbb{P}[A_i] \times \frac{1}{\mathbb{P}[A_i]} \int_{A_i} Xd\mathbb{P} = \int_{A_i} Xd\mathbb{P}$$
ce qu'il fallait démontrer.

Proposition 5.5 (*Cas à densité*)

Soient X et T deux variables aléatoires réelles, X étant intégrable. On suppose que (X,T) admet une densité $f : \mathbb{R}^2 \to \mathbb{R}$ et on note f_T la densité marginale de T. Alors $\mathbb{E}[X\,|T] = g(T)$ avec
$$g(t) = \mathbf{1}_{[f_T \neq 0]}(t) \frac{1}{f_T(t)} \int_{\mathbb{R}} xf(x,t)dx \text{ pour tout } t \in \mathbb{R}.$$

Démonstration : On rappelle que
$$f_T(t) = \int_{\mathbb{R}} f(x,t)dx \text{ pour tout } t \in \mathbb{R}.$$
On pose $Z = g(T)$ avec g définie dans la proposition. Alors g est borélienne (application du théorème de Fubini), donc Z est bien $\sigma(T)$-mesurable. Soit $A \in \sigma(T)$. Alors il existe $B \in \mathcal{B}(\mathbb{R})$ tel que $A = T^{-1}(B)$. On en déduit
$$\int_A Zd\mathbb{P} = \int_{\Omega} \mathbf{1}_{T^{-1}(B)} Zd\mathbb{P} = \int_{\Omega} \mathbf{1}_B(T)g(T)d\mathbb{P}.$$
On utilise alors la définition de g et la formule de transfert. Il vient, après simplification par $f_T(t)$ (un peu délicate mais il suffit de remarquer que $f_T(t) = 0 \Rightarrow f(t,x) = 0$ pour presque tout x),
$$\int_A Zd\mathbb{P} = \int_{\mathbb{R}} \mathbf{1}_B(t) \int_{\mathbb{R}} xf(x,t)dxdt = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_B(t)xf(x,t)dxdt.$$
Mais en utilisant à nouveau la formule de transfert, on trouve
$$\int_A Zd\mathbb{P} = \int_{\Omega} \mathbf{1}_B(T)Xd\mathbb{P} = \int_A Xd\mathbb{P}$$
ce qu'il fallait démontrer.

5.1.4 Principaux résultats

Il existe un certain nombre de résultats importants qui sont utilisés régulièrement dans le calcul des espérances conditionnelles. Chaque preuve d'un de ces résultats constitue un exercice facile et intéressant, nous laissons au lecteur le soin de rédiger ces preuves. Seuls le théorème 5.1 qui est le plus important pour les applications et le moins évident à manipuler, et le corollaire qui traite le cas des les vecteurs gaussiens sont démontrés.

Un résultat de type "barycentre".

Proposition 5.6

Soit X une v.a. intégrable définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ et soient $\mathcal{G}$ et $\mathcal{G}'$ deux sous-tribus de $\mathcal{F}$ telles que $\mathcal{G}' \subset \mathcal{G}$. Alors

$$\mathbb{E}\left[\mathbb{E}\left[X\,|\mathcal{G}\right]|\mathcal{G}'\right](= \mathbb{E}\left[\mathbb{E}\left[X\,|\mathcal{G}'\right]|\mathcal{G}\right]) = \mathbb{E}\left[X\,|\mathcal{G}'\right] \textit{ presque sûrement.}$$

L'espérance d'une espérance conditionnelle.

Corollaire 5.1

Sous les hypothèses sur X et $\mathcal{G}$ de la proposition précédente, $\mathbb{E}\left[\mathbb{E}\left[X\,|\mathcal{G}\right]\right] = \mathbb{E}\left[X\right]$.

Conditionnement sous hypothèse de mesurabilité.

Proposition 5.7

Soient X et Y deux v.a.r. définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On suppose XY et Y intégrables. Alors, si X est $\mathcal{G}$-mesurable,

$$\mathbb{E}\left[XY\,|\mathcal{G}\right] = X\mathbb{E}\left[Y\,|\mathcal{G}\right] \textit{ presque sûrement.}$$

Conditionnement sous hypothèse d'indépendance.

Proposition 5.8

Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On suppose X indépendante de $\mathcal{G}$. Alors

$$\mathbb{E}\left[X\,|\mathcal{G}\right] = \mathbb{E}\left[X\right] \textit{ presque sûrement.}$$

Corollaire 5.2

Soit (X, T) un vecteur gaussien, X et T étant de dimension quelconque. Alors, si $\mathbf{Cov}\left[T\right]$ inversible, on a

$$\mathbb{E}\left[X\,|T\right] = \mathbb{E}\left[X\right] + \mathbf{Cov}\left[X, T\right]\mathbf{Cov}\left[T\right]^{-1}(T - \mathbb{E}\left[T\right]).$$

Démonstration : On suppose pour commencer que X et T sont centrés. Il suffit de montrer que

$$\mathbb{E}\left[X\,|T\right] = \mathbf{Cov}\left[X, T\right]\mathbf{Cov}\left[T\right]^{-1}T.$$

