

AKTAR YALÇIN

INTRODUCTION AUX PRO-  
BABILITÉS  
EISTI ING1-GM 2018-2019

EISTI



## *Table des matières*

1. *Axiomes de Probabilités*      5
  
2. *Probabilités conditionnelles et indépendance et Probabilités sur un espace fini ou dénombrable*      11
  
3. *Variables aléatoires sur un espace fini ou dénombrable et Construction d'une mesure de probabilité*      15
  
4. *Probabilités sur  $\mathbb{R}$  et fonctions de répartition*      21
  
5. *Intégration par rapport à une mesure de probabilité (espérance, variance)*  
25
  
6. *Variables aléatoires indépendantes et Lois de probabilités sur  $\mathbb{R}$*       29
  
7. *Fonctions caractéristiques et quelques propriétés*      39
  
8. *Sommes de variables aléatoires indépendantes et Variables aléatoires gaussiennes*      41
  
9. *Convergence des variables aléatoires, convergence en loi et applications avec fonctions caractéristiques*      45

*10. Loi des grands nombres et le théorème-limite central* 51

*Bibliographie* 53

# 1. Axiomes de Probabilités

Soit  $\Omega$  un espace abstrait, i.e. sans structure particulière.

Rappelons que  $\mathcal{P}(\Omega)$  ou  $2^\Omega$  désigne l'ensemble de tous les sous-ensembles de  $\Omega$ , y compris le sous-ensemble vide noté  $\emptyset$ .

**Définition 1 (tribu (ou  $\sigma$ -algèbre))** Une partie  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  est appelée tribu (ou  $\sigma$ -algèbre) sur  $\Omega$  si elle vérifie les propriétés suivantes :

- (1)  $\Omega \in \mathcal{A}$
- (2)  $\mathcal{A}$  est stable par complémentaire i.e. si  $A \in \mathcal{A}$  alors  $A^c \in \mathcal{A}$ , où  $A^c = \Omega \setminus A$ .
- (3)  $\mathcal{A}$  est stable par union dénombrable i.e. si pour tout  $i \in \mathbb{N}^*$ ,  $A_i \in \mathcal{A}$ , alors  $\bigcup_{i \in \mathbb{N}^*} A_i \in \mathcal{A}$ .

Dans la définition d'une algèbre d'ensembles, "dénombrable" est remplacé par "finie". En théorie des probabilités, le couple  $(\Omega, \mathcal{A})$  est appelé espace probabilisable (contrairement à "espace mesurable").

**Exemple 1** 1) La tribu dite grossière :  $\mathcal{A} = \{\emptyset, \Omega\}$ .

2) La tribu discrète :  $\mathcal{A} = \mathcal{P}(\Omega)$ .

3) Si  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$  alors  $\mathcal{A} = \{\emptyset, \{\omega_1\}, \{\omega_2, \omega_3, \omega_4\}, \Omega\}$  est une tribu sur  $\Omega$ .

4)  $\mathcal{A} = \{A \in \mathcal{P}(\Omega) \text{ t.q. } A \text{ ou } A^c \text{ est dénombrable}\}$  est une tribu sur  $\Omega$

dans ce dernier exemple, si on change "dénombrable" en "fini" cela devient une algèbre mais pas une  $\sigma$ -algèbre.

**Définition 2 (tribu engendrée)** Si  $C \subset \mathcal{P}(\Omega)$ , la tribu engendrée par  $C$  et notée par  $\sigma(C)$  est la plus petite tribu contenant  $C$ . C'est-à-dire

$$\sigma(C) = \bigcap_{\substack{\mathcal{A} \text{ tribu t.q.} \\ C \subset \mathcal{A}}} \mathcal{A}.$$

Remarque : l'intersection d'une famille quelconque de tribus est encore une tribu et l'intersection ci-dessus n'est pas vide, car il y a au moins  $\mathcal{A} = \mathcal{P}(\Omega)$ .

**Exemple 2** Si  $A \subset \Omega$ , alors  $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$ .

Si  $\Omega = \mathbb{R}$  (ou plus généralement si  $\Omega$  est un espace topologique), la tribu borélienne, ou de Borel est la tribu  $\mathcal{B}(\mathbb{R})$  engendrée par la classe des ouverts (ou de manière équivalente, par la classe des fermés); une partie de  $\mathbb{R}$  qui est dans la tribu borélienne s'appelle un borélien.

**Théorème 1**  $\mathcal{B}(\mathbb{R}) = \sigma(\{]-\infty, a], a \in \mathbb{Q}\})$ .

**Définition 3 (Mesure de Probabilité)** Une mesure de probabilité, ou simplement une probabilité, définie sur une tribu  $\mathcal{A}$  d'un espace  $\Omega$ , est une application  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  qui vérifie :

1.  $\mathbb{P}(\Omega) = 1$ .
2. Pour toute suite  $A_1, A_2, A_3, \dots$  d'éléments de  $\mathcal{A}$  (des événements) qui sont deux à deux disjoints (i.e.  $A_n \cap A_m = \emptyset$  si  $n \neq m$ ), on a

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Le nombre  $\mathbb{P}(A)$  s'appelle la probabilité de l'événement  $A$ .

Le point 2. s'appelle  $\sigma$ -additivité de  $\mathbb{P}$ .

**Théorème 2** Si  $\mathbb{P}$  est une probabilité sur la tribu  $\mathcal{A}$ , alors :

- (i)  $\mathbb{P}(\emptyset) = 0$ ;
- (ii) si  $A, B \in \mathcal{A}$  tels que  $A \cap B = \emptyset$ , alors  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$  (condition d'additivité).

Voici deux conséquences immédiates de l'additivité et de l'axiome  $\mathbb{P}(\Omega) = 1$  :

- Propriété 1**
1.  $A \in \mathcal{A} \Rightarrow \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
  2.  $A, B \in \mathcal{A}, A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

*Preuve:*

1. Utiliser l'additivité avec  $A, A^c$  et  $\mathbb{P}(\Omega) = 1$ .
2. Utiliser l'additivité avec  $A$  et  $C := A^c \cap B$ .

□

**Définition 4** Soient  $A$  un ensemble et  $(A_n)_{n \in \mathbb{N}^*}$  une suite d'ensembles.

On note  $(A_n)_{n \in \mathbb{N}^*} \searrow A$  et on dit que  $(A_n)_{n \in \mathbb{N}^*}$  **décroit** (ou "tend en décroissant") vers  $A$  si  $A_{n+1} \subset A_n$  pour tout  $n \in \mathbb{N}^*$  et si  $\bigcap_{n \in \mathbb{N}^*} A_n = A$ .

On note  $(A_n)_{n \in \mathbb{N}^*} \nearrow A$  et on dit que  $(A_n)_{n \in \mathbb{N}^*}$  **croît** (ou "tend en croissant") vers  $A$  si  $A_n \subset A_{n+1}$  pour tout  $n \in \mathbb{N}^*$  et si  $\bigcup_{n \in \mathbb{N}^*} A_n = A$ .

Remarque : si les  $A_n, n \in \mathbb{N}^*$  sont dans la tribu  $\mathcal{A}$ , il en est de même de  $A$  dans les deux cas ci-dessus (propriété de  $\sigma$ -algèbre de  $\mathcal{A}$ ).

Le théorème ci-dessous exprime ce qu'il faut exactement ajouter à l'additivité pour obtenir la  $\sigma$ -additivité :

**Théorème 3** Soit  $\mathcal{A}$  une tribu sur  $\Omega$ . Supposons que  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  satisfasse l'axiome 1 (i.e.  $\mathbb{P}(\Omega) = 1$ ) et soit additive.

Les assertions suivantes sont alors équivalentes :

- (i) on a l'axiome 2 (i.e.  $\sigma$ -additivité);  
(ii) si  $(A_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$  et décroît vers  $\emptyset$  alors  $\mathbb{P}(A_n) \searrow 0$ .  
(iii) si  $(A_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$  et décroît vers  $A$  alors  $\mathbb{P}(A_n) \searrow \mathbb{P}(A)$ .  
(iv) si  $(A_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$  et croît vers  $\Omega$  alors  $\mathbb{P}(A_n) \nearrow 1$ .  
(v) si  $(A_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$  et croît vers  $A$  alors  $\mathbb{P}(A_n) \nearrow \mathbb{P}(A)$ .

*Preuve:* Observons que  $A_n \searrow A$  implique  $A_n^c \nearrow A^c$ . Comme  $\mathbb{P}(A_n^c) = 1 - \mathbb{P}(A_n)$  et  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ , on voit que (ii) et (iv) sont équivalents, ainsi que (iii) et (v). L'implication (v)  $\Rightarrow$  (iv) est évidente.

Supposons maintenant (iv). Soit  $A_n \in \mathcal{A}$  avec  $A_n \nearrow A$ . Posons  $B_n = A_n \cup A^c$ . Alors  $(B_n)$  croît vers  $\Omega$ , de sorte que  $\mathbb{P}(B_n) \nearrow 1$ . Comme  $A_n \subset A$ , on a  $A_n \cap A^c = \emptyset$ , donc  $\mathbb{P}(A_n \cup A^c) = \mathbb{P}(A_n) + \mathbb{P}(A^c)$  et il vient  $1 = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n) = \lim_{n \rightarrow +\infty} (\mathbb{P}(A_n) + \mathbb{P}(A^c))$ . Par suite,  $\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = 1 - \mathbb{P}(A^c) = \mathbb{P}(A)$ , et on a (v).

Il reste à montrer que (i) et (v) sont équivalents. Supposons d'abord (v). Soit des  $A_n \in \mathcal{A}$  deux à deux disjoints, et définissons  $B_n = \bigcup_{p=1}^n A_p$  et  $B = \bigcup_{n=1}^{\infty} A_n$ . L'additivité implique  $\mathbb{P}(B_n) = \sum_{p=1}^n \mathbb{P}(A_p)$  qui croît vers  $\sum_{p=1}^{\infty} \mathbb{P}(A_p)$  quand  $n \rightarrow \infty$  et croît aussi vers  $\mathbb{P}(B)$  par (v), donc  $\mathbb{P}(B) = \mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ , et (i) est prouvé.

Finalement, supposons (i). Soit  $A_n \in \mathcal{A}$ , avec  $A_n$  croissant vers  $A$ . On construit une nouvelle suite d'éléments de  $\mathcal{A}$  avec  $B_1 = A_1$ ,  $B_n = A_n \setminus A_{n-1} = A_n \cap A_{n-1}^c$ . On a  $\bigcup_{n=1}^{\infty} B_n = A$  et les  $(B_n)_{n \geq 1}$  sont deux à deux disjoints. Donc (i) entraîne  $\mathbb{P}(A) = \lim_{n \rightarrow \infty} \sum_{p=1}^n \mathbb{P}(B_p)$ . On a aussi  $\sum_{p=1}^n \mathbb{P}(B_p) = \mathbb{P}(A_n)$ , donc  $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$  et on a (v).  $\square$

Si  $A$  est une partie de  $\Omega$ , on définit sa fonction indicatrice par  $\mathbf{1}_A(\omega) = 1$  si  $\omega \in A$  et  $\mathbf{1}_A(\omega) = 0$  si  $\omega \notin A$ . Notez qu'une fonction indicatrice  $\mathbf{1}_A$  est  $\mathcal{F}$ -mesurable si et seulement si  $A \in \mathcal{F}$ . On associe aussi à toute suite  $(A_n)_{n \in \mathbb{N}^*}$  de parties de  $\Omega$  les ensembles suivants :

$$\limsup_{n \rightarrow \infty} \stackrel{\text{déf}}{=} \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m =: A^* ; \quad \liminf_{n \rightarrow \infty} \stackrel{\text{déf}}{=} \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m =: A_*$$

**Propriété 2**  $A_* \subset A^*$  et si  $(A_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$  alors  $A_*, A^* \in \mathcal{A}$ .

Les règles élémentaires sur  $\cup, \cap$  et la complémentarité, donnent sans difficulté

$$(A^*)^c = (A^c)_*$$

De plus,  $A^*$  et  $A_*$  ont les interprétations suivantes :

**Propriété 3** Soit  $(A_i)_{i \in \mathbb{N}}$  une collection infinie d'ensembles. Alors

- i)  $\omega \in A_* \Leftrightarrow$  "à partir d'un certain rang,  $\omega$  est dans tous les  $A_i$ ".  
ii)  $\omega \in A^* \Leftrightarrow$  " $\omega$  est dans une infinité de  $A_m$ ".

On écrit parfois  $\{A_m \text{ i.s.}\}$  où i.s. signifie "infiniment souvent".

*Preuve:*

- i) Soit  $\omega$  qui, à partir d'un certain rang, est dans tous les  $A_m$ . On traduit cela de la façon suivante (qui est une équivalence) : il existe un rang  $n \geq 1$  tel que pour tout  $m \geq n$ ,  $\omega$  est dans  $A_m$ . D'après la signification des symboles  $\forall, \exists, \cap$  et  $\cup$ , cela revient à écrire

$$\omega \in \underbrace{\bigcup_{n \geq 1}}_{\substack{\text{il existe} \\ n \geq 1}} \underbrace{\bigcap_{m \geq n}}_{\substack{\text{pour tout} \\ m \geq n}} \underbrace{A_m}_{\substack{\omega \text{ est} \\ \text{dans } A_m}} .$$

- ii) Dire que  $\omega$  est dans une infinité de  $A_m$  est équivalent à dire que

" pour tout  $n$ , il existe  $m \geq n$  avec  $\omega$  dans  $A_m$ ".

En effet, si tel est le cas,  $\omega$  est bien dans une infinité de  $A_m$  car, d'après cette propriété,

- avec  $n = 1$ , il existe  $m_1 \geq n$  tel que  $\omega$  est dans  $A_{m_1}$ ,  $m_1 \geq 1$ .
- avec  $n = m_1 + 1$ , il existe  $m_2 \geq n$  tel que  $\omega$  est dans  $A_{m_2}$ ,  $m_2$  strictement plus grand que  $m_1$ ,
- avec  $n = m_2 + 1$ , il existe  $m_3 \geq n$  tel que  $\omega$  est dans  $A_{m_3}$ ,  $m_3$  strictement plus grand que  $m_2$ ,
- ... ,
- avec  $n = m_k + 1$ , il existe  $m_{k+1} \geq n$  tel que  $\omega$  est dans  $A_{m_{k+1}}$ ,  $m_{k+1}$  strictement plus grand que  $m_k$ ,
- ... ,

et finalement,  $\omega$  est dans chaque  $A_{m_k}$  avec  $(m_k)_{k \in \mathbb{N}^*}$  strictement croissante, c'est-à-dire dans une infinité de  $A_m$ .

Réciproquement, s'il est dans une infinité de  $A_i$ , alors il existe une suite d'entiers positifs  $(i_k)_{k \geq 1}$  tel que  $\omega \in A_{i_k}$  pour tout  $k \in \mathbb{N}^*$ . Donc sachant que pour tout  $k \in \mathbb{N}^*$ ,  $i_k \geq k$ , il vient que pour tout  $n \in \mathbb{N}^*$  il existe  $m \geq n$  tel que  $\omega \in A_m$ , et ce  $m$  peut être choisi tel que  $m = i_p$  avec  $p \geq n$ .

Donc, pour ce deuxième point, pour tout  $n$ , on trouve  $m \geq n$ , tel que  $\omega \in A_m$ , en langage  $\forall, \exists$ , cela s'écrit

$$\omega \in \underbrace{\bigcap_{n \geq 1}}_{\substack{\text{pour tout} \\ n \geq 1}} \underbrace{\bigcup_{m \geq n}}_{\substack{\text{il existe} \\ m \geq n}} \underbrace{A_m}_{\substack{\omega \text{ est} \\ \text{dans } A_m}} .$$

□

**Lien entre lim inf, lim sup d'ensembles et de fonctions**

Rappel : Pour une suite réelle  $u = (u_n)_{n \in \mathbb{N}}$ , on définit ses limites supérieure et inférieure par

$$u_* := \liminf_{n \rightarrow +\infty} u_n = \sup_{n \geq 0} \inf_{k \geq n} u_k$$

$$u^* := \limsup_{n \rightarrow +\infty} u_n = \inf_{n \geq 0} \sup_{k \geq n} u_k$$

Ce sont les plus petite et plus grande valeurs d'adhérence de la suite  $u$ . On a toujours

$$u_* \leq u^*$$

et il y a égalité ssi la suite  $u$  converge ; de plus si tel est le cas  $\lim_{n \rightarrow +\infty} u_n = u_* = u^*$ . En plus, en changeant le signe, les limites inférieure et supérieure s'échangent :

$$(-u)^* = -u_*$$

$$(-u)_* = -u^*.$$

Pour une suite de fonctions  $(f_n)_{n \geq 0}$ , on définit des fonctions limites inférieure et supérieure de la façon suivante :

$$f_*(x) = \liminf_{n \rightarrow +\infty} f_n(x), \quad f^*(x) = \limsup_{n \rightarrow +\infty} f_n(x).$$

**Propriété 4** Soit  $(A_n)_{n \in \mathbb{N}}$  une suite d'ensemble mesurables, on a

$$(\mathbf{1}_A)_* = \mathbf{1}_{A_*} \text{ et } (\mathbf{1}_A)^* = \mathbf{1}_{A^*}.$$

*Preuve:* la démonstration complète est à faire soi-même. Pour vous aider : montrer que  $(\mathbf{1}_A)^* = \mathbf{1}_{A^*}$  revient, par définition et notation, à montrer que  $(\mathbf{1}_A(\omega))^* = \mathbf{1}_{A^*}(\omega)$ , qui est équivalent à  $\inf_{n \geq 1} \sup_{m \geq n} \mathbf{1}_{A_m}(\omega) = \mathbf{1}_{(\bigcap_{n \geq 1} \bigcup_{m \geq n} A_m)}(\omega)$ .

