

Licence 3ème année (L3 S6)  
UFR Sciences  
Université Picardie Jules Verne, Amiens

année 2008-2009

# Probabilités

Barbara SCHAPIRA

4 juin 2009

# Table des matières

<b>1</b>	<b>Probabilités et variables aléatoires</b>	<b>4</b>
1.1	Espaces de probabilité . . . . .	4
1.1.1	Le paradoxe de Banach-Tarski . . . . .	4
1.1.2	Tribus . . . . .	4
1.1.3	Mesures de probabilité . . . . .	5
1.1.4	Le vocabulaire des probabilités . . . . .	5
1.1.5	Rappels à connaître du $S_4$ . . . . .	6
1.2	Retour sur les classes monotones . . . . .	7
1.3	Variables aléatoires . . . . .	8
1.3.1	Variable aléatoire, loi d'une variable aléatoire . . . . .	8
1.3.2	Fonction de répartition d'une variable aléatoire à valeurs dans $\mathbb{R}$ (ou $\mathbb{N}$ , $\mathbb{Z}$ ...) . . . . .	10
1.4	Lois et variables aléatoires classiques . . . . .	11
1.4.1	Lois discrètes . . . . .	12
1.4.2	Lois continues . . . . .	13
1.5	Variables aléatoires à densité . . . . .	15
1.5.1	Définitions et premières propriétés . . . . .	15
1.5.2	Changement de variables . . . . .	17
<b>2</b>	<b>Les grands outils probabilistes</b>	<b>19</b>
2.1	Espérance, variance, écart-type . . . . .	19
2.1.1	Définition . . . . .	19
2.1.2	Exemples . . . . .	20
2.1.3	Variance, écart-type . . . . .	20
2.1.4	Corrélation de deux variables aléatoires . . . . .	21
2.1.5	Moments . . . . .	21
2.1.6	Inégalités classiques . . . . .	22
2.2	Fonctions génératrices . . . . .	22
2.3	Fonctions caractéristiques . . . . .	23
2.3.1	Fonction caractéristique d'une variable aléatoire réelle . . . . .	23
2.3.2	Fonction caractéristique d'une variable aléatoire vectorielle . . . . .	27
<b>3</b>	<b>Indépendance</b>	<b>29</b>
3.1	Produit de lois . . . . .	29
3.2	Couple de variables aléatoires, densité marginale . . . . .	30
3.3	Variables aléatoires indépendantes . . . . .	30
3.3.1	Définitions . . . . .	30
3.3.2	Variables aléatoires indépendantes et produit de lois . . . . .	31
3.4	Le lemme de Borel-Cantelli . . . . .	32
3.5	Espérance et indépendance . . . . .	33
3.6	Indépendance et fonctions génératrices . . . . .	34
3.7	Indépendance et fonctions caractéristiques . . . . .	35
3.8	Somme de variables aléatoires indépendantes. . . . .	35

<b>4</b>	<b>Convergences et théorèmes limites</b>	<b>37</b>
4.1	Convergences de variables aléatoires . . . . .	37
4.1.1	Convergence presque sûre . . . . .	37
4.1.2	Convergence en probabilité . . . . .	38
4.1.3	Convergence de mesures, Convergence en loi . . . . .	38
4.1.4	Convergence dans $L^p$ . . . . .	41
4.2	Loi faible et forte des grands nombres . . . . .	42
4.2.1	Loi faible des grands nombres . . . . .	42
4.2.2	Loi forte des grands nombres . . . . .	43
4.3	Théorème limite central . . . . .	45

## L'alphabet grec

Beaucoup d'entre vous n'ont jamais fait de grec ancien. Or les mathématicien-ne-s usent et abusent des lettres grecques. Les voici donc, dans l'ordre, avec minuscule, majuscule, et nom.

$\alpha$ ,  $A$ , Alpha

$\beta$ ,  $B$ , Bêta

$\gamma$ ,  $\Gamma$ , Gamma

$\delta$ ,  $\Delta$ , Delta

$\varepsilon$  (ou parfois  $\epsilon$ ),  $E$ , epsilon

$\zeta$ ,  $Z$ , zêta (prononcer "dzeta")

$\eta$ ,  $H$ , êta

$\theta$   $\Theta$  Thêta

$\iota$   $I$  iota

$\kappa$   $K$  Kappa

$\lambda$   $\Lambda$  lambda

$\mu$   $M$  mu

$\nu$   $V$  nu

$\xi$   $\Xi$  xi

$o$   $O$  omicron (prononcer "omicrone")

$\pi$   $\Pi$  pi

$\rho$   $P$  rho

$\sigma$   $\Sigma$  Sigma

$\tau$   $T$  tau

$v$   $U$  upsilon

$\varphi$  (ou parfois  $\phi$ )  $\Phi$  Phi

$\chi$   $X$  Chi (prononcer "ki")

$\psi$   $\Psi$  psi

$\omega$   $\Omega$  omega.

# Chapitre 1

## Probabilités et variables aléatoires

### 1.1 Espaces de probabilité

#### 1.1.1 Le paradoxe de Banach-Tarski

Soit  $X$  un ensemble. Idéalement, on voudrait dire a priori qu'une mesure  $\mu$  sur  $X$ , ce serait une application  $\mu : \mathcal{P}(X) \rightarrow \overline{\mathbb{R}}_+$  définie sur l'ensemble de toutes les parties de  $X$ , additive ( $\mu(A \cup B) = \mu(A) + \mu(B)$  si  $A$  et  $B$  sont disjoints, avec  $\mu(\emptyset) = 0$ ). (L'intuition de ce qu'est une mesure peut être guidée par la notion de volume et/ou de probabilité.)

Cela existe. Par exemple, sur  $\mathbb{R}$ , pour toute partie  $A \subset \mathbb{R}$ , on peut définir  $\mu_n(A) = \frac{1}{n} \# \{0, \dots, n-1\} \cap A$ . On peut passer à la limite et obtenir ce qu'on appelle une "moyenne"  $\mu$  qui est additive, positive et vérifie  $\mu(\mathbb{R}) = 1$ . Toutefois, elle vérifie aussi  $\mu(A) = 0$  pour toute partie  $A$  bornée. Ennuyeux... On aimerait dire quand même que  $\mu(\mathbb{R}) = \sum_{n \in \mathbb{Z}} \mu([n, n+1])$ .

C'est ce qu'on appelle la  $\sigma$ -additivité. Le problème est alors différent. On ne peut alors pas supposer que toutes les parties d'un ensemble sont mesurables. En effet, c'est le paradoxe suivant, dit de *Banach-Tarski*. On peut prendre une sphère disons de rayon 1, la couper en un nombre fini de morceaux, et les recoller pour faire DEUX sphères de rayon 1 ! Autrement dit, quelque chose d'absolument anti-intuitif. Ces morceaux ne sont pas mesurables, c'est ce qui empêche de dire que les deux sphères ont même volume...

En effet, imaginons qu'on puisse calculer la mesure de tout ensemble de la sphère (i.e. ici l'aire). Dans ce cas, à l'aide du paradoxe ci-dessus, on obtiendrait que la surface d'une sphère de rayon 1 est égale à celle d'une sphère de rayon 2. Contradiction ! Autrement dit, il n'est pas possible que les ensembles intervenant dans le découpage de la sphère ci-dessus soient mesurables.

Notons que la démonstration de ce résultat fait appel à l'axiome du choix. Cet axiome, refusé par certains mathématiciens, dit en gros qu'étant donnée une collection non dénombrable d'ensembles, on peut choisir un élément de chaque ensemble. (Ceci est toujours possible pour des collections finies ou dénombrables en procédant par récurrence.) Plus précisément, si  $(E_i)_{i \in I}$  est une collection d'ensembles, il existe un élément  $x = (x_i)_{i \in I}$  dans le produit infini  $\prod_{i \in I} E_i$ .

**Remarque** En TD d'intégration (exo ? feuille ?) nous avons vu un exo montrant qu'à l'aide de l'axiome du choix, on peut construire un ensemble non mesurable.

#### 1.1.2 Tribus

**Définition 1.1.1** Soit  $X$  un ensemble. Une tribu, ou  $\sigma$ -algèbre est une collection  $\mathcal{A} \subset \mathcal{P}(X)$  de parties de  $X$  qui contient  $\emptyset$  et  $X$ , est stable par union dénombrable et passage au complémentaire.

Remarquons qu'une tribu est stable par intersection dénombrable (exo).

Une *algèbre* est une collection de parties qui contient l'ensemble vide et est stable par union finie et passage au complémentaire.

Les exemples classiques et les plus utiles en pratique sont les suivants. Si  $X$  est (au plus) dénombrable,  $\mathcal{P}(X)$  est une tribu. Sur  $\mathbb{R}^d$ ,  $d \geq 1$  (et en particulier sur  $\mathbb{R}$ ), on considère la *tribu borélienne* engendrée par les ouverts. C'est la plus petite tribu qui contient tous les ouverts, mais donc aussi tous les fermés...

Autres tribus : sur  $X$  la famille  $\{\emptyset, X\}$  est toujours une tribu. Si  $A \subset \mathcal{P}(X)$  est un ensemble, la tribu engendrée par  $A$  est  $\{\emptyset, X, A, X \setminus A\}$ .

En Master, vous aurez l'occasion de rencontrer d'autres tribus, et même des familles de tribus, utiles. Cette année, les tribus  $\mathcal{P}(\mathbb{N})$  et  $\mathcal{B}(\mathbb{R})$  sont les plus importantes.

**Exercice 1.1.2** Trouver toutes les tribus de  $X = \{1, 2\}$ , puis de  $X = \{1, 2, 3\}$ .

### 1.1.3 Mesures de probabilité

Rappelons que si  $\mathcal{A}$  est une tribu sur  $X$ , un ensemble *mesurable* est simplement un ensemble  $A \in \mathcal{P}(X)$  qui est dans la tribu  $\mathcal{A}$ . Autrement dit,  $A \in \mathcal{A} \subset \mathcal{P}(X)$ .

**Définition 1.1.3** Soit  $(X, \mathcal{A})$  un espace mesurable. Une (mesure de) probabilité est une application  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$  qui vérifie  $\mathbf{P}(X) = 1$  et si  $(A_n)_{n \in \mathbb{N}}$  est une suite de sous-ensembles mesurables de  $X$  deux à deux disjoints, alors  $\mathbf{P}(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbf{P}(A_n)$ .

Comparez avec la définition d'une mesure.

Voici quelques propriétés utiles en pratique, à savoir démontrer (certaines ne sont que des reformulations de la définition).

1.  $\mathbf{P}(\emptyset) = 0$
2. Si  $(A, B) \in \mathcal{A}^2$  et  $A \subset B$  alors  $\mathbf{P}(A) \leq \mathbf{P}(B)$ .
3. Si  $(A, B) \in \mathcal{A}^2$  et  $A \cap B = \emptyset$  alors  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$
4.  $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$
5.  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$
6.  $\mathbf{P}(A \cap B) \leq \min(\mathbf{P}(A), \mathbf{P}(B))$
7.  $\mathbf{P}(A \cup B) \geq \max(\mathbf{P}(A), \mathbf{P}(B))$
8. Si  $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$  est une suite croissante pour l'inclusion (i.e.  $A_n \subset A_{n+1}$  pour tout  $n \in \mathbb{N}$ ) alors la suite  $(\mathbf{P}(A_n))_{n \in \mathbb{N}}$  est croissante et  $\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \mathbf{P}(\cup_{n \in \mathbb{N}} A_n)$
9. Si  $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$  est une suite décroissante pour l'inclusion alors la suite  $(\mathbf{P}(A_n))_{n \in \mathbb{N}}$  est décroissante et  $\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \mathbf{P}(\cap_{n \in \mathbb{N}} A_n)$ .

**Exercice 1.1.4** Démontrer ces propriétés. Indication : Considérer  $B_n = A_n \setminus A_{n-1}$ .

**Définition 1.1.5** Un espace de probabilité, ou espace probabilisé,  $(X, \mathcal{A}, \mathbb{R})$  est la donnée d'un ensemble  $X$  muni d'une tribu  $\mathcal{A}$  et d'une mesure de probabilité  $\mathbf{P}$ .

En pratique, les espaces sur lesquels on travaillera le plus sont  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  muni d'une des mesures de probabilité introduites au paragraphe 1.4, ou bien  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  et plus généralement  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ ,  $d \geq 1$ , muni d'une mesure de probabilité de 1.4.

### 1.1.4 Le vocabulaire des probabilités

La théorie des probabilités permet de fournir des modèles mathématiques pertinents à des phénomènes faisant intervenir du hasard, de l'aléa (lancer de dé, cours d'une action en bourse, durée de vie d'une ampoule, temps d'attente d'un bus, fluctuations de certaines quantités : variations des différentes mesures d'une grandeur physique autour de la valeur théorique attendue, variation de la température autour de la valeur moyenne saisonnière attendue, etc

On modélise un ensemble de situations possibles par un espace de probabilité  $(\Omega, \mathcal{A}, \mathbf{P})$ . L'ensemble  $\Omega$  est l'*espace des états*. Par exemple, si on tire une pièce à pile ou face, on peut considérer  $\Omega = \{\text{pile}, \text{face}\}$ . Pour un lancer de dé, on utilisera volontiers  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Si on tire à pile ou face trois fois de suite, on utilisera sans doute  $\Omega = \{\text{pile}, \text{face}\}^3 = \{ppp, ppf, pfp, pff, fpp, fpf, ffp, fff\} \dots$

Rappelons que quand  $\Omega$  est fini ou dénombrable, on utilise  $\mathcal{A} = \mathcal{P}(\Omega)$ . Un *événement* est un ensemble  $A \in \mathcal{A}$ . Par exemple, si on lance un dé, l'événement « obtenir un nombre pair » s'écrit encore  $\{2, 4, 6\} \subset \Omega$ .

En lançant deux dés, avec  $\Omega = \{1, \dots, 6\}^2$ , l'événement « obtenir un 7 » s'écrit  $A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$ .

On appelle  $\mathbf{P}(A)$  la probabilité que l'événement  $A$  se réalise. Si  $\mathbf{P}(A) = 1$  on dit que l'événement  $A$  est certain, ou presque sûr (p.s.).

Intuitivement (historiquement) si on répète une expérience un grand nombre  $n$  de fois, et qu'on mesure la fréquence  $f_n(A)$  d'un événement  $A$ , ( $f_n(A) = \frac{1}{n} \times \text{nb de réalisations de } A$ ), on souhaite définir la probabilité de  $A$  comme étant la limite de  $f_n(A)$  quand le nombre  $n$  d'expériences tend vers  $+\infty$ . Dans la théorie moderne des probabilités, on se donne a priori une mesure de probabilité  $\mathbf{P}$  et on démontre (voir le dernier chapitre de ce cours, lois des grands nombres) que les fréquences  $f_n(A)$  convergent vers  $\mathbf{P}(A)$ .

Une *variable aléatoire* est une fonction mesurable de l'espace mesurable  $(\Omega, \mathcal{A})$  dans un autre espace mesurable  $(E, \mathcal{E})$ . En pratique, on verra qu'on ne connaît que peu de choses sur l'espace  $\Omega$ , qu'on ne définit même pas forcément précisément lorsqu'on modélise un phénomène. Et toute l'étude porte sur l'espace d'arrivée  $(E, \mathcal{E})$  qui est souvent fini ou  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  ou  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ .

Par exemple, lorsque je lance un dé, je peux considérer  $\Omega$  l'ensemble de toutes les trajectoires possibles du dé de ma main au sol, et  $X$  la fonction définie sur  $\Omega$  à valeurs dans  $\{1, 2, 3, 4, 5, 6\}$  qui à une trajectoire associe le numéro sur la face du dessus à l'arrivée du dé sur le sol. Pour une modélisation correcte, on supposera que  $X$  est mesurable (ce qui revient à choisir sur  $\Omega$  la tribu adaptée pour cela). On pourrait aussi choisir  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .

Autre exemple : lancer de deux dés,  $\Omega = \{1, \dots, 6\}^2$  et  $X : (i, j) \mapsto i + j$ .

### 1.1.5 Rappels à connaître du S4

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

**Définition 1.1.6** Soit  $B \in \mathcal{A}$  un ensemble mesurable t.q.  $\mathbf{P}(B) > 0$ . La fonction  $\mathbf{P}(\cdot|B)$  définie par  $A \in \mathcal{A} \mapsto \mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$  est une probabilité sur  $(\Omega, \mathcal{A})$ , et la quantité  $\mathbf{P}(A|B)$  est appelée la probabilité conditionnelle de  $A$  sachant  $B$ .

**Exercice 1.1.7** Vérifier que  $\mathbf{P}(\cdot|B)$  est bien une probabilité.

$\mathbf{P}(A|B)$  représente la probabilité que  $A$  se réalise sachant que  $B$  se réalise.

**Définition 1.1.8** Deux événements  $A$  et  $B$  sont indépendants si  $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$ .

Une famille  $(A_i)_{i \in I}$  finie ou dénombrable est indépendante si pour tout sous-ensemble fini d'indices  $J \subset I$   $\mathbf{P}(\bigcap_{j \in J} A_j) = \prod_{j \in J} \mathbf{P}(A_j)$ .

Les événements  $(A_i)_{i \in I}$  sont deux à deux indépendants si pour tout couple  $(i, j)$  d'indices avec  $i \neq j$ ,  $A_i$  et  $A_j$  sont indépendants.

Bien sûr, si la famille  $(A_i)$  est indépendante dans son ensemble, alors les  $(A_i)$  sont deux à deux indépendants.

Vocabulaire : en l'absence de précision, « des événements indépendants » signifie des événements indépendants dans leur ensemble.

Un exemple : on joue deux fois de suite de pile ou face.  $A$  est l'événement « obtenir pile au premier lancer »,  $B$  « obtenir pile au deuxième lancer » et  $C$  « obtenir deux lancers identiques ». Ces événements sont-ils deux à deux indépendants ? Indépendants dans leur ensemble ?

**Exercice 1.1.9** On lance un dé, et on note  $i$  le chiffre obtenu. (Ici  $\Omega = \{1, \dots, 6\}$ ). Donner un exemple de deux événements  $A$  et  $B$  indépendants.

Remarque : Si  $A$  et  $B$  sont indépendants alors  $\mathbf{P}(A|B) = \mathbf{P}(A)$ .

**Exercice 1.1.10** Si  $A$  et  $B$  sont indépendants, alors  $A$  et  $B^c$  aussi,  $A^c$  et  $B$  aussi,  $A^c$  et  $B^c$  aussi.

**Proposition 1.1.11 (Probabilités composées)** Soit  $(A_1, \dots, A_n) \in \mathcal{A}^n$  tel que  $\mathbf{P}(A_1 \cap \dots \cap A_n) > 0$ . Alors  $\mathbf{P}(A_1 \cap A_2 \dots \cap A_n) = \mathbf{P}(A_1)\mathbf{P}(A_2|A_1)\mathbf{P}(A_3|A_1 \cap A_2) \dots \mathbf{P}(A_n|A_1 \cap \dots \cap A_{n-1})$

**Démonstration :** Exercice très facile à savoir faire! □

**Définition 1.1.12** Une partition mesurable de  $X$  est une famille d'ensembles  $(E_i)_{i \in I}$  mesurables deux à deux disjoints tels que  $\bigcup_{i \in I} E_i = \Omega$ . On suppose souvent implicitement que  $\mathbf{P}(E_i) > 0$  pour tout  $i \in I$ . (Pourquoi est-ce possible sans difficulté ?)

**Proposition 1.1.13 (Probabilités totales)** Soit  $(E_i)_{i \in I} \subset \mathcal{A}^I$  une partition mesurable finie ou dénombrable de  $\Omega$ . Alors pour tout  $A \in \mathcal{A}$ , on a  $\mathbf{P}(A) = \sum_{i \in I} \mathbf{P}(A \cap E_i) = \sum_{i \in I} \mathbf{P}(A|E_i)\mathbf{P}(E_i)$

**Démonstration :** Exercice très facile à savoir faire! □

**Mise en garde** Cette formule élémentaire provoque un nombre incalculable d'erreurs scandaleuses. Débrouillez-vous mais apprenez la, sachez la démontrer et l'utiliser sans écrire des choses du style  $\mathbf{P}(A) \dots \sum_i \mathbf{P}(A|E_i)$   
Exemples : voir TD et compléter.

**Proposition 1.1.14 (Formule de Bayes)** Soit  $(E_i)_{i \in I} \subset \mathcal{A}^I$  une partition mesurable finie ou dénombrable de  $\Omega$ . Si  $A \in \mathcal{A}$  vérifie  $\mathbf{P}(A) > 0$  alors

$$\mathbf{P}(E_n|A) = \frac{\mathbf{P}(A|E_n)\mathbf{P}(E_n)}{\sum_{m \in I} \mathbf{P}(A|E_m)\mathbf{P}(E_m)}$$

**Démonstration :** Exercice très facile à savoir faire! □

Remarque : ce résultat complètement élémentaire ne mérite pas le nom de proposition par sa difficulté, mais par son utilité pratique. (Voir TD).

## 1.2 Retour sur les classes monotones

Ce paragraphe est tiré de [R1].

Rappelons la définition

**Définition 1.2.1** Une famille  $\mathcal{M}$  de parties d'un ensemble  $E$  est appelée classe monotone si

- $E \in \mathcal{M}$ ,
- si  $A \in \mathcal{M}$ ,  $B \in \mathcal{M}$ , et  $A \subset B$ , alors  $B \setminus A \in \mathcal{M}$ ,
- Si  $(A_n)_{n \in \mathbb{N}}$  est une suite croissante d'éléments de  $\mathcal{M}$ , alors  $\cup_{n \in \mathbb{N}} A_n$  est dans  $\mathcal{M}$ .

La Classe monotone  $\mathcal{M}(\mathcal{C})$  engendrée par une famille  $\mathcal{C}$  de parties de  $E$  est la plus petite classe monotone contenant  $\mathcal{C}$ , ou encore l'intersection de toutes les classes monotones contenant  $\mathcal{C}$ .

Une tribu est une classe monotone. La classe monotone  $\mathcal{M}(\mathcal{C})$  engendrée par une famille  $\mathcal{C}$  de parties est donc incluse dans la tribu  $\sigma(\mathcal{C})$  engendrée par  $\mathcal{C}$ .

**Remarque 1.2.2** Une classe monotone stable par intersection finie est une tribu. En effet, les deux premières propriétés garantissent alors que  $\emptyset$  est dans la classe monotone, et que la classe est stable par union finie. La dernière propriété assure enfin que la classe est stable par union dénombrable.

La notion de classe monotone prend toute son utilité dans le théorème d'apparence technique suivant, dont on verra l'intérêt dans le corollaire qui suit.

**Théorème 1.2.3 (Théorème de la classe monotone)** Si  $\mathcal{C}$  est une famille stable par intersection finie, alors  $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$ .

**Démonstration :** Comme  $\mathcal{M}(\mathcal{C}) \subset \sigma(\mathcal{C})$ , il suffit de montrer que  $\mathcal{M}(\mathcal{C})$  est une tribu. Pour cela, vue la remarque ci-dessus, il suffit de montrer qu'elle est stable par intersection finie.

• Posons  $\mathcal{M}_1 = \{A \in \mathcal{M}(\mathcal{C}), \forall C \in \mathcal{C}, A \cap C \in \mathcal{M}(\mathcal{C})\}$ . Alors  $\mathcal{C} \subset \mathcal{M}_1$  (par hypothèse sur  $\mathcal{C}$ ). Et  $\mathcal{M}_1$  est une classe monotone contenant  $\mathcal{C}$  et incluse dans  $\mathcal{M}(\mathcal{C})$ , donc  $\mathcal{M}_1 = \mathcal{M}(\mathcal{C})$ . En effet,  $E \in \mathcal{M}_1$  bien sûr. Si  $A$  et  $B$  sont dans  $\mathcal{M}_1$  et  $A \subset B$ , et  $C \in \mathcal{C}$ , alors  $(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C)$ ;  $B \cap C$  et  $A \cap C$  sont dans  $\mathcal{M}(\mathcal{C})$  par définition de  $\mathcal{M}_1$ , et leur différence aussi car  $\mathcal{M}(\mathcal{C})$  est une classe monotone. La stabilité de  $\mathcal{M}_1$  par union dénombrable croissante découle du fait que  $(\cup_{n \in \mathbb{N}} A_n) \cap C = \cup_{n \in \mathbb{N}} (A_n \cap C)$ .

• Posons  $\mathcal{M}_2 = \{A \in \mathcal{M}(\mathcal{C}), \forall C \in \mathcal{M}(\mathcal{C}), A \cap C \in \mathcal{M}(\mathcal{C})\}$ . De la même façon, on vérifie que  $\mathcal{M}_2$  est une classe monotone, incluse dans  $\mathcal{M}(\mathcal{C})$  et contenant  $\mathcal{C}$ , donc égale à  $\mathcal{M}(\mathcal{C})$ . Donc  $\mathcal{M}(\mathcal{C})$  est stable par intersection finie. □

**Corollaire 1.2.4 (Unicité des mesures)** Si  $\mathcal{C}$  est une classe d'ensembles stable par intersection finie et engendrant une tribu  $\mathcal{A}$  sur  $E$ , alors deux mesures de probabilité sur  $\mathcal{A}$  qui sont égales sur  $\mathcal{C}$  sont égales sur  $\mathcal{A}$ .

Ce corollaire nous sera utile par la suite.

**Démonstration :** Soient  $\mu$  et  $\nu$  deux mesures de probabilité égales sur  $\mathcal{C}$ ; notons

$$\mathcal{M} = \{A \in \mathcal{A}, \mu(A) = \nu(A)\}.$$

On vérifie aisément que  $\mathcal{M}$  est une classe monotone qui contient  $\mathcal{C}$ . Le théorème précédent permet de conclure.  $\square$

Fin cours 1 (2h)

## 1.3 Variables aléatoires

### 1.3.1 Variable aléatoire, loi d'une variable aléatoire

debut cours 2 (2h)

**Définition 1.3.1** Une variable aléatoire est une application mesurable  $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$ . (Souvent,  $E = \mathbb{N}$  ou  $E = \mathbb{R}^d$ ). Lorsque  $E = \mathbb{R}$ , ou  $E \subset \mathbb{R}$  on parle de variable aléatoire réelle (v.a.r.) Lorsque  $E = \mathbb{R}^d$  avec  $d \geq 2$  on parle de vecteur aléatoire.

Lorsque  $E$  est un ensemble discret, fini ou dénombrable, on parle de variable aléatoire discrète.

Lorsque  $E \subset \mathbb{N}$  ou  $\mathbb{Z}$ , on parle de variable aléatoire entière.

Bien évidemment, une v.a. à valeurs dans  $\mathbb{N}$  est une v.a.r. mais on parle de variable aléatoire discrète.

**Remarque 1.3.2** « Rappelons » qu'un ensemble discret est un espace topologique sur lequel les singletons sont ouverts.