Soit donc $Z = \mathbf{Cov}\left[X, T\right]\mathbf{Cov}\left[T\right]^{-1}T$. Z est linéaire en T donc $\sigma(T)$-mesurable. Par ailleurs, on a :

$$\mathbf{Cov}\left[Z, T\right] = \mathbf{Cov}\left[X, T\right]\mathbf{Cov}\left[T\right]^{-1}\mathbf{Cov}\left[T, T\right] = \mathbf{Cov}\left[X, T\right]$$

donc $\mathbf{Cov}[Z-X,T]=0$ et $Z-X$ et T sont décorrélés. Mais $(Z-X,T)$ est un vecteur gaussien (comme transformée linéaire de (X,T)), donc $Z-X$ et T sont indépendants. On en déduit en particulier que

$$\mathbb{E}[Z-X\,|T]=\mathbb{E}[Z-X]=0 \text{ puisque } X \text{ et } T \text{ sont centrés.}$$

Cela implique finalement le résultat cherché $\mathbb{E}[X\,|T]=\mathbb{E}[Z\,|T]=Z$ presque sûrement puisque Z est $\sigma(T)$-mesurable.

Si X et T sont quelconque, on applique le résultat que l'on vient de montrer à $X-\mathbb{E}[X]$ et $T-\mathbb{E}[T]$ centrés pour obtenir le résultat souhaité.

Théorème principal qui combine les résultats précédents.

Théorème 5.1

Soient X et T deux v.a.r. définies sur $(\Omega,\mathcal{F},\mathbb{P})$ et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. Soit $\psi:\mathbb{R}^2\to\mathbb{R}$ une application borélienne. On suppose $\psi(X,T)$ intégrable et $\psi(x,T)$ intégrable pour tout $x\in\mathbb{R}$.

On suppose que X est $\mathcal{G}$-mesurable et que T est indépendante de $\mathcal{G}$. Alors

$$\mathbb{E}[\psi(X,T)\,|\mathcal{G}]=\phi(X) \text{ p.s., où } \phi(x)=\mathbb{E}[\psi(x,T)] \text{ pour tout } x\in\mathbb{R}.$$

Ce théorème donne en fait la seule façon d'écrire proprement et conjointement le fait que X, $\mathcal{G}$-mesurable, n'est pas modifié par le conditionnement et que T, indépendante de $\mathcal{G}$, doit être intégrée.

Démonstration : On pose $Z=\phi(X)$. Alors ϕ est borélienne par Fubini et donc Z est $\sigma(X)$-mesurable donc $\mathcal{G}$-mesurable. Z est également intégrable (toujours par Fubini) et on va montrer pour conclure la propriété (5.2). Soit donc Y $\mathcal{G}$-mesurable de sorte que YZ et YX soient intégrables. On veut montrer que

$$\int_\Omega YZd\mathbb{P}=\int_\Omega Y\psi(X,T)d\mathbb{P}.$$

On a, par la formule de transfert,

$$\int_\Omega Y\psi(X,T)d\mathbb{P}=\int_{\mathbb{R}^3} y\psi(x,t)\mathbb{P}_{(X,Y,T)}(dx,dy,dt),$$

mais (X,Y) et T sont indépendants donc $\mathbb{P}_{(X,Y,T)}=\mathbb{P}_{(X,Y)}\otimes\mathbb{P}_T$. On en déduit par Fubini

$$\int_\Omega Y\psi(X,T)d\mathbb{P}=\int_{\mathbb{R}^2} y\int_{\mathbb{R}}\psi(x,t)d\mathbb{P}_T(dt)d\mathbb{P}_{(X,Y)}(dx,dy).$$

Mais pour tout $x\in\mathbb{R}$,

$$\int_{\mathbb{R}}\psi(x,t)d\mathbb{P}_T(dt)=\int_\Omega\psi(x,T)d\mathbb{P}=\mathbb{E}[\psi(x,T)]=\phi(x).$$

Il vient

$$\int_\Omega Y\psi(X,T)d\mathbb{P}=\int_{\mathbb{R}^2} y\phi(x)d\mathbb{P}_{(X,Y)}(dx,dy)=\int_\Omega Y\phi(X)d\mathbb{P},$$

ce qu'il fallait démontrer.

Remarque : L'espérance conditionnelle satisfait la plupart des propriétés de l'espérance (linéarité, convergence monotone, convergence dominée, inégalité de Jensen...).

5.2 Filtrations, temps d'arrêt

5.2.1 Définitions

Dans ce chapitre, on introduit les filtrations de tribus qui permettent de faire dépendre du temps la quantité d'information connue. Les temps d'arrêt sont des temps aléatoires dont la définition s'appuie sur les filtrations pour introduire la notion de non-anticipativité.

Définition 5.3 (*Filtration*)

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On appelle *filtration de* $\mathcal{F}$ toute suite croissante de sous-tribus de $\mathcal{F}$.

On appelle *espace probabilisé filtré* le quadruplet $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ où $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ est une filtration de $\mathcal{F}$.

Définition 5.4 (*Processus adapté*)

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}, \mathbb{P})$ un espace probabilisé filtré. Un *processus* $(\mathcal{F}_n)_{n\in\mathbb{N}}$*-adapté* est une suite de v.a. X_n définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ telles que X_n est $\mathcal{F}_n$-mesurable pour tout $n \in \mathbb{N}$.