□

**Définition 5 (Convergence d'ensembles)** On dit qu'une suite d'ensemble  $(A_n)_{n \in \mathbb{N}}$  converge vers  $A$  et on note  $A_n \rightarrow A, n \rightarrow +\infty$ , si

$$A^* = A_* = A.$$

On montre alors aussi :

**Propriété 5** Soit  $\mathbb{P}$  une probabilité et  $(A_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$  qui converge vers  $A$ . Alors  $A \in \mathcal{A}$  et  $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$ .

*Preuve:* D'après la Propriété 2, on a  $A_* = A^* \in \mathcal{A}$ , et donc d'après la définition précédente  $A \in \mathcal{A}$ .

Ensuite, soit  $B_n = \bigcap_{m \geq n} A_m$  et  $C_n = \bigcup_{m \geq n} A_m$ . Alors  $(B_n)_n$  croît vers  $A$  et  $(C_n)_n$  décroît vers  $A$ , donc  $\lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(C_n)$  d'après le Théorème 3. Mais  $B_n \subset A_n \subset C_n$ , donc  $\mathbb{P}(B_n) \leq \mathbb{P}(A_n) \leq \mathbb{P}(C_n)$ , de sorte que  $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$ .

□



## 2. Probabilités conditionnelles et indépendance et Probabilités sur un espace fini ou dénombrable

Soient  $A, B \in (\Omega, \mathcal{A}, \mathbb{P})$ . La fréquence  $f_n(A)$  de réalisation de  $A$  lorsqu'on répète  $n$  fois la même expérience, converge (en un sens à préciser!) vers la probabilité  $\mathbb{P}(A)$ , qui quantifie "les chances de voir  $A$  réalisé".

Supposons maintenant qu'on sache que  $B$  est réalisé. Les chances de voir  $A$  réalisé vont changer et être quantifiées par un nouveau nombre  $\mathbb{P}(A|B)$ , "la probabilité de  $A$  sachant  $B$ " : on peut à nouveau considérer la fréquence de  $A$  lorsque l'expérience est répétée  $n$  fois, sauf qu'il faut calculer cette fréquence sur l'ensemble des expériences où  $B$  se réalise.

Il est ainsi naturel de considérer le nombre de fois où  $A$  et  $B$  sont réalisés, c'est-à-dire  $nf_n(A \cap B)$ ; pour obtenir une fréquence il convient de diviser ce nombre par le nombre d'occurrences de  $B$ , i.e.  $nf_n(B)$ , de sorte qu'en définitive on doit avoir

$$\mathbb{P}(A|B) \approx \frac{nf_n(A \cap B)}{nf_n(B)} = \frac{f_n(A \cap B)}{f_n(B)}.$$

La définition ci-dessous est tirée de cette relation en "prenant la limite en  $n$ ".

En particulier, il se peut que le fait de savoir que  $B$  est réalisé ne donne aucune information sur la réalisation de  $A$  : on dit alors que  $A$  est "indépendant" de  $B$ , et cela se traduit par le fait que  $\mathbb{P}(A|B) = \mathbb{P}(A)$ . Et donc  $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A)$ , donc  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . Nous allons maintenant introduire les définitions proprement dites en commençant par l'indépendance :

**Définition 6 (indépendance)** a)  $A$  et  $B$  sont dit indépendants si  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ .

b) Soit  $(A_i)_{i \in I}$  une famille d'événements, où  $I \subset \mathbb{N}$ . Les  $(A_i)_{i \in I}$  sont dits

indépendants, ou parfois "mutuellement indépendants" si

$$\forall J \subset I \text{ fini, } \mathbb{P} \left( \bigcap_{j \in J} A_j \right) = \prod_{j \in J} \mathbb{P}(A_j).$$

Remarque : Si les  $(A_i)_{i \in I}$  sont indépendants, ils sont aussi deux à deux indépendants, ce qui signifie que  $A_i$  et  $A_j$  sont indépendants pour tous  $i, j$  avec  $i \neq j$ , mais la réciproque est fautive.

**Théorème 4** Si  $A$  et  $B$  sont indépendants, il en est de même de  $A$  et  $B^c$ , de  $A^c$  et  $B$ , et de  $A^c$  et  $B^c$ .

*Preuve:* Pour  $A$  et  $B^c$ , on a  $\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(B^c)$ . Les autres propriétés se montrent de manière analogue.  $\square$

**Exemple 3** 1. On choisit une carte au hasard parmi 52 cartes.  $A = \{\text{la carte est un cœur}\}$ , et  $B = \{\text{la carte est une dame}\}$ . Un modèle naturel pour cette expérience consiste à affecter la probabilité  $\frac{1}{52}$  au choix de chacune des cartes. Par additivité,  $\mathbb{P}(A) = \frac{13}{52}$  et  $\mathbb{P}(B) = \frac{4}{52}$  et  $\mathbb{P}(A \cap B) = \frac{1}{52}$ , donc  $A$  et  $B$  sont indépendants.

2. Soit  $\Omega = \{1, 2, 3, 4\}$  et  $\mathcal{A} = \mathcal{P}(\Omega)$ . Soit  $\mathbb{P}(i) = \frac{1}{4}$  pour  $i \in \{1, 2, 3, 4\}$ . Soit enfin  $A = \{1, 2\}$ ,  $B = \{1, 3\}$  et  $C = \{2, 3\}$ . Alors  $A, B, C$  sont deux à deux indépendants, mais pas (mutuellement) indépendants.

**Définition 7 (Probabilité conditionnelle)** Soit  $A, B$  deux événements, avec  $\mathbb{P}(B) > 0$ . La probabilité conditionnelle de  $A$  sachant  $B$  est  $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ .

**Théorème 5 (Probabilité conditionnelle sachant un événement)**

Supposons que  $\mathbb{P}(B) > 0$ .

(a)  $A$  et  $B$  sont indépendants  $\Leftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A)$ .

(b) L'application  $\mathbb{P}_B : \mathcal{A} \ni A \mapsto \mathbb{P}(A|B) \in [0, 1]$  définit une nouvelle mesure de probabilité sur  $\mathcal{A}$ , appelée la "probabilité conditionnelle sachant  $B$ ".

*Preuve:* (a) résulte d'un calcul immédiat. Pour (b), on doit montrer que  $\mathbb{P}_B$  vérifie les deux axiomes (1) et (2) d'une probabilité de mesure. (1) provient de  $\mathbb{P}_B(\Omega) = \mathbb{P}(\Omega|B) = \mathbb{P}(\Omega \cap B)/\mathbb{P}(B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$ .

Quant à (2),  $\mathbb{P}_B(\cup_{n \in \mathbb{N}^*} A_n) = \mathbb{P}(\cup_{n \in \mathbb{N}^*} (A_n \cap B))/\mathbb{P}(B) = \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n \cap B)/\mathbb{P}(B)$ , car les  $(A_n \cap B)_{n \in \mathbb{N}^*}$  sont deux à deux disjoints, d'où  $\mathbb{P}_B(\cup_{n \in \mathbb{N}^*} A_n) = \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n|B) = \sum_{n \in \mathbb{N}^*} \mathbb{P}_B(A_n)$ .  $\square$

**Théorème 6 (Théorème des probabilités composées)** Soit  $A_0 = \Omega$ . Soit  $n \in \mathbb{N}^* \setminus \{1\}$ . On pose  $I_n = \{1, 2, \dots, n\}$ . Si  $(A_i)_{i \in I} \in \mathcal{A}^I$  et si  $\mathbb{P}(\cap_{i \in I_{n-1}}) > 0$ , alors

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}).$$

*Preuve:* En utilisant la définition de la probabilité conditionnelle la preuve se fait par récurrence.  $\square$

**Théorème 7 (Formule des probabilités totales)** Soit  $(E_n)_{n \in I} \in \mathcal{A}^I$  une partition dénombrable de  $\Omega$ . Pour tout  $A \in \mathcal{A}$ , on a

$$\mathbb{P}(A) = \sum_{n \in I} \mathbb{P}(A|E_n)\mathbb{P}(E_n).$$

*Preuve:* On a  $A = A \cap \Omega = \bigcup_{n \in I} (A \cap E_n)$  et comme les  $(A \cap E_n)_{n \in I}$  sont deux à deux disjoints, on obtient par  $\sigma$ -additivité de  $\mathbb{P}$ ,  $\mathbb{P}(A) = \sum_{n \in I} \mathbb{P}(A \cap E_n) = \sum_{n \in I} \mathbb{P}(A|E_n)\mathbb{P}(E_n)$ .  $\square$

**Théorème 8 (Théorème de Bayes, ou de "probabilité des causes".)**

Soit  $(E_n)_{n \in I} \in \mathcal{A}^I$  une partition dénombrable de  $\Omega$ . Pour tout  $A \in \mathcal{A}$  tel que  $\mathbb{P}(A) > 0$  on a

$$\mathbb{P}(E_m|A) = \frac{\mathbb{P}(A|E_m)\mathbb{P}(E_m)}{\sum_{n \in I} \mathbb{P}(A|E_n)\mathbb{P}(E_n)}.$$

*Preuve:* Au vu du précédent théorème le dénominateur vaut  $\mathbb{P}(A)$ , la formule devient donc  $\frac{\mathbb{P}(A|E_m)\mathbb{P}(E_m)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap E_m)}{\mathbb{P}(A)} = \mathbb{P}(E_m|A)$ .  $\square$

Ce théorème a des conséquences profondes en probabilités et en statistique, voir par exemple les exercices en TD.

#### Probabilités sur un espace fini ou dénombrable :

Dans cette section nous supposons que l'espace d'états  $\Omega$  est fini ou dénombrable, et nous choisissons pour la tribu la classe  $\mathcal{A} = \mathcal{P}(\Omega)$ .

**Théorème 9 a)** Une probabilité sur l'ensemble fini ou dénombrable  $\Omega$  est caractérisée par ses valeurs sur les singletons :  $p_\omega := \mathbb{P}(\{\omega\})$ ,  $\omega \in \Omega$ .

b) Si  $(p_\omega)_{\omega \in \Omega}$  est une famille de réels indexée par l'ensemble fini ou dénombrable  $\Omega$ , il existe une probabilité  $\mathbb{P}$  (nécessairement unique par a)) sur  $\Omega$  telle que  $\mathbb{P}(\{\omega\}) = p_\omega$  pour tout  $\omega \in \Omega$  si et seulement si  $p_\omega \geq 0$  et  $\sum_{\omega \in \Omega} p_\omega = 1$ .

*Preuve:* Soit  $A \in \mathcal{A}$ ; on a  $A = \bigcup_{\omega \in A} \{\omega\}$ , réunion finie ou dénombrable de singletons deux à deux disjoints. Si  $\mathbb{P}$  est une probabilité sur  $\Omega$ , la  $\sigma$ -additivité implique  $\mathbb{P}(A) = \dots = \sum_{\omega \in A} p_\omega$ . On a donc a). Pour la condition nécessaire de b), i.e. l'implication  $\Rightarrow$ , on remarque sur si  $\mathbb{P}(\{\omega\}) = p_\omega$ , alors par définition  $p_\omega \geq 0$ , et aussi

$1 = \mathbb{P}(\Omega) = \mathbb{P}(\bigcup_{\omega \in \Omega} \{\omega\}) = \sum_{\omega \in \Omega} \mathbb{P}(\omega) = \sum_{\omega \in \Omega} p_\omega$ . Supposons inversement que les  $p_\omega$  vérifient  $p_\omega \geq 0$  et  $\sum_{\omega \in \Omega} p_\omega = 1$ . On a posé  $\mathbb{P}(A) := \sum_{\omega \in A} p_\omega$ , avec la convention qu'une somme "vide" est nulle. On a donc  $\mathbb{P}(\Omega) = 1$ . Pour la  $\sigma$ -additivité, elle est évident lorsque  $\Omega$  est fini, et lorsque  $\Omega$  est dénombrable elle provient de ce qu'on peut "sommer par paquets" pour obtenir que  $\sum_{i \in I} \sum_{\omega \in A_i} p_\omega = \sum_{\omega \in \bigcup_{i \in I} A_i} p_\omega$  si les  $A_i$  sont deux à deux disjoints.  $\square$

**Exemple 4** • Prenons  $\Omega = \mathbb{N}$  et  $\mathcal{A} = \mathcal{P}(\mathbb{N})$ , et  $\mathbb{P}(\{\omega\}) = e^{-\lambda} \frac{\lambda^\omega}{\omega!}$  pour tout  $\omega \in \Omega$ .  $\mathbb{P}$  est une mesure de probabilité et est appelée la loi de Poisson de paramètre  $\lambda > 0$ .

• Prenons  $\Omega = \mathbb{N}$  et  $\mathcal{A} = \mathcal{P}(\mathbb{N})$ , et  $\mathbb{P}(\{\omega\}) = (1 - \alpha)\alpha^\omega$ , pour tout  $\omega \in \Omega$ .  $\mathbb{P}$  est une mesure de probabilité et est appelée La loi géométrique de paramètre  $\alpha \in [0, 1[$ .

**Définition 8 (Probabilité uniforme)** On dit que la probabilité  $\mathbb{P}$  sur l'espace fini  $\Omega$  est uniforme si  $p_\omega = \mathbb{P}(\{\omega\})$  ne dépend pas de  $\omega$ .

Dans ce cas il est immédiat que  $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$  où  $\text{card}(A)$  désigne le "cardinal", ou nombre de points, de  $A$ .

### 3. Variables aléatoires sur un espace fini ou dénombrable et Construction d'une mesure de probabilité

Dans ce chapitre nous supposons encore l'espace d'états  $\Omega$  dénombrable, et nous choisissons pour tribu la classe  $\mathcal{A} = \mathcal{P}(\Omega)$  de toutes les parties de  $\Omega$ .

**Définition 9** Une variable aléatoire  $X$  est définie comme une application de  $\Omega$  dans un ensemble  $E$  (a priori arbitraire).

Une variable aléatoire représente une quantité qui dépend de l'issue de l'expérience.

Noter que l'espace d'arrivée  $E$  n'est pas nécessairement dénombrable. Par exemple  $E = \mathbb{R}$  si l'expérience consiste à choisir une personne dans une salle et si  $X(\omega)$  représente la taille de la personne  $\omega$ ; mais l'image  $E'$  de  $\Omega$  par  $X$ , i.e.  $E' := \{i \in E, \exists \omega \in \Omega \mid X(\omega) = i\}$ , est, quant à elle, nécessairement dénombrable.

On peut alors définir la loi de  $X$ , notée par  $\mathbb{P}_X$ , appelée aussi la distribution de  $X$ , par

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)), \forall A \in \mathcal{P}(E').$$

Le fait que  $\mathbb{P}_X$  définisse une mesure de probabilité sur  $E'$  est évident (muni de la tribu de toutes ses parties  $\mathcal{P}(E')$ ). Comme  $E'$  est dénombrable,  $\mathbb{P}_X$  est complètement déterminé par les nombres

$$p_j^X := \mathbb{P}(X = j) = \sum_{\{\omega \in \Omega : X(\omega) = j\}} p_\omega,$$

et la famille  $(p_j^X)_{j \in E'}$  est aussi appelée la loi de la variable  $X$ . On a bien sûr  $\mathbb{P}_X(A) = \sum_{j \in A} p_j^X$  pour tout  $A \in \mathcal{P}(E')$ .

Si  $\mathbb{P}_X$  est une loi munie d'un nom, par exemple une loi de Poisson, on dit aussi que  $X$  est une variable aléatoire de Poisson.

**Définition 10** Soit  $X$  une variable aléatoire à valeurs dans  $E = \mathbb{R}$ , définie sur un espace dénombrable. On appelle espérance, ou espérance mathéma-

tique, de  $X$  le nombre

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) p_{\omega},$$

lorsqu'il est bien défini; c'est le cas si  $\Omega$  est fini; c'est aussi le cas, lorsque  $\Omega$  est infini dénombrable, si la série ci-dessus est absolument convergente, ou si  $X(\omega) \geq 0$  pour tout  $\omega \in \Omega$  (dans ce dernier cas on peut avoir  $\mathbb{E}[X] = +\infty$ ).

Notons  $\mathcal{L}^1$  l'espace de toutes les variables aléatoires réelles qui ont une espérance finie. Il est alors immédiat que :

- (1)  $\mathcal{L}^1$  est un espace vectoriel et la fonctionnelle espérance  $\mathbb{E}$  est linéaire.
- (2) La fonctionnelle espérance  $\mathbb{E}$  est positive : si  $X \in \mathcal{L}^1$  et  $X \geq 0$ , alors  $\mathbb{E}[X] \geq 0$ . Plus généralement si  $X, Y \in \mathcal{L}^1$  et  $X \leq Y$  alors  $\mathbb{E}[X] \leq \mathbb{E}[Y]$ .
- (3)  $\mathcal{L}^1$  contient toutes les variables aléatoires bornées. Si  $X \equiv a$ ,  $a \in \mathbb{R}$ , alors  $\mathbb{E}[X] = a$ .
- (4) Si  $X \in \mathcal{L}^1$ , son espérance dépend seulement de sa loi, et si  $E'$  est l'image de  $\Omega$  par  $X$ , on a

$$\mathbb{E}[X] = \sum_{j \in E'} j \mathbb{P}(X = j).$$

$$\begin{aligned} \sum_{\omega \in \Omega} X(\omega) p_{\omega} &= \sum_{\substack{\omega \in \cup_{j \in E'} X^{-1}(\{j\}) \\ j \in E'}} X(\omega) p_{\omega} = \sum_{j \in E'} \sum_{\omega \in X^{-1}(\{j\})} X(\omega) p_{\omega} \\ &= \sum_{j \in E'} \sum_{\omega \in X^{-1}(\{j\})} j p_{\omega} = \sum_{j \in E'} j \sum_{\omega \in X^{-1}(\{j\})} p_{\omega} = \sum_{j \in E'} j \mathbb{P}(X^{-1}(\{j\})) = \sum_{j \in E'} j \mathbb{P}(X = j) \end{aligned}$$

- (5) Si  $X = \mathbf{1}_A$  est l'indicatrice d'un événement  $A$ , on a  $\mathbb{E}[X] = \mathbb{P}(A)$ .