**Définition 1.3.3** La tribu engendrée par une v.a.  $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$  est la plus petite tribu  $\sigma(X) \subset \mathcal{A}$  rendant  $X$  mesurable. Elle peut encore être définie comme  $\sigma(X) = \{X^{-1}(B), B \in \mathcal{E}\}$ .

**Démonstration :** On commence par remarquer que  $\sigma(X)$  contient  $X^{-1}(\mathcal{E})$ , puis que  $X^{-1}(\mathcal{E})$  est une tribu, enfin que  $\sigma(X)$  est par définition la plus petite tribu rendant  $X$  mesurable, i.e. la plus petite tribu contenant  $X^{-1}(\mathcal{E})$ .  $\square$

**Définition 1.3.4** Soit  $\mathbf{P}$  une probabilité sur  $(\Omega, \mathcal{A})$ . La loi image ou mesure image de  $\mathbf{P}$  par  $X$  est la mesure  $\mathbf{P}_X$  sur  $(E, \mathcal{E})$  définie par  $\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A))$  pour tout  $A \in \mathcal{E}$ . La mesure de probabilité  $\mathbf{P}_X$  est appelée la loi de  $X$ .

Notation : on écrit souvent (abus de notation utile pour l'intuition)  $\mathbf{P}(X \in A)$  à la place de  $\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(\{\omega \in \Omega, X(\omega) \in A\})$ .

Lorsqu'on modélise une situation réelle avec des outils probabilistes, en général, on ne connaît pas  $(\Omega, \mathcal{A}, \mathbf{P})$ . Lors d'une expérience aléatoire  $\omega \in \Omega$  on ne fait qu'observer la valeur  $X(\omega) \in \mathbb{N}$  ou  $\mathbb{R}$ . On est alors intéressé par la probabilité que  $X(\omega)$  prenne telle ou telle valeur. Autrement dit, on veut savoir étudier  $\mathbf{P}_X$  sur  $E$  et pas du tout  $\mathbf{P}$  sur  $\Omega$ . L'espace de probabilité  $(\Omega, \mathcal{A}, \mathbf{P})$  est important du point de vue théorique, mais n'est pas explicite en général.

Vous rencontrerez donc fréquemment des phrases du type «  $X : \Omega \rightarrow \{0, 1\}$  suit une loi de Bernoulli de paramètre  $p$  », alors qu'on n'a même pas précisé  $\Omega$ . On dit même  $X$  suit une loi de Bernoulli  $\mathcal{B}(p)$  sur  $\{0, 1\}$  sans évoquer l'espace  $\Omega$  quand seule la loi  $\mathbf{P}_X$  de  $X$  est utile.

Un exemple de loi image :  $\Omega = \{(i, j), 1 \leq i, j \leq 6\}$ ;  $\mathbf{P}(i, j) = \frac{1}{36}$  et  $X : \Omega \rightarrow \mathbb{N}$  définie par  $X(i, j) = i + j$ . Pourquoi  $X$  est-elle mesurable? Sur  $\mathbb{N}$  la loi de  $X$  est définie par  $\mathbf{P}_X(\{n\}) = \mathbf{P}(X^{-1}(n)) = \mathbf{P}(\{\omega \in \Omega, X(\omega) = n\})$ . Et on a  $\mathbf{P}_X(0) = 0 = \mathbf{P}_X(n)$  pour  $n \geq 13$ .  $\mathbf{P}_X(2) = \frac{1}{36}$ ,  $\mathbf{P}_X(3) = \frac{1}{18}, \dots$

Autre exemple :  $\Omega = [0, 1]$  muni de sa tribu borélienne et de la mesure de Lebesgue. Soit  $X$  la variable aléatoire définie par  $X(\omega) = 1$  si  $\omega \in [0, 1/2]$  et 0 sinon ( $X = \mathbf{1}_{[0, 1/2]}$ ). Alors  $\mathbf{P}_X$  vérifie  $\mathbf{P}_X(0) = \mathbf{P}_X(1) = \frac{1}{2}$  et  $\mathbf{P}_X(\mathbb{R} \setminus \{0, 1\}) = 0$ . Autrement dit,  $\mathbf{P}_X = \frac{1}{2}(\delta_0 + \delta_1)$  où  $\delta_x$  désigne la masse de Dirac en  $x$ .



**Exercice 1.3.5** On considère  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  où  $\mathbf{P} = \frac{1}{3}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{6}\delta_1$ . C'est une probabilité sur  $\mathbb{R}$ . Soit  $X : \omega \mapsto |\omega|$ . Montrer que  $\mathbf{P}_X = \frac{1}{2}(\delta_0 + \delta_1)$ .

**Remarque 1.3.6** Une mesure de probabilité  $\mathbf{P}$  sur un espace mesurable  $(\Omega, \mathcal{A})$  est toujours la loi d'une variable aléatoire. Considérer  $X = Id : (\Omega, \mathcal{A}) \rightarrow (\Omega, \mathcal{A})$ . Si au départ  $\Omega$  est muni de la loi de probabilité  $\mathbf{P}$ , alors à l'arrivée la loi image de  $\mathbf{P}$  par  $X$ , i.e. la loi  $\mathbf{P}_X$  de  $X$  est égale à  $\mathbf{P}$ .

**Remarque 1.3.7 (Important !)** Il existe des variables aléatoires différentes de même loi. Par exemple sur  $[0, 1]$  muni de la tribu borélienne et de la mesure de Lebesgue, considérer  $X = \mathbf{1}_{[0,2/3]}$  et  $Y = \mathbf{1}_{[1/4,7/12]} + \mathbf{1}_{]2/3,1]}$ . Considérer même  $Z : \{1, \dots, 6\} \rightarrow \{0, 1\}$  défini par  $Z(\omega) = 1$  si  $\omega$  n'est pas divisible par trois. Alors  $Z$  a la même loi que  $X$  et  $Y$  (exo) alors qu'il n'est pas défini sur le même espace !

**Théorème 1.3.8** Soit  $X : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (E, \mathcal{E})$  une variable aléatoire. Alors pour toute fonction  $h : E \rightarrow \mathbb{R}$  mesurable bornée ou mesurable positive, on a

$$\int_{\Omega} h \circ X(\omega) d\mathbf{P}(\omega) = \int_E h(x) d\mathbf{P}_X(x)$$

Plus généralement,  $h$  est  $\mathbf{P}_X$ -intégrable ssi  $h \circ X$  est  $\mathbf{P}$ -intégrable et la formule ci-dessus est encore vraie dans ce cas.

**Remarque 1.3.9 (Cas particuliers importants)** Lorsque  $E = \mathbb{N}$  le terme de droite se réécrit  $\sum_{n \in \mathbb{N}} h(n) \mathbf{P}(X = n)$ .

Lorsque  $\Omega$  est discret (souvent  $\Omega = \mathbb{N}$  ou  $\Omega$  fini en pratique), alors  $\int_{\Omega} h \circ X(\omega) d\mathbf{P}(\omega) = \sum_{\omega \in \Omega} p(\omega) h \circ X(\omega)$ .

Cette formule dont la démonstration est élémentaire est fondamentale. En effet, répétons encore une fois qu'en pratique, on ne connaît pas  $\Omega$  ni  $\mathbf{P}$ . C'est donc sur  $E$  qu'on effectuera tous les calculs. Le terme de gauche est important théoriquement, le terme de droite sert en pratique. (Ce résultat est appelé principe de transfert dans certains ouvrages.)

**Remarque 1.3.10** Pourquoi ces hypothèses sur  $h$ ? Rappelons que l'intégrale  $\int_E h(x) d\mathbf{P}_X(x)$  est définie pour les fonctions mesurables positives ou les fonctions  $\mathbf{P}_X$ -intégrables. Or en théorie des probabilités, les variables aléatoires bornées sont des fonctions intégrables (Si  $|f| \leq C$ , alors  $\int_E f d\mathbf{P}_X \leq C \mathbf{P}_X(E) = C < \infty$ ).

Cela dit, le bon cadre de ce théorème est  $f$  mesurable positive ou intégrable.

**Attention :** C'est faux en analyse ! Sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  muni de la mesure de Lebesgue, les fonctions bornées ne sont pas toutes intégrables ! Exemple, les fonctions constantes...

**Démonstration :** Principe de preuve très classique, à connaître. On passe progressivement des indicatrices de boréliens à toutes les fonctions mesurables. On reverra ce type de preuve à plusieurs reprises.

- Si  $A \in \mathcal{E}$  et  $h = 1_A$  il s'agit simplement de la définition de  $\mathbf{P}_X : 1_A \circ X = 1_{X^{-1}(A)}$  et  $\mathbf{P}(X^{-1}(A)) = \mathbf{P}_X(A)$ .
- Si  $h = \sum_i \alpha_i \mathbf{1}_{A_i}$  est une fonction simple (somme finie d'indicatrices pondérées par les  $\alpha_i$ ), alors le résultat souhaité est vrai par linéarité de l'intégrale.
- Si  $h$  est mesurable positive, alors il existe une suite  $(h_n)$  croissante de fonctions simples positives telles que  $h_n \rightarrow h$ , et donc aussi  $h_n \circ X \rightarrow h \circ X$  presque sûrement. Le théorème de convergence monotone (Beppo-Levi) donne le résultat souhaité.
- Si  $h$  est mesurable bornée, elle est limite de fonctions simples uniformément bornées ; le théorème de convergence dominée de Lebesgue s'applique.
- Si  $h$  est  $\mathbf{P}_X$ -intégrable (plus faible que mesurable bornée car  $\mathbf{P}_X$  est une probabilité), alors il existe  $h_+$  et  $h_-$  mesurables positives telles que  $h = h_+ - h_-$ . On a  $|h| = h_+ + h_-$ , et  $\int_{\mathbb{R}} |h| d\mathbf{P}_X < \infty$  implique  $\int_{\mathbb{R}} h_{\pm} d\mathbf{P}_X < \infty$ . Encore une fois, la linéarité de l'intégrale va mener à l'égalité voulue, et donc en particulier au fait que  $h \circ X$  est  $\mathbf{P}$ -intégrable. Le même raisonnement en partant de cette dernière hypothèse donnerait  $h \mathbf{P}_X$  intégrable. □

La proposition suivante est un cas particulier très utile.

**Proposition 1.3.11** La v.a.r.  $X$  est intégrable ssi  $\int_{\mathbb{R}} |x| d\mathbf{P}_X(x) < \infty$ .

**Démonstration :** C'est un corollaire du théorème 1.3.8 avec  $h(x) = |x|$ . □

### 1.3.2 Fonction de répartition d'une variable aléatoire à valeurs dans $\mathbb{R}$ (ou $\mathbb{N}$ , $\mathbb{Z}$ ...)

Dans ce paragraphe, sauf mention du contraire,  $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$  désigne une variable aléatoire à valeurs dans un sous-ensemble de  $\mathbb{R}$ .

**Définition 1.3.12** Soit  $\mathbf{P}$  une mesure de probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  ou  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ . La fonction de répartition de  $\mathbf{P}$  est la fonction  $F : \mathbb{R} \rightarrow [0, 1]$  définie par  $F(x) = \mathbf{P}(]-\infty, x])$ .

Exemples :  $\delta_2, \frac{\delta_1 + \delta_2}{2}, \dots$

**Exercice 1.3.13** Tracer le graphe de la fonction de répartition de  $\delta_0$ , de  $\frac{1}{3}\delta_{-2} + \frac{1}{2}\delta_1 + \frac{1}{6}\delta_{13}$ , de la mesure uniforme sur  $[-1, 1] : \mathbf{P} = \frac{1}{2}\mathbf{1}_{[-1, 1]}\lambda$ , où  $\lambda$  est la mesure de Lebesgue...

**Définition 1.3.14** La fonction de répartition de  $X$ , notée  $F_X$ , est la fonction de répartition de sa loi  $\mathbf{P}_X$ . Autrement dit,

$$F_X(x) = \mathbf{P}_X(]-\infty, x]) = \mathbf{P}(X \leq x) = \mathbf{P}(\{\omega \in \Omega, X(\omega) \leq x\})$$

**Proposition 1.3.15** Soit  $F_X$  la fonction de répartition de  $\mathbf{P}_X$ . Alors on a

1.  $F_X$  est croissante.
2.  $F_X$  est à valeurs dans  $[0, 1]$
3.  $F_X$  est continue à droite.
4.  $F_X$  a des limites à gauche en tout point, et  $\lim_{x \rightarrow a^-} F_X(x) = F_X(a^-) = \mathbf{P}(X < a) \leq \mathbf{P}(X \leq a) = F_X(a)$ .
5.  $F_X(x) \rightarrow 0$  quand  $x \rightarrow -\infty$  et  $F_X(x) \rightarrow 1$  quand  $x \rightarrow +\infty$ .
6.  $F_X(b) - F_X(a) = \mathbf{P}(a < X \leq b)$
7.  $\mathbf{P}(X = a) = F_X(a) - \lim_{x \rightarrow a^-} F_X(x) = F_X(a) - F_X(a^-)$

**Remarque 1.3.16** On verra plus loin des cas agréables dans lesquels  $F_X$  est continue sur  $\mathbb{R}$ .

**Démonstration :** 1. Si  $x \leq y$ ,  $]-\infty, x] \subset ]-\infty, y]$  d'où  $F_X(x) = \mathbf{P}_X(]-\infty, x]) \leq \mathbf{P}_X(]-\infty, y]) = F_X(y)$ .

2. Pour tout  $x \in \mathbb{R}$ ,  $F_X(x)$  est la probabilité d'un ensemble.

3. Soit  $x \in \mathbb{R}$ . On veut montrer que  $\lim_{y \rightarrow x^+} F_X(y) = F_X(x)$ . Comme  $F_X$  est croissante, il suffit de montrer que  $\lim_{n \rightarrow \infty} F_X(x + \frac{1}{n}) = F_X(x)$ . (En effet, supposons que  $F_X(x + \frac{1}{n}) \rightarrow F_X(x)$  quand  $n \rightarrow \infty$ . Alors pour tout  $\varepsilon > 0$ , il existe  $N > 0$  tq pour tout  $n \geq N$ ,  $F_X(x) \leq F_X(x + \frac{1}{n}) \leq F_X(x) + \varepsilon$ . Comme  $F_X$  est croissante, pour tout  $x \leq y \leq x + \frac{1}{n}$ ,  $F_X(x) \leq F_X(y) \leq F_X(x + \frac{1}{n}) \leq F_X(x) + \varepsilon$ . Donc  $\lim_{y \rightarrow x^+} F_X(y) = F_X(x)$ .)

Remarquons que  $F_X(x + \frac{1}{n}) = \mathbf{P}_X(]-\infty, x + \frac{1}{n}])$ . La suite d'intervalles  $]-\infty, x + \frac{1}{n}]$  est décroissante, la probabilité  $\mathbf{P}_X$  vérifie donc  $\lim_{n \rightarrow \infty} \mathbf{P}_X(]-\infty, x + \frac{1}{n}]) = \mathbf{P}_X(\bigcap_{n \in \mathbb{N}} ]-\infty, x + \frac{1}{n}]) = \mathbf{P}_X(]-\infty, x]) = F_X(x)$ .

4. Considérons cette fois la suite d'ensembles  $]-\infty, x - \frac{1}{n}]$ . C'est une suite croissante d'intervalles. On a donc  $\lim_{n \rightarrow \infty} F_X(x - \frac{1}{n}) = \mathbf{P}_X(]-\infty, x])$ . (Alors que  $F_X(x) = \mathbf{P}_X(]-\infty, x])$ .) Comme  $F_X$  est croissante, on en déduit que  $F_X$  admet  $\mathbf{P}_X(]-\infty, x]) \leq F_X(x)$  comme limite à gauche en  $x$ . (Remarquons que l'existence de la limite découle immédiatement du fait que si  $y < x$  l'application  $y \mapsto F_X(y)$  est croissante et majorée par  $F_X(x)$ .)

5. On a  $\emptyset = \bigcap_{n \in \mathbb{N}} ]-\infty, -n]$ . Le même raisonnement que ci-dessus donne  $\lim_{n \rightarrow \infty} F_X(-n) = 0$ . Comme  $F_X$  est croissante, ceci implique (vérifier!) que  $\lim_{x \rightarrow -\infty} F_X(x) = 0$ . Pour la limite en  $+\infty$  considérer les intervalles  $]-\infty, n]$ .

6 et 7. Exo immédiat. □

**Proposition 1.3.17** Soient  $P_1$  et  $P_2$  deux mesures de probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Si elles ont même fonction de répartition  $F_1 = F_2 = F$ , alors elles sont égales.

**Démonstration :** Pour tout intervalle  $]a, b]$ , on a  $\mathbf{P}_1(]a, b]) = F(b) - F(a) = \mathbf{P}_2(]a, b])$ . La classe des intervalles du type  $]a, b]$ ,  $a \in \mathbb{R}$ ,  $b \in \mathbb{R}$  est stable par intersection finie et engendre la tribu borélienne (voir les TD du module intégration. Par intersection/union dénombrable, il est clair que la tribu engendrée contient tous les intervalles, et donc tous les boréliens.)

Le corollaire 1.2.4 permet de conclure que  $\mathbf{P}_1 = \mathbf{P}_2$ . □

**Proposition 1.3.18** Soient  $X, Y$  deux variables aléatoires définies respectivement sur  $(\Omega, \mathcal{A}, \mathbf{P})$  et  $(\Omega', \mathcal{A}', \mathbf{P}')$ , à valeurs réelles ou entières. Si  $X$  et  $Y$  ont même fonction de répartition, alors  $X$  et  $Y$  ont même loi. □

**Démonstration :** Reformulation immédiate de la proposition précédente. □

**Proposition 1.3.19** Si  $F$  est une fonction de  $\mathbb{R}$  dans  $[0, 1]$  qui est croissante, continue à droite, et admet pour limite 0 en  $-\infty$  et 1 en  $+\infty$ , alors il existe une mesure de probabilité  $\mathbf{P}$  sur  $\mathbb{R}$  dont  $F$  est la fonction de répartition.

**Démonstration :** L'idée de la preuve est élémentaire. On définit  $\mathbf{P}$  sur tout intervalle  $]a, b]$  par  $\mathbf{P}(]a, b]) = F(b) - F(a)$ . On a bien  $\mathbf{P}(\emptyset) = 0$  et  $\mathbf{P}(\mathbb{R}) = 1$ . La famille des intervalles  $]a, b]$  est une algèbre : stable par intersection et union finie, passage au complémentaire, et contient  $\emptyset$  et  $\mathbb{R}$ . Cette algèbre engendre la tribu borélienne. Pour conclure la preuve, il suffit d'utiliser le théorème ci-dessous 1.3.20. □

**Théorème 1.3.20 (Théorème de prolongement de Carathéodory)** Si  $m$  est une mesure bornée définie sur une algèbre  $\mathcal{B}$  de parties de  $E$ , il existe une unique mesure prolongeant  $m$  sur la tribu  $\sigma(\mathcal{B})$  engendrée par  $\mathcal{B}$ .

**Démonstration :** L'unicité découle du corollaire 1.2.4. L'existence sera démontrée en Master I. □

**Proposition 1.3.21** Une fonction de répartition a au plus un nombre dénombrable de discontinuités.

**Démonstration :** Voir TD □

**Proposition 1.3.22** Soit  $F$  la fonction de répartition de la probabilité  $\mu$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . On définit un « pseudo-inverse » de  $F$  par  $F^{-1}(t) = \inf\{x \in \mathbb{R}, F(x) \geq t\}$ . Soit  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ . Alors  $Y = F^{-1}(U)$  est une variable aléatoire de loi  $\mu$ .

Ceci sert en simulation : si on sait simuler une v.a.r. de loi uniforme, alors on sait simuler n'importe quelle loi sur  $\mathbb{R}$ .

**Démonstration :** Voir TD □

**Définition 1.3.23** Soit  $X = (X_1, \dots, X_d)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ . Sa fonction de répartition est définie par

$$F_X(x_1, \dots, x_d) = \mathbf{P}_X(]-\infty, x_1] \times \dots \times ]-\infty, x_d]) = \mathbf{P}(X_1 \leq x_1 \text{ et } \dots X_d \leq x_d).$$

**Proposition 1.3.24** Soient  $X_1$  et  $X_2$  deux vecteurs aléatoires réels, à valeurs dans  $\mathbb{R}^d$ . S'ils ont même fonction de répartition, alors ils ont même loi.

**Démonstration :** Ceci se fait exactement comme la proposition 1.3.18. □

## 1.4 Lois et variables aléatoires classiques

En pratique, on modélise souvent les phénomènes aléatoires de la vie réelle par quelques lois qu'on connaît bien, et que nous allons revoir ici. Les lois discrètes ont été vues au S4, et doivent être connues.

## 1.4.1 Loïs discrètes

### Le pile ou face, loi uniforme

On lance une pièce et lorsqu'elle retombe, on regarde si la face visible est "face" (la face de Marianne, sur les francs d'autrefois) ou "pile" (la semeuse, de l'autre côté des francs). On suppose que la pièce a autant de chances de tomber sur pile que sur face. On introduit alors  $\Omega = \{\text{pile, face}\}$ , la tribu est  $\mathcal{P}(\Omega)$  et la probabilité  $\mathbf{P}$  définie par  $\mathbf{P}(\text{pile}) = \mathbf{P}(\text{face}) = \frac{1}{2}$ .

Une variante mathématique est de regarder l'espace  $\{0, 1\}$  muni de la tribu  $\mathcal{P}(\{0, 1\})$  et de la probabilité  $\mathbf{P} = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ . Si une variable aléatoire  $X$  est à valeurs dans  $\{0, 1\}$  muni de  $\mathbf{P}$  on dit que  $X$  suit une loi uniforme sur  $\{0, 1\}$ .

Sa fonction de répartition vérifie  $F_X(x) = 0$  si  $x < 0$ ,  $F_X(x) = \frac{1}{2}$  si  $x \in [0, 1[$  et  $F_X(x) = 1$  si  $x \geq 1$ .

Plus généralement, si  $X$  est une v.a. à valeurs dans un ensemble fini  $E$ , on dit qu'elle suit une loi uniforme sur  $E$  si pour tout  $\omega \in E$ , on a  $P_X(\{\omega\}) = \frac{1}{\#E}$ . On dit alors que chacun des événements élémentaires  $\{\omega\}$  est équiprobable.

Un autre exemple classique est celui du lancer de dé équilibré. L'ensemble à considérer est  $\{1, 2, 3, 4, 5, 6\}$  muni de la probabilité uniforme qui assigne à chaque événement élémentaire la probabilité  $\frac{1}{6}$ .

**Exercice 1.4.1** Tracer la fonction de répartition d'une loi de pile ou face sur  $\{0, 1\}$  puis d'une loi uniforme sur  $\{1, \dots, 6\}$ .

### Marches aléatoires

Imaginons un ivrogne qui se promène le long d'une rue (orientée est ouest), et tous les 10 mètres, il tire à pile ou face pour savoir s'il va à l'ouest (pile) ou à l'est (face). Cette marche dite aléatoire est modélisée de la façon suivante : on imagine que l'ivrogne se promène sur  $\mathbb{Z}$  ; à l'instant  $n \geq 0$ , il se trouve à la position  $S_n \in \mathbb{Z}$  (avec  $S_0 = 0$  pour simplifier). Entre l'instant  $n$  et l'instant  $n + 1$  il tire à pile ou face, on note  $X_{n+1}$  la variable aléatoire qui vaut  $+1$  s'il tire face (il va à droite) et  $-1$  sinon. Sa position à l'instant  $n + 1$  est alors  $S_{n+1} = S_n + X_{n+1}$ . Autrement dit, on modélise cette marche par une suite  $(X_n)_{n \geq 1}$  de variables aléatoires à valeurs dans  $\{+1, -1\}$ , de loi uniforme sur  $\{+1, -1\}$ . La position à l'instant  $n$  est donnée par  $S_n = \sum_{k=1}^n X_k$ . Le dernier chapitre nous donnera des outils pour étudier le comportement de  $S_n$  quand  $n$  grand (retour en 0 ? divergence en  $\pm\infty$ ...)

Plus généralement, on peut effectuer des marches aléatoires sur  $\mathbb{Z}^d$ ,  $d \geq 2$ . Alors  $X_n$  est une variable aléatoire à valeurs dans  $\{\pm 1\}^d$ , et  $S_n$  est une v.a. à valeurs dans  $\mathbb{Z}^d$ .

**Remarque 1.4.2** Les marches aléatoires sur  $\mathbb{Z}^d$  ont un comportement bien connu depuis longtemps, mais les chercheurs étudient encore très activement les marches aléatoires sur des groupes  $G$  non commutatifs, comme des groupes de matrices par exemple. On parle alors de produits de matrices aléatoires. À chaque instant  $n$ , au lieu d'ajouter à  $S_{n-1}$  un élément  $X_n$  qui fait partie des générateurs de  $\mathbb{Z}^d$  et de leurs inverses, on multiplie (à gauche ou à droite, ce n'est pas pareil, il faut choisir !)  $S_{n-1}$  par un générateur  $X_n$  du groupe non commutatif  $G$  (noté multiplicativement au lieu d'additivement pour  $\mathbb{Z}^d$ ). On se promène alors aléatoirement dans  $G$ , et on voudrait bien savoir si on va revenir à l'origine ou pas (ce qui dépend de la nature et de la taille du groupe !)

Cette théorie des matrices aléatoires intéresse les mathématiciens pour sa beauté intrinsèque d'une part, mais aussi pour ses intérêts en modélisation. Plusieurs chercheurs-euses du LAMFA travaillent par exemple à modéliser le développement du Prunus en forêt de Compiègne à l'aide de produits aléatoires de matrices.

Fin cours 2 (2h, le 28 janvier 2009)

### Loi de Bernoulli

debut cours 3 (2h, le 18 février 2009)

La loi de Bernoulli de paramètre  $p$ , notée parfois  $\mathcal{B}(p)$  est la loi sur  $\{0, 1\}$  (ou autre ensemble à deux éléments) définie par  $\mathbf{P}(1) = p$  et  $\mathbf{P}(0) = 1 - p$ ,  $p$  étant un paramètre fixé dans  $[0, 1]$ . Autrement dit  $\mathbf{P} = p\delta_1 + (1 - p)\delta_0$ .

Par exemple, on tire une boule dans une urne contenant  $R$  boules rouges et  $N$  noires, soit une proportion  $p = \frac{R}{R+N}$  de boules rouges, et  $1 - p$  de boules noires. On considère  $\Omega = \{1, \dots, R + N\}$  l'ensemble des boules numérotées, et  $X$  la variable aléatoire définie sur  $\Omega$  qui vaut 1 si la boule est rouge et 0 sinon. En supposant que chaque boule a une probabilité identique d'être attrapée, la v.a.  $X$  suit une loi de Bernoulli de paramètre  $p$ .