Définition 5.5 (*Filtration naturelle*)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On appelle *filtration naturelle* du processus $(X_n)_{n\in\mathbb{N}}$ la plus petite filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ qui rend le processus adapté. Elle est définie par

$$\mathcal{F}_n = \sigma - \{X_k,\ k \leq n\},\ \text{pour tout } n \in \mathbb{N}.$$

Définition 5.6 (*Temps d'arrêt*)

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et soit $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ une filtration de $\mathcal{F}$. Un $\mathbb{F}$*-temps d'arrêt* est une v.a. T à valeurs dans $\overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$ qui vérifie

$$[T \leq n] \in \mathcal{F}_n \text{ pour tout } n \in \mathbb{N}.$$

Voici une définition équivalente des temps d'arrêt qui est couramment utilisée dans le cas discret mais qui n'a pas d'équivalent en temps continu au contraire de la définition ci-dessus.

Proposition 5.9

Sous les conditions de la définition précédente, T est un $\mathbb{F}$-temps d'arrêt si et seulement si

$$[T = n] \in \mathcal{F}_n \textit{ pour tout } n \in \mathbb{N}.$$

Démonstration : On suppose T temps d'arrêt. Alors $[T = 0] = [T \leq 0] \in \mathcal{F}_0$ et pour tout $n \geq 1$

$$[T = n] = [T \leq n] \backslash [T \leq n-1] \in \mathcal{F}_n$$

d'après les propriétés des tribus. Réciproquement,

$$[T \leq n] = \bigcup_{k\leq n} [T = k] \text{ pour tout } n \geq 0,$$

ce qui permet de conclure.

On remarque quand on lit ces définitions, que ces temps d'arrêt sont non anticipatifs, au sens où on n'utilise que les informations sur les événements survenus avant le temps n pour déterminer si le temps d'arrêt est survenu ou non à cet instant-là.

5.2.2 Exemples

Par exemple, parmi les temps aléatoires, on trouve le premier temps (ou le second ...) où un événement particulier se produit. La production ou non de cet événement au temps n est généralement modélisée par une v.a. X_n $\mathcal{F}_n$-mesurable, ce qui assure la non anticipativité.

On suppose maintenant que $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ est un espace probabilisé filtré et que $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$. Les variables aléatoires sont définies sur cet espace.

Exemple :

1. T constante. Alors $[T = n]$ est vide ou égale à Ω quelque soit n et donc $[T = n] \in \mathcal{F}_n$ pour tout n et T est un $\mathbb{F}$-temps d'arrêt.
2. Premier temps d'entrée d'un processus adapté dans un ensemble. Soit $(X_n)_{n\in\mathbb{N}}$ un processus adapté à valeurs dans $(E, \mathcal{B}(E))$un espace vectoriel normé muni de sa tribu borélienne. Soit enfin $B \in \mathcal{B}(E)$. Alors la v.a. T définie par
$$T = \inf\{n \geq 0/\ X_n \in B\} \text{ (et } T = +\infty \text{ si } X_n \notin B \text{ pour tout } n)$$
est un $\mathbb{F}$-temps d'arrêt. On l'appelle le premier temps d'atteinte de B ou premier temps d'entrée dans B.
En effet, pour tout $n \in \mathbb{N}$, $T \leq n$ si et seulement si il existe $k \leq n$ tel que $X_k \in B$, i.e.
$$[T \leq n] = \bigcup_{k=0}^{n} [X_k \in B] \in \mathcal{F}_n.$$
3. Dans ce même contexte, le dernier temps de passage dans B n'est pas un $\mathbb{F}$-temps d'arrêt car, pour savoir s'il s'agit du dernier passage, il faut connaître les événements futurs. Précisément, soit T' ce temps.
$$T' = \sup\{n \geq 0/\ X_n \in B\} \text{ (et } T' = +\infty \text{ si } X_n \notin B \text{ pour tout } n).$$
Alors $T' \leq n$ si et seulement si il existe $k \leq n$ tel que $X_k \in B$ et pour tout $k \geq n+1$, $X_k \notin B$, i.e.
$$[T' \leq n] = \left(\bigcup_{k=0}^{n} [X_k \in B]\right) \cap \left(\bigcap_{k\geq n+1} [X_k \notin B]\right)$$
et cet événement n'est pas dans $\mathcal{F}_n$ en général.

Question : Définir le second temps d'atteinte de B. Est-ce un temps d'arrêt ? Justifier votre réponse.

5.2.3 Propriétés

On suppose toujours que $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ est un espace probabilisé filtré, que $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ et que les variables aléatoires sont définies sur cet espace.

Proposition 5.10 (*Premières propriétés*)

1. *Soient S et T deux temps d'arrêt. Alors, $S + T$, $S \vee T$, $S \wedge T$ (et en particulier, pour une constante $n \in \mathbb{N}$, $T + n$, $T \vee n$, $T \wedge n$) sont des temps d'arrêt.*
2. *Si $(T_n)_{n\in\mathbb{N}}$ est une suite de temps d'arrêt alors $\sup_n T_n$ et $\inf_n T_n$ également*

Cette proposition peut-être démontrée en exercice.