En effet  $\mathbb{E}[X]$  est bien défini car  $X \geq 0$ , et on a  $\mathbb{E}[X] = \sum_{\omega \in \Omega} \mathbf{1}_A(\omega) p_{\omega} = \sum_{\omega \in A} \mathbf{1} \times p_{\omega} = \mathbb{P}(A)$ .

**Théorème 10** Soit  $h : \mathbb{R} \rightarrow [0, \infty]$  une fonction positive et  $X$  une variable aléatoire réelle. On alors

$$\mathbb{P}(\{\omega \in \Omega : h(X(\omega)) \geq a\}) \leq \frac{\mathbb{E}[h(X)]}{a}$$

pour tout  $a > 0$ .

*Preuve:* Comme  $X$  est une variable aléatoire, il en est de même de  $Y = h(X)$ ; soit  $A = Y^{-1}([a, \infty[) = \{\omega : h(X(\omega)) \geq a\} = \{h(X) \geq a\}$ , alors on a  $h(X) \geq a \mathbf{1}_A$ , donc  $\mathbb{E}[h(X)] \geq \mathbb{E}[a \mathbf{1}_A] = a \mathbb{E}[\mathbf{1}_A] = a \mathbb{P}(A)$  et le résultat suit.  $\square$

**Corollaire 1 (Inégalité de Markov)**

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a}.$$

*Preuve:* Prendre  $\mathbb{R} \ni x \mapsto h(x) := |x|$  dans le Théorème 10.  $\square$

Remarque : Pour tout réel  $x$ ,  $|x| \leq 1 + x^2$ , donc  $X^2 \in \mathcal{L}^1 \Rightarrow X \in \mathcal{L}^1$  d'où

**Définition 11** Soit  $X$  une variable aléatoire réelle telle que  $X^2 \in \mathcal{L}^1$ . On appelle variance de  $X$  le nombre

$$\sigma_X^2 = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

et l'écart-type le nombre  $\sigma_X$  la racine carrée "positive" de la variance.

L'espérance de  $X$ , est également appelé la moyenne de  $X$ , parfois notée par  $\mu_X$ .

**Corollaire 2 (Inégalités de Bienaymé-Chebyshev)** Si  $X^2 \in \mathcal{L}^1$ , on a

$$a) \mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[X^2]}{a^2} \text{ pour tout } a > 0,$$

$$b) \mathbb{P}(|X - \mu_X| \geq a) \leq \frac{\sigma_X^2}{a^2} \text{ pour tout } a > 0.$$

*Preuve:* Pour a), on applique le Théorème 10 avec  $h : x \mapsto x^2$ , ce qui donne  $\mathbb{P}(|X| \geq a) = \mathbb{P}(h(X) \geq a^2) \leq \frac{\mathbb{E}[X^2]}{a^2}$ . Pour b), on applique a) à la variable aléatoire  $Y = |X - \mu_X|$ .  $\square$

**Exemple 5** 1)  $X$  suit une loi de Poisson de paramètre, que l'on note par  $X \sim \text{Poisson}(\lambda)$ , si pour tout  $A \subset \mathbb{N}$ , on a  $\mathbb{P}(X \in A) = \sum_{j \in A} \mathbb{P}(X = j) = \sum_{j \in A} \frac{\lambda^j}{j!} e^{-\lambda}$ .

L'espérance de  $X$  est donnée par  $\mathbb{E}[X] = \sum_{j=0}^{\infty} j \mathbb{P}(X = j) = \sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} = \lambda$ .

2)  $X$  est de Bernoulli, i.e. elle prend seulement les deux valeurs 1 et 0 avec les probabilités respectives  $p$  et  $q = 1 - p$  (avec  $p \in ]0, 1[$ ). Alors  $\mathbb{E}[X] = 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) = p$ .

3)  $X$  est binomiale, que l'on écrit  $X \sim \mathcal{B}(p, n)$ , i.e.  $X$  peut prendre les valeurs  $0, 1, \dots, n$ , et  $\mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}$ .

**Construction d'une mesure de probabilité**

Dans ce chapitre  $\Omega$  est un espace totalement quelconque, muni d'une tribu  $\mathcal{A}$ .

Nous nous proposons de construire, ou de caractériser, une mesure de probabilité sur  $\mathcal{A}$  (on dit aussi sur  $(\Omega, \mathcal{A})$ ) à partir de la connaissance des probabilités  $\mathbb{P}(A)$  pour les événements  $A$  appartenant à une certaine sous-classe de  $\mathcal{A}$ , qu'on aimerait être "aussi petite que possible".

Quand  $\Omega$  est dénombrable nous avons vu que c'était une chose simple à réaliser, à partir par exemple de la probabilité  $p_\omega = \mathbb{P}(\{\omega\})$  des singletons.

Mais dans le cas général la même méthode ne marche plus : une probabilité "typique"  $\mathbb{P}$  vérifiera  $\mathbb{P}(\{\omega\}) = 0$  pour tout  $\omega$  et donc les nombres  $\mathbb{P}(\{\omega\})$  ne caractérisent pas  $\mathbb{P}$ .

Pour construire  $\mathbb{P}$  il convient d'abord de préciser les ingrédients de base : on verra dans des cas "concrets" et notamment au chapitre suivant, qu'il est souvent aisé de construire une "probabilité" sur une algèbre qui engendre la tribu  $\mathcal{A}$ , et le problème qu'on se pose ci-dessous est l'extension de cette probabilité à la tribu  $\mathcal{A}$  elle-même.

Soit donc une algèbre  $\mathcal{A}_0$  qui engendre la tribu  $\mathcal{A}$ , et une probabilité  $\mathbb{P} : \mathcal{A}_0 \rightarrow [0, 1]$  sur  $\mathcal{A}_0$ , et donc :

1)  $\mathbb{P}(\Omega) = 1$  et 2) pour toute suite  $(A_n)_{n \in \mathbb{N}^*}$  d'éléments de  $\mathcal{A}_0$  deux à deux disjoints et telle que  $\bigcup_{n \in \mathbb{N}^*} A_n \in \mathcal{A}_0$ , on a  $\mathbb{P}(\bigcup_{n \in \mathbb{N}^*} A_n) = \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n)$ .

**Théorème 11** *Toute probabilité  $\mathbb{P}$  définie sur une algèbre  $\mathcal{A}_0$  admet une unique extension (notée aussi  $\mathbb{P}$ ) à la tribu  $\mathcal{A} = \sigma(\mathcal{A}_0)$ .*

Nous allons démontrer l'unicité de  $\mathbb{P}$  dont l'existence est assez difficile.

Note : pour l'existence vous pouvez vous référer par exemple aux livres suivants :

- Théorie de l'intégration ; Vuibert ; Paris - M. Briane, G. Pagès (1998)
- Bases mathématiques du calcul des probabilités. Paris - J. Neveu (1964).
- Introduction to Measure and Integration ; Cambridge University Press ; Cambridge, Grande-Bretagne - S.J. Taylor (1973).

**Définition 12** • *Une classe  $\mathcal{C}$  de parties de  $\Omega$  est stable par intersection si pour tous  $A, B \in \mathcal{C}$  on a  $A \cap B \in \mathcal{C}$  (et alors, si  $A_1, \dots, A_n \in \mathcal{C}$ , on a  $A_1 \cap A_2 \cap \dots \cap A_n \in \mathcal{C}$  pour  $n$  arbitraire fini).*

• *Une classe  $\mathcal{C}$  est stable par limite croissante si pour toute suite croissante  $(A_n)_{n \in \mathbb{N}^*} \in \mathcal{C}^{\mathbb{N}^*}$  on a  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{C}$ .*

• *Une classe  $\mathcal{C}$  est stable par différence si pour tous  $A, B \in \mathcal{C}$  avec  $A \subset B$  on a  $B \setminus A \in \mathcal{C}$ .*

**Théorème 12 (Théorème des classes monotones)** *Soit  $\mathcal{C}$  une classe de parties de  $\Omega$ , contenant  $\Omega$ , stable par intersection.*

*Soit  $\mathcal{D}$  la plus petite classe contenant  $\mathcal{C}$  et stable par limite croissante et par différence. Alors  $\mathcal{D} = \sigma(\mathcal{C})$ .*

*Preuve:* Remarquons d'abord que l'intersection d'une famille quelconque de classes qui sont stables par limite croissante et par différence est une classe également stable par limite croissante et par différence. Ainsi en prenant l'intersection de toutes les classes ayant ces propriétés et contenant  $\mathcal{C}$  on voit qu'il existe une plus petite classe contenant  $\mathcal{C}$  et stable par limite croissante et par différence.

Pour tout  $B \subset \Omega$  notons  $\mathcal{D}_B$  la classe de toutes les parties de  $\mathcal{A}$  vérifiant  $A \in \mathcal{D}$  et  $A \cap B \in \mathcal{D}$ . Étant données les propriétés de  $\mathcal{D}$ , on voit facilement que  $\mathcal{D}_B$  est stable par limite croissante et par différence.

Soit  $B \in \mathcal{C}$  : pour chaque  $C \in \mathcal{C}$  on a  $B \cap C \in \mathcal{C} \subset \mathcal{D}$  et  $C \in \mathcal{D}$ , donc  $C \in \mathcal{D}_B$ . Par suite  $\mathcal{C} \subset \mathcal{D}_B \subset \mathcal{D}$ . Donc  $\mathcal{D} = \mathcal{D}_B$ , grâce aux propriétés de  $\mathcal{D}$  et  $\mathcal{D}_B$ .

Maintenant, soit  $B$  dans  $\mathcal{D}$ . Pour chaque  $C \in \mathcal{C}$  on a  $B \in \mathcal{D}_C$ , et grâce à ce qui précède il vient  $B \cap C \in \mathcal{D}$ , donc  $C \in \mathcal{D}_B$ , donc  $\mathcal{C} \subset \mathcal{D}_B \subset \mathcal{D}$ .

Comme  $\mathcal{D} = \mathcal{D}_B$  pour tout  $B \in \mathcal{D}$ , on conclut que  $\mathcal{D}$  est stable par intersection. De plus  $\Omega \in \mathcal{D}$ , et  $\mathcal{D}$  est stable par différence, donc aussi par complémentation. Comme  $\mathcal{D}$  est aussi stable par limite croissante, il s'ensuit que  $\mathcal{D}$  est une tribu, et c'est clairement la plus petite tribu contenant  $\mathcal{C}$ .  $\square$  La preuve de l'unicité dans le Théorème 11 est une conséquence immédiate du Corollaire 3 ci-dessous, qui est lui-même une conséquence du Théorème 12.

**Corollaire 3** Soient  $\mathbb{P}$  et  $\mathbb{Q}$  deux probabilités définies sur  $\mathcal{A}$ . Si  $\mathbb{P}$  et  $\mathbb{Q}$  coïncident sur une classe  $\mathcal{C} \subset \mathcal{A}$  qui est stable par intersection et si  $\sigma(\mathcal{C}) = \mathcal{A}$ , alors  $\mathbb{P} = \mathbb{Q}$ .

*Preuve:*  $\Omega \in \mathcal{A}$  parce que  $\mathcal{A}$  est une tribu sur  $\Omega$  et comme  $\mathbb{P}(\Omega) = \mathbb{Q}(\Omega) = 1$  on peut supposer sans perte de généralité que  $\Omega \in \mathcal{C}$ . Soit  $\mathcal{B} = \{A \in \mathcal{A} : \mathbb{P}(A) = \mathbb{Q}(A)\}$ . Par définition d'une mesure de probabilité et par le Théorème 3,  $\mathcal{B}$  est stable par différence et par limite croissante. De plus  $\mathcal{B}$  contient  $\mathcal{C}$  par hypothèse. Par suite, comme  $\sigma(\mathcal{C}) = \mathcal{A}$ , on a  $\mathcal{B} = \mathcal{A}$  par application du Théorème des classes monotone 12.  $\square$  Si en général

on ne peut pas étendre la probabilité  $\mathbb{P}$  sur la tribu  $\mathcal{A}$  à la tribu  $\mathcal{P}(\Omega)$ , on peut néanmoins l'étendre à une tribu un peu plus grande que  $\mathcal{A}$  :

**Définition 13** Soit  $\mathbb{P}$  une probabilité sur  $\mathcal{A}$ . Un ensemble négligeable pour  $\mathbb{P}$  est une partie  $A$  de  $\Omega$  telle qu'il existe un élément  $B$  de  $\mathcal{A}$  vérifiant  $A \subset B$  et  $\mathbb{P}(B) = 0$ .

On dit qu'une propriété est vraie *presque sûrement* (en abrégé : p.s.) si elle est vraie en dehors d'un ensemble négligeable ; cette notion dépend bien sûr de la probabilité, de sorte qu'on dit parfois  $\mathbb{P}$ -presque sûrement, ou  $\mathbb{P}$ -p.s.

**Théorème 13** Soit  $\mathbb{P}$  une probabilité sur  $\mathcal{A}$  et  $\mathcal{N}$  la classe de tous les ensembles  $\mathbb{P}$ -négligeables. Alors  $\mathcal{A}' = \{A \cup N : A \in \mathcal{A}, N \in \mathcal{N}\}$  est une tribu, appelée  $\mathbb{P}$ -complétion de  $\mathcal{A}$ . C'est la plus petite tribu contenant  $\mathcal{A}$  et  $\mathcal{N}$ , et  $\mathbb{P}$  s'étend de manière unique en une mesure de probabilité (notée aussi  $\mathbb{P}$ ) sur  $\mathcal{A}'$ , en posant  $\mathbb{P}(A \cup N) = \mathbb{P}(A)$  pour  $A \in \mathcal{A}$  et  $N \in \mathcal{N}$ .

*Preuve:* L'unicité de l'extension, si elle existe est triviale; de même comme  $\emptyset \in \mathcal{A}$  et  $\emptyset \in \mathcal{N}$ , le fait que  $\mathcal{A}'$ , si c'est une tribu, soit la plus petite tribu contenant  $\mathcal{A}$  et  $\mathcal{N}$  est trivial. Il nous suffit donc de montrer que  $\mathcal{A}'$  est une tribu et que si on pose  $Q(B) = P(A)$  pour  $B = A \cup N$  (avec  $A \in \mathcal{A}$  et  $N \in \mathcal{N}$ ), alors  $Q(B)$  ne dépend pas de la décomposition  $B = A \cup N$  et  $Q$  est une probabilité sur  $\mathcal{A}'$ .

D'abord,  $\Omega \in \mathcal{A} \subset \mathcal{A}'$  et on a déjà vu que  $\emptyset \in \mathcal{A}'$ . Ensuite comme  $\mathcal{A}$  et  $\mathcal{N}$  sont stables par les unions dénombrables,  $\mathcal{A}'$  l'est également. Soit  $B = A \cup N \in \mathcal{A}'$ , avec  $A \in \mathcal{A}$  et  $N \in \mathcal{N}$ . Il existe  $C \in \mathcal{A}$  avec  $P(C) = 0$  et  $N \subset C$ . Posons  $A' = A^c \cap C^c \in \mathcal{A}$  et  $N' = N^c \cap A^c \cap C \subset C$  (donc  $\in \mathcal{N}$ ); comme  $B^c = A' \cup N'$ , on a  $B^c \in \mathcal{A}'$ , et  $\mathcal{A}'$  est stable par complémentation : ainsi  $\mathcal{A}'$  est une tribu.

Supposons maintenant que  $A_1 \cup N_1 = A_2 \cup N_2$  avec  $A_i \in \mathcal{A}$  et  $N_i \in \mathcal{N}$ . La différence symétrique  $A_1 \Delta A_2 = (A_1 \cap A_2^c) \cup (A_1^c \cap A_2)$  est contenue dans  $N_1 \cup N_2$ , lui-même contenu dans un élément  $C$  de  $\mathcal{A}$  de probabilité nulle : par suite  $P(A_1) = P(A_2)$ , ce qui montre que  $Q$  est défini sans ambiguïté, et coïncide de manière évidente avec  $P$  sur  $\mathcal{A}$ . Enfin le fait que  $Q$  soit une probabilité sur  $\mathcal{A}'$  est évident.  $\square$

## 4. Probabilités sur $\mathbb{R}$ et fonctions de répartition

### 0.0.1 Probabilités sur $\mathbb{R}$

Dans ce chapitre nous travaillerons sur un espace mesurable  $(\Omega, \mathcal{F})$  avec  $\Omega = \mathbb{R}$  et  $\mathcal{F} = \mathcal{B}(\mathbb{R})$ .

**Définition 14** Si  $\mathbb{P}$  est une probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , sa fonction de répartition est la fonction  $F(x) = \mathbb{P}(] - \infty, x])$ .

**Théorème 14** La fonction de répartition  $F$  caractérise la probabilité, i.e. que si  $\mathbb{Q}$  est une autre probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , de fonction de répartition  $G(x) = \mathbb{Q}(] - \infty, x])$  et si  $F = G$ , alors  $\mathbb{P} = \mathbb{Q}$ .

*Preuve:* la preuve est admise, voir page 41, Théorème 7.1 de <sup>1</sup>. □ 1.

La signification de ce théorème est que la probabilité est entièrement connue si on connaît seulement  $F$  : on peut donc en principe retrouver  $\mathbb{P}(A)$  pour n'importe quel borélien  $A$  à partir de la fonction  $F$  (déterminer "explicitement"  $\mathbb{P}(A)$  en fonction de  $F$  est une autre chose!).