**Exercice 1.4.3** Tracer la fonction de répartition d'une loi de Bernoulli de paramètre  $\frac{1}{3}$ .

## Loi binomiale

Une urne contient  $R$  boules rouges,  $N$  boules noires, soit encore une proportion  $p = \frac{R}{R+N}$  de boules rouges. Si on tire  $n$  boules avec remise (on remet la boule dans l'urne à chaque fois), l'espace des états est alors  $\Omega = \{1, \dots, R+N\}^n$ ; soit  $X$  la v.a. qui à un tirage aléatoire  $\omega \in \Omega$  de  $n$  boules avec remise associe le nombre de boules rouges obtenues. Un calcul élémentaire montre que  $\mathbf{P}_X(k) = \mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$  pour tout  $0 \leq k \leq n$ . On dit que  $X$  suit une loi binomiale de paramètres  $(n, p)$  parfois encore notée  $\mathcal{B}(n, p)$ .

**Exercice 1.4.4** Vérifier que  $\mathbf{P}_X$  est une probabilité sur  $\{0, 1, \dots, n\}$ .

**Exercice 1.4.5** Tracer le graphe de la fonction de répartition d'une loi binomiale de paramètres  $n = 4$  et  $p = \frac{1}{3}$ .

## Loi hypergéométrique

C'est la loi de la variable aléatoire  $X$  comptant le nombre de boules rouges obtenues lors d'un tirage de  $n$  boules sans remise dans une urne contenant  $R$  rouges et  $N$  noires. On la note parfois  $\mathcal{H}(n, N, R)$ , et on vérifie que  $\mathbf{P}_X(k) = \mathbf{P}(X = k) = \frac{C_R^k C_N^{n-k}}{C_{R+N}^n}$ .

## Loi de Poisson

Notée parfois  $\mathcal{P}(\lambda)$ . Loi de Poisson de paramètre  $\lambda \in \mathbb{R}$ . c'est la mesure de probabilité sur  $\mathbb{N}$  définie par  $\mathbf{P}(n) = \frac{\lambda^n}{n!} e^{-\lambda}$ .

**Exercice 1.4.6** Vérifier que  $\mathbf{P}$  est une probabilité sur  $\mathbb{N}$

**Exercice 1.4.7** Tracer le graphe de la fonction de répartition d'une loi de Poisson de paramètre 1.

La loi de Poisson est la loi des événements rares.

Par exemple, on sait que le temps moyen d'attente à un guichet (en nombre de minutes) vaut  $\lambda > 0$ . Alors la probabilité d'attendre  $k$  minutes est bien modélisée par une loi de Poisson de paramètre  $\lambda$ .

De même, si on sait qu'une ampoule dure en moyenne  $\lambda$  jours, alors on peut modéliser le temps entre deux pannes par une loi de Poisson de paramètre  $\lambda$ .

Autre exemple pratique : on prélève (sans remise donc)  $n$  individus dans une population qui contient des individus de deux types  $A$  et  $B$  (malades/sains, bactéries mutantes/ordinaires, etc). Le type  $A$  est présent en proportion  $p$  très faible, le type  $B$  est en proportion  $1 - p$  proche de 1. Le nombre moyen d'individus  $A$  prélevés vaut  $\lambda = np$ . Si  $n$  est grand,  $p$  petit et  $1 \leq \lambda = np \leq 10$ , la v.a. représentant le nombre d'individus de type  $A$  obtenus suit théoriquement une loi hypergéométrique, et en pratique on peut supposer que c'est une loi binomiale (quand  $n$  est très grand, un tirage avec ou sans remise importe peu). En fait, on peut même considérer que c'est une loi de Poisson de paramètre  $\lambda = np$ . Ceci se démontre en utilisant la formule de Stirling  $n! \sim \sqrt{2\pi n} (n/e)^n$  quand  $n \rightarrow \infty$ .

## Loi géométrique

La loi géométrique de paramètre  $\alpha \in ]0, 1[$  est la mesure de probabilité sur  $\mathbb{N}$  définie par  $\mathbf{P}(n) = (1-\alpha)\alpha^n$ . C'est la probabilité d'obtenir une boule noire après  $n$  boules rouges lorsque la probabilité d'obtenir une boule rouge est  $\alpha$ . (NB : la suite  $(\mathbf{P}(n))_{n \in \mathbb{N}}$  est une suite géométrique).

Il arrive (dans certains livres) que les paramètres  $\alpha$  et  $1 - \alpha$  soient échangés, ou bien même la loi soit définie sur  $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$  par  $\mathbf{P}(n) = \alpha(1-\alpha)^{n-1}$ ,  $n \geq 1$ . Par conséquent, on précise toujours les notations lorsqu'on parle de loi géométrique.

**Exercice 1.4.8** Vérifier que  $\mathbf{P}$  est une probabilité sur  $\mathbb{N}$  (ou  $\mathbb{N}^*$ ) pour chacune des conventions de notation ci-dessus.

**Exercice 1.4.9** Tracer la fonction de répartition d'une loi géométrique de paramètre  $\frac{1}{2}$ .

## 1.4.2 Lois continues

**Définition 1.4.10** Une v.a.r. a une loi continue si sa fonction de répartition est continue.

Les v.a. ci-dessous sont continues. Vérifiez-le sur chaque exemple.

## Loi uniforme

La loi uniforme sur l'intervalle  $[a, b] \subset \mathbb{R}$  muni de sa tribu borélienne) est la mesure de probabilité définie pour tout  $A \in \mathcal{B}([a, b])$  par

$$\mathbf{P}(A) = \frac{1}{b-a} \int_A dx = \frac{1}{b-a} \lambda(A)$$

où  $\lambda$  est la mesure de Lebesgue sur  $\mathbb{R}$ .

Sa fonction de répartition vérifie  $F(x) = \frac{1}{b-a} \int_{-\infty, x] \cap [a, b]} dx$ . D'où  $F(x) = 0$  si  $x \leq a$ ,  $F(x) = \frac{x-a}{b-a}$  si  $a \leq x \leq b$ ,  $F(x) = 1$  après. Dessinez le graphe de  $F$ . Qu'observez-vous ?

Notez que rien ne change si on considère l'intervalle  $(a, b)$  ouvert en  $a$  ou en  $b$ .

Plus généralement, on peut considérer la loi uniforme sur un ensemble  $E$  borélien quelconque de  $\mathbb{R}$ , définie pour tout  $A \in \mathcal{B}(E) \subset \mathcal{B}(\mathbb{R})$  par  $\mathbf{P}(A) = \frac{\lambda(A)}{\lambda(E)}$ .

La même définition est valide dans  $\mathbb{R}^d$ ,  $d \geq 2$ , en remplaçant la mesure de Lebesgue sur  $\mathbb{R}$  par la mesure de Lebesgue sur  $\mathbb{R}^d$ .

## Loi gaussienne, ou loi normale

D'une certaine façon, c'est la loi de probabilité la plus importante qui existe. Vous comprendrez pourquoi au dernier chapitre, dans le théorème limite central (TLC). Cette loi « universelle » modélise les écarts à la moyenne. Typiquement lorsqu'on mesure une grandeur physique qui a une valeur théorique attendue, il y a toujours une erreur, dont les causes peuvent être multiples et dues à des sommes de nombreuses causes microscopiques ou macroscopiques (instrument de mesure, précision de l'expérience, etc) et qu'on modélise souvent par une variable aléatoire de loi normale. En d'autres termes, la loi normale modélise des erreurs macroscopiques dues à une somme d'erreurs microscopiques. Vous comprendrez pourquoi lorsqu'on étudiera le TLC.

La loi normale centrée réduite notée  $\mathcal{N}(0, 1)$  sur  $\mathbb{R}$  donne à tout borélien  $A$  de  $\mathbb{R}$  la probabilité

$$\mathbf{P}(A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2} dx$$

**Exercice 1.4.11** Vérifier que  $\mathbf{P}$  est une probabilité sur  $\mathbb{R}$ .

La loi normale de moyenne  $m$ , d'écart-type  $\sigma$  (variance  $\sigma^2$ ) est notée  $\mathcal{N}(m, \sigma)$ . Elle est définie par

$$\mathbf{P}(A) = \frac{1}{\sqrt{2\pi}\sigma} \int_A e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

## Loi exponentielle

La loi exponentielle de paramètre  $a$ , notée parfois  $\mathcal{E}(a)$ , vérifie pour tout  $A \in \mathcal{B}(\mathbb{R}_+)$

$$\mathbf{P}(A) = \int_A a e^{-ax} \mathbf{1}_{\mathbb{R}_+}(x) dx.$$

**Exercice 1.4.12** Vérifier que  $\mathbf{P}$  est une probabilité sur  $\mathbb{R}$ .

**Exercice 1.4.13** Tracer le graphe de la fonction de répartition d'une loi exponentielle de paramètre 1.

Cette loi permet de modéliser les processus sans mémoire, qui ne vieillissent pas, les durées de vie. Par exemple, c'est la loi de la variable aléatoire  $X$  qui décrit le temps que met une particule instable à se désintégrer. Elle vérifie la propriété suivante.

**Proposition 1.4.14** Soit  $X$  une v.a. réelle. Alors  $X$  suit une loi exponentielle si et seulement si pour tous  $x > 0$  et  $y > 0$ ,

$$\mathbf{P}(X > x + y | X > y) = \mathbf{P}(X > x).$$

**Démonstration :** EXO. Voir TD

□

La durée de vie d'un néon suit une loi exponentielle. Par conséquent, inutile de changer un néon qui fonctionne encore...

## Loi de Cauchy

Elle est définie pour tout  $A \in \mathcal{B}(\mathbb{R}_+)$  par

$$\mathbf{P}(A) = \frac{1}{\pi} \int_A \frac{1}{1+x^2} dx.$$

**Exercice 1.4.15** Vérifier que  $\mathbf{P}$  est une probabilité sur  $\mathbb{R}$ .

Nous nous en servons souvent comme contre-exemple.

Si  $U$  suit une loi uniforme sur  $]-\frac{\pi}{2}, \frac{\pi}{2}[$  alors  $\tan U$  suit une loi de Cauchy.

## 1.5 Variables aléatoires à densité

### 1.5.1 Définitions et premières propriétés

**Définition 1.5.1** Soit  $\mu$  une probabilité définie sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Elle est à densité, ou absolument continue par rapport à la mesure de Lebesgue  $\lambda$  s'il existe une fonction  $f : \mathbb{R} \rightarrow \mathbb{R}_+$  Lebesgue-intégrable et positive, telle que pour tout borélien  $A$  de  $\mathbb{R}$ ,

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

On dit que  $f$  est une densité de  $\mu$  par rapport à  $\lambda$ .

Remarquons qu'une densité  $f$  vérifie

$$\int_{\mathbb{R}} f d\lambda = 1.$$

**Remarque 1.5.2** Une mesure de probabilité peut avoir plusieurs densités. Il n'y a donc pas unicité. Cela dit, si  $f$  et  $g$  sont deux densités de  $\mu$  alors  $f = g$   $\mu$ -presque sûrement. Par exemple, la loi uniforme sur  $[a, b]$  a pour densités (entre autres)  $\frac{1}{b-a} \mathbf{1}_{[a,b]}$  mais aussi  $\frac{1}{b-a} \mathbf{1}_{]a,b]}$ , ou  $\frac{1}{b-a} \mathbf{1}_{[a,b[}$ , ou ... Cela dit, il y a unicité de la densité dans  $L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ .

En fait, cette définition a parfaitement un sens pour des mesures positives  $\mu$  quelconques, et pas nécessairement des mesures de probabilité. Dans ce cas, la seule chose qui change est que la densité  $f$  est positive mesurable mais non nécessairement Lebesgue-intégrable.

**Remarque 1.5.3** La définition est inchangée dans  $\mathbb{R}^d$ ,  $d \geq 2$ . Relisez tout le paragraphe en vérifiant ce qui s'applique dans  $\mathbb{R}^d$ ,  $d \geq 2$ .

**Exercice 1.5.4** Vérifier que toutes les lois du paragraphe 1.4.2 sont à densité, et donner leur densité.

**Proposition 1.5.5** Si  $\mathbf{P}$  est une mesure de probabilité à densité  $f$  sur  $\mathbb{R}$ , alors sa fonction de répartition  $F$  est continue sur  $\mathbb{R}$  et vérifie pour tous  $a \leq b$  réels

$$F(b) - F(a) = \mathbf{P}(]a, b]) = \int_{]a, b]} f d\lambda$$

**Démonstration :** Par définition,  $F(b) - F(a) = \mathbf{P}(]-\infty, b]) - \mathbf{P}(]-\infty, a])$ . L'additivité de  $\mathbf{P}$  donne  $F(b) - F(a) = \mathbf{P}(]a, b]) = \int_{]a, b]} f d\lambda$  (la dernière égalité résultant du fait que  $\mathbf{P}$  est à densité).

Cette expression permet de déduire la continuité de  $F$ . En effet,  $F(x) = F(a) + \int_{]a, x]} f d\lambda$  pour n'importe quel  $a$  fixé. Et votre cours d'intégration vous permet d'affirmer qu'une fonction définie par une intégrale est continue. (Exo : vérifier cela dans votre cours d'intégration).  $\square$

**Proposition 1.5.6** Si  $X$  a une fonction de répartition  $F$  qui est  $C^1$  (respectivement  $C^1$  par morceaux, i.e.  $C^1$  en dehors d'un nombre fini de discontinuités), alors  $X$  a pour densité  $f = F'$ , et  $f$  est continue (resp. Continue par morceaux).

**Démonstration :** Soit  $f = F'$ . Cette fonction est définie partout sauf au plus en un nombre fini de points, elle est continue par morceaux, positive (car  $F$  est croissante), et  $\int_{\mathbb{R}} f d\lambda = F(\infty) - F(-\infty) = 1$ . Soit donc  $\mu$  la mesure de probabilité de densité  $f = F'$ . Cette mesure coïncide avec  $\mathbf{P}_X$  sur tous les intervalles  $]a, b]$ . Le corollaire 1.2.4 donne  $\mu = \mathbf{P}_X$ .  $\square$

**Exercice 1.5.7** Donner l'allure du graphe des densités et des fonctions de répartition de tous les exemples du paragraphe 1.4.2.

**Remarque 1.5.8** Si  $\mu$  est à densité,  $f$  est mesurable mais pas nécessairement continue. C'est  $F$  qui est continue.

**Remarque 1.5.9** La réciproque de la proposition 1.5.5 est fautive : il existe des lois  $\mu$  dont la fonction de répartition est continue mais qui n'ont pas de densité. Autrement dit, il existe des lois continues mais pas absolument continues.

Une façon d'obtenir de telles lois est d'utiliser la construction ci-dessous. On construit une fonction  $F$  continue croissante sur  $[0, 1]$ , avec  $F(0) = 0$  et  $F(1) = 1$ , mais qui est constante presque partout ! Si la loi correspondante était à densité, sa densité serait nulle, donc la loi serait nulle, contradiction avec le fait que c'est une probabilité.

L'un de ces monstres, appelé *escalier du diable*, est construit de la manière suivante.

À l'étape 0, on définit  $F_0(x) = 0$  si  $x \leq 0$ ,  $F_0(x) = 1$  si  $x \geq 1$  et  $F_0(x) = x$  sur  $[0, 1]$ .  $F_0$  est continue, croissante, et a pour limites 0 en  $-\infty$  et 1 en  $+\infty$ .

À l'étape 1, on découpe  $[0, 1]$  en trois intervalles égaux de longueur  $\frac{1}{3}$ . Sur l'intervalle du milieu  $[1/3, 2/3]$  on pose  $F_1(x) = \frac{1}{2}$ . Sur les deux intervalles  $[0, 1/3]$  et  $[2/3, 1]$ , on prolonge  $F_1$  en une fonction affine par morceaux et continue :  $F_1(x) = \frac{3}{2}x$  sur  $[0, 1/3]$  et  $F_1(x) = \frac{3}{2}x - \frac{1}{2}$  sur  $[2/3, 1]$ .

À l'étape 2, on recommence sur chacun des intervalles  $[0, 1/3]$  et  $[2/3, 1]$ . (En dehors, sur  $]-\infty, 0] \cup [1/3, 2/3] \cup [1, +\infty[$ , on garde  $F_2(x) = F_1(x)$ .) Sur l'intervalle  $[1/9, 2/9]$  on pose  $F_2(x) = \frac{1}{4}$ , sur l'intervalle  $[7/9, 8/9]$   $F_2(x) = \frac{3}{4}$  et sur les quatre intervalles  $[0, \frac{1}{9}]$ ,  $[\frac{2}{9}, \frac{1}{3}]$ ,  $[\frac{2}{3}, \frac{7}{9}]$ ,  $[\frac{8}{9}, 1]$ , on prolonge  $F_2$  en une fonction affine par morceaux continue croissante.

Et on continue la construction... Par récurrence, on construit une suite  $(F_n)$  de fonctions continues sur  $\mathbb{R}$ , croissantes, valant 0 avant 0 et 1 après 1, et qui sont constantes par morceaux sur une union d'intervalles  $I_n$  de longueur de plus en plus grande.

À la limite,  $(F_n)$  converge. En effet, il est élémentaire de vérifier que pour tout  $n \geq 1$ , on a  $\|F_{n+1} - F_n\| \leq \frac{1}{2^n}$ . On en déduit aisément que pour tout  $p \geq 1$ , on a  $\|F_{n+p} - F_n\| \leq \frac{1}{2^{n-1}}$ . En particulier,  $(F_n)_{n \in \mathbb{N}}$  est une suite de Cauchy pour la topologie de la norme infinie sur les fonctions continues sur  $[0, 1]$ . Elle converge donc et sa limite, notée  $F$  est continue. On l'appelle *l'escalier du diable* : elle croît sur  $[0, 1]$  de 0 à 1, de façon continue, en étant constante presque partout ! C'est la fonction de répartition d'une mesure de probabilité  $\mathbf{P}$  qui n'admet pas de densité.

Voici une proposition peu difficile dont l'intérêt apparaîtra au paragraphe 1.5.2.

**Proposition 1.5.10** La variable aléatoire réelle  $X$  a pour densité la fonction mesurable positive  $f$  si et seulement si pour toute fonction  $h : \mathbb{R} \rightarrow \mathbb{R}$  mesurable bornée, on a

$$\int_{\mathbb{R}} h(x) d\mathbf{P}_X(x) = \int_{\mathbb{R}} h(x) f(x) d\lambda(x)$$

**Démonstration :** Si pour toute  $h : \mathbb{R} \rightarrow \mathbb{R}$  mesurable bornée,  $\int_{\mathbb{R}} h(x) d\mathbf{P}_X(x) = \int_{\mathbb{R}} h(x) f(x) d\lambda(x)$ , alors pour  $h = \mathbf{1}_A$ ,  $A$  borélien quelconque, on obtient  $P_X(A) = \int_A f(x) d\lambda(x)$ . Donc  $\mathbf{P}_X$  est une loi à densité  $f$ .

Réciproquement, si  $\mathbf{P}_X$  a pour densité  $f$ , l'égalité cherchée est vraie pour toutes les fonctions  $h = \mathbf{1}_A$  indicatrices de boréliens  $A$  (par définition d'une loi à densité). Par linéarité, l'égalité est encore vraie pour toutes les fonctions simples  $h = \sum_i c_i \mathbf{1}_{A_i}$  (somme finie). (Par convergence monotone, l'égalité est vraie pour toutes les fonctions  $h$  mesurables positives. Inutile ici, mais vrai.) Par convergence dominée, l'égalité est vraie pour toutes les fonctions intégrables, et en particulier toutes les fonctions mesurables bornées.  $\square$

**Exercice 1.5.11** Soit  $X : \Omega = [0, 1] \rightarrow \mathbb{R}$  définie par  $X(\omega) = 2\omega + 1$ . Donner la loi (et la densité) de  $X$ .

On calcule  $\int_{\mathbb{R}} h d\mathbf{P}_X = \int_{[0,1]} h(2\omega + 1) d\omega = \int_{[0,3]} h(w) \frac{dw}{2} = \int_{\mathbb{R}} h(x) \frac{\mathbf{1}_{[0,3]}(x)}{2} dx$ . La densité de  $X$  est donc  $\frac{\mathbf{1}_{[0,3]}}{2}$ .

Finissez ce paragraphe par une question que se posent souvent les mathématicien-nes : qu'est-ce qui caractérise une densité ? Autrement dit, quelles fonctions sont des densités de lois de probabilité ?



**Proposition 1.5.12** Si  $f : (\mathbb{R}, \mathcal{B}(\mathbb{R}) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$  est une fonction mesurable positive avec  $\int_{\mathbb{R}} f d\lambda = 1$  alors il existe une mesure de probabilité  $\mu$  sur  $\mathbb{R}$  qui admet  $f$  pour densité.

**Démonstration :** Fait dans le module intégration. L'idée est complètement élémentaire : si  $A$  est un borélien de  $\mathbb{R}$ , on définit  $\mu(A) = \int_A f d\lambda$ . Vérifiez en exercice que cela définit bien une mesure de probabilité (à densité évidemment) sur  $\mathbb{R}$ .  $\square$

fin cours 3 (fait aussi le paragraphe 3.2)

## 1.5.2 Changement de variables

Dans ce paragraphe, nous partons d'une variable aléatoire  $X$  à densité, et nous cherchons à savoir si la variable aléatoire  $Y = \varphi(X)$  est à densité, et si oui, quelle est sa densité ?

Nous utiliserons la proposition 1.5.10. La stratégie est la suivante. On sait que  $X$  a pour densité  $f_X$ . On applique cette proposition 1.5.10. Soit  $h$  une fonction mesurable bornée ; alors  $h \circ \varphi$  est encore mesurable bornée. On déduit de 1.5.10 que pour toute fonction mesurable bornée du type  $h \circ \varphi$ , on a

$$\int_{\Omega} h \circ \varphi \circ X(\omega) d\mathbf{P}(\omega) = \int_{\mathbb{R}} h \circ \varphi(x) d\mathbf{P}_X(x) = \int_{\mathbb{R}} h \circ \varphi(x) f_X(x) d\lambda(x).$$

Remarquons que le premier terme vaut encore  $\int_{\Omega} h \circ Y(\omega) d\mathbf{P}(\omega) = \int_{\mathbb{R}} h(y) d\mathbf{P}_Y(y)$ .

On essaie alors d'appliquer le théorème du changement de variables (en posant  $y = \varphi(x)$ ) au terme  $\int_{\mathbb{R}} h \circ \varphi(x) f_X(x) d\lambda(x)$ , en espérant le mettre sous la forme  $\int_{\mathbb{R}} h(y) f_Y(y) d\lambda(y)$ , avec  $f_Y$  une fonction mesurable positive. La proposition 1.5.10 permet alors de conclure que  $f_Y$  est la densité de  $\mathbf{P}_Y$ .

Voyons cela sur un exemple très détaillé. Cherchons la densité de la loi de  $Y = X + 1$ , si  $X$  est une v.a. à densité  $f_X$ . On veut utiliser la proposition 1.5.10. On étudie donc la quantité  $\int_{\mathbb{R}} h \circ Y(y) d\mathbf{P}_Y(y)$  qui vaut  $\int_{\Omega} h \circ Y(\omega) d\mathbf{P}(\omega)$  (théorème 1.3.8), puis  $\int_{\Omega} h \circ \varphi \circ X(\omega) d\mathbf{P}(\omega)$  avec  $\varphi(x) = x + 1$ . Cette intégrale vaut encore (théorème 1.3.8 appliqué à  $X$ )  $\int_{\mathbb{R}} h \circ \varphi(x) d\mathbf{P}_X(x)$ , puis  $\int_{\mathbb{R}} h(x + 1) f_X(x) dx$  (car  $X$  est à densité). Le théorème de changement de variables donne  $\int_{\mathbb{R}} h(x + 1) f_X(x) dx = \int_{\mathbb{R}} h(x) f_X(x - 1) dx$ . Finalement, pour toute fonction  $h$  mesurable bornée, on a montré que  $\int_{\mathbb{R}} h(y) d\mathbf{P}_Y(y) = \int_{\mathbb{R}} h(y) f_X(y - 1) dy$ . La proposition 1.5.10 nous permet de conclure que  $Y = X + 1$  est à densité et que  $f_Y(y) = f_X(y - 1)$ .

Rappelons le théorème du changement de variables :

**Théorème 1.5.13 (Changement de variable)** Soit  $\varphi : U \subset \mathbb{R}^d \rightarrow V \subset \mathbb{R}^d$  un difféomorphisme de classe  $C^1$  entre deux ouverts, et  $g : V \rightarrow \mathbb{R}$  Lebesgue intégrable. Alors

$$\int_V g(y) dy = \int_U g(\varphi(x)) |\text{Jac } \varphi(x)| dx.$$

Pour la démonstration, voir votre cours d'intégration.

### Un exemple fondamental : le passage en coordonnées polaires

Soit  $U = ]0, 2\pi[ \times \mathbb{R}_+^*$  et  $V = \mathbb{R}^2 \setminus \{(x, 0), x \geq 0\}$ . Et  $\varphi(r, \theta) = (r \cos \theta, r \sin \theta)$ . Bien sûr,  $|\text{Jac } (\varphi)(r, \theta)| = r$ . Si  $g : V \rightarrow \mathbb{R}$  est intégrable ou positive, alors

$$\int_V g(x, y) dx dy = \int_U g(r \cos \theta, r \sin \theta) dr r d\theta.$$

Application : calcul de  $\int_{\mathbb{R}^2} e^{-x^2/2} dx$ . Indication : Passer en coordonnées polaires dans l'intégrale  $\int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy$ .

Remarque très importante en pratique : ci-dessus, l'application  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  n'est pas un difféomorphisme (car pas bijective) mais quitte à enlever des ensembles de mesure nulle, l'application  $\varphi|_U$  est un difféomorphisme.

## Reformulation de ce qui précède

On peut donner un énoncé de changement de variable probabiliste, comme ci-dessous. Néanmoins, il est sans doute plus utile de retenir et savoir refaire sa preuve que de retenir son énoncé.

**Proposition 1.5.14** Soit  $\varphi$  un difféomorphisme d'un ouvert  $U$  de  $\mathbb{R}^d$  dans un ouvert  $V$  de  $\mathbb{R}^d$ . Si  $X$  est une v.a. qui prend p.s. ses valeurs dans  $U$  et qui a pour densité  $p$ , alors la v.a.  $Y = \varphi \circ X$  a pour densité  $p \circ \varphi^{-1} \times |J_{\varphi^{-1}}|$ .