Il est important de pouvoir donner des informations sur une suite à un instant aléatoire. Si $(X_n)_{n\in\mathbb{N}}$ est une suite et T un temps aléatoire tel que $T < +\infty$, on peut définir la valeur de la suite au temps T comme la v.a. X_T définie par

$$(X_T)(\omega) = X_{T(\omega)}(\omega) \text{ pour tout } \omega \in \Omega.$$

Définition 5.7 (*Suite arrêtée*)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite $\mathbb{F}$-adaptée et soit T un $\mathbb{F}$-temps d'arrêt. On appelle suite $(X_n)_{\in\mathbb{N}}$ arrêtée au temps T la suite $(X_{T\wedge n})_{n\in\mathbb{N}}$.

Proposition 5.11

Une suite $\mathbb{F}$-adaptée $(X_n)_{n\in\mathbb{N}}$ arrêtée au $\mathbb{F}$-temps d'arrêt T est une suite $\mathbb{F}$-adaptée.

Démonstration : En effet, on remarque que, pour tout $n \in \mathbb{N}$

$$X_{T\wedge n} = \left(\sum_{k=0}^{n} \mathbf{1}_{[T=k]} + \mathbf{1}_{[T>n]}\right) X_{T\wedge n} = \sum_{k=0}^{n} \mathbf{1}_{[T=k]} X_k + \mathbf{1}_{[T>n]} X_n.$$

Pour tout $k \leq n$, $[T = k] \in \mathcal{F}_n$ et X_k $\mathcal{F}_n$-mesurable. De plus $[T > n] = [T \leq n]^c \in \mathcal{F}_n$ donc $X_{T\wedge n}$ est $\mathcal{F}_n$-mesurable.

Avant de définir les martingales qui synthétisent les éléments introduits dans les deux premières sections de ce chapitre, nous allons encore introduire la notion d'événements antérieurs à un temps aléatoire. Cette notion nous est connue pour un temps déterministe puisque les événements antérieurs à n sont les événements de $\mathcal{F}_n$. On veut donc définir une tribu $\mathcal{F}_T$ pour un temps aléatoire T.

Définition 5.8

On appelle tribu des événements antérieurs à T, v.a. à valeurs dans $\overline{\mathbb{N}}$, la famille

$$\mathcal{F}_T = \{A \in \mathcal{F} /\ \forall n \in \mathbb{N}, A \cap [T \leq n] \in \mathcal{F}_n\}.$$

Remarque : Pour tout $n \in \mathbb{N}$ et pour tout $A \in \mathcal{F}$,

$$A \cap [T \leq n] = \{\omega \in A /\ T(\omega) \leq n\}.$$

Proposition 5.12

Si T est un $\mathbb{F}$-temps d'arrêt alors $\mathcal{F}_T$ est une sous-tribu de $\mathcal{F}$.

Démonstration :

. $\Omega \in \mathcal{F}_T$. En effet, pour tout n, $\Omega \cap [T \leq n] = [T \leq n] \in \mathcal{F}_n$.

. Soit $A \in \mathcal{F}_T$. Montrons que $A^c \in \mathcal{F}_T$. Soit donc $n \in \mathbb{N}$, on a

$$A^c \cap [T \leq n] = [T \leq n] \backslash A = [T \leq n] \backslash (A \cap [T \leq n]) \in \mathcal{F}_n$$

donc $A^c \in \mathcal{F}_T$.

. Soit $(A_k)_{k\geq 1}$ une suite de $\mathcal{F}_T$ et soit $n \in \mathbb{N}$. Alors

$$\left(\bigcup_{k\geq 1} A_k\right) \cap [T \leq n] = \bigcup_{k\geq 1} (A_k \cap [T \leq n]) \in \mathcal{F}_n.$$

Donc $\mathcal{F}_T$ est stable par union dénombrable.

On a donc bien vérifié que $\mathcal{F}_T$ est une tribu.

Proposition 5.13

Soient $(X_n)_{n\in\mathbb{N}}$ une suite $\mathbb{F}$-adaptée à valeur dans $(E,\mathcal{B}(E))$, un espace vectoriel normé muni de sa tribu borélienne, et T un $\mathbb{F}$-temps d'arrêt. Alors la v.a. $X_T\mathbf{1}_{[T<+\infty]}$ est $\mathcal{F}_T$-mesurable.

Démonstration : Soit $B\in\mathcal{B}(E)$. On définit

$$A=[X_T\mathbf{1}_{[T<+\infty]}\in B].$$

On doit montrer que, pour tout $n\in\mathbb{N}$,

$$A\cap[T\le n]\in\mathcal{F}_n.$$

Or

$$A\cap[T\le n]=\bigcup_{k=0}^{n}(A\cap[T=k])=\bigcup_{k=0}^{n}([X_k\in B]\cap[T=k])\in\mathcal{F}_n.$$

5.3 Martingales à temps discret

Soit $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$, avec $\mathbb{F}=(\mathcal{F}_n)_{n\in\mathbb{N}}$, un espace probabilisé filtré. On considère des suites à valeurs dans $(\mathbb{R},\mathcal{B}(\mathbb{R}))$.

5.3.1 Définition et propriétés

Définition 5.9 (*Martingale*)

Une suite aléatoire $(X_n)_{n\ge0}$ définie sur $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ est une $\mathbb{F}$-*martingale* si, pour tout $n\ge 0$,

1. X_n et $\mathcal{F}_n$-mesurable,
2. X_n est intégrable,
3. $\mathbb{E}\left[X_{n+1}\,|\mathcal{F}_n\right]=X_n$ presque sûrement.