Il est donc important de savoir caractériser les fonctions  $F$  qui sont des fonctions de répartition : cela permet en principe de caractériser toutes les probabilités  $\mathbb{P}$ , et cela fait l'objet du théorème suivant.

**Théorème 15** Une fonction  $F$  est la fonction de répartition d'une (unique) probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  si et seulement si elle vérifie les propriétés suivantes :

- (i)  $F$  est croissante ;
- (ii)  $F$  est continue à droite ;
- (iii)  $\lim_{x \rightarrow -\infty} F(x) = 0$  et  $\lim_{x \rightarrow \infty} F(x) = 1$ .

*Preuve:* la preuve est admise, voir page 42, Théorème 7.2 de <sup>2</sup>. □ 2.

**Propriété 6** Soit  $F$  la fonction de répartition de la probabilité  $\mathbb{P}$  sur  $\mathbb{R}$ . En notant  $F(x^-)$  la limite à gauche de  $F$  au point  $x$  (qui existe puisque  $F$  est croissante), pour tous  $x < y$  on a

- (i)  $\mathbb{P}(]x, y]) = F(y) - F(x)$ ,
- (ii)  $\mathbb{P}([x, y]) = F(y) - F(x^-)$ ,
- (iii)  $\mathbb{P}([x, y[) = F(y^-) - F(x^-)$ ,
- (iv)  $\mathbb{P}(]x, y[) = F(y^-) - F(x)$ ,
- (v)  $\mathbb{P}(\{x\}) = F(x) - F(x^-)$ ,

et en particulier  $\mathbb{P}(\{x\}) = 0$  pour tout  $x$  si et seulement si la fonction  $F$  est continue.

*Preuve:* (i) se démontrer avec  $F(x) + \mathbb{P}(]x, y]) = \mathbb{P}(]-\infty, x]) + \mathbb{P}(]x, y]) = \mathbb{P}(]-\infty, x] \cup ]x, y])$  car  $\mathbb{P}$  est une mesure et on a deux ensembles disjoints, et donc  $F(x) + \mathbb{P}(]x, y]) = \mathbb{P}(]-\infty, y]) = F(y)$ .

Pour (ii) on écrit que  $\mathbb{P}(]x - 1/n, y]) = F(y) - F(x - 1/n)$  par (i). Le membre de droite converge quand  $n \rightarrow \infty$  vers  $F(y) - F(x^-)$  par définition de la limite à gauche de  $F$ ; quant au membre de gauche  $\mathbb{P}(]x - 1/n, y])$ , il converge vers  $\mathbb{P}([x, y])$  par application du théorème vu en cours, en effet  $\cap_{n \in \mathbb{N}^*} ]x - 1/n, y] = [x, y]$  et  $(]x - 1/n, y])_{n \in \mathbb{N}^*}$  décroît vers  $[x, y]$ , d'où  $\mathbb{P}(\cap_{n \in \mathbb{N}^*} ]x - 1/n, y]) = \lim_{n \rightarrow \infty} \mathbb{P}(]x - 1/n, y]) = \mathbb{P}([x, y])$ .

Les assertions (iii), (iv) et (v) se montrent de manière analogue.  $\square$

**Exemple 6** Si  $f$  est positive et Riemann-intégrable et  $\int_{\mathbb{R}} f(x) dx = 1$ , la fonction  $F(x) = \int_{-\infty}^x f(y) dy$  définit la fonction de répartition d'une probabilité  $\mathbb{P}$  sur  $\mathbb{R}$  (voir le théorème précédent), en effet  $F$  est croissante, càd (car continue),  $\lim_{x \rightarrow -\infty} F(x) = 0$  et  $\lim_{x \rightarrow \infty} F(x) = 1$ .

La fonction  $f$  est appelée sa densité. (Il n'est pas vrai que toute probabilité sur  $\mathbb{R}$  admet une densité; en effet s'il existe une densité alors la fonction  $F$  est continue, alors qu'il existe beaucoup de fonctions de répartition discontinues, ainsi que le montre l'exemple suivant.)

**Exemple 7** Soit  $\alpha \in \mathbb{R}$ . La "masse de Dirac" au point  $\alpha$  est la probabilité sur  $\mathbb{R}$  qui est donné par  $\mathbb{P}(A) = \mathbf{1}_A(\alpha)$ . Sa fonction de répartition est  $F(x) = \mathbf{1}_{]a, \infty[}(x)$ .

Dans les exemples suivants nous définissons la probabilité (ou "loi")  $\mathbb{P}$  sur  $\mathbb{R}$  par sa densité  $f$ ; c'est-à-dire que nous définissons la fonction positive  $f$  d'intégrale 1, et la fonction de répartition  $F$  de  $\mathbb{P}$  est définie par  $F(x) = \int_{-\infty}^x f(u) du$ . On commet un léger abus de langage (habituel) en identifiant  $f$  avec la loi  $\mathbb{P}$  correspondante, puisqu'elle détermine en fait de manière unique cette loi.

**Exemple 8**  $f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$  pour tout  $x \in \mathbb{R}$ , la loi est appelée la loi uniforme sur  $[a, b]$ .

**Exemple 9**  $f(x) = \beta e^{-\beta x} \mathbf{1}_{\mathbb{R}_+}(x)$  pour tout  $x \in \mathbb{R}$ , la loi exponentielle de paramètre  $\beta > 0$ .

**Exemple 10**  $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbf{1}_{\mathbb{R}_+}(x)$  pour tout  $x \in \mathbb{R}$ , la loi gamma de paramètre  $\alpha$  et  $\beta$  avec  $\alpha, \beta \in \mathbb{R}_+^*$ .  $\Gamma$  désigne la fonction gamma, définie par  $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$  pour tout  $\alpha > 0$ ; remarquons que  $\Gamma(\alpha) = (\alpha - 1)!$  pour tout  $\alpha \in \mathbb{N}^*$  et  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ .

**Exemple 11**  $f(x) = \alpha \beta^\alpha x^{\alpha-1} e^{-(\beta x)^\alpha} \mathbf{1}_{\mathbb{R}_+}(x)$  pour tout  $x \in \mathbb{R}$ , la loi de Weibull, avec  $\alpha, \beta \in \mathbb{R}_+^*$ .

**Exemple 12**  $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$  pour tout  $x \in \mathbb{R}$ , la loi Normale (ou distribution Gaussienne) avec  $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$ ; la notation standard est  $\mathcal{N}(\mu, \sigma^2)$ .

**Exemple 13** En posant  $g_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$  pour tout  $x \in \mathbb{R}$  et en posant  $f(x) = \frac{1}{x} g_{\mu, \sigma^2}(\ln(x)) \mathbf{1}_{\mathbb{R}_+^*}(x)$  pour tout  $x \in \mathbb{R}$ , la loi lognormale, avec  $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$ .

**Exemple 14**  $f(x) = \frac{\beta}{2} e^{-\beta|x-\alpha|}$  pour tout  $x \in \mathbb{R}$ , la loi double exponentielle, avec  $\alpha, \beta \in \mathbb{R}$ .

**Exemple 15**  $f(x) = \frac{1}{\beta\pi} \frac{1}{1+(x-\alpha)/\beta)^2}$  pour tout  $x \in \mathbb{R}$ , la loi de Cauchy, avec  $(\alpha, \beta) \in \mathbb{R} \times \mathbb{R}_+^*$ .

## 0.0.2 Variables aléatoires

Dans un chapitre précédent nous avons considéré des variables aléatoires définies sur un espace dénombrable  $\Omega$ .

Nous voulons maintenant considérer un espace  $\Omega$  quelconque, éventuellement non dénombrable, cet espace étant muni d'une tribu  $\mathcal{A}$  d'événements et d'une probabilité  $\mathbb{P}$  (sur  $\mathcal{A}$ ).

Comme nous l'avons vu, une variable aléatoire est une application  $X : \Omega \rightarrow E$ . Ce qui nous intéresse est la détermination de la probabilité  $\mathbb{P}(\{X \in A\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}(X^{-1}(A))$  (trois manières équivalent d'écrire le même événement). Pour cela, il faut évidemment que l'ensemble  $\{X \in A\}$  soit dans la tribu  $\mathcal{A}$ , ce qui a priori n'est pas vrai pour une partie arbitraire  $A$  de  $F$ . Cela motive la définition suivante :

**Définition 15 (Variable aléatoire)** Soit  $(\Omega, \mathcal{A})$  un espace mesurable, toute fonction mesurable  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  est appelée variable aléatoire.

**Remarque** : en probabilités avancées on étudiera les v.a. à valeurs  $\mathbb{R}^d$ , i.e. des v.a.  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . Rappelons un théorème issu de Mesure et Intégration utile :

**Théorème 16** Soit  $\mathcal{C}$  une classe de parties de  $F$  telle que  $\sigma(\mathcal{C}) = \mathcal{F}$  tribu sur  $F$ . Pour qu'une application  $X : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$  soit mesurable ssi  $X^{-1}(\mathcal{C}) \subset \mathcal{E}$ , i.e.  $X^{-1}(C) \in \mathcal{E}$  pour tout  $C \in \mathcal{C}$ .

Un cas particulier de ce théorème permet de caractériser de manière particulièrement simple les v.a. à valeurs réelles, i.e. quand  $F = \mathbb{R}$ .

**Corollaire 4** Soit  $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  et  $X, X_n : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$ . On a l'équivalence suivante :

$X$  est mesurable ssi  $\{X \leq a\} \in \mathcal{E}$  pour tout  $a \in \mathbb{R}$  ssi  $\{X \leq a\} \in \mathcal{E}$  pour tout  $a \in \mathbb{Q}$ .

*Preuve:* Il suffit d'appliquer le théorème précédent avec le fait que  $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C}) = \sigma(\mathcal{C}')$ , où  $\mathcal{C} = \{] - \infty, a], a \in \mathbb{R}\}$  et  $\mathcal{C}' = \{] - \infty, a], a \in \mathbb{Q}\}$ .  $\square$

**Remarque :** grâce au cours de Mesure et Intégration on sait que  $\sup_n X_n, \inf_n X_n, \limsup_{n \rightarrow \infty} X_n, \liminf_{n \rightarrow \infty} X_n$  sont toutes mesurables et que si  $(X_n)_n$  converge simplement vers  $X$ , alors  $X$  est mesurable.

Soit  $X$  une v.a. définie sur un espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$ , à valeurs dans un espace mesurable  $(E, \mathcal{E})$ . La loi de  $X$  est alors définie par

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B), \forall B \in \mathcal{E}.$$

$\mathbb{P}_X$  est appelée l'image de  $\mathbb{P}$  par  $X$ , comme  $X^{-1}$  commute avec les réunions et les intersections (finies, dénombrables ou infinies non dénombrables!), et comme  $X^{-1}(E) = \Omega$ , on a de manière immédiate :

**Théorème 17** La loi  $\mathbb{P}_X$  de  $X$  est une mesure de probabilité sur  $(E, \mathcal{E})$ .

Lorsque  $X$  est une variable aléatoire réelle, sa loi  $\mathbb{P}_X$  est une probabilité sur  $\mathbb{R}$ , dont on sait qu'elle est entièrement caractérisée par sa fonction de répartition :

$$F_X(x) = \mathbb{P}_X(] - \infty, x]) = \mathbb{P}(X \leq x), \forall x \in \mathbb{R}.$$

La fonction  $F_X$  s'appelle la fonction de répartition de la v.a.  $X$ . Lorsque  $F_X$  admet une densité  $f_X$  (i.e.  $F_X(x) = \int_{-\infty}^x f_X(y) dy$  pour tout  $x \in \mathbb{R}$ ), on dit aussi que la fonction  $f_X$  est la densité de la v.a.  $X$ .

## 5. Intégration par rapport à une mesure de probabilité (espérance, variance)

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité. Nous voulons définir l'espérance d'une variable aléatoire réelle (puis la variance) sur cet espace de probabilité, en utilisant la notion d'intégrale d'une fonction mesurable (variable aléatoire).

**Définition 16** a) Soit  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  variable aléatoire réelle positive. L'espérance de  $X$  (ou "espérance mathématique", ou "intégrale" par rapport à  $\mathbb{P}$  de  $X$ ), notée par  $\mathbb{E}[X]$ , est définie par  $\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$ , d'après Mesure et Intégration  $\mathbb{E}[X]$  existe et est à valeurs  $[0, +\infty]$ .

b) On dit qu'une v.a.  $X$  réelle (pas forcément positive) a une espérance finie, et on la note par  $\mathbb{E}[X]$ , si  $\mathbb{E}[X^+]$  et  $\mathbb{E}[X^-]$  sont toutes deux finies et on la définit par  $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$  (remarque  $X = X^+ - X^-$ ), qu'on écrit également sous la forme  $\int_{\Omega} X(\omega) d\mathbb{P}(\omega)$ .

c) On note  $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$  (lorsqu'il n'y a pas d'ambiguïté on note cela par  $\mathcal{L}^1$ ) l'ensemble de toutes les v.a.  $X$  réelles intégrables (i.e.  $\mathbb{E}[X]$  est fini).

d) Une v.a. réelle  $X$  admet une espérance si  $\mathbb{E}[X^+]$  et  $\mathbb{E}[X^-]$  ne sont pas toutes deux simultanément égales à  $+\infty$ . Dans ce cas, son espérance  $\mathbb{E}[X]$  est encore définie par  $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$  (avec les conventions  $+\infty + \alpha = +\infty$  et  $-\infty + \alpha = -\infty$  pour tout  $\alpha \in \mathbb{R}$ ).

Noter que si  $\mathbb{E}[X]$  admet une espérance, on a  $\mathbb{E}[X] \in [-\infty, +\infty]$ , et  $X$  est intégrable si et seulement si son espérance existe et est finie (en effet  $|X| = X^+ + X^-$ , voir Mesure et Intégration).

Le théorème suivant présente les propriétés les plus importantes de la fonctionnelle espérance. Les preuves des parties se trouvent dans le cours de Mesure et Intégration.

Dans son énoncé, l'assertion " $X = Y$  presque sûrement", ou en abrégé "p.s.", signifie que l'événement  $\{X \neq Y\}$  est négligeable ( $\mathbb{P}$ -négligeable), i.e.  $\mathbb{P}(X \neq Y) = 0$  ou de manière équivalente  $\mathbb{P}(X = Y) = 1$ ; de même  $X \geq Y$  p.s. signifie que  $\mathbb{P}(X \geq Y) = 1$ , et  $(X_n)_n$  converge p.s. vers  $X$ , signifie que  $\mathbb{P}(\{\omega \in \Omega, \lim_{n \rightarrow \infty} X_n(\omega) = X\}) = 1$ .

**Théorème 18**

- a)  $\mathcal{L}^1$  est un espace vectoriel et l'espérance est une forme linéaire sur  $\mathcal{L}^1$ , et aussi une fonctionnelle positive (i.e.  $X \geq 0$  p.s. implique  $\mathbb{E}[X] \geq 0$ ). Si de plus  $X$  et  $Y$  sont deux v.a. avec  $0 \leq X \leq Y$  et  $Y \in \mathcal{L}^1$ , alors  $X \in \mathcal{L}^1$  et  $\mathbb{E}[X] \leq \mathbb{E}[Y]$ .
- b)  $X \in \mathcal{L}^1$  si et seulement si  $|X| \in \mathcal{L}^1$ , ou de manière équivalent si  $\mathbb{E}[|X|] < +\infty$ , et dans ce cas  $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$ . En particulier toute v.a. bornée est intégrable.
- c) Si  $X = Y$  p.s., alors  $\mathbb{E}[X] = \mathbb{E}[Y]$ .
- d) (Théorème de convergence monotone.) Si les v.a.  $X_n, n \in \mathbb{N}$ , sont positives et si la suite  $(X_n)_n$  croît p.s. vers  $X$ , alors  $\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X]$  (même si  $\mathbb{E}[X] = \infty$ ).
- e) (Lemme de Fatou.) Si les v.a.  $X_n, n \in \mathbb{N}$  vérifient  $X_n \geq Y$  p.s. avec  $Y$  intégrable, les v.a.  $X_n$  et  $\liminf_{n \rightarrow \infty} X_n$  admettent une espérance, et on a  $\mathbb{E}[\liminf_{n \rightarrow \infty} X_n] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n]$ . Cela est vraie en particulier si  $X_n \geq 0$  p.s. pour tout  $n$ .
- f) (Théorème de convergence dominée de Lebesgue.) Si  $(X_n)_n$  converge p.s. vers  $X$  et si  $|X_n| \leq Y$  p.s. pour tout  $n$ , où  $Y$  est une v.a. intégrable, alors  $X_n \in \mathcal{L}^1, X \in \mathcal{L}^1$ , et  $\mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$ .

Rappelons que comme en Mesure et Intégration, l'égalité p.s. entre v.a. est clairement une relation d'équivalence, et deux v.a. équivalentes (i.e. p.s. égales) ont la même espérance : On peut donc définir l'espace  $L^1$  comme " $\mathcal{L}^1$  modulo cette relation d'équivalence".

En d'autres termes un élément de  $L^1$  est une classe d'équivalence, c'est-à-dire l'ensemble de toutes les v.a. qui sont p.s. égales à une v.a. particulière. Vu (c) ci-dessus, on peut parler de "l'espérance" de cette classe d'équivalence, qui est l'espérance de n'importe laquelle des v.a. appartenant à cette classe. Comme de plus l'addition de deux v.a. et la multiplication d'une v.a. par une constante préservent l'égalité p.s., l'ensemble  $L^1$  est aussi un espace vectoriel. Par suite on commet l'abus (anodin) d'identifier une v.a. avec sa classe d'équivalence, et on écrit d'habitude  $X \in L^1$  au lieu de  $X \in \mathcal{L}^1$ .