**Démonstration :** Soit  $h$  une fonction mesurable bornée définie sur  $V$ . A l'aide du théorème de transfert et du théorème de changement de variable, on montre que

$$\begin{aligned} \int_V h(y) d\mathbf{P}_Y &= \int_{\Omega} h \circ Y(\omega) d\mathbf{P}(\omega) = \int_{\Omega} h \circ \varphi \circ X(\omega) d\mathbf{P}(\omega) = \int_U h \circ \varphi(x) d\mathbf{P}_X(x) \\ &= \int_U h \circ \varphi(x) p(x) dx = \int_V h(y) p \circ \varphi^{-1}(y) |J_{\varphi^{-1}}(y)| dy \end{aligned}$$

La proposition 1.5.10 permet de conclure. □

**Exercice 1.5.15** Soit  $X$  une v.a.r. de loi exponentielle et  $Y = X^2$ . Donner la densité de  $Y$ .

Soit  $\varphi : x \mapsto x^2$ . C'est un difféo de  $\mathbb{R}_+^*$  dans  $\mathbb{R}_+^*$ , et de  $\mathbb{R}_-^*$  dans  $\mathbb{R}_+^*$ . Et on a

$$\begin{aligned} \int_{\mathbb{R}} h(y) d\mathbf{P}_Y(y) &= \int_{\Omega} h \circ Y d\mathbf{P} = \int_{\Omega} h \circ \varphi \circ X d\mathbf{P} = \int_{\mathbb{R}} h \circ \varphi(x) f_X(x) dx \\ &= \int_{\mathbb{R}_+^*} + \int_{\mathbb{R}_-^*} \dots = \int_{\mathbb{R}_+^*} h(y) \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) dy \end{aligned}$$

**Exercice 1.5.16** Exemple à détailler. Soit  $X$  une variable aléatoire réelle de densité  $f_X(x) = e^{-x} 1_{\mathbb{R}_+}(x)$ ,  $Y = |\log X|$ , et  $\varphi(x) = |\log x|$ . On pose  $\varphi_1(x) = \log x$  de  $U_1 = ]1, +\infty[$  dans  $\mathbb{R}_+^*$ , et  $\varphi_2(x) = -\log x$  de  $U_2 = ]0, 1[$  dans  $\mathbb{R}_+^*$ . On a bien sûr  $\varphi_1^{-1}(y) = e^y$  et  $\varphi_2^{-1}(y) = e^{-y}$ . La proposition ci-dessus donne

$$f_Y(y) = \left( e^{-e^y} e^y + e^{-e^{-y}} e^{-y} \right) 1_{\mathbb{R}_+^*}(y).$$

# Chapitre 2

## Les grands outils probabilistes

Dans ce chapitre, nous introduisons les notions d'espérance, variance, fonction caractéristique, et nous apprenons à les manipuler (calculs). Nous verrons leur utilité aux chapitres suivants.

### 2.1 Espérance, variance, écart-type

#### 2.1.1 Définition

L'espérance d'une variable aléatoire  $X : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  ou  $X : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (\mathbb{N}, \mathcal{P}(\mathbb{N}))$  est la moyenne de toutes ses valeurs possibles pondérées par la probabilité  $\mathbf{P}$ , i.e. la valeur qu'on peut espérer que  $X$  prenne.

Commençons par les variables aléatoires à valeurs entières.

**Définition 2.1.1** Si  $X : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (\mathbb{N}, \mathcal{P}(\mathbb{N}))$  est une v.a.  $\mathbf{P}$ -intégrable, son espérance est définie par

$$E(X) := \int_{\Omega} X(\omega) d\mathbf{P}(\omega) = \sum_{\mathbb{N}} k \mathbf{P}_X(k) = \sum_{\mathbb{N}} k \mathbf{P}(X = k).$$

Si  $X$  est une v.a. positive, son espérance est encore définie comme ci-dessus, mais peut éventuellement être infinie. Si  $\Omega$  est discret (cas fréquent), ceci se réécrit

$$E(X) := \sum_{\omega \in \Omega} X(\omega) \mathbf{P}(\omega) = \sum_{\mathbb{N}} k \mathbf{P}_X(k) = \sum_{\mathbb{N}} k \mathbf{P}(X = k).$$

Un exemple : on lance deux dés équilibrés.  $\Omega = \{1, \dots, 6\}^2$  muni de la probabilité uniforme  $\mathbf{P}$ . On regarde la somme des deux chiffres :  $X(i, j) = i + j$ . Alors  $E(X) = \dots = 7$ . (Exo).

Traitons maintenant le cas d'une v.a. à valeurs réelles.

**Définition 2.1.2** Si  $X : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  est une v.a.  $\mathbf{P}$ -intégrable, son espérance est définie par

$$E(X) := \int_{\Omega} X(\omega) d\mathbf{P}(\omega) = \int_{\mathbb{R}} x d\mathbf{P}_X(x).$$

Si  $X$  est une v.a. positive, son espérance est encore définie comme ci-dessus, mais peut éventuellement être infinie.

Exemple :  $X : \Omega = [0, 1] \rightarrow \Omega$  définie par  $X(\omega) = \omega$ ,  $\mathbf{P} = \mathbf{P}_X$  loi uniforme sur  $[0, 1]$ . Alors  $E(X) = \frac{1}{2}$ .

L'espérance n'existe pas toujours, ou n'est pas forcément finie. Dans ce cas, on dit que  $X$  n'est pas intégrable. Dans le vocabulaire des analystes, l'espérance de  $X$  existe et est finie ssi  $X \in L^1(\Omega, \mathbf{P})$ . Notez aussi que  $\|X\|_{L^1(\Omega)} = E(|X|)$ .

Lorsque  $X$  est une v.a. positive,  $E(X)$  a toujours un sens, mais peut être infinie.

**Exercice 2.1.3** Montrez qu'une v.a.  $X$  qui suit une loi de Cauchy sur  $\mathbb{R}$  n'est pas intégrable. Son espérance n'est pas définie.

**Remarque 2.1.4** On parle d'espérance de v.a. car les v.a. sont ce qui intéresse le plus les probabilistes, mais en réalité, l'espérance est une quantité qui ne dépend que de la loi de  $X$ , et pas de la v.a. Autrement dit, deux v.a. de même loi ont même espérance, et on peut parler de la valeur moyenne d'une mesure de probabilité  $\mu$  (définie comme  $\int_{\mathbb{R}} x d\mu(x)$ ) plutôt que de l'espérance d'une v.a.  $X$  de loi  $\mathbf{P}_X = \mu$ .

Les propriétés élémentaires de l'espérance sont essentiellement celles de l'intégrale, en ajoutant le fait qu'on intègre par rapport à une mesure de masse 1. Soient donc des v.a. à valeurs dans  $\mathbb{R}$  ou  $\mathbb{N}$ , définies sur  $(\Omega, \mathcal{A}, \mathbf{P})$ .

1.  $E(X + Y) = E(X) + E(Y)$
2.  $E(\lambda X) = \lambda E(X)$
3.  $E(\mathbf{1}_A) = \mathbf{P}(A)$
4.  $X$  intégrable ssi  $|X|$  intégrable.
5. Si  $X$  intégrable et  $\mathbf{P}_X(\mathbb{R}_+) = \mathbf{P}(X \geq 0) = 1$  alors  $E(X) \geq 0$

**Exercice 2.1.5** *Démontrer ces propriétés. Indication pour la dernière propriété : intégrer sur les ensembles  $\{\omega \in \Omega, X(\omega) \geq 0\}$  et  $\{\omega \in \Omega, X(\omega) < 0\}$ .*

Reformulons la proposition 1.5.10.

**Proposition 1.5.10** *La variable aléatoire réelle  $X$  a pour densité la fonction mesurable positive  $f$  si et seulement si pour toute fonction  $h : \mathbb{R} \rightarrow \mathbb{R}$  mesurable bornée, on a*

$$E(h(X)) = \int_{\mathbb{R}} h(x)f(x)d\lambda(x)$$

## 2.1.2 Exemples

### v.a. discrètes

Si  $X$  suit une loi uniforme sur  $\{0, 1\}$ ,  $E(X) = \frac{1}{2}$ .

Si  $X$  suit une loi de Bernoulli  $\mathcal{B}(p)$  alors  $E(X) = p$ .

Si  $X$  suit une loi binomiale  $\mathcal{B}(n, p)$ , alors  $E(X) = np$ .

Si  $X$  suit une loi de Poisson  $\mathcal{P}(\lambda)$  alors  $E(X) = \lambda$ .

Si  $X$  suit une loi géométrique ( $\mathbf{P}(X = k) = p^k(1 - p)$ ) alors  $E(X) = \frac{p}{1-p}$ .

### v.a. continues

Si  $X$  suit une loi uniforme sur  $[a, b]$ , alors  $E(X) = \frac{a+b}{2}$ .

Si  $X$  suit une loi normale centrée réduite  $E(X) = 0$

Si  $X$  suit une loi normale de moyenne  $m$   $E(X) = m$ .

Si  $X$  suit une loi exponentielle de paramètre  $a$   $E(X) = \frac{1}{a}$ .

Si  $X$  suit une loi de Cauchy, alors  $X$  n'est pas intégrable ( $E(X)$  n'est pas définie.)

NB la loi exp modélise les durées de vie :  $\frac{1}{a}$  représente la durée de vie de la particule considérée.

## 2.1.3 Variance, écart-type

Les quantités introduites dans ce paragraphe servent beaucoup en statistiques.

**Définition 2.1.6** *Soit  $X$  une v.a. intégrable, à valeurs entières ou réelles. Sa variance est la quantité  $\text{Var}(X) = E((X - E(X))^2)$ . Lorsque  $X$  est intégrable, la variance de  $X$  existe toujours dans  $\mathbb{R}_+ \cup \{+\infty\}$ . L'écart-type de  $X$  est  $\sigma(X) = \sqrt{\text{var}(X)}$ .*

Justifier si nécessaire cette définition.

**Remarque 2.1.7** Si  $X$  est une v.a.r. de carré intégrable, alors par l'inégalité de Cauchy-Schwartz, on déduit  $E(|X|) \leq \sqrt{E(X^2)}\sqrt{E(1)}$ . Or  $E(1) = \mathbf{P}_X(\mathbb{R}) = 1$ . Donc  $E(X^2) < \infty$  implique  $E(|X|) < \infty$ , i.e.  $X$  intégrable (et aussi bien sûr  $\text{Var}(X) < \infty$ ). Autrement dit, en théorie des probabilités, on a l'inclusion  $L^2(\Omega, \mathcal{A}, \mathbf{P}) \subset L^1(\Omega, \mathcal{A}, \mathbf{P})$ . Cette inclusion bien pratique n'est pas vérifiée en analyse sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ , ni plus généralement sur les espaces mesurés de mesure infinie.

Cette inclusion est stricte. En effet, il existe des exemples de variables aléatoires intégrables mais dont la variance est infinie.

La variance représente la moyenne des carrés des écarts à la moyenne. On mesure ainsi la façon dont  $X$  prend des valeurs lointaines ou proches de la moyenne  $E(X)$ , souvent ou non. On aurait pu s'intéresser à la quantité  $E(|X - E(X)|)$  a priori, mais les carrés sont plus aisés à manipuler dans les calculs que les  $|\cdot|$ .

On vérifie que

$$\text{Var}(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

**Proposition 2.1.8**  $\text{Var}(X) = 0$  ssi  $X = E(X)$  p.s. ssi  $X$  est constante p.s.

**Démonstration :** Exo

□

## Variance des lois classiques

**Exercice 2.1.9** Calculer les variances de v.a. suivant les lois classiques introduites plus haut. Ce type d'exercice calculatoire fera à coup sûr l'objet de questions considérées (par moi) comme très faciles le jour du partiel ou de l'examen.

Indication : pour certaines lois, il est utile de calculer d'abord  $E(X(X - 1))$  puis  $\text{Var}(X) = E(X(X - 1) + E(X) - (E(X))^2$

$X$  suit une loi uniforme sur  $\{0, 1\}$  :  $\text{Var}(X) = \frac{1}{4}$

$X$  suit une loi  $\mathcal{B}(p)$  sur  $\{0, 1\}$  :  $\text{Var}(X) = p(1 - p)$ .

$X$  suit une loi binomiale  $\mathcal{B}(n, p)$  :  $\text{Var}(X) = np(1 - p)$ .

$X$  suit une loi de Poisson  $\mathcal{P}(\lambda)$  :  $\text{Var}(X) = \lambda$ .

$X$  suit une loi géométrique  $\mathcal{G}(p)$  :  $\text{Var}(X) = \frac{p}{(1-p)^2}$

$X$  suit une loi uniforme sur l'intervalle  $[a, b]$  :  $\text{Var}(X) = \frac{(a-b)^2}{12}$

$X$  suit une loi normale  $\mathcal{N}(m, \sigma^2)$  :  $\text{Var}(X) = \sigma^2$

$X$  suit une loi exponentielle de paramètre  $a$  :  $\text{Var}(X) = \frac{1}{a^2}$

## 2.1.4 Corrélation de deux variables aléatoires

**Définition 2.1.10** Soient  $X$  et  $Y$  deux v.a. réelles intégrables. La covariance de  $X$  et  $Y$ , ou coefficient de corrélation, est définie par

$$\text{Cov}(X, Y) := E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Vérifiez l'égalité de droite.

Lorsque  $\text{Cov}(X, Y) = 0$ , on dit que  $X$  et  $Y$  sont non corrélées.

**Lemme 2.1.11** Si  $X$  et  $Y$  sont indépendantes, alors elles sont non corrélées.

**Démonstration :** Voir chapitre 3.

□

**Remarque 2.1.12** La réciproque est fautive. Voir ch 3

## 2.1.5 Moments

**Définition 2.1.13** On dit que  $X$  admet un moment d'ordre  $k \geq 1$ , si  $|X|^k$  est intégrable, soit encore  $E(|X|^k) < \infty$ . Son moment d'ordre  $k$  est alors  $E(X^k)$ .

En analyse, on dirait que  $X$  est dans  $L^k(\Omega, \mathbf{P})$ .

### 2.1.6 Inégalités classiques

Les inégalités de ce paragraphe sont fondamentales et élémentaires. Les preuves sont à retenir au moins autant, si ce n'est plus, que les énoncés.

**Lemme 2.1.14 (Inégalité de Markov)** Soient  $X$  une v.a. réelle et  $a > 0$  un réel positif. Alors pour tout réel  $p > 0$  on a

$$P(|X| \geq a) \leq \frac{E(|X|^p)}{a^p}.$$

**Démonstration :**  $E(|X|) = \int_{\Omega} |X| d\mathbf{P} = \int_{\{|X| \geq a\}} |X| d\mathbf{P} + \int_{\{|X| < a\}} |X| d\mathbf{P}$ . Or vu l'ensemble sur lequel on intègre, on a  $\int_{\{|X| \geq a\}} |X| d\mathbf{P} \geq a\mathbf{P}(|X| \geq a)$ . Et le dernier terme  $\int_{\{|X| < a\}} |X| d\mathbf{P}$  est positif. Autrement dit on a  $E(|X|) \geq a\mathbf{P}(|X| \geq a)$ .

Pour l'inégalité avec  $p$ , le raisonnement est identique (exo).  $\square$

**Lemme 2.1.15 (Inégalité de Bienaymé-Tchebychev)** Soit  $X$  une v.a.r. de carré  $X^2$  intégrable. Alors pour tout  $a > 0$ , on a

$$P(|X - E(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}$$

**Démonstration :** Appliquer l'inégalité de Markov à  $Y = X - E(X)$  avec  $p = 2$ .  $\square$

Cette inégalité illustre bien comment la variance d'une variable aléatoire permet de contrôler les écarts à la moyenne de la v.a.

## 2.2 Fonctions génératrices

Dans cette section, on présente un outil élémentaire très utile dans l'étude des v.a. à valeurs dans  $\mathbb{N}$ . On verra une généralisation aux v.a. r. au paragraphe suivant.

Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{N}$ . Remarquons que par définition de l'espérance, on a

$$E(s^X) = \sum_{n=0}^{\infty} \mathbf{P}(X = n) s^n \quad \text{pour tout } s \geq 0$$

**Définition 2.2.1** La fonction génératrice de  $X$  est la fonction  $g$  définie par

$$g_X(s) = \sum_{n=0}^{\infty} \mathbf{P}(X = n) s^n = E(s^X) \quad \text{pour tout } 0 \leq s \leq 1$$

Posons  $q_n = \mathbf{P}(X = n)$ . On a alors  $g_0 = g_X(0)$ ,  $g_X(1) = 1$ , et  $\lim_{s \rightarrow 1} g_X(s) = g_X(1) = 1$  par convergence monotone. On vérifie que sur  $[0, 1]$ ,  $g_X$  est convexe. La série entière  $g_X$  a un rayon de convergence  $R \geq 1$ , puisque  $g_X(1) = 1$ . Donc  $g_X$  est  $C^\infty$  sur  $[0, 1[$ . Ses dérivées successives vérifient  $g'_X(s) = \sum_{n=1}^{\infty} n q_n s^{n-1}$ ,  $g''_X(s) = \sum_{n=2}^{\infty} n(n-1) q_n s^{n-2}$  ... On en déduit aisément

**Proposition 2.2.2** La fonction génératrice vérifie pour tout  $n \geq 0$ ,  $\mathbf{P}(X = n) = \frac{1}{n!} g_X^{(n)}(0)$ . En particulier, la fonction génératrice d'une variable aléatoire détermine la loi de  $X$ .

Donnons des exemples de calculs de fonctions génératrices.

Loi binomiale  $B(n, p)$ .

$$g(s) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} s^k = (ps + (1-p))^n$$

Loi de Poisson  $\mathcal{P}(\lambda)$

$$g(s) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k s^k}{k!} = e^{\lambda(s-1)}$$

Loi géométrique  $\mathcal{G}(a)$

$$g(s) = \sum_{k=0}^{\infty} (1-a) a^k s^k = \frac{1-a}{1-as}$$

## Calcul des moments

Rappelons que  $E(|X|^p) < \infty$  implique  $E(|X|^q) < \infty$  pour tout  $q \leq p$ . (Pourquoi? Exo!) Notons aussi qu'une v.a. à valeurs entières est positive!

**Proposition 2.2.3** Soit  $X$  une v.a. à valeurs dans  $\mathbb{N}$ .

- $E(X) < \infty$  ssi  $g_X$  est dérivable à gauche en 1, et dans ce cas,  $E(X) = g'_X(1)$ .
- $E(X^2) < \infty$  ssi  $g_X$  est deux fois dérivable à gauche en 1, et dans ce cas,  $E(X(X-1)) = g''_X(1)$ .

**Démonstration :** voir TD ou ouvrages classiques □

**Exercice 2.2.4** Calculer espérance et variance de v.a. de loi binomiale, Poisson, ou géométrique en utilisant la proposition ci-dessus.

**Remarque 2.2.5** Un grand intérêt des fonctions génératrices est que la fonction génératrice d'une somme de variables aléatoires indépendantes est le produit des fonctions génératrices. Ceci sera détaillé dans le cas des fonctions caractéristiques, plus générales.

## 2.3 Fonctions caractéristiques

### Intégrale de fonctions à valeurs complexes

Vous avez vu en intégration la définition et les propriétés de l'intégrale de fonctions à valeurs réelles. Mais tout s'étend sans aucun problème aux fonctions à valeurs complexes.

Soit  $f : \Omega \rightarrow \mathbb{C}$  une fonction qui s'écrit donc  $f = u + iv$ , avec  $u$  et  $v$  deux fonctions de  $\Omega$  dans  $\mathbb{R}$ . Si  $\mu$  est une mesure sur  $\Omega$ , on dit que  $f$  est intégrable pour  $\mu$  si  $\int_{\Omega} |f(\omega)| d\mu < +\infty$ . Remarquons que  $f$  est intégrable ssi  $\operatorname{Re} f$  et  $\operatorname{Im} f$  sont intégrables, puisque

$$\max(|\operatorname{Re} f|, |\operatorname{Im} f|) \leq |f| \leq |\operatorname{Re} f| + |\operatorname{Im} f|$$

On définit alors l'intégrale de  $f$  par

$$\int_{\Omega} f d\mu = \int u d\mu + i \int v d\mu.$$

Cette intégrale est un nombre complexe.

Les calculs se font... comme on pense!

$$\int_0^1 e^{iax} dx = \int_0^1 \cos ax + i \sin ax dx = \left[ \frac{e^{iax}}{ia} \right]_0^1 = \frac{e^{ia} - 1}{ia}.$$

Si  $f : \mathbb{R} \rightarrow \mathbb{C}$  est continue, alors  $u = \operatorname{Re} f$  et  $v = \operatorname{Im} f$  sont aussi continues. Elles ont donc des primitives, qu'on note  $U$  et  $V$ . Alors  $F = U + iV$  est une primitive de  $f$ . Et bien sûr  $\int_a^b f(x) dx = \int_a^b u(x) dx + i \int_a^b v(x) dx = [U + iV]_a^b = [F]_a^b = F(b) - F(a)$ .

### 2.3.1 Fonction caractéristique d'une variable aléatoire réelle

**Définition 2.3.1** Soit  $\mu$  une mesure finie sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  (i.e.  $\mu(\mathbb{R}) < +\infty$ ). Alors la transformée de Fourier de la mesure  $\mu$  est la fonction  $\hat{\mu} : \mathbb{R} \rightarrow \mathbb{C}$  définie pour tout  $t \in \mathbb{R}$  par

$$\hat{\mu}(t) = \int_{\mathbb{R}} e^{itx} d\mu(x).$$

La fonction  $\hat{\mu}$  est bien définie car  $x \rightarrow e^{itx}$  est continue donc mesurable, et bornée donc intégrable (car  $\mu$  est supposée de masse totale finie).

**Définition 2.3.2** Soit  $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  une v.a. réelle. Sa fonction caractéristique, notée  $\varphi_X$ , est la transformée de Fourier de la loi  $P_X$  de  $X$ . Autrement dit pour tout  $t \in \mathbb{R}$

$$\varphi_X(t) = \hat{P}_X(t) = \int_{\mathbb{R}} e^{itx} dP_X(x) = E(e^{itX}).$$

Remarquez que la convention est légèrement différente de celle utilisée pour la transformée de Fourier de fonctions en analyse.

**Remarque 2.3.3** Si  $X$  est une v.a. à densité  $f$ , alors sa fonction caractéristique n'est rien d'autre, au signe près, que la transformée de Fourier de  $f$  :  $\varphi_X(t) = \hat{f}(-t)$ .

**Proposition 2.3.4** Soit  $X$  une v.a. réelle.

1.  $\varphi_X(0) = 1$
2.  $|\varphi_X(t)| \leq 1$  pour tout  $t \in \mathbb{R}$ .
3.  $\varphi_{aX+b}(t) = e^{itb} \varphi_X(at)$
4.  $\varphi_X$  est à valeurs réelles si et seulement si  $P_X = P_{-X}$ .
5.  $\varphi_X$  est uniformément continue.
6. Soit  $p \geq 1$ . Si  $E(|X|^p) < +\infty$ , alors  $\varphi_X$  est de classe  $C^p$  et  $\varphi_X^{(k)}(0) = i^k E(X^k)$  pour tout  $1 \leq k \leq p$ .

**Démonstration :** \* D'abord  $\varphi_X(0) = E(1) = 1$ .

\*  $|\varphi_X(t)| \leq \int_{\mathbb{R}} |e^{itx}| dP_X(x) = 1$ .

\* Si  $Y = aX + b$ , avec  $a \neq 0$  et  $b \in \mathbb{R}$ ,

$$\varphi_Y(t) = E(e^{itY}) = E(e^{itaX+itb}) = e^{itb} E(e^{itaX}) = e^{itb} \varphi_{aX}(t) = e^{itb} \varphi_X(at).$$

\*  $\overline{\varphi_X(t)} = \overline{E(e^{itX})} = E(e^{-itX}) = E(e^{it(-X)}) = \varphi_{-X}(t)$ . Par conséquent, si  $\overline{\varphi_X} = \varphi_X$ , on a  $\varphi_{-X} = \varphi_X$ , d'où  $P_X = P_{-X}$  (d'après le théorème 2.3.6 ci-dessous). Réciproquement si  $P_X = P_{-X}$  alors bien sûr  $\varphi_X = \varphi_{-X} = \overline{\varphi_X}$ .

\* La fonction  $\varphi_X$  est uniformément continue si pour tout  $\varepsilon > 0$  il existe  $\alpha > 0$  tq pour tous  $s, t$ , si  $|t - s| < \alpha$  alors  $|\varphi_X(t) - \varphi_X(s)| \leq \varepsilon$ . Calculons donc

$$|\varphi_X(t) - \varphi_X(s)| \leq \int_{\mathbb{R}} |e^{itx} - e^{isx}| dP_X(x) = \int_{\mathbb{R}} |e^{i(t-s)x} - 1| dP_X(x).$$

Quand  $u \rightarrow 0$ ,  $|e^{iu} - 1| \rightarrow 0$ . Donc  $|e^{i(t-s)x} - 1| \rightarrow 0$  quand  $|s - t| \rightarrow 0$ , uniformément en  $s$  et  $t$ . Vous pouvez détailler le "passage aux  $\varepsilon$ ". En intégrant par rapport à la probabilité  $P_X$ , le résultat en découle.

\* Soit  $p \geq 1$  et supposons que  $E(|X|^p) < \infty$ . On a  $\frac{d}{dt} e^{itx} = ix e^{itx}$  d'où  $|\frac{d}{dt} e^{itx}| = |x|$ . Or  $|x| \leq \max(1, |x|^p)$ , donc l'hypothèse nous dit que  $x \rightarrow |x|$  est intégrable. Le théorème de dérivation sous le signe  $\int$  implique que  $\varphi_X(t)$  est dérivable, de dérivée  $\varphi_X'(t) = E(ix e^{itX})$ . Pour les dérivées d'ordre supérieur, le raisonnement est le même. Par récurrence, on obtient  $\varphi_X^{(k)}(t) = i^k E(X^k e^{itX})$ . En particulier,  $\varphi_X^{(k)}(0) = i^k E(X^k)$ .  $\square$

Rappelons cette proposition classique utile pour les lois à densité.

**Proposition 2.3.5 (Riemann-Lebesgue)** Si  $f : \mathbb{R} \rightarrow \mathbb{R}$  est intégrable, alors sa transformée de Fourier  $\mathcal{F}f$  tend vers 0 à l'infini.

En particulier, si  $f$  est la densité de la loi  $P_X$  d'une variable aléatoire, alors  $\lim_{t \rightarrow \pm\infty} \varphi_X(t) = \lim_{t \rightarrow \mp\infty} \hat{f}(t) = 0$ .

**Démonstration :** Ceci a été fait en intégration. Principe de la preuve : on vérifie d'abord que c'est vrai pour les fonctions indicatrices d'intervalles bornés. Par linéarité, c'est vrai pour les fonctions en escalier, qui sont denses dans l'ensemble des fonctions intégrables. On en déduit le résultat par densité (à justifier proprement).  $\square$

Exemples (à détailler en exercice)

Si  $X$  est une v.a.r. qui suit une loi de Bernoulli  $\mathcal{B}(p)$ , alors  $\varphi_X(t) = E(e^{itX}) = pe^{it} + 1 - p$ .