Définition 5.10 (*Sur- et sous-martingales*)

Une suite définie comme ci-dessus satisfaisant, pour tout n, les hypothèses 1. et 2. et

4. $\mathbb{E}\left[X_{n+1}\,|\mathcal{F}_n\right]\le X_n$ (resp. $\mathbb{E}\left[X_{n+1}\,|\mathcal{F}_n\right]\ge X_n$) presque sûrement

est appelée une $\mathbb{F}$-*sur-martingale* (resp. $\mathbb{F}$-*sous-martingale*).

Remarque :

(i) Lorsqu'il n'y a pas d'ambigüité sur la filtration concernée, on peut parler plus simplement de martingale.

(ii) Une suite à la fois sur- et sous- martingale relativement à une même filtration est une martingale (toujours relativement à cette filtration).

(iii) Toute $\mathbb{F}$-martingale est martingale relativement à sa filtration naturelle.

(iv) La relation 3. pour tout n dans la définition de la martingale peut être remplacée par

$$\mathbb{E}\left[X_{n+k}\,|\mathcal{F}_n\right]=X_n \text{ presque sûrement, pour tout } n\ge0 \text{ et pour tout } k\ge1.$$

Question : Démontrer les deux derniers points ci-dessus.

Proposition 5.14 (*Propriétés des martingales*)

1. *Si $(X_n)_{n\in\mathbb{N}}$ est une martingale, alors $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ pour tout $n \geq 0$.*
2. *Les constantes sont des martingales par rapport à toute filtration.*
3. *Tout combinaison linéaire de martingales est une martingale (par rapport à une même filtration donnée).*
4. *$(X_n)_{n\in\mathbb{N}}$ est une sous-martingale si et seulement si $(-X_n)_{n\in\mathbb{N}}$ est une sur-martingale.*
5. *Si $(X_n)_{n\in\mathbb{N}}$ est une martingale et φ une application convexe telle que $\varphi(X_n)$ est intégrable pour tout $n \in \mathbb{N}$, alors $(\varphi(X_n))_{n\in\mathbb{N}}$ est une sous-martingale. En particulier, si $(X_n)_{n\in\mathbb{N}}$ est une martingale alors $(|X_n|)_{n\in\mathbb{N}}$ est une sous-martingale et si $(X_n)_{n\in\mathbb{N}}$ est une martingale de carré intégrable alors $(X_n^2)_{n\in\mathbb{N}}$ est une sous-martingale.*
6. *Si $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ sont des sous-martingales alors $(X_n \vee Y_n)_{n\in\mathbb{N}}$ est une sous-martingale.*

Question : Démontrer la proposition précédente, comme exercice d'application sur l'espérance condtionnelle.

Exemple :

1. **Marche aléatoire :** Soit $(\xi_n)_{n\in\mathbb{N}}$ une suite i.i.d. de v.a. d'espérance m. Pour tout $n \in \mathbb{N}$, on pose $S_n = \xi_1 + ... + \xi_n$. $(S_n)_{n\in\mathbb{N}}$ est appelée *marche aléatoire*. C'est un processus adapté à la filtration naturelle de la suite $(\xi_n)_{n\in\mathbb{N}}$ que l'on note $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

   Question :

   (a) Montrer que, pour tout $n \geq 1$, $\mathbb{E}[S_{n+1} | \mathcal{F}_n] = S_n + m$ presque sûrement et en déduire la nature de la suite dans les cas $m = 0$, $m > 0$, $m < 0$.

   (b) Etudier le cas particulier où les ξ_n sont de loi $\mathcal{B}(p)$ en fonction de $p \in [0, 1]$.

2. **Martingale fermée :** Soit X une v.a. intégrable et soit $(\mathcal{F}_n)_{n\in\mathbb{N}}$ une filtration. Pour tout n, on pose $X_n = \mathbb{E}[X | \mathcal{F}_n]$. Montrer alors que $(X_n)_{n\in\mathbb{N}}$ est une martingale.
3. **Changement de probabilité :** Soient $\mathbb{P}$ et $\mathbb{Q}$ deux probabilités sur $(\Omega, \mathcal{F})$ et soir $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ une filtration de $\mathcal{F}$. Pour tout $n \in \mathbb{N}$, on définit $\mathbb{P}_n$ et $\mathbb{Q}_n$ les restrictions de $\mathbb{P}$ et $\mathbb{Q}$ à $\mathcal{F}_n$. On suppose que, pour tout n, $\mathbb{Q}_n << \mathbb{P}_n$ et on note X_n la densité de $\mathbb{Q}_n$ par rapport à $\mathbb{P}_n$. Alors $(X_n)_{n\in\mathbb{N}}$ est une $\mathbb{F}$-martingale.

   Démonstration : La suite $(X_n)_{n\in\mathbb{N}}$ est $\mathbb{F}$-adaptée et intégrable par construction. De plus, pour tout $n \in \mathbb{N}$ et pour tout $A \in \mathcal{F}_n$, comme $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, on a $\mathbb{Q}_n(A) = \mathbb{Q}_{n+1}(A)$, soit

   $$\int_A X_n d\mathbb{P} = \int_A X_{n+1} d\mathbb{P}$$

   i.e. $\mathbb{E}[X_n + 1 | \mathcal{F}_n] = X_n$, ce qu'il fallait démontrer.