Si  $1 < p < \infty$ , on définit  $\mathcal{L}^p$  comme l'ensemble des v.a.  $X$  telles que  $|X|^p \in \mathcal{L}^1$ ;  $L^p$  est défini de manière analogue à  $L^1$  : c'est  $\mathcal{L}^p$  modulo la relation d'équivalence "égalité p.s."; ou de manière équivalente, deux éléments de  $\mathcal{L}^p$  qui sont p.s. égaux sont considérés comme deux représentants du même élément de  $L^p$ .

Une conséquence utile du théorème de convergence dominée de Lebesgue est le résultat suivant, qui permet d'invertir sommation et espérance.

**Théorème 19** Soit  $(X_n)_n$  une suite de v.a.

a) Si les  $X_n$  sont toutes positives, alors

$$\mathbb{E} \left[ \sum_{n=1}^{\infty} X_n \right] = \sum_{n=1}^{\infty} \mathbb{E}[X_n].$$

b) Si  $\sum_{n=1}^{\infty} \mathbb{E}[|X_n|] < \infty$ , alors la série  $\sum_{n=1}^{\infty} X_n$  converge p.s., sa somme est intégrable et on a l'égalité  $\mathbb{E}[\sum_{n=1}^{\infty} X_n] = \sum_{n=1}^{\infty} \mathbb{E}[X_n]$ .

*Preuve:*  $S_n := \sum_{k=1}^n |X_k|$  et  $T_n := \sum_{k=1}^n X_k$ .

$(S_n)_n$  croît vers  $S := \sum_{k=1}^{\infty} |X_k|$  donc par TCM on a  $\mathbb{E}[S] = \sum_{k=1}^{\infty} \mathbb{E}[|X_k|] < \infty$ .

Si les  $X_n$  sont toutes positives alors  $S_n = T_n$  et donc (a) est prouvé.

Si les  $X_n$  ne sont pas nécessairement positives, mais si on a  $\sum_{k=1}^{\infty} \mathbb{E}[|X_k|] < \infty$ , on obtient aussi que  $\mathbb{E}[S] < \infty$ .

Maintenant pour chaque  $\varepsilon > 0$ ,  $\mathbb{P}(S = \infty) \leq \varepsilon \mathbb{E}[S]$ , et donc  $\mathbb{E}[S] < \infty$  et  $\varepsilon$  arbitraire donnent que  $\mathbb{P}(S = \infty) = 0$ , donc  $\sum_{k=1}^{\infty} X_k$  est p.s. une série absolument convergente, et sa somme disons  $T$  est la limite des  $T_n$ . De plus  $|T_n| \leq S_n \leq S$  et  $S \in L^1$ . Par suite le TCD implique  $\mathbb{E}[\sum_{k=1}^{\infty} X_k] = \mathbb{E}[\lim_{n \rightarrow \infty} T_n] = \mathbb{E}[T]$ .  $\square$

Rappelons que  $L^1$  et  $L^2$  sont les ensembles de classes d'équivalences de v.a. intégrables (respectivement de carré intégrable), pour la relation d'équivalence "égalité p.s."

**Théorème 20**

a) Si  $X$  et  $Y$  sont dans  $L^2$ , on a  $XY \in L^1$  et l'inégalité de Cauchy-Schwarz :

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.$$

b) On a  $L^2 \subset L^1$  et si  $X \in L^2$  alors  $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$ .

c) L'espace  $L^2$  est un espace vectoriel : si  $X, Y \in L^2$  et  $\alpha, \beta \in \mathbb{R}$ , alors  $\alpha X + \beta Y \in L^2$  (en fait vous verrez que  $L^2$  est un espace de Hilbert).

*Preuve:* a) On a  $|XY| \leq X^2/2 + Y^2/2$  d'où la réponse. La fonction  $a \mapsto \mathbb{E}[(aX + Y)^2]$  est du second degré en  $a$  qui est positive pour tout  $a \in \mathbb{R}$ , donc il est de discriminant négatif, d'où la réponse.

b) Utiliser a) avec  $X = X \times 1$  et  $1 \in L^2$ .

c) Pour tous  $X, Y \in L^2$  et pour tous  $\alpha, \beta \in \mathbb{R}$ , on a  $(\alpha X + \beta Y)^2 \leq \alpha^2 X^2/2 + \beta^2 Y^2/2$  d'où la réponse.  $\square$

**Définition 17** Si  $X \in L^2$ , la variance de  $X$ , notée aussi  $\sigma^2(X)$  ou  $\sigma_X^2$ , est définie par  $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ .

**Remarque :**  $X \in L^2 \Rightarrow X \in L^1$ , donc  $\mathbb{E}[X]$  existe. On a  $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ .

**Théorème 21 (Inégalité de Bienaymé-Chebyshev)**

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[X^2]}{a^2}, \forall a \in \mathbb{R}_+^*.$$

*Preuve:*  $a^2 \mathbf{1}_{\{|X| \geq a\}} \leq X^2$  pour tout  $a \in \mathbb{R}_+^*$  d'où la réponse.  $\square$

**Corollaire 5**

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\sigma^2(X)}{a^2}, \forall a \in \mathbb{R}_+^*.$$

Le théorème suivant est très utile, de même que le corollaire qui le suit, pour les calculs effectifs d'espérance. Il montre que l'espérance d'une v.a. ne dépend que de sa loi, et il s'agit d'une version de la formule "de changement de variable" dans les intégrables, vue sous un aspect "abstrait".

**Théorème 22** Soit  $X$  une v.a. définie sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $(E, \mathcal{E})$ , et de loi  $\mathbb{P}_X$ . Soit  $h : (E, \mathcal{E}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  une application mesurable.

a) On a  $h(X) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P}) \Leftrightarrow h \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ .

b) Si  $h$  est positive satisfaisant à a) alors on a

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

*Preuve:* Rappelons que par définition  $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$  pour tout  $B \in \mathcal{B}(\mathbb{R})$ . Or  $\mathbb{E}[\mathbf{1}_B(X)] = \mathbb{P}(X^{-1}(B)) = \mathbb{P}_X(B) = \int_{\mathbb{R}} \mathbf{1}_B(x) d\mathbb{P}_X(x)$  donc la formule est vraie pour  $h = \mathbf{1}_B$  simple  $n = 1$ , pour  $h$  simple (étagé) quelconque on a la formule valide par linéarité de l'espérance. Si  $h$  est positive il existe (d'après Mesure et Intégration) une suite croissante  $(h_n)_n$  de fonctions mesurables positives simples telles que  $\lim_{n \rightarrow \infty} h_n = h$ . Donc  $\mathbb{E}[h(X)] = \mathbb{E}[\lim_n h_n(X)] = \lim_n \mathbb{E}[h_n(X)] = \lim_n \int_{\mathbb{R}} h_n(x) d\mathbb{P}_X(x) = \int_{\mathbb{R}} \lim_n h_n(x) d\mathbb{P}_X(x) = \int_{\mathbb{R}} h(x) d\mathbb{P}_X(x)$ , où on a utilisé deux fois le TCM. Cela prouve b) quand  $h$  est positive. En appliquant à  $|h|$ , cela prouve aussi a) (rappelons que  $X \in \mathcal{L}^1$  ssi  $\mathbb{E}[|X|] < \infty$ ). Enfin en écrivant  $h = h^+ - h^-$  on déduit la formule demandée pour  $h$  de signe quelconque de la même formule pour  $h$  positive par soustraction.  $\square$

Le résultat suivant peut se déduire de ce théorème, cela sera prouvée plus tard.

**Corollaire 6** Soit  $X$  une v.a. réelle admettant une densité  $f_X$  (i.e.  $\mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(u) du$  pour tout  $x \in \mathbb{R}$ ). Si  $\mathbb{E}[|h(X)|] < \infty$  ou si  $h$  est positive, on a  $\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x) f_X(x) dx$ .

## 6. Variables aléatoires indépendantes et Lois de probabilités sur $\mathbb{R}$

### 0.1 Variables aléatoires indépendantes

Rappelons que deux événements  $A$  et  $B$  sont indépendants si la connaissance du fait que  $B$  est réalisé ne change pas la probabilité de  $A$ , ce qui revient à dire que  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . De même une famille quelconque  $(A_i)_{i \in I}$  d'événements est indépendante si  $\mathbb{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j)$  pour tout sous-ensemble fini  $J$  de  $I$ .

De manière analogue, deux v.a.  $X$  et  $Y$  seront indépendantes si  $\{Y \in A\}$  et  $\{X \in B\}$  sont indépendants pour tous  $A$  et  $B$  dans les tribus correspondantes.

Si  $X$  est mesurable de  $(\Omega, \mathcal{A})$  dans  $(E, \mathcal{E})$ , on rappelle (voir Mesure et Intégration) que la tribu engendrée par  $X$  est  $X^{-1}(\mathcal{E}) = \{X^{-1}(A) : A \in \mathcal{E}\}$ .

**Définition 18** a) Soit  $(\mathcal{A}_i)_{i \in I}$  une famille de sous-tribus de  $\mathcal{A}$ . Les tribus  $\mathcal{A}_i$  sont indépendantes si pour tout partie finie de  $J$  de  $I$  et tous  $A_i \in \mathcal{A}_i$  on a

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

b) Soit  $(X_i)_{i \in I}$  une famille de variables aléatoires, à valeurs respectives dans les  $(E_i, \mathcal{E}_i)$ . Les variables aléatoires  $X_i$  sont indépendantes si les tribus engendrées  $X_i^{-1}(\mathcal{E}_i)$  sont indépendantes.

Dans la suite de ce chapitre nous considérons essentiellement des couples de v.a., de façon à garder des notations simples. Tous les résultats s'étendent de manière évidente aux familles finies de v.a.

Noter que  $X$  et  $Y$  prennent leurs valeurs dans  $(E, \mathcal{E})$  et  $(F, \mathcal{F})$  respectivement : les espace  $E$  et  $F$  peuvent être différents.

**Théorème 23** Pour que  $X$  et  $Y$  soient indépendantes, il faut et il suffit que l'une des conditions équivalentes ci-dessous soit satisfaite :

a)  $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$  pour tous  $A \in \mathcal{E}, B \in \mathcal{F}$ .

- b)  $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$  pour tous  $A \in \mathcal{C}$ ,  $B \in \mathcal{D}$ , où  $\sigma(\mathcal{C}) = \mathcal{E}$  et  $\sigma(\mathcal{D}) = \mathcal{F}$  avec  $\mathcal{E}$  et  $\mathcal{F}$  des classes de parties de  $\mathbb{E}$  et  $\mathbb{F}$ , stables par intersection finie.
- c)  $f(X)$  et  $g(Y)$  sont indépendantes pour tout couple  $(f, g)$  de fonctions mesurables réelles sur  $E$  et  $F$  respectivement.
- d)  $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$  pour tout couple  $(f, g)$  de fonctions mesurables réelles, bornées (ou positives) sur  $E$  et  $F$  respectivement.
- e) Si de plus  $E$  et  $F$  sont des espaces métriques avec leurs tribus boréliennes respectives  $\mathcal{E}$  et  $\mathcal{F}$ , alors  $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$  pour tout couple  $(f, g)$  de fonctions mesurables continues réelles bornées.

*Preuve:* Comme  $X^{-1}(A) = \{X \in A\}$  et  $X^{-1}(B) = \{X \in B\}$  pour tous  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$  d'où a).

L'implication a)  $\Rightarrow$  b) est triviale.

Pour l'implication b)  $\Rightarrow$  a) : La classe  $\mathcal{C}_B$  des ensembles  $A \in \mathcal{E}$  vérifiant  $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$  pour  $B \in \mathcal{D}$  donné est stable par limites croissantes et par différence, et contient  $\mathcal{C}$  par hypothèse, tandis que  $\mathcal{C}$  est stable par intersection. Le théorème des classes monotones entraîne alors que  $\mathcal{C}_B = \mathcal{E}$ . En d'autres termes on a b) avec  $\mathcal{C} = \mathcal{E}$ . De manière analogue (en fixant  $A \in \mathcal{E}$ ) on obtient qu'on a b) avec de plus  $\mathcal{D} = \mathcal{F}$  : en d'autres termes, on a a).

Pour l'implication c)  $\Rightarrow$  a) : il suffit de prendre  $f(x) = x$  et  $g(y) = y$  (les applications identité sur  $E$  et  $F$  respectivement).

Pour l'implication a)  $\Rightarrow$  c) : Étant données  $f$  et  $g$ , on voit que  $f(X)^{-1}(\mathcal{E}) = X^{-1}(f^{-1}(\mathcal{E})) \subset X^{-1}(\mathcal{E})$ , et de même  $g(Y)^{-1}(\mathcal{F}) \subset Y^{-1}(\mathcal{F})$  : le résultat est alors évident.

L'implication d)  $\Rightarrow$  a) : Prendre  $f(x) = \mathbf{1}_A(x)$  et  $g(x) = \mathbf{1}_B(x)$ .

a)  $\Rightarrow$  d) : Par hypothèse on a d) pour  $f$  et  $g$  fonctions indicatrices, et donc pour  $f$  et  $g$  fonctions simples par linéarité. Si  $f$  et  $g$  sont positives il existe des suites  $(f_n)_n$  et  $(g_n)_n$  de fonctions simples positives croissant vers  $f$  et  $g$  respectivement. Noter que les produits  $f_n(X)g_n(Y)$  croissent vers  $f(X)g(Y)$ . En appliquant le TCM il vient alors  $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[\lim_{n \rightarrow \infty} f_n(X)g_n(Y)] = \lim_{n \rightarrow \infty} \mathbb{E}[f_n(X)g_n(Y)] = \lim_{n \rightarrow \infty} \mathbb{E}[f_n(X)]\mathbb{E}[g_n(Y)] = \lim_{n \rightarrow \infty} \mathbb{E}[f_n(X)] \lim_{n \rightarrow \infty} \mathbb{E}[g_n(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ . On a donc le résultat lorsque  $f$  et  $g$  sont positives. Lorsque  $f$  et  $g$  sont bornées on écrit  $f = f^+ - f^-$  et  $g = g^+ - g^-$ , et on conclut par linéarité.

d)  $\Rightarrow$  e) : Évident.

e)  $\Rightarrow$  b) : Il suffit de montrer b) lorsque  $\mathcal{C}$  et  $\mathcal{D}$  sont les classes de tous les fermés de  $E$  et  $F$  (ces classes sont stables par intersection). Soit par exemple  $A$  un fermé de  $E$ ; si  $f_n(x) = \min(1, nd(x, A))$ , où  $d(x, A)$  désigne la distance du point  $x$  au fermé  $A$ , alors  $f_n$  est continue, et vérifie  $0 \leq f_n \leq 1$  et décroît vers l'indicatrice  $\mathbf{1}_A$ . Pour

$B$  fermé de  $F$  soit de même une suite  $(g_n)_n$  de fonctions continues  $0 \leq g_n \leq 1$  et convergeant vers  $\mathbf{1}_B$ . On peut reproduire la fin de la preuve de l'implication  $a) \Rightarrow d)$  en remplaçant le TCM par celui de la convergence dominée, pour obtenir  $b)$ .  $\square$

Rappel d'un théorème de Mesure et Intégration :

**Théorème 24** Soit  $f : (E \times F, \mathcal{E} \otimes \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  mesurable. Pour tout  $x \in E$ ,  $y \mapsto f(x, y)$  est une fonction  $\mathcal{F}$ -mesurable; de même pour tout  $y \in F$  la fonction  $x \mapsto f(x, y)$  est  $\mathcal{E}$ -mesurable.

**Théorème 25 (Tonelli-Fubini)** Soient  $\mathbb{P}$  et  $\mathbb{Q}$  des probabilités sur  $(E, \mathcal{E})$  et  $(F, \mathcal{F})$  respectivement.

(a) Posons  $R(A \times B) = \mathbb{P}(A)\mathbb{Q}(B)$  pour tous  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$ . Alors  $R$  s'étend de manière unique en une probabilité sur  $(E \times F, \mathcal{E} \otimes \mathcal{F})$ , notée  $\mathbb{P} \otimes \mathbb{Q}$ .

(b) Pour toute fonction  $f : (E \times F, \mathcal{E} \otimes \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  positive ou  $\mathbb{P} \otimes \mathbb{Q}$ -intégrable, la fonction  $x \mapsto \int_F f(x, y)\mathbb{Q}(dy)$  est  $\mathcal{E}$ -mesurable, la fonction  $y \mapsto \int_E f(x, y)\mathbb{P}(dx)$  est  $\mathcal{F}$ -mesurable, et on a

$$\begin{aligned} \int_{E \times F} f d(\mathbb{P} \otimes \mathbb{Q}) &= \int_E \left( \int_F f(x, y)\mathbb{Q}(dy) \right) \mathbb{P}(dx) \\ &= \int_F \left( \int_E f(x, y)\mathbb{P}(dx) \right) \mathbb{Q}(dy). \end{aligned}$$

*Preuve:* Voir Théorème 10.3 p.71 de<sup>3</sup>.  $\square$

**Corollaire 7** Soient  $X$  et  $Y$  deux v.a. sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs respectives dans  $(E, \mathcal{E})$  et  $(F, \mathcal{F})$ . Le couple  $Z := (X, Y)$  peut être considéré comme une v.a. à valeurs dans  $(E \times F, \mathcal{E} \otimes \mathcal{F})$  et les deux v.a.  $X$  et  $Y$  sont indépendantes si et seulement si la loi  $\mathbb{P}_{(X, Y)}$  du couple égale le produit  $\mathbb{P}_X \otimes \mathbb{P}_Y$ .