Si  $X$  est une v.a. uniforme sur  $[-a, a]$  alors

$$\varphi_X(t) = \int_{-a}^a \frac{e^{itx}}{2a} dx = \frac{1}{a} \left[ \frac{e^{itx}}{it} \right]_{-a}^a = \frac{\sin ta}{ta}.$$

Si  $X$  est une v.a. dont la loi est une loi normale centrée réduite, elle a pour densité  $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ .

On a donc

$$\varphi_X(t) = \int_{\mathbb{R}} \frac{e^{itx-x^2/2}}{\sqrt{2\pi}} dx.$$

Comment calculer cette intégrale? D'abord on remarque que  $\frac{d}{dt} e^{itx-x^2/2} \frac{1}{\sqrt{2\pi}} = ix e^{itx-x^2/2} \frac{1}{\sqrt{2\pi}}$  est intégrable. La proposition précédente nous permet de déduire que  $\varphi'_X(t) = \int_{\mathbb{R}} \frac{ix e^{itx-x^2/2}}{\sqrt{2\pi}} dx$ . Par intégration par parties, on obtient (détaillez)  $\varphi'_X(t) = -t\varphi_X(t)$  et  $\varphi_X(0) = 1$ . Cette équation différentielle (cf cours sur les équations différentielles) a pour unique solution

$$\varphi_X : t \mapsto \varphi_X(t) = e^{-t^2/2}.$$

Si  $X$  est une v.a. dont la loi est une loi normale de moyenne  $m$  et de variance  $\sigma^2$ , sa loi  $P_X$  a pour densité  $f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-m)^2/2\sigma^2}$ . Un changement de variable nous ramène au cas précédent, et on trouve

$$\varphi_X(t) = \exp(itm - \frac{\sigma^2 t^2}{2}).$$

Autre méthode (Revuz [?]).

Si  $X$  est une v.a. de loi  $\mathcal{N}(m, \sigma^2)$ , alors  $Y = X - m$  a pour loi une loi normale centrée  $\mathcal{N}(0, \sigma^2)$ . (démonstration immédiate à l'aide des fonctions de répartition ou du changement de variable  $y = x - m$ ). En particulier, ce changement de variable donne immédiatement  $\varphi_X(t) = e^{itm} E(e^{itY})$ . Il nous reste à calculer  $E(e^{itY})$ . Si  $z \in \mathbb{R}$ , alors un changement de variable élémentaire donne

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{zx} e^{-x^2/\sigma^2} dx = \exp(\sigma^2 z^2/2).$$

Maintenant, si  $z \in \mathbb{C}$ , on a  $|e^{zx}| \leq e^{|z||x|}$ . On en déduit (détailler un peu) que l'intégrale  $\frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{zx} e^{-x^2/\sigma^2} dx$  est bien définie et holomorphe en  $z \in \mathbb{C}$ . Cette fonction de  $z \in \mathbb{C}$  coïncide sur l'axe réel avec la fonction holomorphe  $z \mapsto \exp(\sigma^2 z^2/2)$ . Le théorème de prolongement analytique assure qu'elles sont égales sur  $\mathbb{C}$ . Prendre  $z = it$  donne alors le résultat voulu.

**Théorème 2.3.6** Soient  $\mu$  et  $\nu$  deux mesures de probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Si  $\hat{\mu} = \hat{\nu}$ , alors  $\mu = \nu$ . Soient  $X$  et  $Y$  deux v.a. réelles. Alors leurs lois  $P_X$  et  $P_Y$  sont égales si et seulement si  $\varphi_X = \varphi_Y$ .

Ce théorème nous donne l'unicité d'une loi de fonction caractéristique donnée. Autrement dit la fonction caractéristique d'une v.a. caractérise la loi de la v.a. Attention, je vous rappelle qu'il existe plein de v.a. qui n'ont rien à voir entre elles (pas le même espace  $\Omega$  de départ par exemple), et qui ont même loi.

Ce théorème peut se démontrer de diverses façons. Il découle par exemple du lemme et du théorème ci-dessous (cf Revuz Probas).

Introduisons une notation. Si  $m$  est une mesure de probabilité sur  $\mathbb{R}$ , pour tous  $a < b$ , on pose

$$H_m(a, b) = m(]a, b]) + \frac{m(\{a\}) + m(\{b\})}{2}.$$

**Lemme 2.3.7** Soient  $m_1$  et  $m_2$  deux probabilités sur  $\mathbb{R}$  telles que  $H_{m_1} = H_{m_2}$ . Alors  $m_1 = m_2$ .

**Démonstration :**  $m_1$  et  $m_2$  étant finies, elles ont au plus un nombre dénombrable d'atomes (i.e. de singletons de mesure  $m_i(\{x\}) > 0$ ). Considérons la classe des intervalles (ouverts, fermés, semi-ouverts) dont les extrémités ne sont pas des atomes ni de  $m_1$  ni de  $m_2$ . Cette classe est stable par intersection finie, et engendre la tribu. Le corollaire 1.2.4 donne le résultat voulu.  $\square$

**Théorème 2.3.8 (Inversion de Fourier)** Soit  $m$  une mesure de probabilité sur  $\mathbb{R}$ . Alors pour tout  $a < b$ , on a

$$\lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \hat{m}(t) dt = H_m(a, b).$$

Si  $\hat{m}$  est intégrable, alors  $m$  a une densité continue  $p$  donnée par

$$p(x) = \frac{1}{2\pi} \int \hat{m}(t) e^{-itx} dt.$$

**Démonstration :** Remarquons avant toute chose que la fonction  $(t, x) \in \mathbb{R}^* \times \mathbb{R} \mapsto \frac{e^{-ita} - e^{-itb}}{it} \hat{m}(t)$  est continue, se prolonge par continuité en  $t = 0$  par  $(b - a)$ , et est donc bornée sur  $\mathbb{R}^2$ .

Le théorème de Fubini s'applique donc, et donne :

$$\int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \hat{m}(t) dt = \int_{\mathbb{R}} \int_{-T}^T \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt dm(x)$$

Décomposons l'exponentielle en  $\cos + i \sin$ . Remarquons qu'une application impaire a une intégrale sur  $[-T, T]$  nulle. L'intégrale qui nous intéresse vaut donc encore

$$\int_{\mathbb{R}} \int_{-T}^T \frac{\sin(t(x-a)) - \sin(t(x-b))}{t} dt dm(x) = \int_{\mathbb{R}} \left( \int_{-T(x-a)}^{T(x-a)} \frac{\sin t}{t} dt - \int_{-T(x-b)}^{T(x-b)} \frac{\sin t}{t} dt \right) dm(x)$$

L'égalité, après séparation des deux termes sous l'intégrale, s'obtient par un changement de variable élémentaire.

A noter qu'il s'agit d'intégrales de Riemann, avec  $\int_{-T(x-a)}^{T(x-a)} \frac{\sin t}{t} dt = - \int_{T(x-a)}^{-T(x-a)} \frac{\sin t}{t} dt$ .

La fonction  $x \mapsto \int_0^x \frac{\sin t}{t} dt$  est bornée et tend vers  $\pi/2$  en  $+\infty$  (**EXO classique intégration**). Par parité, on vérifie que  $\int_{-T}^T \frac{\sin t}{t} dt$  tend vers  $\pi$  lorsque  $T \rightarrow +\infty$ , et vers  $-\pi$  lorsque  $T \rightarrow -\infty$ . Si  $x > a$ , alors  $T(x-a) \rightarrow +\infty$  quand  $T \rightarrow +\infty$ , et si  $x < a$ , alors  $T(x-a) \rightarrow -\infty$ .

Avec les remarques ci-dessus, après étude du signe de  $x - a$  et de  $x - b$ , on vérifie que la quantité  $\int_{-T(x-a)}^{T(x-a)} \frac{\sin t}{t} dt - \int_{-T(x-b)}^{T(x-b)} \frac{\sin t}{t} dt$  tend vers  $\pi - \pi = 0$  si  $x > \max(a, b)$ , vers  $\pi + \pi = 2\pi$  si  $a < x < b$ , vers  $0 = -\pi + \pi$  si  $x < a$ , vers  $\pi$  si  $x = a$  ou  $x = b$ . Autrement dit, la fonction de  $x$  entre parenthèses converge vers  $\pi(\mathbf{1}_{\{a\}} + \mathbf{1}_{\{b\}} + 2\mathbf{1}_{]a,b[})$ , en restant uniformément bornée. Le théorème de convergence dominée de Lebesgue s'applique alors pour donner le résultat de convergence voulu.

Montrons maintenant la deuxième partie du théorème. Le théorème de Fubini permet de voir que l'intégrale de l'énoncé vaut encore

$$\int_a^b \left( \int_{-T}^T e^{-itx} \hat{m}(t) dt \right) dx$$

Si  $\hat{m}$  est intégrable, alors  $\left| \int_{[-T, T]} e^{-itx} \hat{m}(t) dt \right| \leq \int_{\mathbb{R}} |\hat{m}(t)| dt < \infty$ , d'où l'on voit que lorsque  $T \rightarrow \infty$ , par le théorème de convergence dominée, l'intégrale entre parenthèses converge vers  $\int_{\mathbb{R}} e^{-itx} \hat{m}(t) dt$ , en restant majorée par  $\|\hat{m}\|_1$ . Le théorème de convergence dominée de Lebesgue appliqué une deuxième fois donne alors, à la limite,

$$H_m(a, b) = \lim_{T \rightarrow \infty} \int_a^b \left( \int_{-T}^T e^{-itx} \hat{m}(t) dt \right) dx = \frac{1}{2\pi} \int_a^b \int_{\mathbb{R}} e^{-itx} \hat{m}(t) dt dx,$$

ce qu'il fallait démontrer. □

**Corollaire 2.3.9** Soit  $X$  une variable aléatoire. Si sa fonction caractéristique  $\varphi_X$  est intégrable alors  $X$  est à densité (continue).

**Démonstration :** La deuxième partie du théorème donne l'implication : si  $\varphi_X$  est intégrable, alors  $X$  a une densité (continue). □

**Remarque 2.3.10** En fait, si la fonction caractéristique  $\varphi_X$  de  $X$  est intégrable, alors  $X$  a une densité qui est même uniformément continue. (Adapter la preuve de l'uniforme continuité de la fonction caractéristique, proposition 2.3.4, pour le voir. )

**Remarque 2.3.11** Une variable aléatoire à densité n'a pas forcément une fonction caractéristique intégrable. Exemple : si  $X$  suit une loi uniforme sur  $[-1, 1]$ , sa fonction caractéristique vaut

$$\varphi_X(t) = \frac{1}{2} \int_{-1}^1 e^{itx} dx = \frac{\sin t}{t}$$

Cette fonction est continue, bornée, mais n'est pas intégrable sur  $\mathbb{R}$ .

On peut construire une variable aléatoire à densité continue, et de fonction caractéristique pas intégrable, en utilisant la remarque ci-dessus, et en construisant une densité positive, continue, non uniformément continue, d'intégrale 1. Ceci se fait par exemple, à une constante multiplicative près, en prenant  $f$  nulle en dehors des intervalles  $[n, n + \frac{2}{n^3}]$ , sur lesquels son graphe est un triangle, de hauteur  $n$ , et de base  $2/n^3$ .

### 2.3.2 Fonction caractéristique d'une variable aléatoire vectorielle

Rappelons que si  $x$  et  $y$  sont deux vecteurs de  $\mathbb{R}^d$ , leur produit scalaire, noté de façons très variées, vérifie

$$\langle x, y \rangle = (x|y) = x \cdot y = x_1 y_1 + \dots + x_d y_d.$$

**Définition 2.3.12** Soit  $\mu$  une mesure finie sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . Sa transformée de Fourier  $\hat{\mu} : \mathbb{R}^d \rightarrow \mathbb{C}$  est définie pour tout  $u \in \mathbb{R}^d$  par

$$\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{iu \cdot x} d\mu(x).$$

Si  $X$  est une v.a. vectorielle à valeurs dans  $\mathbb{R}^d$ , sa fonction caractéristique  $\varphi_X : \mathbb{R}^d \rightarrow \mathbb{C}$  est la transformée de Fourier de sa loi  $P_X$ .

Soient  $\mu$  une mesure sur  $\mathbb{R}^p$  et  $\nu$  une mesure sur  $\mathbb{R}^q$ . Remarquez que

$$\widehat{(\mu \times \nu)} = \hat{\mu} \hat{\nu}.$$

En effet, en utilisant le théorème de Fubini, on a pour  $u = (u_1, u_2)$  avec  $u_1 \in \mathbb{R}^p$  et  $u_2 \in \mathbb{R}^q$

$$\begin{aligned} \widehat{(\mu \times \nu)}(u) &= \int_{\mathbb{R}^{p+q}} e^{iu \cdot x} d(\mu \times \nu)(x_1, x_2) = \int_{\mathbb{R}^p} \int_{\mathbb{R}^q} e^{iu_1 x_1} e^{iu_2 x_2} d\nu(x_2) d\mu(x_1) \\ &= \int_{\mathbb{R}^p} e^{iu_1 x_1} d\mu(x_1) \int_{\mathbb{R}^q} e^{iu_2 x_2} d\nu(x_2) = \hat{\mu}(u_1) \hat{\nu}(u_2) \end{aligned}$$

Ceci sera très important par la suite.

**Théorème 2.3.13** Soient  $\mu$  et  $\nu$  deux mesures de probabilité sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . Si  $\hat{\mu} = \hat{\nu}$ , alors  $\mu = \nu$ . Soient  $X$  et  $Y$  deux v.a. à valeurs dans  $\mathbb{R}^d$ . Alors leurs lois  $P_X$  et  $P_Y$  sont égales si et seulement si  $\varphi_X = \varphi_Y$ .

**Démonstration :** Admis. Voir Master I, ou Revuz. □

Exemple de calcul d'une fonction caractéristique vectorielle :  $\Omega = [0, 1]$ ,  $\mathcal{A} = \mathcal{B}([0, 1])$  et  $P =$  Lebesgue sur  $[0, 1]$ . Soit  $X : \Omega \rightarrow \mathbb{R}^2$  définie par  $X(\omega) = (1_{[0, 1/2]}(\omega), \omega)$ . Notons  $X_1(\omega) = 1_{[0, 1/2]}(\omega)$  et  $X_2(\omega) = \omega$  ses coordonnées. Si  $u = (u_1, u_2) \in \mathbb{R}^2$ , on calcule (détaillez)

$$\begin{aligned} \varphi_X(u) &= E(e^{iu \cdot X}) = E(e^{i(u_1, u_2) \cdot (X_1(\omega), X_2(\omega))}) = E(e^{iu_1 X_1(\omega) + iu_2 X_2(\omega)}) \\ &= \int_{\Omega} e^{iu_1 X_1(\omega) + iu_2 X_2(\omega)} dP(\omega) = \dots = \frac{e^{iu_1}}{iu_2} (e^{iu_2/2} - 1) + \frac{e^{iu_2} - e^{iu_2/2}}{iu_2} \end{aligned}$$

**Proposition 2.3.14** Soit  $X$  une v.a. à valeurs dans  $\mathbb{R}^d$ . Alors

1.  $\varphi_X(0) = 1$
2.  $|\varphi_X(u)| \leq 1$  pour tout  $u \in \mathbb{R}^d$
3. Si  $A \in M_{d,n}(\mathbb{R})$  et  $b \in \mathbb{R}^n$  alors  $\varphi_{A \cdot X + b}(v) = e^{iv \cdot b} \varphi_X(tAv)$  pour  $v \in \mathbb{R}^n$ .
4.  $\varphi_X$  est uniformément continue
5. Si  $n = (n_1, \dots, n_d)$  et  $E(|X_1|^{n_1} \dots |X_d|^{n_d}) < +\infty$  alors  $\frac{\partial^n \varphi_X(0)}{\partial u^n} = i^{|n|} E(X_1^{n_1} \dots X_d^{n_d})$

6. Si  $\hat{\mu} = \hat{\nu}$  alors  $\mu = \nu$  et  $X$  et  $Y$  ont même loi si et seulement si  $\varphi_X = \varphi_Y$ .

**Démonstration :** Démontrons seulement 3). Si  $Y = A.X + b \in \mathbb{R}^n$  et  $v \in \mathbb{R}^n$ , alors  $\varphi_Y(v) = E(e^{iv \cdot Y}) = E(e^{iv \cdot (AX+b)}) = e^{iv \cdot b} E(e^{iv \cdot AX}) = e^{iv \cdot b} E(e^{i(A^t v) \cdot X}) = e^{iv \cdot b} \varphi_X(A^t v)$ .  $\square$

**Théorème 2.3.15 (Théorème de Levy)** Soit  $(P_n)_{n \in \mathbb{N}}$  une suite de mesures de probabilité sur  $\mathbb{R}^d$  telles que  $\hat{P}_n$  converge simplement vers une fonction  $\phi$  continue en 0. Alors la suite  $(P_n)_{n \in \mathbb{N}}$  converge étroitement vers une mesure  $m$  telle que  $\hat{m} = \phi$ .

Ce théorème est du niveau M1. Nous commenterons cet énoncé à nouveau lorsque nous aurons vu ce que signifie la convergence étroite.

# Chapitre 3

## Indépendance

### 3.1 Produit de lois

Ce paragraphe a été traité en détail dans le module intégration.

Soient  $(X_1, \mathcal{A}_1, \mathcal{P}_1)$  et  $(X_2, \mathcal{A}_2, \mathcal{P}_2)$  deux espaces de probabilité.

La tribu produit sur  $X_1 \times X_2$  est la tribu engendrée par les ensembles  $A_1 \times A_2$ ,  $A_i \in \mathcal{A}_i$ . En particulier, elle contient  $\mathcal{A}_1 \times \mathcal{A}_2$ , mais elle est plus grosse. Elle est notée  $\mathcal{A}_1 \otimes \mathcal{A}_2$ . Par exemple, la tribu borélienne  $\mathcal{B}(\mathbb{R}^2)$  est également la tribu produit  $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ .

La mesure produit  $\mathbf{P}_1 \otimes \mathbf{P}_2$  est l'unique mesure qui vérifie pour tous  $A_1 \in \mathcal{A}_1$ ,  $A_2 \in \mathcal{A}_2$

$$\mathbf{P}_1 \otimes \mathbf{P}_2(A_1 \times A_2) = \mathbf{P}_1(A_1)\mathbf{P}_2(A_2).$$

$\mathbf{P}_1$  et  $\mathbf{P}_2$  sont les mesures marginales de  $\mathbf{P}_1 \otimes \mathbf{P}_2$ . L'existence et l'unicité de la mesure produit (i.e. le fait qu'elle peut être prolongée, et de manière unique, en une mesure sur tous les ensembles de la tribu produit, ont été démontrées en cours d'intégration. Il s'agit encore d'arguments de type classes monotones.

Par définition de  $\mathbf{P}_1 \otimes \mathbf{P}_2$ , si  $g(x, y) = \mathbf{1}_{A_1 \times A_2}(x, y) = \mathbf{1}_{A_1}(x)\mathbf{1}_{A_2}(y)$  alors les applications  $x \mapsto \int_{\mathbb{R}} g(x, y)d\mathbf{P}_2(y) = \mathbf{1}_{A_1}(x)\mathbf{P}_2(A_2)$  et  $y \mapsto \int_{\mathbb{R}} g(x, y)d\mathbf{P}_1(x) = \mathbf{1}_{A_2}(y)\mathbf{P}_1(A_1)$  sont mesurables, et on a

$$\int_{\mathbb{R}^2} g d(\mathbf{P}_1 \otimes \mathbf{P}_2) = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} g(x, y)d\mathbf{P}_2(y) \right) d\mathbf{P}_1(x) = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} g(x, y)d\mathbf{P}_1(x) \right) d\mathbf{P}_2(y).$$

On peut montrer que ceci se généralise à toutes les fonctions  $g(x, y) = \mathbf{1}_A(x, y)$ , avec  $A \in \mathcal{B}_1 \times \mathcal{B}_2$  (voir cours intégration 2). Ensuite, cela se généralise aisément à toutes les fonctions simples, puis par un argument de convergence monotone à toutes les fonctions mesurables positives. Le théorème de convergence dominée permet d'obtenir cela pour toutes les fonctions  $g$  qui sont  $\mathbf{P}_1 \otimes \mathbf{P}_2$  intégrables. En résumé, on a :

**Théorème 3.1.1 (Théorèmes de Tonelli et Fubini)** Soient  $\mathbf{P}_1$  et  $\mathbf{P}_2$  deux probabilités sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Soit  $\mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2$  la mesure produit. Soit  $g : (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) \rightarrow \mathbb{R}$  une fonction mesurable qui est positive ou  $\mathbf{P}$ -intégrable. Alors les fonctions  $g_2 : y \mapsto \int_{\mathbb{R}} g(x, y)d\mathbf{P}_1(x)$  et  $g_1 : x \mapsto \int_{\mathbb{R}} g(x, y)d\mathbf{P}_2(y)$  sont mesurables, et on a

$$\int_{\mathbb{R}^2} g d(\mathbf{P}_1 \otimes \mathbf{P}_2) = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} g(x, y)d\mathbf{P}_2(y) \right) d\mathbf{P}_1(x) = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} g(x, y)d\mathbf{P}_1(x) \right) d\mathbf{P}_2(y).$$

En particulier,  $g_1$  est  $\mathbf{P}_1$ -intégrable et  $g_2$  est  $\mathbf{P}_2$  intégrable lorsque  $g$  est  $\mathbf{P}$ -intégrable.

**Remarque 3.1.2** Rappelons que si  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  est mesurable, on applique d'abord le théorème de Tonelli à  $|g|$  pour savoir si  $g$  est intégrable, puis le théorème de Fubini à  $g$  si elle est intégrable.

**Démonstration :** Voir cours intégration. □

En pratique, lorsqu'on travaille avec des fonctions mesurables positives, ce théorème dit qu'intégrer par rapport à la mesure produit  $\mathbf{P}_1 \otimes \mathbf{P}_2$ , c'est pareil que d'intégrer d'abord suivant une coordonnée, puis suivant l'autre. C'est comme cela qu'on calcule en pratique  $\int_{\mathbb{R}^2} g d(\mathbf{P}_1 \otimes \mathbf{P}_2)$ .

## 3.2 Couple de variables aléatoires, densité marginale

Paragraphe traité au cours 3

Soit  $X : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (\mathbb{R}^{d_1}, \mathcal{B}(\mathbb{R}^{d_1}))$  et  $Y : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (\mathbb{R}^{d_2}, \mathcal{B}(\mathbb{R}^{d_2}))$ . Le couple  $(X, Y)$  est simplement la variable aléatoire  $Z = (X, Y) : \omega \in \Omega \mapsto Z(\omega) = (X_1(\omega), X_2(\omega)) \in \mathbb{R}^{d_1+d_2}$ . La loi conjointe de  $X$  et  $Y$  est la loi du couple  $Z = (X, Y)$ .

**Attention** En général, la loi du couple  $(X, Y)$  n'est pas le produit des lois de  $X$  et  $Y$ . On verra que c'est le cas ssi  $X$  et  $Y$  sont indépendantes (voir ch 3). Par exemple, si  $X = 1_A$ , avec  $A \in \mathcal{A}$  et  $Y = 1 - X$ , alors la loi de  $(X, Y)$  vérifie  $\mathbf{P}_{(X,Y)}(\{(0, 1)\}) = 1 - \mathbf{P}(A)$ ,  $\mathbf{P}_{(X,Y)}(\{(1, 0)\}) = \mathbf{P}(A)$ , et  $\mathbf{P}_{(X,Y)}(0, 0) = \mathbf{P}_{(X,Y)}(1, 1) = 0$ . (Vérifier cela, puis en déduire que la loi de  $(X, Y)$  n'est pas le produit  $\mathbf{P}_X \otimes \mathbf{P}_Y$ .)

Si  $Z = (X, Y)$  est une variable aléatoire, on appelle lois marginales les lois de  $X$  et  $Y$ .

Le théorème suivant est utilisé en pratique avec  $d_1 = d_2 = 1$ , i.e.  $X$  et  $Y$  sont des v.a. réelles.

**Théorème 3.2.1** Soit  $Z = (X, Y)$  une v.a. vectorielle sur  $\mathbb{R}^{d_1+d_2} = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$  de densité  $f_Z$ . Alors  $X$  et  $Y$  sont à densité,  $f_X(x) = \int_{\mathbb{R}^{d_2}} f_Z(x, y) dy$  et  $f_Y(y) = \int_{\mathbb{R}^{d_1}} f_Z(x, y) dx$ . Les densités  $f_X$  et  $f_Y$  sont appelées densités marginales.

**Démonstration :**  $\mathbf{P}_X(A) = \mathbf{P}_{(X,Y)}(A \times \mathbb{R}^{d_2}) = \int_{A \times \mathbb{R}^{d_2}} f_Z(x, y) dx dy = \int_A (\int_{\mathbb{R}^{d_2}} f_Z(x, y) dy) dx$  par Fubini.  $\square$

**Remarque 3.2.2 Attention** La réciproque est fautive en général : si  $X$  et  $Y$  sont à densité,  $f_Z$  n'est pas nécessairement à densité. Par exemple, si  $X$  est une v.a. de loi uniforme sur  $[0, 1]$  et  $Y = X$ , alors le couple  $Z = (X, X)$  vérifie  $\mathbf{P}_Z(A) = 0$  pour tout ensemble borélien  $A$  n'intersectant pas la diagonale  $\{(x, x) \in \mathbb{R}^2, x \in [0, 1]\}$ . Ceci implique (détailler) que  $Z$  ne peut pas être à densité. (Le fait que  $X$  suive une loi uniforme n'est pas important ici, ce qui est important, c'est  $Y = X$  !)

Un exemple de calcul de densité marginale :  $Z = (X, Y)$  de densité  $f_Z(x, y) = \frac{1}{2}(x+y)e^{-(x+y)} \mathbf{1}_{\mathbb{R}_+}(x) \mathbf{1}_{\mathbb{R}_+}(y)$ , alors  $f_X(x) = ?$  et  $f_Y(y) = ?$ .

## 3.3 Variables aléatoires indépendantes

### 3.3.1 Définitions

**Définition 3.3.1** Deux v.a.  $X$  et  $Y$  de  $(\Omega, \mathcal{A}, \mathbf{P})$  dans  $(E, \mathcal{E})$  sont indépendantes ssi pour tous ensembles mesurables  $A, B \in \mathcal{E}$  on a

$$\mathbf{P}(X \in A, Y \in B) = \mathbf{P}(X \in A) \mathbf{P}(Y \in B).$$

**Définition 3.3.2** Les variables aléatoires  $(X_i)_{i \in I}$  ( $I$  au plus dénombrable) sont indépendantes dans leur ensemble ssi pour tout sous-ensemble  $J \subset I$  fini, on a  $\mathbf{P}(\forall j \in J, X_j \in A_j) = \prod_{j \in J} \mathbf{P}(X_j \in A_j)$ . On dit qu'elles sont deux à deux indépendantes si pour tout couple d'indices  $(i, j) \in I^2$  avec  $i \neq j$ , les v.a.  $X_i$  et  $X_j$  sont indépendantes au sens de la définition ci-dessus.