5.3.2 Théorème d'arrêt

La propriété de martingale peut s'étendre aux temps d'arrêt bornés.

Proposition 5.15

Soit $(M_n)_{n\in\mathbb{N}}$ une $\mathbb{F}$-martingale. Soient T et S deux $\mathbb{F}$-temps d'arrêt majorés par une constante c presque sûrement et tels que $S \leq T$. Alors

$$\mathbb{E}[M_T | \mathcal{F}_S] = M_S \text{ presque sûrement.}$$

Démonstration : En premier lieu, M_S est bien $\mathcal{F}_S$-mesurable. Soit $B \in \mathcal{F}_S$. D'une part

$$\int_B M_S d\mathbb{P} = \sum_{j=0}^{c} \int_{B\cap[S=j]} M_j d\mathbb{P} = \sum_{j=0}^{c} \int_{B\cap[S=j]} M_c d\mathbb{P} = \int_B M_c d\mathbb{P}$$

car $B \cap [S = j] = (B \cap [S \leq j]) \cap [S > j-1] \in \mathcal{F}_j$ pour $j \leq c$ et $(M_n)_{n\in\mathbb{N}}$ est une martingale. D'autre part, comme $\mathcal{F}_S \subset \mathcal{F}_T$, $B \in \mathcal{F}_T$ et on a également

$$\int_B M_T d\mathbb{P} = \int_B M_c d\mathbb{P}.$$

On en déduit

$$\int_B M_S d\mathbb{P} = \int_B M_T d\mathbb{P},$$

ce qu'il fallait démontrer.

On déduit de cette proposition le théorème d'arrêt suivant.

Théorème 5.2 (*Théorème d'arrêt*)

Soient $(M_n)_{n\in\mathbb{N}}$ une $\mathbb{F}$-martingale et T un $\mathbb{F}$-temps d'arrêt. Alors $(M_{T\wedge n})_{n\in\mathbb{N}}$ est une $\mathbb{F}$-martingale.

Démonstration : On a vu que $(M_{T\wedge n})_{n\in\mathbb{N}}$ est une suite $\mathbb{F}$-adaptée. En outre,

$$M_{T\wedge n} = \sum_{k=0}^{n} M_k \mathbf{1}_{[T=k]}$$

avec

$$|M_k \mathbf{1}_{[T=k]}| \leq |M_k|$$

pour tout $k \leq n$. Donc $M_{T\wedge n}$ est intégrable. Soit alors $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}\left[M_{T\wedge(n+1)} \,|\mathcal{F}_n\right] &= \mathbb{E}\left[M_{T\wedge n}\mathbf{1}_{[T\leq n]} + M_{n+1}\mathbf{1}_{[T>n]} \,|\mathcal{F}_n\right] \\ &= \mathbb{E}\left[M_{T\wedge n}\mathbf{1}_{[T\leq n]} \,|\mathcal{F}_n\right] + \mathbb{E}\left[M_{n+1}\mathbf{1}_{[T>n]} \,|\mathcal{F}_n\right]. \end{aligned}$$

La première variable conditionnée est $\mathcal{F}_n$-mesurable donc

$$\mathbb{E}\left[M_{T\wedge n}\mathbf{1}_{[T\leq n]} \,|\mathcal{F}_n\right] = M_{T\wedge n}\mathbf{1}_{[T\leq n]} \text{ presque sûrement,}$$

et comme $[T > n] \in \mathcal{F}_n$,

$$\mathbb{E}\left[M_{n+1}\mathbf{1}_{[T>n]} \,|\mathcal{F}_n\right] = \mathbf{1}_{[T>n]}\mathbb{E}\left[M_{n+1} \,|\mathcal{F}_n\right] = M_n\mathbf{1}_{[T>n]} \text{ p.s.}$$

puisque $(M_n)_{n\in\mathbb{N}}$ est une martingale. D'où, comme sur $[T > n]$, $M_n = M_{T\wedge n}$,

$$\mathbb{E}\left[M_{T\wedge(n+1)} \,|\mathcal{F}_n\right] = M_{T\wedge n}\mathbf{1}_{[T\leq n]} + M_{T\wedge n}\mathbf{1}_{[T>n]} = M_{T\wedge n} \text{ p.s.}$$

ce qu'il fallait démontrer.

5.3.3 Convergence des martingales

Exemple : [Urne de Polya] Ce processus est un exemple de processus de branchement.

Une urne contient r boules rouges et b boules blanches. Soit c un entier positif fixé. A tout temps n, on extrait une boule au hasard dans l'urne et on l'y replace avec c boules de la même couleur. On note X_n la proportion de boules blanches dans l'urne après n tirages.

☐ On modélise la suite $(X_n)_{n\in\mathbb{N}}$.

1. Donner X_0.

2. Exprimer X_{n+1} en fonction de X_n.
3. Montrer que $(X_n)_{n\in\mathbb{N}}$ est une martingale (par rapport à sa filtration naturelle).