*Preuve:* Il est évident que  $Z^{-1}(A \times B) = X^{-1}(A) \cap Y^{-1}(B) \in \mathcal{A}$  si  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$ , de sorte que la mesurabilité de  $Z$  découle de la définition de la tribu  $\mathcal{E} \otimes \mathcal{F}$  et du Théorème 16. L'indépendance de  $X$  et  $Y$  revient au fait que pour tous  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$  on ait  $\mathbb{P}((X, Y) \in A \times B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$ , ce qui équivaut à  $\mathbb{P}_{(X, Y)}(A \times B) = \mathbb{P}_X(A)\mathbb{P}_Y(B)$ , de sorte que l'unicité de la probabilité produit dans le théorème de Fubini nous donne la seconde assertion.  $\square$

- Nous faisons maintenant une légère digression en construisant un modèle pour des variables aléatoires indépendantes, ce qui permet en particulier de vérifier que cette notion n'est pas vide.

- Soit d'abord  $\mu$  une probabilité sur un espace mesurable  $(E, \mathcal{E})$ . Il est facile de construire une v.a.  $X$  à valeurs dans  $E$  et de loi  $\mu$  :

Rappelons (voir Mesure et Intégration) qu'en général  $\mathcal{E}, \mathcal{F}$  tribus respectives sur  $E$  et  $F$  n'impliquent pas que le produit cartésien  $\mathcal{E} \times \mathcal{F}$  est une tribu sur  $E \times F$ . La plus petite tribu de  $E \times F$  contenant  $\mathcal{E} \times \mathcal{F}$ , soit  $\sigma(\mathcal{E} \times \mathcal{F})$  est très fréquemment rencontrée et est notée ainsi :  $\mathcal{E} \otimes \mathcal{F} = \sigma(\mathcal{E} \times \mathcal{F})$ .

*Preuve:* (du théorème 24) Supposons d'abord  $f$  de la forme  $f(x, y) = \mathbf{1}_C(x, y)$ , pour  $C \in \mathcal{E} \otimes \mathcal{F}$ . Soit  $\mathcal{H} = \{C \in \mathcal{E} \otimes \mathcal{F} : y \mapsto \mathbf{1}_C(x, y) \text{ est } \mathcal{F}\text{-mesurable pour chaque } x \in E \text{ fixé}\}$ . Alors  $\mathcal{H}$  est une tribu (facile) contenant  $\mathcal{E} \times \mathcal{F}$ , donc  $\sigma(\mathcal{E} \times \mathcal{F}) \subset \mathcal{H}$ . Comme par construction  $\mathcal{H} \subset \sigma(\mathcal{E} \times \mathcal{F})$ , il vient  $\mathcal{H} = \mathcal{E} \otimes \mathcal{F}$ . On a donc montré le résultat pour les fonctions indicatrices, donc par linéarité pour les fonctions mesurables simples. Par limite croissante, on l'obtient aussi pour les fonctions mesurables positives. Enfin par différence  $f = f^+ - f^-$ , on l'obtient pour les fonctions mesurables quelconques.  $\square$

3.

prendre simplement  $\Omega = E, \mathcal{A} = \mathcal{E}, \mathbb{P} = \mu$ , et prendre pour  $X$  l'identité ( $X(x) = x$ ).

- Un peu plus compliquée est la construction d'un couple de deux v.a. indépendantes  $X$  et  $Y$ , à valeurs dans  $(E, \mathcal{E})$  et  $(F, \mathcal{F})$ , et de lois respectives  $\mu$  et  $\nu$  (probabilités données a priori sur  $E$  et  $F$ ) : on peut prendre  $\Omega = E \times F, \mathcal{A} = \mathcal{E} \otimes \mathcal{F}, \mathbb{P} = \mu \otimes \nu$ , et  $X(x, y) = x$  et  $Y(x, y) = y$ , où  $(x, y) \in E \times F$ .

Il est malheureusement bien plus compliqué, mais indispensable pour les applications (loi loi des grands nombres par exemple qu'on verra plus tard, ou même l'étude d'une suite infinie de jets de dés, etc. . .) de construire une suite infinie de v.a. indépendantes de lois données. Plus précisément, pour chaque entier  $n$  on se donne une v.a.  $X_n$  définie sur un espace de probabilité  $(\Omega, \mathcal{A}_n, \mathbb{P}_n)$ , à valeurs dans  $(E_n, \mathcal{E}_n)$  et de loi  $\mu_n$  (pour construire chaque  $X_n$  on peut opérer comme ci-dessus).

Ensuite, on pose  $\Omega = \prod_{n=1}^{\infty} \Omega_n$  (produit cartésien dénombrable),  $\mathcal{A} = \otimes_{n=1}^{\infty} \mathcal{A}_n$  où  $\otimes_{n=1}^{\infty} \mathcal{A}_n$  désigne la plus petite tribu de  $\Omega$  à laquelle appartiennent tous les ensembles suivants :  $A_1 \times A_2 \times \dots \times A_k \times \Omega_{k+1} \times \Omega_{k+2} \times \dots$ ,  $A_i \in \mathcal{A}_i, k \in \mathbb{N}^*$ .

Le théorème suivant constitue un résultat non trivial de la théorie de la Mesure et généralise le théorème de Fubini.

**Théorème 26 (dû à Cassius Ionescu-Tulcea)** *Avec les notations ci-dessus, il existe une probabilité  $\mathbb{P}$  et une seule sur  $(\Omega, \mathcal{A})$  telle que  $\mathbb{P}(A_1 \times \dots \times A_k \times \Omega_{k+1} \times \Omega_{k+2} \times \dots) = \prod_{i=1}^k \mathbb{P}_i(A_i)$  pour tous  $k \in \mathbb{N}^*$ , et  $A_i \in \mathcal{A}_i$ .*

*Maintenant, on note  $\tilde{X}_n$  l'extension naturelle de  $X_n$  à  $\Omega$ , c'est-à-dire que pour chaque  $\omega := (\omega_1, \omega_2, \dots, \omega_n, \dots) \in \Omega$  avec  $\omega_i \in \Omega_i$  on pose  $\tilde{X}_n(\omega) = X_n(\omega_n)$ .*

*Preuve:* Pour une preuve voir par exemple Theorem 16.6 (p. 150) de 4. □ 4.

**Corollaire 8 (Lomnick-Ulam)** *Les variables aléatoires  $\tilde{X}_n$  ci-dessus sont indépendantes, et chaque  $\tilde{X}_n$  a la loi  $\mu_n$ .*

*Preuve:* Soit  $B_n \in \mathcal{E}_n$ . On a  $\tilde{X}_n^{-1}(B_n) = \Omega_1 \times \dots \times \Omega_{n-1} \times X_n^{-1}(B_n) \times \Omega_{n+1} \times \Omega_{n+2} \times \dots$  et le théorème ci-dessus nous donne que pour tout  $k \in \mathbb{N}^*$  :

$$\begin{aligned} \mathbb{P}(\cap_{n=1}^k \tilde{X}_n^{-1}(B_n)) &= \mathbb{P}(X_1^{-1}(B_1) \times \dots \times X_k^{-1}(B_k) \times \Omega_{k+1} \times \dots) \\ &= \prod_{n=1}^k \mathbb{P}_n(X_n \in B_n) = \prod_{n=1}^k \mu_n(B_n). \end{aligned}$$

En particulier  $\mathbb{P}(\tilde{X}_n \in B_n) = \mu_n(B_n)$ .

P.S. Le résultat original est dans <sup>5</sup> (1934). □ 5.

**Théorème 27 (Lemme de Borel-Cantelli)** Soit  $(A_n)_{n \in \mathbb{N}^*}$  une suite d'événements sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Alors on a :

- a) Si  $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ , alors  $\mathbb{P}(A^*) = 0$ .  
 b) Si  $\mathbb{P}(A^*) = 0$  et si les  $A_n$ ,  $n \in \mathbb{N}^*$  sont indépendants, alors on a  $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ .

*Preuve:* a) : Soit  $a_n = \mathbb{P}(A_n) = \mathbb{E}[\mathbf{1}_{A_n}]$ , donc d'après b) du Théorème 19 on a  $\sum_{n=1}^{\infty} a_n < \infty$  implique  $\sum_{n=1}^{\infty} \mathbf{1}_{A_n} < \infty$  p.s. Comme  $\sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega) = \infty$  ssi  $\omega \in A^*$  d'où a) prouvé.

b) :  $\mathbb{P}(A^*) = \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \mathbb{P}(\cup_{m=n}^k A_m) = 1 - \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \prod_{m=n}^k (1 - a_m)$   
 par indépendance, donc  $\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \prod_{m=n}^k (1 - a_m) = 1$ , d'où avec  $\ln$  on obtient  $\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \sum_{m=n}^k \ln(1 - a_m) = 0$ , i.e.  $\lim_{n \rightarrow \infty} \sum_{m \geq n} \ln(1 - a_m) = 0$ , or  $|\ln(1 - x)| \geq x$  pour  $0 \leq x < 1$ , d'où  $\sum_m a_m$  converge i.e.  $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ .  $\square$

Soit maintenant des v.a.  $X_n$  définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Soient les tribus  $\mathcal{B}_n = \sigma(X_n)$ ,  $\mathcal{C}_n = \sigma(\cup_{p \geq n} \mathcal{B}_p)$  et  $\mathcal{C}_\infty = \cap_{n=1}^{\infty} \mathcal{C}_n$ .

$\mathcal{C}_\infty$  est appelée la tribu asymptotique. Un événement est dans cette tribu s'il ne dépend que du comportement de la suite  $X_n$  lorsque  $n \rightarrow \infty$  (d'où son nom) : voici des exemples de v.a.  $\mathcal{C}_\infty$ -mesurables juste après le théorème suivant.

**Théorème 28 (Loi 0 – 1)** Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de v.a. indépendantes définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , et soit  $\mathcal{C}_\infty$  la tribu asymptotique associée. Pour tout  $C \in \mathcal{C}_\infty$ , on a  $\mathbb{P}(C) \in \{0, 1\}$ .

*Preuve:* Soit  $\mathcal{D}_n = \sigma(\cup_{p < n} \mathcal{B}_p)$ . Par hypothèse les tribus  $\mathcal{C}_n$  et  $\mathcal{D}_n$  sont indépendantes, donc si  $A \in \mathcal{C}_n$  et  $B \in \mathcal{D}_n$  on a  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . Si  $A \in \mathcal{C}_\infty$  nous avons donc  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$  pour tout  $B \in \mathcal{D}_n$  donc aussi pour tout  $B \in \mathcal{D} = \sigma(\cup_n \mathcal{D}_n)$ , par le théorème des classes monotones. Cependant  $\mathcal{C}_\infty \subset \mathcal{D}$ , donc on a  $B = A \in \mathcal{C}_\infty$ , qui implique  $\mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A) = \mathbb{P}(A)^2$ , donc  $\mathbb{P}(A) \in \{0, 1\}$ .  $\square$

**Conséquence.** Si les v.a.  $X_1, X_2, \dots, X_n, \dots$  sont indépendantes, on a

1.  $\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) \text{ existe}\} \in \mathcal{C}_\infty$ , donc soit la suite  $(X_n)_n$  converge p.s. soit elle diverge p.s.
2. Chaque v.a. qui est  $\mathcal{C}_\infty$ -mesurable (pour une preuve voir le TD) est p.s. constante. En particulier pour  $Y \in \{X, \bar{X}\}$ ,  $\limsup_{n \rightarrow \infty} Y_n$ ,  $\liminf_{n \rightarrow \infty} Y_n$ , sont toutes p.s. constantes, où  $\bar{X}_n := \frac{1}{n} \sum_{p=1}^n X_p$ .

$$\{\omega \in A^*\} = \{A_m \text{ i.s.}\}, A^* = \limsup_{n \rightarrow \infty} A_n = \cap_{n=1}^{\infty} \cup_{m \geq n} A_m.$$

Nous verrons plus bas (comme cas particulier de la "loi 0 – 1") que si les  $A_n$  sont indépendants,  $\mathbb{P}(A^*) \in \{0, 1\}$ . Par suite si la série  $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$  converge on a  $\mathbb{P}(A^*) = 0$ , et si elle diverge on a  $\mathbb{P}(A^*) = 1$ .

## 0.2 Lois de probabilités sur $\mathbb{R}$

**Rappel 1** : Les intégrales (dites impropres) sur des intervalles non bornés sont définies de la façon suivante. Soit  $f$  une fonction définie sur  $[a, +\infty[$ , elle est dite localement Riemann intégrable, si elle est Riemann intégrable sur tout compact  $[a, x]$ , avec  $x \geq a$ . L'intégrale  $\int_a^\infty f(t) dt$  est dite convergente (resp. absolument convergente) si  $\int_a^x f(t) dt$  (resp.  $\int_a^x |f(t)| dt$ ) existe et est finie pour tout  $x \geq a$  et admet une limite finie lorsque  $x \rightarrow \infty$ .

**Rappel 2** :

**Théorème 29** Soit  $f$  une fonction Riemann intégrable sur  $[a, b]$ . Alors, il existe  $g \in \mathcal{L}_{\mathbb{R}}^1([a, b], \mathcal{B}([a, b]), \lambda)$  telle que  $f = g$   $\lambda$ -p.p., et telle que  $\int_{[a, b]} f(x) d\lambda(x) = \int_{[a, b]} f(x) dx$ .

En particulier, si  $f$  est en plus mesurable, alors  $f \in \mathcal{L}_{\mathbb{R}}^1([a, b], \mathcal{B}([a, b]), \lambda)$  et les deux intégrales coïncident.

Ce théorème implique que les intégrales de Lebesgue et de Riemann de fonctions admettant une limite à gauche et une limite à droite en tout point (où cela a un sens) de  $[a, b]$ , coïncident.

**Théorème 30 ( Riemann-Lebesgue (1902) )** Soit  $f$  une fonction localement Riemann intégrable, d'intégrale absolument convergente sur  $[a, +\infty[$ .

Alors, il existe  $g \in \mathcal{L}_{\mathbb{R}}^1([a, +\infty[, \mathcal{B}([a, +\infty[, \lambda)$  telle que  $f = g$   $\lambda$ -p.p., et telle que  $\int_a^\infty f(x) d\lambda(x) = \int_a^\infty f(x) dx$ .

Une fonction borélienne est intégrable au sens de Lebesgue si et seulement si  $f^+$  et  $f^-$  sont intégrables, où  $f = f^+ - f^-$  avec  $f^\pm := \max(\pm f, 0)$ .

**Définition 19** La densité d'une probabilité  $\mathbb{P}$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  est une fonction borélienne positive  $f$  qui vérifie pour tout  $x \in \mathbb{R}$  :

$$\mathbb{P}(\cdot - \infty, x] = \int_{-\infty}^x f(y) dy = \int_{\mathbb{R}} f(y) \mathbf{1}_{]-\infty, x]}(y) dy.$$

Si  $\mathbb{P} = \mathbb{P}_X$ , la loi d'une v.a. réelle  $X$ , on dit alors que  $X$  admet la densité  $f$  (on dit aussi  $f$  est une densité de  $X$ ).

**Remarque** : Comme on l'a déjà vu, il n'est pas vrai que toute probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  admette une densité, puisque dans ce cas la fonction de répartition  $F$  est continue, ce qui n'est pas vrai de toutes les fonctions de répartition. Il existe même des fonctions de répartition continues qui n'admettent pas de densité.

Un exemple de fonction Lebesgue intégrable qui n'est pas Riemann intégrable : soit  $f(x) = \mathbf{1}_{\mathbb{Q} \cap [0, 1]}(x)$  pour tout  $x \in \mathbb{R}$ . Montrer que  $\int_{\mathbb{R}} f(x) d\lambda(x) = 0$ , mais que  $f$  n'est pas Riemann intégrable sur  $[0, 1]$ .

**Théorème 31** a) Une fonction borélienne positive  $f$  sur  $\mathbb{R}$  est la densité d'une probabilité  $\mathbb{P}$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  si et seulement si elle vérifie  $\int_{\mathbb{R}} f(x) dx = 1$ . Dans ce cas elle détermine la probabilité de manière unique, et toute autre fonction borélienne positive  $g$  vérifiant  $\lambda(f \neq g) = 0$  est aussi une densité de  $\mathbb{P}$ .

b) Inversement si une probabilité admet une densité, celle-ci est déterminée de manière unique à un ensemble de mesure de Lebesgue nulle près (i.e. si  $f$  et  $g$  sont deux densités pour la même probabilité, on a  $\lambda(f \neq g) = 0$ ).

*Preuve:* Voir le Théorème 11.3 (p.78) de <sup>6</sup>. □ 6.

**Remarque :** La densité  $f$  et la fonction de répartition  $F$  étant liées par  $F(x) = \int_{-\infty}^x f(y) dy$ , on est tenté de conclure que  $F$  est dérivable et que sa dérivée vaut  $F'(x) = f(x)$ . C'est vrai en tout point  $x$  où la fonction est continue. On peut montrer (c'est un résultat difficile dû à Lebesgue, Théorème de différentiation de Lebesgue (1904)) que c'est vrai pour  $\lambda$ -presque tout  $x$ . Comme  $f$  n'est définie de manière unique qu'à un ensemble de mesure de Lebesgue nulle près, "concrètement" si on connaît  $F$  et si on sait que  $f$  existe (par exemple si  $F$  est continue partout et dérivable par morceaux), on peut prendre pour  $f$  la dérivée de  $F$  partout où celle-ci existe, et des valeurs arbitraires (par exemple 0) ailleurs.