Question : l'une de ces deux notions est-elle plus forte que l'autre ? Laquelle ? pourquoi ?

**Définition 3.3.3** Soit  $(\mathcal{A}_i)_{i \in I}$  une famille de sous-tribus de  $\mathcal{E}$ . Elles sont indépendantes dans leur ensemble si pour tout sous-ensemble  $J \subset I$  fini, et tout  $(A_j)_{j \in J} \in (\mathcal{A}_j)_{j \in J}$ , on a  $\mathbf{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbf{P}(A_j)$ .

**Proposition 3.3.4** Soit  $X_i : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (E_i, \mathcal{E}_i)$ ,  $i \in I$ , une famille de v.a. Les v.a.  $(X_i)_{i \in I}$  sont indépendantes dans leur ensemble ssi les tribus  $X_i^{-1}(\mathcal{E}_i) \subset \mathcal{A}$  le sont.

**Démonstration :** Par définition de l'intersection d'ensembles, on a  $\{\omega \in \Omega, \forall j \in J, X_j \in A_j\} = \cap_{j \in J} \{\omega \in \Omega, X_j(\omega) \in A_j\}$  On en déduit  $\mathbf{P}(\forall j \in J, X_j \in A_j) = \mathbf{P}(\cap_{j \in J} X_j^{-1}(A_j))$ . Les  $(X_j)_{j \in J}$  sont indépendantes ssi le premier terme vaut  $\prod_{j \in J} \mathbf{P}(X_j \in A_j)$ . Les tribus  $X_j^{-1}(\mathcal{E}_j)$  sont indépendantes ssi le deuxième terme vaut  $\prod_{j \in J} \mathbf{P}(X_j^{-1}(A_j))$ .  $\square$

**Proposition 3.3.5** Soient  $X_i : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (\mathbb{R}^{n_i}, \mathcal{B}(\mathbb{R}^{n_i}))$ ,  $i \in I$  des v.a. indépendantes et  $f_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R}^{p_i}$ ,  $i \in I$  des applications mesurables. Soient  $Y_i = f_i(X_i)$  pour tout  $i \in I$ . Alors les  $(Y_i)$  sont indépendantes.

Vous pouvez lire ce résultat avec  $n_i = p_i = 1$  et considérer que toutes les v.a. sont des v.a. réelles.

**Démonstration :** Soit  $J \subset I$  une partie finie. Soient  $B_j \subset \mathcal{B}(\mathbb{R}^{p_i})$ . Alors  $Y_j(\omega) \in B_j$  ssi  $X_j(\omega) \in f_j^{-1}(B_j)$ . En particulier,  $\mathbf{P}(\forall j \in J, Y_j \in B_j) = \mathbf{P}(\forall j \in J, X_j \in f_j^{-1}(B_j))$ . Les  $(X_j)$  étant indépendantes, ce dernier terme vaut encore  $\prod_{j \in J} \mathbf{P}(X_j \in f_j^{-1}(B_j)) = \prod_{j \in J} \mathbf{P}(Y_j \in B_j)$ . Les  $Y_j$  sont donc mutuellement indépendantes.  $\square$

### 3.3.2 Variables aléatoires indépendantes et produit de lois

**Proposition 3.3.6** Soient  $X_1, X_2$  deux v.a. de  $(\Omega, \mathcal{A}, \mathbf{P})$  dans  $(E_i, \mathcal{E}_i)$ ,  $i = 1, 2$ . Soit  $\mathbf{P}_{X_i}$ ,  $i = 1, 2$  la loi de  $X_i$  sur  $E_i$ . Alors  $X_1$  et  $X_2$  sont indépendantes ssi  $\mathbf{P}_{(X_1, X_2)} = \mathbf{P}_{X_1} \otimes \mathbf{P}_{X_2}$ .

**Démonstration :** Par définition de la mesure produit  $\mathbf{P}_{X_1} \otimes \mathbf{P}_{X_2}$ , on a  $\mathbf{P}_{(X_1, X_2)} = \mathbf{P}_{X_1} \otimes \mathbf{P}_{X_2}$  ssi pour tous boréliens  $A_1 \in \mathcal{E}_1$  et  $A_2 \in \mathcal{E}_2$ ,  $\mathbf{P}(X_1 \in A_1 \text{ et } X_2 \in A_2) = \mathbf{P}_{X_1}(A_1)\mathbf{P}_{X_2}(A_2)$ . Ce dernier terme étant égal à  $\mathbf{P}(X_1 \in A_1)\mathbf{P}(X_2 \in A_2)$ , ceci équivaut donc à l'indépendance de  $X_1$  et  $X_2$ .  $\square$

**Proposition 3.3.7** Soient  $X_1$  et  $X_2$  deux v.a.r. Alors  $X_1$  et  $X_2$  sont indépendantes ssi pour tous  $(x_1, x_2) \in \mathbb{R}^2$ , on a  $F_{(X_1, X_2)}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$ .

**Démonstration :** Si  $X_1$  et  $X_2$  sont indépendantes, alors en considérant les ensembles  $A_1 = ]-\infty, x_1]$  et  $A_2 = ]-\infty, x_2]$  on obtient le résultat voulu.

Réciproquement, si pour tous  $(x_1, x_2) \in \mathbb{R}^2$ , on a  $F_{(X_1, X_2)}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$ , alors sur l'algèbre engendrée par les produits d'intervalles  $]a_1, b_1] \times ]a_2, b_2]$ , la loi de  $(X_1, X_2)$  est bien le produit des lois de  $X_1$  et de  $X_2$ . Le corollaire 1.2.4 permet de conclure que  $\mathbf{P}_{(X_1, X_2)} = \mathbf{P}_{X_1} \otimes \mathbf{P}_{X_2}$ , i.e.  $X_1$  et  $X_2$  sont indépendantes.  $\square$

Remarque : les deux propositions ci-dessus restent vraies pour  $n$  variables aléatoires indépendantes dans leur ensemble.

**Proposition 3.3.8** Soient  $X$  et  $Y$  deux v.a. réelles à densités respectives  $f_X$  et  $f_Y$ . Alors  $X$  et  $Y$  sont indépendantes ssi le produit  $Z = (X, Y)$  est à densité, et que cette densité vérifie

$$f_{(X, Y)}(x, y) = f_X(x)f_Y(y) \quad p.s.$$

Remarque : le couple  $Z = (X, Y)$  peut être à densité sans que  $X$  et  $Y$  soient indépendantes. Exemple : si  $f_Z(x, y) = \frac{1}{2}(x+y)e^{-(x+y)}$ .

Si  $Z = (X, Y)$  est à densité alors  $X$  et  $Y$  aussi. En revanche,  $X$  et  $Y$  peuvent être à densité et pas  $Z = (X, Y)$ . Donnez des exemples (le plus simple étant si  $X$  est à densité et  $Y = X$ ).

**Démonstration :** Supposons que pour presque tous  $(x, y) \in \mathbb{R}^2$ , on ait  $f_{(X, Y)}(x, y) = f_X(x)f_Y(y)$ . Calculons

$$\begin{aligned} P(X \in A, Y \in B) &= \int_{A \times B} f_{(X, Y)}(x, y) dx dy = \int_{A \times B} f_X(x)f_Y(y) dx dy \\ &= \int_A f_X(x) \left( \int_B f_Y(y) dy \right) dx = \left( \int_A f_X(x) dx \right) \left( \int_B f_Y(y) dy \right) = P(X \in A)P(Y \in B) \end{aligned}$$

Si  $X$  et  $Y$  sont indépendantes et à densité, montrons que  $f_{(X, Y)}(x, y) = f_X(x)f_Y(y)$  pour tout  $(x, y) \in \mathbb{R}^2$ . On sait que  $\mathbf{P}_{(X, Y)} = \mathbf{P}_X \otimes \mathbf{P}_Y$ . Si  $h$  est une fonction mesurable bornée, donc intégrable, par le théorème de Fubini, on en déduit que  $\int_{\mathbb{R}^2} h(x, y) d\mathbf{P}_{(X, Y)}(x, y) = \int_{\mathbb{R}^2} h(x, y) d\mathbf{P}_X(x) d\mathbf{P}_Y(y) = \int_{\mathbb{R}^2} h(x, y) f_X(x) f_Y(y) dx dy$ . La proposition 1.5.10 d permet de conclure que  $f_{(X, Y)}(x, y) = f_X(x)f_Y(y)$  pour tout  $(x, y) \in \mathbb{R}^2$ .  $\square$

### 3.4 Le lemme de Borel-Cantelli

**Définition 3.4.1** Soit  $(\Omega, \mathcal{A})$  un espace mesurable. Soit  $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$  une suite d'ensembles mesurables. La limite supérieure des  $A_n$  est l'ensemble (ou événement)

$$\begin{aligned} \limsup_{n \rightarrow \infty} A_n &:= \bigcap_{n \in \mathbb{N}} \left( \bigcup_{k \geq n} A_k \right) \\ &= \{ \omega \in \Omega, \forall n \in \mathbb{N}, \exists k \geq n, \omega \in A_k \} \\ &= \ll \text{l'événement } A_n \text{ se réalise pour une infinité de } n \in \mathbb{N} \gg \end{aligned}$$

C'est un ensemble (événement) qui ne dépend pas de  $n$ . De la même manière, la limite inférieure de la suite  $(A_n)$  est l'ensemble

$$\begin{aligned} \liminf_{n \rightarrow \infty} A_n &:= \bigcup_{n \in \mathbb{N}} \left( \bigcap_{k \geq n} A_k \right) \\ &= \{ \omega \in \Omega, \exists n \in \mathbb{N}, \forall k \geq n, \omega \in A_k \} \\ &= \ll \text{l'événement } A_n \text{ se réalise pour tous les entiers } n \text{ à partir d'un certain rang} \gg \end{aligned}$$

**Exercice 3.4.2** Vérifier les propriétés immédiates suivantes :

- \*  $\liminf_{n \rightarrow \infty} A_n \subset \limsup_{n \rightarrow \infty} A_n$
- \*  $(\limsup A_n)^c = \liminf (A_n)^c$
- \*  $\mathbf{1}_{\limsup_{n \rightarrow \infty} A_n} = \limsup_{n \rightarrow \infty} \mathbf{1}_{A_n}$
- \*  $\mathbf{1}_{\liminf_{n \rightarrow \infty} A_n} = \liminf_{n \rightarrow \infty} \mathbf{1}_{A_n}$

Le théorème suivant est d'une grande utilité.

**Théorème 3.4.3 (Borel-Cantelli)** Soit  $(\Omega, \mathcal{A}, \mathbf{P})$  un espace de probabilité, et  $(A_n)_{n \in \mathbb{N}}$  une suite d'ensembles mesurables.

- (a) Si  $\sum_{n \geq 0} \mathbf{P}(A_n) < +\infty$  alors  $\mathbf{P}(\limsup_{n \rightarrow \infty} A_n) = 0$
- (b) Si  $\mathbf{P}(\limsup A_n) = 0$  et si les  $(A_n)$  sont deux à deux indépendants alors  $\sum_{n \geq 0} \mathbf{P}(A_n) < +\infty$
- (c) Si  $\sum_{n \geq 0} \mathbf{P}(A_n) = +\infty$  et si les  $(A_n)$  sont deux à deux indépendants alors  $\mathbf{P}(\limsup A_n) = 1$ .

La morale de ce théorème est la suivante. La partie (a), très facile à démontrer, dit que si les probabilités des événements  $A_n$  décroissent trop vite (la série converge) alors il ne peut pas y en avoir une infinité qui se réalisent en même temps. Du point de vue ensembliste, les ensembles  $A_n$  sont tellement petits que la somme de leurs mesures converge. Alors il ne peut (presque sûrement) pas y en avoir une infinité qui s'intersectent en un ensemble de probabilité non nulle.

Les parties (b) et (c) établissent une sorte de réciproque sous l'hypothèse d'indépendance. Si les ensembles  $A_n$  remplissent beaucoup l'espace  $\Omega$ , et s'ils sont indépendants entre eux, alors il y en a une infinité qui s'intersectent en un ensemble de probabilité non nulle.

Voici un exemple qui illustre tout de suite la nécessité de l'hypothèse d'indépendance. Soit  $\Omega = [0, 1]$  muni de sa tribu borélienne et de la mesure de Lebesgue, et  $A_n = [0, \frac{1}{n}]$ . Cette suite d'intervalles vérifie l'hypothèse de (c). Pourtant,  $\limsup A_n = \{0\}$  est un ensemble de mesure nulle. Les  $(A_n)$  sont ici tout sauf indépendants, puisqu'ils sont inclus les uns dans les autres !

**Démonstration :** (a) Soit  $X$  la variable aléatoire  $X = \sum_{n \in \mathbb{N}} \mathbf{1}_{A_n}$ . On a  $\sum_{n \in \mathbb{N}} \mathbf{P}(A_n) = \int_{\mathbb{R}} X d\mathbf{P} < +\infty$  (détailler ces égalités). Donc  $X$  est intégrable. Donc  $X < +\infty$  p.s. Or  $X(\omega) < +\infty$  ssi  $\omega$  appartient à un nombre fini de  $A_n$  (détailler). Donc presque sûrement,  $\omega$  n'appartient pas à  $\limsup_{n \rightarrow \infty} A_n$ . Soit encore  $\mathbf{P}(\limsup_{n \rightarrow \infty} A_n) = 0$ .

Avant de montrer (c), vérifions que (c) implique (b). En effet, si  $\mathbf{P}(\limsup A_n) = 0$  alors  $\mathbf{P}(\limsup A_n) \neq 1$ , donc la conclusion de (c) est fautive. Si de plus les  $(A_n)$  sont indépendants, alors  $\sum_{n \in \mathbb{N}} \mathbf{P}(A_n) < \infty$ .

Montrons (c). Commençons par une preuve sous l'hypothèse (plus forte) que les  $(A_n)_{n \in \mathbb{N}}$  sont indépendants dans leur ensemble. Supposons  $\sum_{n \in \mathbb{N}} \mathbf{P}(A_n) = +\infty$ . Nous allons montrer que  $\mathbf{P}((\limsup_{n \rightarrow \infty} A_n)^c) = 0$ . On sait que  $(\limsup_{n \rightarrow \infty} A_n)^c = \bigcup_{n \geq 1} \bigcap_{m \geq n} A_m^c$ . Or

$$\begin{aligned} \mathbf{P}(\bigcap_{m=n}^{+\infty} A_m^c) &\leq \mathbf{P}(\bigcap_{m=n}^p A_m^c) = \prod_{m=n}^p \mathbf{P}(A_m^c) \\ &= \prod_{m=n}^p (1 - \mathbf{P}(A_m)) \leq \prod_{m=n}^p \exp(-\mathbf{P}(A_m)) \leq \exp(-\sum_{m=n}^p \mathbf{P}(A_m)) \end{aligned}$$



Dans ce calcul, où a-t-on utilisé l'indépendance? On s'est aussi servi de l'inégalité grossière mais très utile suivante : pour tout  $x > 0$ , on a  $1 - x \leq \exp(-x)$  (démontrez-la).

En faisant tendre  $p$  vers  $+\infty$ , on obtient  $\mathbf{P}(\cap_{m=n}^{+\infty} A_m^c) \leq \exp(-\sum_{m=n}^{+\infty} \mathbf{P}(A_m)) = 0$ . Une union d'ensembles de mesure nulle est encore de mesure nulle, d'où  $\mathbf{P}(\liminf(A_m)^c) = 0$ . D'où le résultat.

Démontrons maintenant le théorème sous l'hypothèse que les  $(A_n)$  sont deux à deux indépendants. Soit  $S_n = \sum_{k=0}^n \mathbf{1}_{A_k}$ .

Commençons par montrer que  $\frac{S_n}{E(S_n)}$  tend vers 1 dans  $L^2$ . Pour cela, notons que l'hypothèse  $\sum_{n \in \mathbb{N}} P(A_n) = +\infty$  équivaut à  $E(S_n) \rightarrow +\infty$ , puis calculons

$$\begin{aligned} \left\| \frac{S_n}{E(S_n)} - 1 \right\|_{L^2} &= E \left( \left( \frac{S_n}{E(S_n)} - 1 \right)^2 \right) = \frac{1}{(E(S_n))^2} E((S_n - E(S_n))^2) = \frac{\text{Var}(S_n)}{(E(S_n))^2} \\ &= \frac{\sum_{k=0}^n \text{Var}(\mathbf{1}_{A_k})}{(\sum_{k=0}^n \mathbf{P}(A_k))^2} \quad (\text{par ind. 2 à 2}) \\ &= \frac{\sum_{k=0}^n \mathbf{P}(A_k)(1 - \mathbf{P}(A_k))}{(\sum_{k=0}^n \mathbf{P}(A_k))^2} \leq \frac{1}{\sum_{k=0}^n \mathbf{P}(A_k)} \rightarrow 0 \quad \text{quand } n \rightarrow \infty \end{aligned}$$

On en déduit ((\*) voir cours intégration ou détails plus bas) qu'il existe une sous-suite  $n_k \rightarrow \infty$ , telle que p.s.,  $\frac{S_{n_k}}{E(S_{n_k})} \rightarrow 1$ . Autrement dit il existe un ensemble mesurable  $\Omega' \subset \Omega$  de probabilité  $\mathbf{P}(\Omega') = 1$ , tel que pour tout  $\omega \in \Omega'$ ,  $\frac{S_{n_k}(\omega)}{E(S_{n_k})} \rightarrow 1$ . Pour un tel  $\omega \in \Omega'$ , bien sûr, vue l'hypothèse  $E(S_n) \rightarrow +\infty$ ,  $S_{n_k}(\omega) \rightarrow +\infty$ . Donc  $\omega$  appartient à une infinité de  $A_i$ . Donc  $\Omega' \subset \limsup_{n \rightarrow \infty} A_n$ , donc  $\mathbf{P}(\limsup_{n \rightarrow \infty} A_n) = 1$ .  $\square$

En guise de rappel, énonçons un

**Lemme 3.4.4** *Si la suite de fonctions  $(f_n)_{n \in \mathbb{N}}$  converge vers  $f$  dans  $L^2(\Omega, \mathcal{A}, \mathbf{P})$ , alors il existe une sous-suite  $(f_{n_k})_{k \in \mathbb{N}}$  qui converge presque sûrement vers  $f$  (et dans  $L^2$ ).*

**Démonstration :** Comme  $\|f_n - f\|_2 \rightarrow 0$ , il est possible de construire une sous-suite  $f_{n_k}$  tq pour tout  $k \in \mathbb{N}$ ,  $\|f_{n_k} - f\|_{L^2} \leq \frac{1}{2^k}$ . Posons alors  $g_k = f_{n_k} - f$ . On en déduit que la série  $\sum_{k \in \mathbb{N}} |g_k|$  est absolument convergente dans  $L^2$ . L'inégalité de Minkowski assure que pour tout  $K \in \mathbb{N}$ ,  $\left\| \sum_{k=1}^K g_k \right\|_{L^2} \leq \sum_{k=1}^K \|g_k\|_{L^2}$ .

En passant à la limite, on en déduit bien sûr que  $(\sum_{k=1}^{\infty} |g_k|)^2$  est intégrable, et donc finie presque sûrement. Donc  $\sum_{k=1}^{\infty} g_k$  converge presque sûrement. Donc  $g_k \rightarrow 0$  presque sûrement quand  $k \rightarrow \infty$ . Donc  $f_{n_k} \rightarrow f$  quand  $k \rightarrow \infty$ , presque sûrement (et dans  $L^2$  bien sûr).  $\square$

## 3.5 Espérance et indépendance

**Proposition 3.5.1** *Soient  $X, Y$  deux variables aléatoires réelles.*

*$X$  et  $Y$  sont indépendantes si et seulement si pour toutes fonctions  $f$  et  $g$  mesurables bornées,*

$$E(f(X)g(Y)) = E(f(X))E(g(Y)).$$

*Le même résultat est vrai en considérant des fonctions mesurables positives.*

**Démonstration :** Si l'égalité est vraie pour toutes fonctions mesurables bornées, on peut prendre  $f = 1_A$  et  $g = 1_B$  avec  $A, B$  boréliens quelconques. Ceci donne

$$P(X \in A \text{ et } Y \in B) = E(1_A(X)1_B(Y)) = E(1_A(X))E(1_B(Y)) = P(X \in A)P(Y \in B)$$

Donc  $X$  et  $Y$  sont indépendantes.

Si  $X$  et  $Y$  sont indépendantes, le résultat est immédiat pour les indicatrices de boréliens. On passe ensuite aux fonctions simples, puis aux fonctions mesurables bornées par convergence dominée, ou aux fonctions mesurables positives par convergence monotone.  $\square$

Voici une variante du théorème précédent :

**Proposition 3.5.2** Si  $X$  et  $Y$  sont indépendantes et intégrables, alors  $XY$  est intégrable et  $E(XY) = E(X)E(Y)$ .

Ce n'est pas un corollaire du théorème précédent. Par ailleurs, la réciproque est fautive. Si  $E(XY) = E(X)E(Y)$  on dit que  $X$  et  $Y$  sont non corrélées, mais ça ne signifie pas qu'elles sont indépendantes.

**Démonstration :** D'abord on se ramène à des variables aléatoires mesurables bornées.

Soit  $n \geq 1$  et  $f_n(x) = x.1_{[-n,n]}(x)$ . Alors les v.a.  $f_n(X)$  et  $f_n(Y)$  sont mesurables bornées. Le théorème précédent montre que  $E(f_n(X)f_n(Y)) = E(f_n(X))E(f_n(Y))$  pour tout  $n \geq 1$ .

Si  $X$  et  $Y$  sont des variables aléatoires positives, alors  $(f_n(X))_{n \in \mathbb{N}}$ ,  $(f_n(Y))_{n \in \mathbb{N}}$ ,  $(f_n(X)f_n(Y))_{n \in \mathbb{N}}$ , sont des suites monotones de v.a. positives qui convergent respectivement vers  $X$ ,  $Y$  et  $XY$ . Le théorème de convergence monotone (Beppo Levi) permet de conclure que  $E(XY) = E(X)E(Y)$ .

Si  $X$  et  $Y$  sont intégrables, on écrit  $X = X^+ - X^-$ , avec  $X^+ = \max(X, 0)$ , et  $X^- = \max(-X, 0)$ . De même, on écrit  $Y = Y^+ - Y^-$ , avec  $Y^+ = \max(Y, 0)$ , et  $Y^- = \max(-Y, 0)$ . Alors les v.a.  $X^\pm$  et  $Y^\pm$  sont positives. On a d'après ci-dessus  $E(X^\pm)E(Y^\pm) = E(X^\pm Y^\pm)$ . On calcule alors

$$\begin{aligned} E(X)E(Y) &= E(X^+ - X^-)E(Y^+ - Y^-) = (E(X^+) - E(X^-)) (E(Y^+) - E(Y^-)) \\ &= E(X^+)E(Y^+) + E(X^-)E(Y^-) - E(X^+)E(Y^-) - E(X^-)E(Y^+) \\ &= \dots \text{complétez} \dots \\ &= E(X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+) = E(XY). \end{aligned}$$

D'où le résultat. □

**Proposition 3.5.3** Si  $X$  et  $Y$  sont indépendantes et intégrables, alors

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

**Démonstration :** C'est équivalent à la proposition précédente. En effet, on a

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2E(XY) - 2E(X)E(Y).$$

□

La réciproque est fautive. Par exemple regardez  $X$  une variable aléatoire de loi uniforme sur  $[-1, 1]$ , et  $Y = X^2$ . Alors  $X$  et  $Y$  sont non corrélées mais pas indépendantes. (A vérifier en EXO)

## 3.6 Indépendance et fonctions génératrices

Cette section ne sera pas développée, vu que les résultats sont analogues à ceux que l'on obtient pour les fonctions caractéristiques. Mentionnons tout de même les résultats suivants.

**Proposition 3.6.1** Soient  $X$  et  $Y$  deux v.a. à valeurs dans  $\mathbb{N}$  indépendantes. Alors, pour tout  $s \in [0, 1]$ , on a

$$g_{X+Y}(s) = g_X(s)g_Y(s).$$

**Démonstration :** Par définition, et indépendance (proposition 3.5.1 avec des fonctions positives), on a

$$g_{X+Y}(s) = E(s^{X+Y}) = E(s^X s^Y) = E(s^X)E(s^Y) = g_X(s)g_Y(s).$$

□

Un exemple d'application : Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes de lois de Poisson de paramètres respectifs  $\lambda$  et  $\mu$ , alors

$$g_{X+Y}(s) = e^{(\lambda+\mu)(s-1)},$$

d'où l'on déduit que  $X + Y$  suit une loi de Poisson de paramètre  $\lambda + \mu$ .