☐ On cherche maintenant le comportement asymptotique de la suite $(X_n)_{n\in\mathbb{N}}$. L'enjeu est ici de voir s'il y a une proportion limite. En fait, on peut montrer (on ne le fera pas ici) que $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers une v.a. de loi bêta de paramètres $(b/c, r/c)$. En outre, si $b = r = c$ alors la limite presque sûre suit une loi uniforme sur $[0, 1]$.

L'étude de la convergence des martingales en temps discret est un chapitre important de la théorie des probabilités. Nous donnons ici quelques résultats sans preuve qui donnent néanmoins un aperçu des problématiques et quelques outils pratiques.

Théorème 5.3

Toute martingale bornée dans L^1 converge presque sûrement vers une v.a. intégrable.

Remarque :

. Ce théorème s'applique par exemple à l'urne de Polya car la suite $(X_n)_{n\in\mathbb{N}}$ est bornée dans L^∞ ($0 \leq X_n \leq 1$ presque sûrement pour tout n).

. Ce résultat est bien sûr faux pour une suite quelconque de L^1.

On peut remplacer la condition de borne dans L^1 par la condition d'équi-intagrabilité suivante.

Définition 5.11 (*Suite équi-intégrable*)

Une suite $(X_n)_{n\in\mathbb{N}}$ est dite équi-intégrable si

$$\lim_{c\to+\infty} \sup_{n\in\mathbb{N}} \int_{[|X_n|\geq c]} |X_n| \, d\mathbb{P} = 0.$$

Proposition 5.16 (*Critères d'équi-intégrabilité*)

(i) Les suites bornées dans L^p, pour $p > 1$ sont équi-intégrables.

(ii) Les suites dominées par une fonction L^1 sont équi-intégrables.

Avec cette condition, on va pouvoir préciser le lien entre la suite et sa limite.

Théorème 5.4

Soit $(M_n)_{n\in\mathbb{N}}$ une martingale.

1. *Si $(M_n)_{n\in\mathbb{N}}$ est équi-intégrable alors $(M_n)_{n\in\mathbb{N}}$ converge presque sûrement et dans L^1 vers une v.a. intégrable M et $M_n = \mathbb{E}[M\,|\mathcal{F}_n]$ pour tout n.*
2. *Si $M_n = \mathbb{E}[Y\,|\mathcal{F}_n]$ pour tout n où $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, alors $(M_n)_{n\in\mathbb{N}}$ converge presque sûrement et dans L^1 vers $\mathbb{E}[Y\,|\mathcal{G}]$ où*

$$\mathcal{G} = \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right).$$

Enfin, ce dernier théorème aborde la problématique de la convergence L^2 (les convergence L^1 et presque sûres découlaient déjà du théorème précédent et des critères d'équi-intégrabilité).

Théorème 5.5

Si $(M_n)_{n\in\mathbb{N}}$ est une martingale bornée dans L^2 alors $(M_n)_{n\in\mathbb{N}}$ converge dans L^2 vers une v.a. M telle que $M_n = \mathbb{E}[M\,|\mathcal{F}_n]$ presque sûrement pour tout n.

5.4 Exercices

5.4.1 Conditionnement

Exercice 5.1.

Soient X et Y deux v.a. réelles définies sur un même espace probabilisé.

1. On suppose X et Y à valeurs dans $\mathbb{N}$ et X intégrable, calculer $\mathbb{E}[X\,|Y]$.
2. On suppose X et Y indépendantes de loi de Poisson de paramètre respectif λ et μ, calculer $\mathbb{E}[X\,|Y+X]$.
3. On suppose X et Y indépendantes de loi uniforme sur $[0,1]$. On pose $U=\max(X,Y)$ et $V=\min(X,Y)$, calculer $\mathbb{E}[V\,|U]$.

Exercice 5.2. [*Espérance conditionnelle*]

Soient X et Y deux v.a indépendantes et de même loi P.

1. Montrer que $\mathbb{E}[X\,|X+Y]=\mathbb{E}[Y\,|X+Y]$ ps.
2. En déduire $\mathbb{E}[X\,|X+Y]$.

Exercice 5.3. [*Vecteurs gaussiens*]

Soit (X,Y) un couple gaussien centré de matrice de covariance $\begin{pmatrix} 4/3 & 1 \\ 1 & 1 \end{pmatrix}$.

Calculer $\mathbb{E}[X\,|Y-X]$ et donner sa loi.

Exercice 5.4. [*Fonction caractéristique et conditionnement*] On suppose que X et Y sont deux v.a telles que $Y \rightsquigarrow \mathcal{N}(m,\alpha^2)$ et $\mathbb{E}\left[e^{itX}|Y\right]=e^{-\sigma^2t^2/2}e^{itY}$.

1. Déterminer la fonction caractéristique de X et en déduire sa loi.
2. Même question pour (X,Y)
3. Montrer que $X-Y$ et Y sont indépendantes
4. Déterminer l'espérance conditionnelle de Y sachant X.