**Corollaire 9** Soit  $X$  une v.a. réelle admettant la densité de  $f$ . Si  $g$  est une fonction borélienne sur  $\mathbb{R}$ , elle est intégrable (resp. admet une intégrale) par rapport à la loi  $\mathbb{P}_X$  de  $X$  si et seulement si la fonction production  $gf_X$  est intégrable (resp. admet une intégrale) par rapport à la mesure de Lebesgue, et on a alors

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) \mathbb{P}_X(dx) = \int_{\mathbb{R}} g(x) f_X(x) dx.$$

*Preuve:* L'égalité est satisfaite pour les fonctions  $g$  indicatrices grâce au Théorème 31, car elle se réduit alors à la forme  $\mathbb{P}(A) = \int_A f(y) dy$ . Par linéarité elle est donc vraie pour les fonctions boréliennes simples, puis par limite croissante pour les fonctions boréliennes positives. Pour les fonctions de signe quelconque, on déduit le résultat par linéarité encore, en utilisant les décompositions  $g = g^+ - g^-$  et  $gf_X = g^+ f_X - g^- f_X$ . □

Soit  $X$  une v.a. réelle de densité  $f_X$ . Soit  $Y = g(X)$  une nouvelle v.a. avec  $g$  borélienne, peut-on exprimer la densité de  $Y$  si elle existe, en termes de  $f_X$  et  $g$ ? On le peut dans les "bons cas", par exemple :

**Théorème 32** Soit  $X$  une v.a. de densité  $f_X$  et  $g$  une fonction borélienne. Soit  $Y = g(X)$ . La fonction de répartition de  $Y$  est

$$F_Y(y) = \mathbb{P}(Y \leq y) = \int_{A_y} f_X(u) dy, \text{ où } A_y = g^{-1}(] - \infty, y]).$$

Si  $F_Y$  est continue partout et dérivable sauf en un nombre fini de points, on peut utiliser le résultat ci-dessus pour obtenir la densité  $f_Y$  de  $Y$ .

**Exemple 16** Soit  $X \sim \mathcal{U}([0,1])$  et  $Y = -\frac{1}{\lambda} \ln(X)$  où  $\lambda > 0$ . Alors  $F_Y(y) = \mathbb{P}(X \geq e^{-\lambda y}) = (1 - e^{-\lambda y}) \mathbf{1}_{\mathbb{R}_+}(y)$  pour tout  $y \in \mathbb{R}$ .

Donc  $f_Y(y) = \frac{d}{dy} F_Y(y) = \lambda e^{-\lambda y} \mathbf{1}_{\mathbb{R}_+^*}(y)$  pour tout  $y \in \mathbb{R}$  et on voit que  $Y \sim \text{Exp}(\lambda)$ .

Attention : cet exemple est très simple car  $g$  est injective. Le résultat général pour  $g$  injective est donné ci-dessous, le cas non injectif étant donné plus bas.

**Corollaire 10** Soit  $X$  une v.a.r. qui admet une densité  $f_X$ . Soit  $g \in C^1(\mathbb{R}; \mathbb{R})$  (on dit aussi continûment dérivable) dont la dérivée ne s'annule pas (elle est donc strictement monotone), et soit  $h = g^{-1}$  sa fonction réciproque, qui est définie sur  $g(\mathbb{R})$ , continûment dérivable sur  $g(\mathbb{R})$ .

Alors la v.a.r.  $Y := g(X)$  admet une densité  $f_Y$  donnée par

$$f_Y = |h'| (f_X \circ h) \mathbf{1}_{g(\mathbb{R})}.$$

*Preuve:* Supposons  $g$  croissante.  $g(\mathbb{R}) = ]a, b[$  avec  $a, b \in \overline{\mathbb{R}}$ . Soit  $F_Y$  la fonction caractéristique de  $Y$ . Si  $a < y < b$  on a par la croissance de  $h$ ,  $F_Y(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(h(g(X)) \leq h(y)) = \mathbb{P}(X \leq h(y)) = F_X(h(y)) = \int_{-\infty}^{h(y)} f_X(x) dx$ . On sait par ailleurs que la dérivée de  $h$  est  $h' = \frac{1}{f'(h(x))}$  (en effet  $g$  est injective, dérivable, de dérivée non nulle, donc  $h$  dérivable), donc  $F_Y$  est dérivable sur  $]a, b[$  et  $F_Y(y) = f(h(y))h'(y) = f(h(y))|h'(y)|$ . Enfin  $F_Y(y) = 0$  si  $y \leq a$ , quand  $a > -\infty$ , et de même  $F_Y(y) = 1$  si  $y \geq b$ , quand  $b < \infty$ , donc  $F_Y$  est également dérivable, et de dérivée nulle, sur  $] -\infty, a[ \cup ]b, \infty[$ . On en déduit le résultat. Si  $g$  est décroissante, pour  $y \in g(\mathbb{R})$ , le même argument conduit à  $F_Y(y) = f(h(y))(-h'(y)) = f(h(y))|h'(y)|$ , et on conclut comme ci-dessus. □

**Corollaire 11** Soit  $X$  une v.a.r. admettant une densité  $f_X$  continue par morceaux. Soit  $g \in C^1(\mathbb{R}; \mathbb{R})$  et strictement monotone par morceaux, i.e. il existe une partition  $(I_i)_{1 \leq i \leq n}$  de  $\mathbb{R}$  constituée d'intervalles et telle que  $g \in C^1(I_i; g(I_i))$  **-difféomorphisme**,  $I_i'$  étant l'intervalle ouvert ayant les mêmes extrémités que  $I_i$ . Pour chaque  $i$  on note  $h_i = (g|_{I_i})^{-1}$  et  $\Lambda_i = g(I_i)$ . Alors la v.a.  $Y := g(X)$  admet une densité  $f_Y$ , définie par

$$f_Y = \sum_{i=1}^n |h_i'| (f_X \circ h_i) \mathbf{1}_{\Lambda_i}.$$

*Preuve:* La preuve, analogue à celle du corollaire précédent, est laissée. Cette méthode de démonstration utilise la continuité par

**Rappel :** Par définition on dit que  $f : I \rightarrow J$  est  $C^k(I; J)$ -difféomorphisme si  $f$  est une bijection de  $I$  sur  $J$ , et si  $f$  et  $f^{-1}$  sont de classe  $C^k$ , i.e.  $f^i$  existe pour tout  $i \leq k$  et  $f^k$  est continue.

Tout simplement l'hypothèse du corollaire est :

$g$  est  $C^1(\mathbb{R}; g(\mathbb{R}))$ -difféomorphisme, en effet une propriété appelée Caractérisation des  $C^k$ -difféomorphismes dit que  $f \in C^k(I; f(I))$ -difféomorphisme ssi  $f \in C^k$  sur  $I$  et  $f' \neq 0$  sur  $I$ .

morceaux de  $f_X$ , mais le résultat reste vrai quand  $f_X$  est borélienne.

□



## 7. Fonctions caractéristiques et quelques propriétés

**Définition 20** Pour une v.a.r.  $X$  on définit  $\mathbb{E}[e^{iX}] := \mathbb{E}[\cos(X)] + i\mathbb{E}[\sin(X)]$  avec  $i^2 = -1$ .

**Définition 21** Soit  $\mu$  une mesure de probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Sa transformée de Fourier, notée  $\widehat{\mu}$ , est la fonction sur  $\mathbb{R}$  définie  $\widehat{\mu}(u) = \int_{\mathbb{R}} e^{iux} d\mu(x)$ .

**Remarque :**  $\widehat{\mu}$  est bien définie puisque  $|e^{iux}| \leq 1$  et  $1 \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ .

**Définition 22** Soit  $X$  une v.a.r. Sa fonction caractéristique  $\phi_X$  est la fonction sur  $\mathbb{R}$  définie par  $\phi_X(u) = \mathbb{E}[e^{iuX}]$ . En d'autres termes  $\phi_X(u) = \int_{\mathbb{R}} e^{iux} d\mathbb{P}_X(x) = \widehat{\mathbb{P}_X}(u)$ .

Ainsi on remarque que la fonction caractéristique de  $X$  est la transformée de Fourier de sa loi  $\mathbb{P}_X$ .

**Théorème 33** Soit  $\mu$  une probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . La fonction  $\widehat{\mu}$  est bornée (de module inférieur ou égale à 1), continue, et vérifie  $\widehat{\mu}(0) = 1$ .

*Preuve:* Voir le Théorème 13.1 de 7 (page 111). □ 7.

**Définition 23** On dit qu'une v.a.r.  $X$  admet des moments d'ordre jusqu'à l'ordre  $m$  si  $X \in L^k(\Omega)$  pour tout  $1 \leq k \leq m$ .

**Théorème 34** Soit  $X$  une v.a.r. admettant des moments jusqu'à l'ordre  $m$ . Alors  $\phi_X$  est  $m$  fois dérivable et

$$\phi_X^{(m)}(u) = i^m \mathbb{E}[X^m e^{iuX}].$$

**Remarque et Application :**

Si  $X \in L^1(\Omega)$  alors  $\mathbb{E}[X] = -i\phi_X'(0)$ . Si  $X \in L^2(\Omega)$  alors  $\mathbb{E}[X^2] = -\phi_X''(0)$ .

**Théorème 35** Soit  $X$  une v.a.r. et  $a \in \mathbb{R}$  avec  $b \in \mathbb{R}$ , alors  $\phi_{a+bX}(u) = e^{iua} \phi_X(bu)$  pour tout  $u \in \mathbb{R}$ .

**Théorème 36 (Théorème d'unicité)** *La transformée de Fourier  $\hat{\mu}$  d'une mesure de probabilité  $\mu$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  caractérise  $\mu$  : i.e. que si deux mesures de probabilités sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  admettent la même transformée de Fourier, alors elles sont égales.*

*Preuve:* Voir le Théorème 14.1. de <sup>8</sup> (page 117) □ 8.

**Corollaire 12** *Deux v.a.r.  $X$  et  $Y$  sont indépendantes si et seulement si  $\mathbb{E}[e^{i(uX+vY)}] = \phi_X(u)\phi_Y(v)$  pour tous  $u, v \in \mathbb{R}$ .*

*Preuve:* Si  $X$  et  $Y$  sont indépendantes, alors c'est vrai par indépendance. L'autre sens sera démontré en probabilité avancées lorsque nous manipulerons les vecteurs aléatoires. □

## 8. Sommes de variables aléatoires indépendantes et Variables aléatoires gaussiennes

### 0.2.1 Somme de v.a. indépendantes

Une part importante des applications des probabilités découle des propriétés des sommes de v.a.r. indépendantes. Un exemple simple apparaît en statistique : si nous répétons  $n$  fois la même expérience indépendamment et si nous appelons  $X_j$  le résultat de la  $j$ -ième expérience, la "valeur moyenne" est donnée par  $\bar{x} = \frac{1}{n} \sum_{j=1}^n X_j$ . La v.a.  $\bar{x}$  est alors appelée un estimateur de la moyenne  $\mu$  de la loi commune des  $X_j$ . La statistique étudiée quand (et comment)  $\bar{x}$  converge vers  $\mu$  quand  $n \rightarrow \infty$ . Même une fois qu'on sait que  $\bar{x}$  tend vers  $\mu$ , on a aussi besoin de savoir quelle taille donner à  $n$  pour être raisonnablement sûr que  $\bar{x}$  est proche de la vraie valeur  $\mu$ , qui est en général inconnue. D'autres problèmes difficiles se posent aussi naturellement : quelle est la loi de  $\bar{x}$ ? Si on ne peut la déterminer, peut-on au moins l'approximer? Quelle taille doit avoir  $n$  pour que cette approximation soit bonne? Si on dispose d'informations complémentaires sur  $\mu$ , comment les utiliser pour améliorer l'estimateur  $\bar{x}$ ? Bien évidemment, avant de commencer à résoudre les plus simples de ces questions, il nous faut étudier les sommes de v.a. indépendantes. Voici un premier théorème

**Théorème 37** Soient  $X$  et  $Y$  deux v.a. réelles. La loi  $\mu_Z$  de  $Z := X + Y$  est le produit de convolution des lois  $\mu_X$  et  $\mu_Y$  de  $X$  et  $Y$ , qui est définie par

$$\mu_Z(A) = (\mu_X * \mu_Y)(A) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_A(x+y) d\mu_X(x) d\mu_Y(y),$$

i.e.  $\mu_Z(A) = \int_{\mathbb{R}^2} g(z) d(\mu_X \otimes \mu_Y)(z)$ , où  $g(z) := \mathbf{1}_A(z_1 + z_2)$  pour tout  $z \in \mathbb{R}^2$ .

*Preuve:* preuve admise, sera donnée en probabilités avancées.  $\square$

**Remarque :** Donc cela monter que pour  $h : \mathbb{R} \rightarrow \mathbb{R}$  borélienne et  $Z =$

$X + Y$ , avec  $X$  et  $Y$  indépendantes :  $\mathbb{E}[h(Z)] = \int_{\mathbb{R}^2} f(z) d(\mu_X \otimes \mu_Y)(z)$   
où  $f(z) := h(z_1 + z_2)$  pour tout  $z \in \mathbb{R}^2$ .

**Théorème 38** Soient  $X$  et  $Y$  deux v.a. réelles indépendantes, et  $Z := X + Y$ .  
La fonction caractéristique  $\phi_Z$  est le produit de  $\phi_X$  et  $\phi_Y$ , i.e.

$$\phi_Z(u) = \phi_X(u)\phi_Y(u) \forall u \in \mathbb{R}.$$

*Preuve:* Prenez  $h(u) = e^{iuz}$  dans la preuve précédente. □

**Théorème 39** Soient  $X$  et  $Y$  deux v.a.r. indépendantes, et  $Z = X + Y$ .  
Alors on a

a) Si  $X$  admet une densité  $f_X$ , alors  $Z$  admet une densité  $f_Z$  donnée par

$$f_Z(z) = \int_{\mathbb{R}} f_X(z - y) d\mu_Y(y).$$

b) Si de plus  $Y$  admet une densité  $f_Y$ , alors

$$f_Z(z) = \int_{\mathbb{R}} f_X(z - y)f_Y(y) dy = \int_{\mathbb{R}} f_X(x)f_Y(z - x) dx.$$

*Preuve:* Pour montrer b), supposons a) démontré, alors on a  $f_Z(z) = \int_{\mathbb{R}} f_X(z - y) d\mu_Y(y)$ , comme  $d\mu_Y(y) = f_Y(y) dy$  d'où la première égalité de b). La seconde s'obtient en échangeant les rôles de  $X$  et  $Y$ .

Pour montrer a), d'après le théorème 37 on a pour tout borélien  $A \in \mathcal{B}(\mathbb{R})$  :

$\mu_Z(A) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_A(x + y) d\mu_X(x) d\mu_Y(y)$ , i.e.  $\mu_Z(A) = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \mathbf{1}_A(x + y) f_X(x) dx \right) d\mu_Y(y)$ , donc en posant  $z = x + y$ , on a  $dz = dx$  et on obtient  $\mu_Z(A) = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \mathbf{1}_A(z) f_X(z - y) dz \right) d\mu_Y(y)$  et avec le théorème de Fubini, on obtient la réponse. □

**Théorème 40** Soient  $X$  et  $Y$  deux v.a.r. indépendantes, de carré intégrable, alors on a  $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$ .

*Preuve:* Facile □

### 0.2.2 Variables aléatoires gaussiennes

Rappelons que la loi normale que l'on désigne par la notation  $\mathcal{N}(\mu, \sigma^2)$ , où  $\mu \in \mathbb{R}$  et  $\sigma^2 > 0$ , est la probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  de densité  $f$  définie par  $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ .

Il est commode d'étendre la classe des lois normales de façon à inclure le cas où  $\sigma^2 = 0$ , de la manière suivante : on note  $\mathcal{N}(\mu, 0)$  la loi égale à la masse de Dirac au point  $\mu$ , i.e. la loi d'une v.a. qui est

Remarquons que  $\phi_Z(u) = \phi_X(u)\phi_Y(u)$  pour tout  $u \in \mathbb{R}$  où  $Z := X + Y$ , n'implique pas que  $X$  et  $Y$  sont indépendantes.

p.s. égale à la constante  $\mu$ . Cette loi n'admet pas de densité, aussi parle-t-on parfois de loi normale "dégénérée". La loi  $\mathcal{N}(0,1)$  s'appelle la loi normale réduite, ou standard.

Si  $X$  est une v.a.r. de loi  $\mathcal{N}(\mu, \sigma^2)$ , sa fonction caractéristique  $\phi_X$  est  $\phi_X(u) = e^{iu\mu - \frac{1}{2}\sigma^2 u^2}$  pour tout  $u \in \mathbb{R}$ .

Rappelons que  $\mathbb{E}[X] = \mu$  et  $\text{Var}(X) = \sigma^2$ .

**Théorème 41** *La fonction caractéristique d'une variable aléatoire  $X$  qui suit la loi  $\mathcal{N}(\mu, \sigma^2)$  est donnée par  $\phi_X(u) = e^{iu\mu - \frac{1}{2}u^2\sigma^2}$  pour tout  $u \in \mathbb{R}$ .*

*Preuve:* à faire via le Théorème de résidu de Cauchy. □

**Théorème 42** *Soit  $X \sim \mathcal{N}(\mu, \sigma^2)$  alors  $X$  et  $\mu + \sigma Y$  où  $Y \sim \mathcal{N}(0,1)$  suivent la même loi.*

*Preuve:* En effet, d'après le Théorème d'unicité, il suffit de prouver que  $\phi_X(u) = \phi_{\mu + \sigma Y}(u)$  pour tout  $u \in \mathbb{R}$ . Or pour tout  $u \in \mathbb{R}$ , on a  $\phi_{\mu + \sigma Y}(u) = \mathbb{E}[e^{iu(\mu + \sigma Y)}] = e^{iu\mu} \mathbb{E}[e^{iu\sigma Y}] = e^{iu\mu} e^{-\frac{1}{2}(u^2\sigma^2)}$ , d'où la réponse. □

**Définition 24** *On dit que deux v.a.r.  $X$  et  $Y$  de carré intégrable sont non corrélées si  $\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = 0$ .*



## 9. Convergence des variables aléatoires, convergence en loi et applications avec fonctions caractéristiques

**Définition 25** On dit qu'une suite  $(X_n)_{n \geq 1}$  de v.a.r. converge presque sûrement vers une v.a.r.  $X$  si l'ensemble  $N$  des  $\omega$  tels que la suite numérique  $(X_n(\omega))_{n \geq 1}$  ne converge pas vers  $X(\omega)$  vérifie  $\mathbb{P}(N) = 0$ . Rappelons qu'on dit que  $N$  est négligeable.