Mentionnons également un critère d'indépendance. Soient  $X_1$  et  $X_2$  deux v.a. à valeurs entières, et  $X = (X_1, X_2)$ , le couple associé à valeurs dans  $\mathbb{N}^2$ . On définit pour tout  $(u, v) \in [0, 1]^2$  la *fonction génératrice du couple*  $(X, Y)$  comme la fonction de deux variables

$$g_{(X,Y)}(u, v) = \sum_{m \in \mathbb{N}, n \in \mathbb{N}} \mathbf{P}(X=m, Y=n) u^m v^n = E(u^X v^Y).$$

Alors on peut démontrer

**Proposition 3.6.2** *Les v.a.  $X$  et  $Y$  à valeurs dans  $\mathbb{N}$  sont indépendantes ssi pour tout couple  $(u, v) \in [0, 1]^2$ ,*

$$g_{(X,Y)}(u, v) = g_X(u)g_Y(v).$$

**Démonstration :** Si  $X$  et  $Y$  sont indépendantes, le résultat est un corollaire de la proposition 3.5.1. Réciproquement, si les deux quantités sont égales, en calculant les dérivées partielles de tous ordres des deux termes de l'égalité en  $(0, 0)$ , on obtient pour tout  $(m, n) \in \mathbb{N} \times \mathbb{N}$   $\mathbf{P}(X=m, Y=n) = \mathbf{P}(X=m)\mathbf{P}(Y=n)$ .  $\square$

### 3.7 Indépendance et fonctions caractéristiques

**Théorème 3.7.1** *Soit  $X_1$  une v.a. à valeurs dans  $\mathbb{R}^p$ ,  $X_2$  une v.a. à valeurs dans  $\mathbb{R}^q$  et  $X = (X_1, X_2)$  v.a. à valeurs dans  $\mathbb{R}^{p+q}$ . Alors les v.a.  $X_1$  et  $X_2$  sont indépendantes si et seulement si*

$$\varphi_X(u_1, u_2) = \varphi_{X_1}(u_1)\varphi_{X_2}(u_2) \quad \text{pour tout } (u_1, u_2) \in \mathbb{R}^2.$$

**Démonstration :** Supposons  $X_1$  et  $X_2$  indépendantes. Remarquons que  $\varphi_X(u) = E(e^{iu \cdot X}) = E(e^{iu_1 \cdot X_1} e^{iu_2 \cdot X_2})$ . Définissons  $f(x) = e^{iu_1 x}$  et  $g(x) = e^{iu_2 x}$ .  $f$  et  $g$  sont mesurables et bornées. La proposition 3.5.1 donne  $E(f(X_1)g(X_2)) = E(f(X_1))E(g(X_2))$ . Soit encore  $\varphi_X(u_1, u_2) = \varphi_{X_1}(u_1)\varphi_{X_2}(u_2)$

Réciproquement, supposons que  $\varphi_X(u_1, u_2) = \varphi_{X_1}(u_1)\varphi_{X_2}(u_2)$  pour tout  $(u_1, u_2) \in \mathbb{R}^2$ . Autrement dit,  $\hat{P}_X = \hat{P}_{X_1} \hat{P}_{X_2} = \widehat{P_{X_1} \times P_{X_2}}$ . Comme la transformée de Fourier d'une mesure la détermine (théorème 2.3.6), on en déduit que  $P_{(X_1, X_2)} = P_X = P_{X_1} \times P_{X_2}$ . D'après le chapitre 2, ceci implique que  $X_1$  et  $X_2$  sont indépendantes.  $\square$

### 3.8 Somme de variables aléatoires indépendantes.

**Exercice 3.8.1** Soit  $(X, Y)$  une v.a. à valeurs dans  $\mathbb{R}^2$ , de densité  $f_{(X,Y)}$ . Trouver la densité de  $Z = X + Y$ . (*Indication : Trouver d'abord la densité du couple  $(Z, W)$ , où  $W = Y$ .*)

**Corrigé :** *D'abord, on regarde le couple  $(Z, W)$ . On introduit l'application  $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  telle que  $g(x, y) = (x + y, y)$ . C'est un difféomorphisme de  $\mathbb{R}^2$  dans  $\mathbb{R}^2$ . On applique le théorème du changement de variable (avec  $\det J_{g^{-1}}(y) = 1$ ), et on obtient  $f_{(Z,Y)}(z, y) = f_{(X,Y)}(z - y, y)$ . Ensuite, la densité de  $Z$  s'obtient comme densité marginale du couple  $(Z, Y)$ , soit encore*

$$f_Z(z) = \int_{\mathbb{R}} f_{(Z,Y)}(z, y) dy = \int_{\mathbb{R}} f_{(X,Y)}(z - y, y) dy.$$

*Le cas particulier à retenir est le suivant. Quand  $X$  et  $Y$  sont indépendantes, alors  $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$  d'où*

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(z - y)f_Y(y) dy = \int_{\mathbb{R}} f_X(y)f_Y(z - y) dy = f_X * f_Y(z),$$

*où  $f_X * f_Y$  est le produit de convolution de  $f_X$  et  $f_Y$  défini ci-dessous.*

On définit le *produit de convolution* de deux fonctions  $f$  et  $g$  de  $\mathbb{R}$  dans  $\mathbb{R}$  par

$$f * g(x) = \int_{\mathbb{R}} f(x - y)g(y) dy$$

On montre que  $f * g = g * f$ . L'exercice ci-dessus montre que si  $X$  et  $Y$  sont des v.a. à densité indépendantes alors

$$f_{X+Y} = f_X * f_Y.$$

On définit le produit de convolution de deux mesures  $\mu$  et  $\nu$  sur  $\mathbb{R}$  par

$$\mu * \nu(A) = \int_{\mathbb{R}^2} \mathbf{1}_A(x+y) d\mu(x) d\nu(y).$$

Si  $X_1$  et  $X_2$  sont indépendantes, alors

$$P_{(X_1, X_2)} = P_{X_1} \times P_{X_2} \quad \text{et} \quad P_{X_1+X_2} = P_{X_1} * P_{X_2}$$

d'où

$$\begin{aligned} \varphi_{(X_1, X_2)}(u_1, u_2) &= \varphi_{X_1}(u_1) \varphi_{X_2}(u_2) \quad \text{pour tout } (u_1, u_2) \in \mathbb{R}^2 \quad \text{et} \\ \varphi_{X_1+X_2}(u) &= \varphi_{X_1}(u) \varphi_{X_2}(u) \quad \text{pour tout } u \in \mathbb{R}. \end{aligned}$$

Plus généralement, si  $X_1, \dots, X_n$  sont  $n$  v.a. indépendantes et de même loi que la va  $Y$ , et si  $S_n = X_1 + \dots + X_n$ , alors (à retenir)

$$\begin{aligned} \varphi_{S_n}(u) &= E(e^{iu \cdot S_n}) = E(e^{iu \cdot (X_1 + \dots + X_n)}) = \varphi_{(X_1, \dots, X_n)}(u, \dots, u) = \varphi_{X_1}(u) \dots \varphi_{X_n}(u) \\ &= \varphi_Y(u)^n \end{aligned}$$

Par exemple, si  $X_1, \dots, X_n$  sont des va indépendantes de loi de Bernoulli  $\mathcal{B}(p)$ ,  $p \in ]0, 1[$ , alors  $S_n$  suit une loi binomiale  $\mathcal{B}(n, p)$ . En effet,  $\varphi_{S_n}(u) = (\varphi_{X_1}(u))^n = (pe^{iu} + (1-p))^n$  et si  $Z$  est une v.a. binomiale  $\mathcal{B}(n, p)$  alors un calcul donne  $\varphi_Z(u) = (pe^{iu} + 1 - p)^n$  (faites le calcul). Donc  $\varphi_Z = \varphi_{S_n}$  donc  $Z$  et  $S_n$  suivent la même loi. Ceci permet de calculer rapidement  $E(Z)$  et  $Var(Z)$ .  $E(Z) = E(S_n) = nE(X_1) = np$  et  $Var(Z) = Var(S_n) = nVar(X_1) = np(1-p)$ .

# Chapitre 4

## Convergences et théorèmes limites

Le but de ce chapitre est de comprendre ce qui se passe lorsqu'on répète une infinité de fois une expérience. L'exemple le plus simple est celui du lancer d'une pièce équilibrée. À la limite, la fréquence observée de pile doit converger vers  $1/2$ . Soit  $P(n)$  le nombre de pile observés en  $n$  lancers. Alors  $P(n)/n$  tend vers  $1/2$ . Mais que peut-on dire de la suite  $P(n)$ ? Est-elle proche de  $n/2$ ? Et proche en quel sens?

Les théorèmes limites, et en particulier les lois des grands nombres et le théorème limite central, répondent à ce type de questions. Encore faut-il préciser les différentes notions de convergence utilisées. C'est le programme de ce chapitre.

Au programme du CAPES (et de l'agreg), il faut connaître tout ce qui concerne les approximations de certaines lois par d'autres. Je vous renvoie pour cela au cours de L2 et aux livres de la BU.

### 4.1 Convergences de variables aléatoires

Étant donnée une suite de variables aléatoires  $X_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , que signifie la phrase : « la suite  $(X_n)_{n \in \mathbb{N}}$  converge ». La première idée qui vient à l'esprit est la suivante, que vous devez connaître.

La suite  $(X_n)_{n \in \mathbb{N}}$  converge simplement vers  $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  si pour tout  $\omega \in \Omega$ , la suite  $(X_n(\omega))_{n \in \mathbb{N}}$  converge vers  $X(\omega)$ .

Vous connaissez également la notion de convergence uniforme. Mais ces types de convergence ne sont pas adaptés en probabilités, où ce qui se passe sur les ensembles de mesure nulle ne nous intéresse pas. Nous définirons donc plusieurs types de convergence plus pratiques.

#### 4.1.1 Convergence presque sûre

La suite  $(X_n)_{n \in \mathbb{N}}$  converge presque sûrement vers  $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  (on note  $X_n \rightarrow X$  p.s.) s'il existe un ensemble de probabilité 1 sur lequel elle converge simplement vers  $X$ .

Autrement dit, il existe un ensemble  $E$  avec  $P(E) = 0$  tel que pour tout  $\omega \in \Omega \setminus E$ , la suite  $(X_n(\omega))_{n \in \mathbb{N}}$  converge vers  $X(\omega)$ .

Dit encore autrement, on a

$$P(\{\omega \in \Omega, X_n(\omega) \rightarrow X(\omega)\}) = P(X_n \rightarrow X) = 1.$$

On parle encore de convergence « presque partout », p.p.

Exemples et contre exemple.

\* Prenez la fonction  $X_n : [0, 1] \rightarrow \mathbb{R}$  qui vaut zéro sauf autour de  $\frac{1}{k}$ , si  $n > 2k(k+1)$ .  $X_n$  est alors affine croissante de  $\frac{1}{k} - \frac{1}{n}$  à  $\frac{1}{k}$ , elle vaut 1 en  $1/k$ , puis décroît symétriquement pour valoir 0 en  $1/k + 1/n$ . Quand  $n \rightarrow \infty$ , si  $\omega \notin \{\frac{1}{k}, k \geq 1\}$ ,  $X_n(\omega) \rightarrow X(\omega) = 0$ . Donc  $(X_n)_{n \in \mathbb{N}}$  converge p.s. vers la fonction nulle sur  $[0, 1]$ . Pourtant plus  $n$  est grand, plus il y a de points en lesquels  $X_n$  vaut 1.

\* La suite de fonctions  $X_n : x \mapsto x^n$  définies sur  $[0, 1]$  converge presque sûrement vers la fonction nulle. (Elle converge simplement vers la fonction  $\mathbf{1}_{\{1\}}$ , qui est presque partout nulle.)

**Remarque 4.1.1** L'exemple ci-dessus montre qu'une limite presque sûre n'est pas unique. Toutefois, si  $f_n \rightarrow f$  p.s. et  $f_n \rightarrow g$  p.s. alors  $f = g$  p.s.

### 4.1.2 Convergence en probabilité

On dit que  $(X_n)$  converge vers  $X$  en probabilité, et on note  $X_n \xrightarrow{P} X$  si pour tout  $\varepsilon > 0$ ,  $P(|X_n - X| > \varepsilon) \rightarrow 0$  quand  $n \rightarrow \infty$ , soit encore  $P(\{\omega \in \Omega, |X_n(\omega) - X(\omega)| > \varepsilon\}) \rightarrow 0$ .

Par exemple, si  $\Omega = [-1, 1]$  et  $(X_n)$  est une suite de variables aléatoires indépendantes, avec  $X_n$  nulle sauf sur un intervalle de taille  $\frac{1}{n}$  où elle vaut 1, alors  $X_n \rightarrow 0$  en probabilité, mais pas p.s. En effet, on a  $\sum_{n \in \mathbb{N}} P(X_n = 1) = \sum \frac{1}{n} = +\infty$  donc (Borel Cantelli)  $P(X_n = 1 \text{ infiniment souvent}) = 1$ . Donc sur un ensemble de probabilité 1,  $(X_n)$  ne converge pas.

**Proposition 4.1.2** *Si  $X_n$  converge vers  $X$  p.s., alors  $X_n$  converge vers  $X$  en probabilité.*

**Démonstration :** Il existe un ensemble  $\Omega_0 \subset \Omega$  tel que  $P(\Omega_0) = 1$  et pour tout  $\omega \in \Omega_0$ ,  $X_n(\omega) \rightarrow X(\omega)$ . Soit  $\varepsilon > 0$ . Pour tout  $\omega \in \Omega_0$  il existe  $n(\omega, \varepsilon)$  tel que pour tout  $n \geq n(\omega, \varepsilon)$ , on ait  $|X_n(\omega) - X(\omega)| \leq \varepsilon$ . En particulier, pour tout  $\omega \in \Omega_0$ , on a  $\lim_{n \rightarrow \infty} 1_{\{|X_n - X| > \varepsilon\}}(\omega) = 0$ . Autrement dit  $1_{\{|X_n - X| > \varepsilon\}} \rightarrow 0$  p.s. De plus cette suite de fonctions est bien évidemment bornée par une fonction intégrable (la fonction constante égale à 1), de sorte que le théorème de convergence dominée s'applique et on a  $P(|X_n - X| > \varepsilon) = E(1_{\{|X_n - X| > \varepsilon\}}) \rightarrow 0$  quand  $n \rightarrow \infty$ . Ceci étant vrai pour tout  $\varepsilon > 0$ , cela signifie exactement que  $X_n$  converge vers  $X$  en probabilité.  $\square$

Ces deux types de convergences concernent vraiment les variables aléatoires. Mais ils ne permettent pas de travailler avec l'exemple suivant : Soit  $(p_n)_{n \in \mathbb{N}}$  une suite de réels de  $[0, 1]$  qui converge vers  $p \in [0, 1]$ . Soit  $X_n = 1_{[0, p_n]}$  et  $X = 1_{[0, p]}$ . Alors  $X_n$  suit une loi de Bernoulli de paramètre  $p_n$  (détailler, exo). De même,  $X$  suit une loi de Bernoulli de paramètre  $p$ , et  $p_n \rightarrow p$ . Pourtant,  $X_n$  ne tend pas p.s. vers  $X$ , ni en proba. (détailler, exo). La bonne notion pour ce type de v.a. est expliquée au paragraphe suivant.

### 4.1.3 Convergence de mesures, Convergence en loi

Dans l'exemple précédent,  $X_n \rightarrow 1_{[0, p]} \neq X$  en probas et p.s. Pourtant, la loi de  $X_n$  (Bernoulli de paramètre  $p_n$ ) tend vers la loi de  $X$  au sens ci-dessous.

#### Convergence étroite, faible, vague des mesures

Notons  $C_b(\mathbb{R})$  l'espace des fonctions continues et bornées de  $\mathbb{R}$  dans  $\mathbb{R}$ . Une suite de mesures  $\sigma$ -finies  $(\nu_n)_{n \in \mathbb{N}}$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  converge étroitement vers  $\nu$  si pour toute fonction  $f \in C_b(\mathbb{R})$  continue bornée sur  $\mathbb{R}$ ,

$$\int_{\mathbb{R}} f(x) d\nu_n(x) \rightarrow \int_{\mathbb{R}} f(x) d\nu(x) \quad (4.1)$$

Exemple : si  $(x_n)$  est une suite qui converge vers  $x \in \mathbb{R}$ , alors la suite de masses de Dirac  $\delta_{x_n}$  converge étroitement vers  $\delta_x$ . (Exo)

Exemple : Si  $(\nu_n)$  est une suite de mesures de probabilité à densité  $f_n : \mathbb{R} \rightarrow \mathbb{R}$ , et que la suite  $f_n$  converge simplement vers une fonction  $f$  bornée d'intégrale 1, alors la suite  $(\nu_n)$  converge étroitement vers la mesure  $\nu$  de densité  $f$ . (Exo)

La suite  $(\nu_n)_{n \in \mathbb{N}}$  converge faiblement vers  $\nu$  si (4.1) est vrai pour toute fonction  $f \in C_0(\mathbb{R})$  continue tendant vers 0 à l'infini. Elle converge vaguement vers  $\nu$  si (4.1) est vrai pour toute fonction  $f \in C_K(\mathbb{R})$  continue à support compact de  $\mathbb{R}$  dans  $\mathbb{R}$ .

**Remarque 4.1.3** Comme  $C_K(\mathbb{R}) \subset C_0(\mathbb{R}) \subset C_b(\mathbb{R})$ , la convergence étroite implique la convergence faible, qui elle-même implique la convergence vague. La réciproque est fautive. Par exemple, la suite  $(\delta_n)_{n \in \mathbb{N}}$  converge vaguement et faiblement vers 0, mais pas étroitement. La suite  $(n\delta_n)_{n \in \mathbb{N}}$  converge vaguement vers 0 mais pas faiblement. (EXO : vérifier ces assertions.)

**Proposition 4.1.4** *Si la suite de probabilités  $(\nu_n)_{n \in \mathbb{N}}$  converge vaguement vers une mesure de probabilité  $\nu$ , alors elle converge aussi faiblement et étroitement vers  $\nu$ .*

**Démonstration :** Remarquons qu'il suffit de montrer la convergence de  $\int_{\mathbb{R}} f d\nu_n$  vers  $\int_{\mathbb{R}} f d\nu$  pour les fonctions continues bornées positives. (EXO).

Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction continue bornée positive. Soit  $\varphi_p : \mathbb{R} \rightarrow \mathbb{R}$  la fonction continue, affine par morceaux, à support compact, valant 1 sur  $[-p, p]$ , 0 hors de  $[-p-1, p+1]$ , et affine continue sur les deux intervalles restant.

Soit  $\varepsilon > 0$ . Par convergence monotone, la suite  $\int_{\mathbb{R}} \varphi_p d\nu$  converge vers 1 quand  $p \rightarrow \infty$ . Pour tout  $p \geq p_0$  assez grand, nous pouvons supposer que  $1 - \varepsilon \leq \int_{\mathbb{R}} \varphi_p d\nu \leq 1$ .

Comme  $(\nu_n)_{n \in \mathbb{N}}$  converge vaguement vers  $\nu$ , pour tout  $p \geq p_0$ , pour  $n \geq N_p$  assez grand, on a aussi  $1 - 2\varepsilon \leq \int_{\mathbb{R}} \varphi_p d\nu_n \leq 1$ . Il en découle immédiatement que  $-2\varepsilon \|f\|_{\infty} + \int_{\mathbb{R}} f d\nu_n \leq \int_{\mathbb{R}} f_p d\nu_n \leq \int_{\mathbb{R}} f d\nu_n$ .

Soit maintenant  $f_p = f \times \varphi_p$ . C'est une fonction continue à support compact, positive, et la suite  $(f_p)_{p \in \mathbb{N}}$  converge en croissant vers  $f$ . Remarquons que pour tout  $p \geq p_0$  fixé, pour  $n \geq N_p$ , on a

$$-2\varepsilon \|f\|_{\infty} + \int_{\mathbb{R}} f d\nu_n \leq \int_{\mathbb{R}} f_p d\nu_n \leq \int_{\mathbb{R}} f d\nu_n.$$

Le terme du milieu converge vers  $\int_{\mathbb{R}} f_p d\nu$  quand  $n \rightarrow \infty$ . On en déduit immédiatement que

$$-2\varepsilon \|f\|_{\infty} + \int_{\mathbb{R}} f d\nu \leq \liminf_{n \rightarrow \infty} \int_{\mathbb{R}} f d\nu_n \leq \limsup_{n \rightarrow \infty} \int_{\mathbb{R}} f d\nu - n \leq \int_{\mathbb{R}} f d\nu + 2\varepsilon \|f\|_{\infty}.$$

En faisant maintenant tendre  $\varepsilon$  vers 0, on obtient le résultat voulu.  $\square$

## Convergence en loi des variables aléatoires

La suite  $(X_n)_{n \in \mathbb{N}}$  converge vers  $X$  en loi si la suite  $P_{X_n}$  converge étroitement vers  $P_X$ . C'est donc une notion bien plus faible que les deux précédentes (mais aussi utile).

En termes plus probabilistes, la suite  $X_n$  tend vers  $X$  en loi si pour toute fonction  $f \in C_b(\mathbb{R})$ ,

$$E(f(X_n)) \rightarrow E(f(X))$$

**Attention** La convergence en loi n'est pas une convergence de la suite des fonctions  $X_n$  mais de la suite des mesures  $P_{X_n}$ . C'est par abus de langage que les probabilistes disent « la suite  $(X_n)$  converge en loi ».

Reprenons l'exemple du paragraphe précédent : Soit  $X_n = 1_{[0, p_n]}$  et  $X = 1_{[0, p]}$ , avec  $p_n \rightarrow p$ . Alors la suite  $X_n$  converge en loi vers  $X$ , i.e. la suite des lois de Bernoulli  $P_{X_n}$  sur  $\{0, 1\}$  de paramètre  $p_n$  converge étroitement vers la loi de Bernoulli  $P_X$  de paramètre  $p$ . (**EXO**)

**Théorème 4.1.5** Soit  $(X_n)$  une suite de v.a. réelles sur  $(\Omega, \mathcal{A}, P)$ , et  $X$  une v.a. réelle sur  $(\Omega, \mathcal{A}, P)$ . Les propriétés suivantes sont équivalentes.

1. La suite  $(X_n)$  converge en loi vers  $X$
2. La suite de fonctions de répartition vérifie  $F_{X_n}(x) \rightarrow F_X(x)$  pour tout  $x \in \mathbb{R}$  en lequel  $F_X$  est continue.
3. La suite de fonctions caractéristiques  $(\varphi_{X_n})$  converge simplement vers  $\varphi_X$ , i.e. pour tout  $t \in \mathbb{R}$   $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$ .

**Démonstration :** 1. implique 3. : Fixons  $t \in \mathbb{R}$ . Par définition,  $\varphi_{X_n}(t) = E(e^{itX_n}) = E(f(X_n))$  avec  $f : x \rightarrow e^{itx}$  continue et bornée de  $\mathbb{R}$  dans  $\mathbb{C}$ . Comme  $X_n \rightarrow X$  en loi,  $E(f(X_n)) \rightarrow E(f(X))$ , i.e.  $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$ .

3. implique 1. \* On sait que pour tout  $t \in \mathbb{R}$ , on a  $\int_{\mathbb{R}} e^{itx} dP_{X_n}(x) \rightarrow \int_{\mathbb{R}} e^{itx} dP_X(x)$  quand  $n \rightarrow \infty$ . Si  $f \in L^1(\mathbb{R})$ , alors par le théorème de Fubini,

$$\int_{\mathbb{R}} f(t) e^{itx} dP_{X_n}(x) dt = \int_{\mathbb{R}} f(t) \varphi_{X_n}(t) dt = \int_{\mathbb{R}} \hat{f}(x) dP_{X_n}(x)$$

De plus,  $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$ . Par convergence dominée, on en déduit que  $\int_{\mathbb{R}} f(t) \varphi_{X_n}(t) dt \rightarrow \int_{\mathbb{R}} f(t) \varphi_X(t) dt$ , et donc que  $E(\hat{f}(X_n)) \rightarrow E(\hat{f}(X))$  quand  $n \rightarrow \infty$ . Autrement dit, pour toute fonction  $g$  qui s'écrit sous la forme  $g = \hat{f}$ , pour une fonction  $f \in L^1(\mathbb{R})$ , on a la convergence souhaitée. Maintenant, il faudrait obtenir  $E(g(X_n)) \rightarrow E(g(X))$  pour toute fonction continue bornée, ou continue tendant vers 0 à l'infini, ou continue à support compact.

\*\* On a  $C_K^2(\mathbb{R}) \subset \{f \in L^1, \hat{f} \in L^1\} \subset \mathcal{F}(L^1) \subset C_0(\mathbb{R})$ .

La première inclusion vient du fait que  $\hat{f}(t) = -\frac{1}{t^2} f''(t)$ . En effet, la fonction  $f$  étant  $C^2$  à support compact,

sa dérivée seconde est continue à support compact. Sa transformée de Fourier  $\hat{f}''$  est donc bornée, et  $\hat{f}$  est donc en  $O(\frac{1}{t^2})$  en  $\pm\infty$ .

La deuxième inclusion vient de la formule d'inversion de Fourier : si  $\hat{f} \in L^1$ , alors  $f = \hat{g}$ , avec  $g(x) = \frac{1}{2\pi} \hat{f}(-x)$ .

La troisième inclusion est classique (lemme de Riemann Lebesgue), voir le cours d'intégration.

\*\*\* L'espace  $C_K^2(\mathbb{R})$  est dense dans  $C_0(\mathbb{R})$  pour la topologie de la norme  $\|\cdot\|_\infty$ , donc l'ensemble  $\mathcal{F}(L^1)$  aussi.

Concluons maintenant. Si  $g \in \mathcal{F}(L^1)$ , alors  $g$  s'écrit comme  $g = \hat{h}$ ,  $h \in L^1(\mathbb{R})$ , et par Fubini, on a

$$E(g(X_n)) = \int_{\mathbb{R}} g d\mathbf{P}_{X_n} = \int_{\mathbb{R}} h \hat{\mathbf{P}}_{X_n} dx$$

Par hypothèse,  $\hat{\mathbf{P}}_{X_n}$  converge simplement vers  $\hat{\mathbf{P}}_X$ . Par convergence dominée, on en déduit  $E(g(X_n)) \rightarrow E(g(X))$  quand  $n \rightarrow \infty$ .

Soit maintenant  $f \in C_0(\mathbb{R})$ . Par densité de  $\mathcal{F}(L^1)$  pour la norme  $\|\cdot\|_\infty$ , pour tout  $\varepsilon > 0$ , il existe une fonction  $g_\varepsilon \in \mathcal{F}(L^1)$ , telle que  $\|f - g_\varepsilon\|_\infty \leq \varepsilon$ . Ceci implique  $|E(f(X_n)) - E(g_\varepsilon(X_n))| \leq \varepsilon$  pour tout  $n \in \mathbb{N}$ . De même,  $|E(g_\varepsilon(X)) - E(f(X))| \leq \varepsilon$ . En faisant tendre  $n \rightarrow \infty$ , on obtient

$$E(f(X)) - 2\varepsilon \leq \liminf_{n \rightarrow \infty} E(f(X_n)) \leq \limsup_{n \rightarrow \infty} E(f(X_n)) \leq E(f(X)) + 2\varepsilon$$

En faisant maintenant tendre  $\varepsilon$  vers 0, on obtient le résultat voulu.

1. implique 2. On veut montrer que  $F_{X_n}(x) = E(1_{]-\infty, x]}(X_n)) \rightarrow F_X(x) = E(1_{]-\infty, x]}(X))$ . Ceci serait immédiat si la fonction  $1_{]-\infty, x]}$  était continue (elle est bornée). L'idée est de l'approcher par des fonctions continues.

Soit  $x \in \mathbb{R}$  un point en lequel  $F_X$  est continue. Rappelons que  $F_X$  a au plus un nombre dénombrable de discontinuités. Soit  $f_\varepsilon$  la fonction qui vaut 1 sur  $] -\infty, x - \varepsilon]$  puis décroît linéairement jusqu'à valoir 0 sur  $[x, +\infty[$ . Soit  $g_\varepsilon$  la fonction qui vaut 1 sur  $] -\infty, x]$  puis décroît linéairement jusqu'à valoir 0 sur  $[x + \varepsilon, +\infty[$ . Alors  $f_\varepsilon \leq 1_{]-\infty, x]} \leq g_\varepsilon$ , d'où  $E(f_\varepsilon(X_n)) \leq F_{X_n}(x) \leq E(g_\varepsilon(X_n))$ .