Exercice 5.5. [*Variance conditionnelle*]

Soit X une variable aléatoire réelle de carré intégrable définie sur un espace probabilisé $(\Omega,\mathcal{F},\mathbb{P})$ et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On pose

$$\mathbf{Var}[X|\mathcal{G}]=\mathbb{E}\left[(X-\mathbb{E}[X\,|\mathcal{G}])^2\,|\mathcal{G}\right].$$

Montrer que

$$\mathbf{Var}[X]=\mathbb{E}[\mathbf{Var}[X|\mathcal{G}]]+\mathbf{Var}[\mathbb{E}[X\,|\mathcal{G}]].$$

Exercice 5.6. [*Somme aléatoire de variables aléatoires*]

Soit $(X_i)_{i\geq 1}$ une suite de v.a. d'espérance μ. Soit N une v.a. à valeurs dans $\mathbb{N}^*$, indépendante des v.a. X_i, d'espérance m. On pose

$$S_N=\sum_{i=1}^{N}X_i.$$

Calculer $\mathbb{E}[S_N\,|N]$ puis en déduire $\mathbb{E}[S_N]$.

5.4.2 Temps d'arrêt

Exercice 5.7. Montrer la proposition 5.10

Exercice 5.8. Soient T et S deux temps d'arrêt. On suppose $T \leq S$. Montrer que $\mathcal{F}_T \subset \mathcal{F}_S$.

Exercice 5.9. Soient T et S deux temps d'arrêt.

1. Montrer que $\mathcal{F}_{T \wedge S} = \mathcal{F}_T \cap \mathcal{F}_S$.
2. Montrer que $\mathcal{F}_{T \vee S} \supset \mathcal{F}_T \bigvee \mathcal{F}_S$ $(= \sigma(\mathcal{F}_T \cup \mathcal{F}_S)$ par définition).

5.4.3 Martingales

Exercice 5.10. [*LA martingale du casino*]

Le casino propose un jeu de Pile ou Face avec une pièce équilibrée. A chaque tour le joueur doit, pour pouvoir jouer, faire une mise. S'il perd, il perd la totalité de sa mise et s'il gagne il remporte deux fois sa mise (donc sa mise plus un gain égal à sa mise).

Un joueur décide d'adopter la stratégie suivante : il mise un euro au premier tour. Tant qu'il perd, il double sa mise et dès qu'il gagne il s'arrête de jouer.

1. Rappeler pourquoi, presque sûrement, le joueur va gagner.
2. Modéliser le gain du joueur par une martingale arrêtée.
3. Quel est le gain du joueur quand finit le jeu ? Quel est le gain moyen à tout temps n ?
4. Soit M la mise totale du joueur au cours du jeu. Exprimer M puis calculer son espérance.

Exercice 5.11. [*Autour du théorème d'arrêt*]

Soient $(X_n)_n$ une $\mathbb{F}$-martingale et T un $\mathbb{F}$-temps d'arrêt.

1. On suppose pour cette question que $T \leq c$. Montrer, en utilisant le théorème d'arrêt, que $\mathbb{E}[X_T] = \mathbb{E}[X_S]$ pour $T, S \leq c$.
2. On suppose pour cette question que T est fini p.s., que X_T est intégrable et que $\mathbb{E}\left[|X_n|\mathbf{1}_{[T>n]}\right]$ tend vers 0 quand n tend vers l'infini. Montrer que $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.
(*Ind : Remarquer que* $X_{T\wedge n} = X_{T\wedge n}\mathbf{1}_{[T\leq n]} + X_{T\wedge n}\mathbf{1}_{[T>n]}$).
3. On suppose que $(X_n)_n$ converge dans L^1 vers une variable X. On définit
$$X_T = \begin{cases} X_n & \text{sur} \quad [T = n] \\ X & \text{sur} \quad [T = +\infty] \end{cases}$$
et l'on suppose que X_T est intégrable. Montrer que $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.
4. Soit $(Y_n)_n$ une suite de variables aléatoires intégrables adaptée à $\mathbb{F}$. On suppose que, pour tout temps d'arrêt T borné, $\mathbb{E}[Y_T] = \mathbb{E}[Y_0]$.
   (a) Soient n fixé et $A \in \mathcal{F}_n$. On définit $T = n\mathbf{1}_{A^c} + (n+1)\mathbf{1}_A$. Montrer que T est un $\mathbb{F}$-temps d'arrêt et en déduire que $\mathbb{E}[Y_{n+1} | \mathcal{F}_n] = Y_n$ presque sûrement.
   (b) Montrer que $(Y_n)_n$ est une martingale.

Exercice 5.12. [*Martingales et convergence p.s.*]

Soit $(X_n)_n$ une suite aléatoire i.i.d de loi exponentielle de paramètre 1. Soit $(\varepsilon_n)_n$ une suite aléatoire i.i.d telle $\mathbb{P}[\varepsilon_1 = 1] = \mathbb{P}[\varepsilon_1 = -1] = 1/2$. On suppose que les deux suites sont indépendantes, et l'on note $\mathcal{F}_n$ la tribu engendrée par $(\varepsilon_1, \ldots, \varepsilon_n, X_1, \ldots, X_n)$ pour $n \geq 1$, avec $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Soit $p \in]1, 2[$. On pose

$$M_n = \sum_{i=1}^{n} \varepsilon_i (X_1 + \cdots + X_i)^{-1/p}, \; n > 0.$$

1. Montrer que $(M_n)_n$ est une $\mathbb{F}$-martingale.
2. Montrer que
$$\mathbb{E}\left[(M_n - M_1)^2\right] \leq \int_{\mathbb{R}_+} x^{-2/p}(1 - e^{-x})dx, \; n \in \mathbb{N}.$$

   En déduire que $(M_n)_n$ converge p.s..