Noter qu'on a aussi  $N^c = \Lambda = \{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}$  vérifie  $\mathbb{P}(\Lambda) = 1$ . On note la convergence presque sûre ainsi :

$$\lim_{n \rightarrow \infty} X_n = X \text{ p.s. ou } X_n \rightarrow X \text{ p.s.}$$

Il apparaît que la convergence presque sûre est parfois forte dans certaines situations, et ce pourquoi nous introduisons deux autres mode de convergence.

**Définition 26** On dit qu'une suite  $(X_n)_{n \geq 1}$  de v.a.r. converge dans  $L^p(\Omega)$ , où  $p \in [1, \infty[$ , vers une v.a.r.  $X$  si les  $X_n$  sont dans  $L^p(\Omega)$ , et si on a

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0.$$

On dit aussi que  $(X_n)_n$  converge en moyenne d'ordre  $p$  vers  $X$ , et on écrit  $X_n \xrightarrow{L^p(\Omega)} X$

Les cas les plus importants sont  $p = 1$  (convergence "en moyenne") et  $p = 2$  (convergence "en moyenne quadratique"). Quand  $p = 1$ , on a  $|\mathbb{E}[X_n - X]| \leq \mathbb{E}[|X_n - X|]$  et  $|\mathbb{E}[|X_n|] - \mathbb{E}[|X|]| \leq \mathbb{E}[|X_n - X|]$  parce que  $||x| - |y|| \leq |x - y|$ . Donc  $X_n \xrightarrow{L^1(\Omega)} X$  implique  $\mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$  et  $\mathbb{E}[|X_n|] \rightarrow \mathbb{E}[|X|]$ .

De même, quand  $X_n \xrightarrow{L^p(\Omega)} X$  pour  $p \in ]1, \infty[$ , on peut montrer que  $\mathbb{E}[|X_n|^p] \rightarrow \mathbb{E}[|X|^p]$ .

**Définition 27** On dit qu'une suite  $(X_n)_{n \geq 1}$  de v.a.r. converge en probabilité vers une v.a.r.  $X$  si

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0,$$

ou plus simplement  $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$  quand  $n \rightarrow \infty$ . On écrit alors  $X_n \xrightarrow{\mathbb{P}} X$

Une autre manière d'écrire les choses est la suivante :

$X_n$  converge en probabilité vers  $X$  lorsque,

$$\forall \varepsilon, \delta > 0, \exists N_{\varepsilon, \delta}, n \geq N \Rightarrow \mathbb{P}(|X_n - X| > \varepsilon) < \delta.$$

Avant d'établir les rapports entre les différents modes de convergence on va donner un critère simple de convergence en probabilité.

**Théorème 43** On a  $X_n \xrightarrow{\mathbb{P}} X$  si et seulement si  $\lim_{n \rightarrow \infty} \mathbb{E} \left[ \frac{|X_n - X|}{1 + |X_n - X|} \right] = 0$ .

*Preuve:* On ne nuit pas à la généralité en supposant que  $X = 0$ , de sorte qu'il nous suffit de montrer que  $X_n \xrightarrow{\mathbb{P}} X$  équivaut à  $\lim_{n \rightarrow \infty} \mathbb{E} \left[ \frac{|X_n|}{1 + |X_n|} \right] = 0$ . Supposons d'abord que  $X_n \xrightarrow{\mathbb{P}} X$ . Pour tout  $\varepsilon > 0$ , on a  $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n| > \varepsilon) = 0$ . Mais  $\frac{|X_n|}{1 + |X_n|} \leq \frac{|X_n|}{1 + |X_n|} \mathbf{1}_{|X_n| > \varepsilon} + \varepsilon \mathbf{1}_{|X_n| \leq \varepsilon} \leq \mathbf{1}_{|X_n| > \varepsilon} + \varepsilon$ , donc  $\mathbb{E} \left[ \frac{|X_n|}{1 + |X_n|} \right] \leq \mathbb{P}(|X_n| > \varepsilon) + \varepsilon$ , en prenant les limites on arrive à  $\limsup_{n \rightarrow \infty} \mathbb{E} \left[ \frac{|X_n|}{1 + |X_n|} \right] \leq \varepsilon$ ; comme  $\varepsilon$  est arbitrairement petit il vient  $\lim_{n \rightarrow \infty} \mathbb{E} \left[ \frac{|X_n|}{1 + |X_n|} \right] = 0$ .

Supposons à l'inverse que  $\lim_{n \rightarrow \infty} \mathbb{E} \left[ \frac{|X_n|}{1 + |X_n|} \right] = 0$ . La fonction  $f(x) = x/(1+x)$  est strictement croissante, donc  $f(\varepsilon) \mathbf{1}_{|X_n| > \varepsilon} \leq f(|X_n|) \mathbf{1}_{|X_n| > \varepsilon} \leq f(|X_n|)$ . En prenant les espérances, puis les limites, on arrive à  $f(\varepsilon) \lim_{n \rightarrow \infty} \mathbb{P}(|X_n| > \varepsilon) \leq \lim_{n \rightarrow \infty} \mathbb{E} [f(|X_n|)] = 0$ . Comme  $\varepsilon > 0$  est fixé, on conclut que  $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n| > \varepsilon) = 0$ .  $\square$

Le théorème suivant nous indique que la convergence en probabilité est le mode convergence le plus faible introduit jusqu'à présent.

**Théorème 44** Soit  $(X_n)_{n \geq 1}$  une suite de v.a.r.

- a) Si  $X_n \xrightarrow{L^p(\Omega)} X$ , alors on a  $X_n \xrightarrow{\mathbb{P}} X$ ,  
 b) Si  $X_n \rightarrow X$  p.s., alors on a  $X_n \xrightarrow{\mathbb{P}} X$ .

Ce théorème nous dit que  $X_n \xrightarrow{\mathbb{P}} X$  ssi  $\mathbb{E}[f(|X_n - X|)] \rightarrow 0$  pour la fonction  $f(x) = |x|/(1+|x|)$ . Un examen soigneux de cette preuve montre que la même équivalence reste vraie pour tout fonction  $f$  sur  $[0, \infty[$  qui est bornée, strictement croissante, continue, nulle en 0. Par exemple,  $X_n \xrightarrow{\mathbb{P}} X$  ssi  $\mathbb{E}[|X_n - X| \wedge 1] \rightarrow 0$ , et aussi ssi  $\mathbb{E}[\arctan(|X_n - X|)] \rightarrow 0$ .

*Preuve:* Pour a) :  $\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^p} \mathbb{E}[|X_n - X|^p] \rightarrow 0$  quand  $n \rightarrow \infty$  (pour  $\varepsilon > 0$  fixé), d'où le résultat.

Pour b) : Posons  $f(x) = |x|/(1+|x|)$ , alors on a  $f(|X_n - X|) \leq 1$ , et donc  $\lim_{n \rightarrow \infty} \mathbb{E}[f(|X_n - X|)] = \mathbb{E}[\lim_{n \rightarrow \infty} f(|X_n - X|)] = \mathbb{E}[0] = 0$ , grâce au

théorème de convergence dominée de Lebesgue.  $\square$  Attention les réciproques sont fausses en général.

Par contre il existe des réciproques "partielles" :

**Théorème 45** Supposons que  $X_n \xrightarrow{\mathbb{P}} X$ . Il existe alors une sous-suite  $(n_k)_{k \in \mathbb{N}^*} \in \mathbb{N}^* \mathbb{N}^*$  telle que  $X_{n_k} \rightarrow X$  p.s. lorsque  $k \rightarrow \infty$ .

*Preuve:*  $f(x) := |x|/(1 + |x|)$ , on a  $\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n - X)] = 0$  par un théorème précédent et donc on peut trouver suite d'entiers  $n_k$  telle que  $\mathbb{E}[f(X_{n_k} - X)] < 1/2^k$ . On a alors  $\sum_{k=1}^{\infty} \mathbb{E}[f(X_{n_k} - X)] < \infty$ , donc  $\sum_{k=1}^{\infty} f(X_{n_k} - X) < \infty$  p.s. et donc  $\lim_{n \rightarrow \infty} |X_{n_k} - X| = 0$  p.s.  $\square$

**Théorème 46** Supposons que  $X_n \xrightarrow{\mathbb{P}} X$  et que  $|X_n| \leq Y$  pour tout  $n$ , où  $Y \in L^p(\Omega)$  pour  $p \in [1, \infty[$ . On a alors  $X \in L^p(\Omega)$  et  $X_n \xrightarrow{L^p(\Omega)} X$

*Preuve:*  $X_n \in L^p$  et pour tout  $\varepsilon > 0$ ,  $\mathbb{P}(|X| > Y + \varepsilon) \leq \mathbb{P}(|X - X_n| > \varepsilon)$  et donc en passant à la limite on obtient  $\mathbb{P}(|X| > Y + \varepsilon) \leq \lim_n \dots = 0$ , ceci étant vrai pour tout  $\varepsilon > 0$  on obtient  $\mathbb{P}(|X| > Y) \leq \lim_{m \rightarrow \infty} \mathbb{P}(|X| > Y + \frac{1}{m}) = 0$ , et donc  $|X| \leq Y$  p.s. et donc  $X \in L^p(\Omega)$ . Supposons maintenant que  $(X_n)_n$  ne converge pas vers  $X$  dans  $L^p$ . Il existe alors une suite d'entier strictement croissante  $n_k$  et un  $\varepsilon > 0$  tels que  $\mathbb{E}[|X_{n_k} - X|^p] \geq \varepsilon$  pour tout  $k$ . La sous-suite  $(X_{n_k})_k$  converge évidemment vers  $X$  en probabilité, donc elle admet (d'après le théorème précédent) une sous-sous-suite  $(X_{n_{k_j}})_j$  qui converge p.s. vers  $X$ . Donc  $(X_{n_{k_j}} - X)_j$  converge p.s. vers 0, tout en restant majorées en valeurs absolue par  $2Y$  : une application du théorème de Lebesgue donne alors que  $\mathbb{E}[|X_{n_{k_j}} - X|^p] \rightarrow 0$ , en contradiction avec le fait que  $\mathbb{E}[|X_{n_k} - X|^p] \geq \varepsilon$  pour tout  $k$ .  $\square$

Le théorème suivant est facile, mais utile

**Théorème 47** Soit  $f$  une fonction continue.

a) Si  $X_n \rightarrow X$  p.s., alors  $f(X_n) \rightarrow f(X)$  p.s.

b) Si  $X_n \xrightarrow{\mathbb{P}} X$ , alors  $f(X_n) \xrightarrow{\mathbb{P}} f(X)$ .

*Preuve:* voir la page 153 de<sup>9</sup>.

$\square$  9.

## 0.2.3 convergence en loi

**Définition 28** Soit  $\mu_n$  et  $\mu$  deux probabilités sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . On dit que  $(\mu_n)_n$  converge étroitement vers  $\mu$  si  $(\int_{\mathbb{R}} f(x) d\mu_n(x))_n$  converge vers  $\int_{\mathbb{R}} f(x) d\mu(x)$  pour toute fonction  $f$  réelle continue et bornée.

**Définition 29** Soit  $(X_n)_n$  et  $X$  des v.a.r. On dit que  $(X_n)_n$  converge en loi vers  $X$  si  $(\mathbb{P}_{X_n})_n$  converge étroitement vers  $\mathbb{P}_X$ , et on écrit  $X_n \xrightarrow{\text{Loi}} X$ .

Remarque : les  $X_n$  et  $X$  ne sont pas forcément définies sur un même espace de probabilité, puisque seul leurs lois jouent le rôle dans la définition.

**Théorème 48** Soit  $(X_n)_n$  et  $X$  des v.a.r. On a  $X_n \xrightarrow{\text{Loi}} X$  ssi

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$$

pour toute fonction réelle  $f$  continue et bornée sur  $\mathbb{R}$ .

*Preuve:* Il suffit de combiner les deux définitions précédentes et d'observer que  $\mathbb{E}[f(X_n)] = \int_{\mathbb{R}} f(x) d\mathbb{P}_{X_n}(x)$  et  $\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) d\mathbb{P}_X(x)$ .

□

Pour avoir la convergence p.s. ou dans  $L^p$  ou en probabilité les v.a.  $X_n$  et  $X$  doivent être définies sur le même espace de probabilité. Donc a priori la convergence en loi n'est pas comparable aux autres. On a cependant le résultat suivant :

**Théorème 49** Soient  $(X_n)_n$  et  $X$  des v.a.r. toutes définies sur le même espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$ . Si  $(X_n)_n$  converge vers  $X$  en probabilité, alors  $X_n$  converge aussi vers  $X$  en loi.

*Preuve:* Soit  $f$  une fonction réelle bornée et continue sur  $\mathbb{R}$ . Le théorème 47 implique que  $f(X_n)$  converge vers  $f(X)$  en probabilité également. Comme  $f$  est bornée, cette convergence a aussi lieu dans  $L^1$  (théorème 46). Donc  $\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ , et le théorème précédent donne le résultat. □

**Théorème 50** Soit  $(X_n)_n$  et  $X$  des v.a. définies sur le même espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$ . Si  $(X_n)_n$  converge en loi vers  $X$  et si de plus  $X$  est p.s. constante, alors  $(X_n)_n$  converge vers  $X$  en probabilité.

*Preuve:* Supposons que  $X = a$  p.s. La fonction  $f(x) = |x - a|/(1 + |x - a|)$  étant bornée et continue, on a  $\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n - a)] = 0$ , et donc

$X_n \xrightarrow{\mathbb{P}} X$  en vertu du théorème 43 □

**Théorème 51** Soient  $(X_n)_n$  et  $X$  des v.a.r.

- a) Si  $X_n \xrightarrow{\text{Loi}} X$  alors  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$  pour tout  $x$  appartenant à un sous-ensemble dense de  $\mathbb{R}$ , donné par  $D = \{x \in \mathbb{R}, F_X(x^-) = F_X(x)\}$ , i.e.  $F_X$  est continue pour  $x \in D$ , qui est appelé l'ensemble des points de continuité de la fonction  $F_X$ .
- b) Supposons que  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$  pour tout  $x$  appartenant à un sous-ensemble dense de  $\mathbb{R}$ . Alors  $X_n \xrightarrow{\text{Loi}} X$ .

*Preuve:* Voir la page 159 de <sup>10</sup>. Remarquez que  $D \subset \mathbb{R}$  est dense dans  $\mathbb{R}$  puisque  $D^c$  est dénombrable. □ 10.

**Théorème 52 (Théorème de Continuité de P. Lévy)** Soit  $(\mu_n)_{n \geq 1}$  une suite de probabilités sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , et  $(\hat{\mu}_n)_n$  la suite de leurs transformées de Fourier (ou fonction caractéristiques).

- a) Si  $(\mu_n)_n$  converge étroitement vers une probabilité  $\mu$  (ou  $X_n \xrightarrow{\text{Loi}} X$ , i.e.  $(\mathbb{P}_{X_n})_n$  converge étroitement vers  $\mathbb{P}_X$ ) alors  $(\hat{\mu}_n)_n$  converge simplement vers la transformée de Fourier  $\hat{\mu}$  (ou  $\phi_{X_n} \rightarrow \phi_X$ )
- b) Si  $(\hat{\mu}_n)_n$  converge simplement vers une fonction  $f$ , et si de plus cette fonction est continue au point 0, alors il existe une probabilité  $\mu$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  t.q.  $\hat{\mu} = f$ , et  $(\mu_n)_n$  converge étroitement vers  $\mu$ .

*Preuve:* voir le Théorème 19.1. qui se trouve à la page 173 de <sup>11</sup>. □ 11.



## 10. Loi des grands nombres et le théorème-limite central

### 0.2.4 La loi des grands nombres

**Théorème 53 (Loi forte des grands nombres)** Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de v.a.r. i.i.d. de carré intégrables avec  $\mu := \mathbb{E}[X_1]$  et  $\sigma^2 = \sigma_{X_1}^2 < \infty$ . Soit  $S_n := \sum_{i=1}^n X_i$ . On a alors

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu \text{ p.s. et dans } L^2(\Omega).$$

*Preuve:* Voir le Théorème 20.1. de la page 179 de<sup>12</sup>. □ 12.

**Théorème 54 (Loi des grands nombres de Kolmogorov)** Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de v.a.r. i.i.d. et soit  $\mu \in \mathbb{R}$ .

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu \text{ p.s.} \Leftrightarrow \mathbb{E}[X_1] = \mu.$$

Dans ce cas, la convergence a aussi lieu dans  $L^1(\Omega)$ .

**Théorème 55 (Théorème-limite central-version 1)** Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de v.a.r. i.i.d. avec  $\mathbb{E}[X_1] = \mu$  et  $\text{Var}(X_1) = \sigma^2 > 0$ .

Alors  $\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right)_n$  converge en loi vers une v.a.  $X \sim \mathcal{N}(0, 1)$ .

*Preuve:* Voir Théorème 21.1. de la page 187 de<sup>13</sup>. □ 13.

**Théorème 56 (Théorème-limite central-version 1)** Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de v.a.r. indépendantes. Supposons que  $\mathbb{E}[X_n] = 0$  pour tout  $n \in \mathbb{N}^*$ . Supposons que  $\sum_j \mathbb{E}[|X_j|^{2+\varepsilon}] < \infty$  et  $\sum_{n=1}^{\infty} \sigma_n^2 = \infty$  pour un  $\varepsilon > 0$ . Alors

$$\frac{S_n}{\sqrt{\sum_{j=1}^n \sigma_j^2}} \xrightarrow[n \rightarrow \infty]{\text{Loi}} \text{vers une v.a. } X \sim \mathcal{N}(0, 1).$$



## *Bibliographie*