Comme  $f_\varepsilon$  et  $g_\varepsilon$  sont continues,  $E(f_\varepsilon(X_n)) \rightarrow E(f_\varepsilon(X))$  et  $E(g_\varepsilon(X_n)) \rightarrow E(g_\varepsilon(X))$  quand  $n \rightarrow \infty$ . On en déduit aisément en faisant tendre  $n \rightarrow \infty$  que

$$E(f_\varepsilon(X)) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq E(g_\varepsilon(X)).$$

On peut maintenant passer à la limite quand  $\varepsilon \rightarrow 0$ . Cela donne

$$E(1_{]-\infty, x]}) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq E(1_{]-\infty, x]}) = F_X(x).$$

Comme  $F_X$  est continue en  $x$ , on a  $E(1_{]-\infty, x]}) = P(X < x) = P(X \leq x) = F_X(x)$ . On en déduit le résultat voulu :  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ .

2. implique 1. Le problème est d'approcher  $f$  par des indicatrices d'intervalles  $]c, d]$ . Pour cela, on va "découper  $f$  en morceaux". Fixons  $\varepsilon > 0$ . D'abord, ramenons nous à un intervalle borné au lieu de  $\mathbb{R}$ , sur lequel  $f$  sera uniformément continue. D'après 4.1.4, on peut supposer que  $f$  est continue à support compact, et que le support de  $f$  est inclus dans un intervalle  $]a, b]$ , tel que  $F_X$  soit continue en  $a$  et en  $b$ .

La fonction  $f$  est uniformément continue sur  $]a, b]$ . Il existe donc  $\delta > 0$  tel que si  $|x - y| < \delta$  alors  $|f(x) - f(y)| < \varepsilon$ .

Découpons  $]a, b]$  en  $k \geq \frac{b-a}{\delta}$  intervalles  $]a_i, a_{i+1}]$ ,  $0 \leq i \leq k-1$ ,  $a_0 = a$ ,  $a_k = b$ , de longueurs  $|a_{i+1} - a_i|$  inférieures à  $\delta$  et tels que  $F_X$  soit continue en  $a_i$ . Sur  $]a_i, a_{i+1}]$  on va approcher  $f$  par  $f(a_i)$ . Et on majore

$$\begin{aligned} |E(f(X_n)) - E(f(X))| &\leq |E(f(X_n)) - E(f(X))| \\ &\leq \|f\|_\infty \varepsilon + \left| E(f(X_n)) - \sum_{i=0}^k f(a_i) P(X_n \in ]a_i, a_{i+1}]) \right| \\ &\quad + \sum_{i=0}^k |f(a_i)| |P(X_n \in ]a_i, a_{i+1}]) - P(X \in ]a_i, a_{i+1}])| \\ &\quad + \left| \sum_{i=0}^k f(a_i) P(X \in ]a_i, a_{i+1}]) - E(f(X)) \right| \end{aligned}$$



Le terme  $\left| E(f(X_n)1_{]a,b]}(X_n)) - \sum_{i=0}^k f(a_i)P(X_n \in ]a_i, a_{i+1}]) \right|$  vaut au plus  $\varepsilon P(X_n \in ]a, b]) \leq \varepsilon$ , car  $f$  est uniformément continue et les intervalles  $]a_i, a_{i+1}]$  sont de taille inférieure à  $\delta$  (qui a été choisi pour que  $f$  varie au plus de  $\varepsilon$  sur ces intervalles). Le terme  $\left| \sum_{i=0}^k f(a_i)P(X \in ]a_i, a_{i+1}]) - E(f(X)1_{]a,b]}(X)) \right|$  se majore exactement de la même façon.

Remarquons que  $P(X_n \in ]a_i, a_{i+1}]) = F_{X_n}(a_{i+1}) - F_{X_n}(a_i) \rightarrow F_X(a_{i+1}) - F_X(a_i) = P_X(]a_i, a_{i+1}])$  car les  $a_i$  ont été choisis pour que  $F_X$  soit continue en  $a_i$ . Donc il existe un certain rang  $N_0$  à partir duquel  $|P(X_n \in ]a_i, a_{i+1}]) - P_X(]a_i, a_{i+1}])| \leq \delta\varepsilon$ . Le terme  $\sum_{i=0}^k |f(a_i)| |P(X_n \in ]a_i, a_{i+1}]) - P(X \in ]a_i, a_{i+1}])|$  est donc majoré par  $\varepsilon(b-a)\|f\|_\infty$  pour tout  $n \geq N_0$ .

Finalement, pour tout  $\varepsilon > 0$ , à partir d'un certain rang  $N = \max(N_0, N_1)$ , on a  $|E(f(X_n)) - E(f(X))| \leq 2\varepsilon + \varepsilon(b-a)\|f\|_\infty$ . On a bien montré ce qu'on voulait, à savoir  $E(f(X_n)) \rightarrow E(f(X))$  quand  $n \rightarrow \infty$ .  $\square$

**Théorème 4.1.6** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a. Si  $(X_n)$  converge presque sûrement vers  $X$ , alors elle converge en probabilité vers  $X$ .

Si  $(X_n)_{n \in \mathbb{N}}$  converge en probabilité vers  $X$ , alors  $(X_n)$  converge en loi vers  $X$ , i.e.  $(P_{X_n})_{n \in \mathbb{N}}$  converge étroitement vers  $P_X$ .

**Remarque 4.1.7** Si  $X_n \rightarrow X$  p.s., et si  $f : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction continue, alors  $f(X_n) \rightarrow f(X)$  p.s. (Démontrez-le, EXO).

**Démonstration :** Seule la deuxième partie du théorème reste à démontrer (cf plus haut). Supposons que  $X_n$  converge vers  $X$  en probabilité. Soit  $x$  un point en lequel  $F_X$  est continue. On veut montrer que  $F_{X_n}(x) \rightarrow F_X(x)$ . Remarquons d'abord que

$$\begin{aligned} F_X(x - \varepsilon) &= P(X \leq x - \varepsilon) = P(X \leq x - \varepsilon \text{ et } |X_n - X| > \varepsilon) + P(X \leq x - \varepsilon \text{ et } |X_n - X| \leq \varepsilon) \\ &\leq P(|X_n - X| > \varepsilon) + F_{X_n}(x) \end{aligned}$$

Par définition de la convergence en probabilité,  $\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$ . On en déduit que pour tout  $\varepsilon > 0$ ,

$$F_X(x - \varepsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x).$$

De la même façon, on a

$$F_{X_n}(x) = P(X_n \leq x) = P(X_n \leq x \text{ et } |X_n - X| > \varepsilon) + P(X_n \leq x \text{ et } |X_n - X| \leq \varepsilon) \leq P(|X_n - X| > \varepsilon) + F_X(x + \varepsilon).$$

On en déduit que  $\limsup F_{X_n}(x) \leq F_X(x + \varepsilon)$ . En passant à la limite quand  $\varepsilon \rightarrow 0$ , comme  $F_X$  est continue en  $x$ , on obtient  $F_X(x) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x)$ . Ceci donne bien le résultat voulu, à savoir que  $F_{X_n}(x) \rightarrow F_X(x)$ .  $\square$

Dans un cas très particulier, qui nous servira plus tard, on a quand même le résultat suivant dans l'autre sens.

**Proposition 4.1.8** Si  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers une variable aléatoire constante  $X = c$ , de loi  $P_X = \delta_c$  égale à la masse de Dirac  $\delta_c$  en  $c$ , alors  $X_n$  converge en probabilité vers  $X = c$ .

**Démonstration :** On a  $P(|X_n - X| < \varepsilon) \geq P(-\varepsilon < X_n - c \leq \frac{\varepsilon}{2}) = F_{X_n}(c + \frac{\varepsilon}{2}) - F_{X_n}(c - \varepsilon)$ . Quand  $n \rightarrow \infty$  on en déduit  $\liminf P(|X_n - X| < \varepsilon) \geq F_X(c + \varepsilon/2) - F_X(c - \varepsilon) = 1$ . Ceci implique bien sûr  $P(|X_n - X| < \varepsilon) \rightarrow 1$  quand  $n \rightarrow \infty$ .  $\square$

#### 4.1.4 Convergence dans $L^p$

On dit que la suite  $X_n$  converge vers  $X$  dans  $L^p$ ,  $p > 0$ , si  $E(|X_n - X|^p) \rightarrow 0$  quand  $n \rightarrow \infty$ . Cette notion de convergence, également très utile, a été étudiée en intégration, et sera revue abondamment en M1 en analyse et en probas.

## 4.2 Loi faible et forte des grands nombres

**Motivation :** On répète une expérience un grand nombre  $n$  de fois. Par exemple lancer d'une pièce (éventuellement truquée)  $n$  fois, On pose  $X_n = 0$  si le  $n$ -ième lancer est un pile et  $X_n = 1$  si c'est un face. Les lois des grands nombres justifient l'intuition que nous avons qui nous dit que la proportion de face obtenus en  $n$  lancers, i.e. la quantité  $\frac{X_1 + \dots + X_n}{n}$ , converge vers l'espérance de chaque v.a.  $X_n$ , à savoir ici  $E(X_n) = P(X_n = 1)$ .

Dans toute cette section,  $(X_n)$  désigne une suite de v.a. *indépendantes et de même loi*. Pour faciliter les notations, on notera  $X$  une v.a. indépendante des  $(X_n)$  et de même loi, de sorte que  $E(|X|)$  désignera l'espérance de chacune des v.a.  $|X_n|$ . On note  $S_n = X_1 + \dots + X_n$ .

### 4.2.1 Loi faible des grands nombres

La loi faible des grands nombres est le résultat (facile) suivant.

**Théorème 4.2.1 (Loi faible des grands nombres)** Si  $E(|X|) < +\infty$  alors  $\frac{S_n}{n}$  converge en probabilité vers  $E(X)$ .

**Remarque 4.2.2** Commençons par une remarque qui servira aussi au paragraphe suivant. On a  $\frac{S_n}{n} - E(X) = \frac{1}{n} \sum_{i=1}^n (X_i - E(X_i))$ . Notons  $Y_i = X_i - E(X_i)$  et  $\Sigma_n = \sum_{i=1}^n Y_i = S_n - nE(X)$ . Alors  $\frac{S_n}{n}$  converge en probabilité (resp. presque sûrement) vers  $E(X)$  si et seulement si  $\frac{\Sigma_n}{n}$  converge en probabilité (resp. presque sûrement) vers  $0 = E(Y)$ . Autrement dit, on peut toujours supposer dans les démonstrations que  $E(X) = 0$ .

**Démonstration :** Vue la remarque ci-dessus, on peut supposer que  $E(X) = 0$ .

- Pour simplifier la preuve, nous supposons d'abord que  $E(|X|^2) < \infty$ . Ceci est plus fort que l'hypothèse  $E(|X|) < \infty$ ; en effet, l'inégalité de Cauchy-Schwartz donne  $E(|X|) = E(1 \cdot |X|) \leq \sqrt{E(1^2)} \sqrt{E(X^2)} = \sqrt{E(X^2)}$ .

D'après l'inégalité de Bienaymé-Chebychev, on a

$$P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{\text{Var}(S_n)}{n^2\varepsilon^2} = \frac{n\text{Var}(X)}{n^2\varepsilon^2} = \frac{E(X^2)}{n\varepsilon^2} \rightarrow 0.$$

- Démontrons maintenant le théorème sous l'hypothèse plus faible  $E(|X|) < \infty$ . Alors la fonction caractéristique  $\varphi_X$  est dérivable et  $\varphi'_X(0) = E(X) = 0$ . La formule de Taylor donne alors  $\varphi_X(t) = 1 + o(t) = 1 + t\varepsilon(t)$  avec  $\varepsilon(t) \rightarrow 0$  en  $O$ . On en déduit qu'à  $t$  fixé quand  $n \rightarrow \infty$   $\varphi_{S_n/n}(t) = \left(\varphi_X\left(\frac{t}{n}\right)\right)^n = \left(1 + \frac{t}{n}\varepsilon(t/n)\right)^n = 1 + o(1) \rightarrow 1$ . Donc  $\varphi_{S_n/n}$  converge vers la fonction constante égale à 1, qui est la fonction caractéristique de la masse de Dirac  $\delta_0$ . Autrement dit,  $X_n$  converge en loi vers 0. D'après la proposition 4.1.8, on en déduit que  $S_n/n$  converge en probabilité vers la variable aléatoire constante égale à 0.  $\square$

**Remarque 4.2.3** Lorsque les  $(X_n)$  sont une suite indépendante de v.a. de Bernoulli de paramètre  $p \in ]0, 1[$ , alors  $E(X_n^2) < \infty$ , et (en utilisant  $p(1-p) \leq \frac{1}{4}$ ), la démonstration précédente donne

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{\text{Var}(S_n)}{n^2\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

Autrement dit,  $S_n/n$  converge en probabilité vers  $p$ , et la convergence est uniforme en  $p$ .

À l'aide de ce résultat (loi faible des grands nombres uniforme en  $p$ ), on peut montrer de façon probabiliste un très joli résultat. Le théorème de Weierstrass dit que les polynômes sont denses dans l'ensemble des fonctions continues sur  $[a, b]$  muni de la topologie de la norme infinie  $\|\cdot\|_\infty$ , soit encore pour toute fonction continue sur un intervalle  $[a, b]$  et pour tout  $\varepsilon > 0$ , il existe un polynôme  $P$  tel que

$$\sup_{x \in [a, b]} |f(x) - P(x)| \leq \varepsilon.$$

Bernstein a démontré cela de la manière suivante (voir [L] par exemple).

**Proposition 4.2.4** Soit  $f : [0, 1] \rightarrow \mathbb{R}$  continue. Alors

$$\sup_{x \in [0, 1]} \left| f(x) - \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} \right| \rightarrow 0 \quad \text{quand } n \rightarrow \infty.$$

**Démonstration :** Considérons une suite  $X_1, \dots, X_n, \dots$  de v.a. Bernoulli  $\mathcal{B}(p)$ ,  $p \in [0, 1]$ . On a  $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$ .

La fonction  $f$  est continue sur un compact, donc uniformément continue.

Fixons  $\varepsilon > 0$ . Il existe  $\eta > 0$  tel que  $|x - y| < \eta$  implique  $|f(x) - f(y)| < \varepsilon$ . On calcule alors

$$\begin{aligned} & \left| f(p) - \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f\left(\frac{k}{n}\right) \right| = |f(p) - E(f\left(\frac{S_n}{n}\right))| \\ & \leq \sum_{k: |p-k/n| < \eta} \binom{n}{k} p^k (1-p)^{n-k} |f(p) - f\left(\frac{k}{n}\right)| + \sum_{k: |p-k/n| \geq \eta} (|f(p)| + |f\left(\frac{k}{n}\right)|) \binom{n}{k} p^k (1-p)^{n-k} \\ & \leq \varepsilon + 2\|f\|_\infty P\left(\left|\frac{S_n}{n} - p\right| \geq \eta\right) \end{aligned}$$

La loi faible des grands nombres uniforme pour des v.a. de Bernoulli dit que

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \eta\right) \leq \frac{p(1-p)}{n\eta^2} \leq \frac{1}{4n\eta^2}.$$

Donc dans le cas particulier d'une loi binomiale, il existe  $N \in \mathbb{N}$  indépendant de  $p \in [0, 1]$  tel que pour tout  $n \geq N$ ,  $P\left(\left|\frac{S_n}{n} - p\right| \geq \eta\right) \leq \varepsilon$ . Donc pour  $n \geq N$ ,  $|f(p) - \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f\left(\frac{k}{n}\right)| \leq \varepsilon + 2\varepsilon\|f\|_\infty$ . Ceci montre bien que  $|f(p) - \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f\left(\frac{k}{n}\right)|$  converge vers 0 quand  $n \rightarrow \infty$ , uniformément en  $p$ .  $\square$

## 4.2.2 Loi forte des grands nombres

**Théorème 4.2.5 (Loi forte des grands nombres)**  $E(|X|) < +\infty$  si et seulement si  $S_n/n$  converge p.s. vers  $E(X)$ .

**Démonstration :** Comme précédemment, on peut supposer que  $E(X) = 0$ . Plus il existe une grande puissance  $p$  telle que  $E(|X|^p) < \infty$ , plus la preuve est facile.

En effet, rappelons au passage que, si  $p_1 < p_2$ , alors  $E(|X|^{p_2}) < \infty$  implique  $E(|X|^{p_1}) < \infty$ . En effet,  $|X|^{p_1} \leq \max(1, |X|^{p_2}) \leq 1 + |X|^{p_2}$ , d'où  $E(|X|^{p_1}) \leq 1 + E(|X|^{p_2})$ .

Nous allons commencer par démontrer le résultat sous l'hypothèse  $E(X^4) < \infty$ ; puis sous l'hypothèse  $E(X^2) < \infty$ . Nous admettrons le cas général.

Supposons d'abord que  $E(|X|^4) < \infty$ .

Le principe de la démonstration est le suivant : on essaie de majorer  $P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right)$  par quelque chose de suffisamment petit pour que la série  $\sum_{n \geq 1} P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right)$  converge (ce qui implique donc la loi faible des grands nombres!) On en déduit (lemme de Borel Cantelli) que  $P(\limsup_{n \rightarrow \infty} \left\{\left|\frac{S_n}{n}\right| \geq \varepsilon\right\}) = 0$ . Autrement dit, il existe un ensemble  $\Omega_\varepsilon$  tel que  $P(\Omega_\varepsilon) = 1$ , tel que pour tout  $\omega \in \Omega_\varepsilon$ , il existe un nombre fini de  $n$  pour lesquels  $\left|\frac{S_n(\omega)}{n}\right| \geq \varepsilon$ . Donc à partir d'un certain rang  $n(\omega, \varepsilon)$ ,  $\left|\frac{S_n(\omega)}{n}\right| < \varepsilon$ .

On considère alors pour tout  $k \in \mathbb{N}^*$   $\varepsilon = 1/k$  l'ensemble  $\Omega_{1/k}$ . L'intersection  $\cap_{k \in \mathbb{N}^*} \Omega_{1/k}$  est de probabilité 1 (EXO : passez au complémentaire pour le démontrer). Et pour tout  $\omega \in \cap_{k \in \mathbb{N}^*} \Omega_{1/k}$ ,  $\frac{S_n(\omega)}{n}$  converge vers 0.

En résumé la preuve consiste maintenant à majorer  $P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right)$  par le terme général d'une série convergente.

Comme  $E(|X|^4) < \infty$ , on a

$$P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{E(S_n^4)}{n^4 \varepsilon^4}$$

Développons  $E(S_n^4)$ . On a

$$E(S_n^4) = \sum_{k=1}^n E(X_k^4) + 4 \sum_{k \neq l} E(X_k X_l^3) + 6 \sum_{k \neq l} E(X_k^2 X_l^2) + 12 \sum_{j \neq k, l, k \neq l} E(X_j X_k X_l^2) + 24 \sum_{i, j, k, l} E(X_i X_j X_k X_l).$$

Les indices sont toujours supposés deux à deux disjoints. À l'aide de l'inégalité de Cauchy Schwarz on montre que  $|E(X_k X_l^3)| \leq E(X^4)$  et  $E(X_k^2 X_l^2) \leq E(X^4)$  et  $E(X_j X_k X_l^2) \leq E(X^4)$  et  $E(X_i X_j X_k X_l) \leq E(X^4)$ .

La première somme est donc majorée par  $nE(X^4)$ , les deux suivantes par cste  $\frac{n(n-1)}{2}E(X^4)$ . En divisant par  $n^4 \varepsilon^4$ , cela donne des termes généraux de suites convergentes (en  $1/n^3$  et  $1/n^2$ ).

Ce raisonnement ne fonctionne pas pour les deux dernières sommes. Mais on utilise alors le fait que  $E(X) = 0$  et l'indépendance des  $X_i$  pour montrer qu'elles sont nulles. (**EXO**). Ceci montre en fait que seules la première et la troisième somme ne sont pas nulles.

On en déduit que pour tout  $\varepsilon > 0$ ,  $\sum_{n \geq 1} P(|\frac{S_n}{n}| \geq \varepsilon)$  converge, et le lemme de Borel-Cantelli permet de conclure, comme mentionné au début de la preuve.

Lorsqu'on suppose seulement (hypothèse plus faible) que  $E(X^2) < +\infty$ , on est obligé d'être plus subtil. Rappelons que puisqu'on a supposé  $E(X) = 0$ , l'inégalité de Bienaymé-Chebychev donne

$$P\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

En remplaçant  $n$  par  $n^2$  et en prenant pour  $\varepsilon = n^{-1/4}$ , on trouve

$$P\left(\left|\frac{S_{n^2}}{n^2}\right| > \frac{1}{n^{1/4}}\right) \leq \frac{\sigma^2}{n^{3/2}}.$$

C'est le terme général d'une série convergente. Donc  $\sum_{n \geq 1} P\left(\left|\frac{S_{n^2}}{n^2} - E(X)\right| > \frac{1}{n^{1/4}}\right) < \infty$ . Le lemme de Borel Cantelli nous permet d'en déduire que  $P(\limsup_{n \rightarrow \infty} A_n) = 0$ , où  $A_n = \{\omega \in \Omega, \left|\frac{S_{n^2}}{n^2} - E(X)\right| > \frac{1}{n^{1/4}}\}$ . Il existe donc un sous-ensemble  $\Omega' \subset \Omega$  de probabilité 1, tel que pour tout  $\omega \in \Omega'$ , on ait  $\left|\frac{S_{n^2(\omega)}}{n^2}\right| \leq \frac{1}{n^{1/4}}$  à partir d'un certain rang  $n(\omega)$ .

Par conséquent, pour tout  $\omega \in \Omega'$ ,  $\frac{S_{n^2(\omega)}}{n^2} \rightarrow 0$  quand  $n \rightarrow \infty$ .

Autrement dit,  $\frac{S_{n^2}}{n^2} \rightarrow 0$  presque sûrement quand  $n \rightarrow \infty$ . C'est presque le résultat voulu : convergence pour une sous-suite au lieu de convergence pour toute la suite.

Notons  $X_n^+ = \max(X_n, 0)$  et  $X_n^- = \max(-X_n, 0)$ . Soit  $S_n^\pm = \sum_{i=1}^n X_n^\pm$ . Le même argument que ci-dessus montre que  $\frac{S_{n^2}^\pm}{n^2}$  converge p.s. quand  $n \rightarrow \infty$ .

Maintenant pour  $n \in \mathbb{N}$  quelconque, il existe un unique  $k \in \mathbb{N}$  tel que  $k^2 \leq n < (k+1)^2$ . On écrit alors

$$\frac{k^2}{(k+1)^2} \frac{S_{k^2}^\pm}{k^2} \leq \frac{k^2}{n} \frac{S_{k^2}^\pm}{k^2} \leq \frac{S_n^\pm}{n} \leq \frac{(k+1)^2}{n} \frac{S_{(k+1)^2}^\pm}{(k+1)^2} \leq \frac{(k+1)^2}{k^2} \frac{S_{(k+1)^2}^\pm}{(k+1)^2}.$$

Les deux termes tout à gauche et tout à droite convergent presque sûrement vers  $E(X)$ . Donc  $S_n^\pm/n$  aussi.  $\square$

**Remarque 4.2.6** Sous l'hypothèse la plus faible possible, à savoir  $E(|X|) < \infty$ , la preuve est plus difficile. Voir par exemple [B-L].

Ce théorème a de nombreuses applications. Citons par exemple celle-ci, qui ne semble pas a priori relever des probabilités : dans l'intervalle  $[0, 1]$ , presque tout nombre est *normal*, i.e. il a autant de 0 que de 1 dans son développement binaire !

Le principe de la démonstration est le suivant. Soit  $X_n : [0, 1] \rightarrow \{0, 1\}$  l'application qui associe à  $\omega$  le  $n$ -ième terme de son développement binaire propre (i.e. qui ne termine pas par une infinité de 1). On montre que  $X_n$  est mesurable, et que la suite  $(X_n)$  est une suite de v.a. indépendantes de Bernoulli de paramètre  $1/2$ .

La loi des grands nombres nous assure alors que  $\frac{1}{n} \sum_{k=1}^n X_n \rightarrow \frac{1}{2}$  quand  $n \rightarrow \infty$ . Et cette quantité est exactement égale à la proportion de 1 dans le développement binaire de  $\omega$ . Autrement dit, cette proportion de 1 tend vers  $1/2$  pour presque tout  $\omega$ , ce qui est le résultat voulu.

Attention : il existe beaucoup de nombres de  $[0, 1]$  n'ayant pas autant de 0 et de 1 dans leur développement. Par exemple, un rationnel a un développement binaire périodique. Le développement  $0, 1110111011101110\dots$  n'a bien évidemment pas autant de 0 et de 1.

### 4.3 Théorème limite central

**Théorème 4.3.1 (Théorème limite central)** *Supposons que  $E(X^2) < +\infty$ . Soit  $\sigma^2 = \text{Var}(X)$ . Alors  $\frac{S_n - nE(X)}{\sigma\sqrt{n}}$  converge en loi vers une v.a. de loi  $\mathcal{N}(0, 1)$  normale centrée réduite.*

La normalisation de ce théorème s'explique de la manière suivante. La loi des grands nombres assure que  $\frac{S_n}{n} - E(X) \rightarrow 0$  p.s. Autrement dit,  $S_n - nE(X) = n\psi(n)$ , où  $\psi(n) \rightarrow 0$  quand  $n \rightarrow \infty$ . Le théorème limite central décrit les variations de  $S_n$  autour de  $nE(X)$ .

Précisément, l'utilisation concrète de ce théorème est la suivante. Pour tous  $a < b$  dans  $\mathbb{R}$ , on a

$$P\left(\frac{S_n - nE(X)}{\sigma\sqrt{n}} \in ]a, b[ \right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

En pratique, on identifie dans les usages le terme de gauche et le terme de droite dès que  $n > 30$ .

Ce théorème illustre l'universalité de la loi normale. En effet, ce théorème est vrai quelle que soit la loi suivie par les  $X_i$ . Autrement dit, si on répète une expérience  $n$  fois avec  $n$  grand, dont le résultat peut être modélisé par une variable aléatoire  $X$  (avec  $E(X^2) < \infty$ ), alors la moyenne  $S_n/n$  des résultats des expériences converge vers  $E(X)$  et les écarts, convenablement normalisés, suivent une loi normale, indépendamment de la loi de  $X$ .

Vous verrez si vous faites des statistiques que ce théorème est crucial en statistiques, en particulier pour construire des intervalles de confiance. Le principe de son utilisation est le suivant. On suppose que  $n$  est assez grand pour avoir le droit de dire  $P\left(\frac{S_n - nE(X)}{\sigma\sqrt{n}} \in ]-a, a[ \right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-a}^a e^{-x^2/2} dx$  (Cette approximation est rarement justifiée en pratique, et on ne parle que rarement de l'erreur commise.) On choisit  $a$  suffisamment grand pour avoir  $\frac{1}{\sqrt{2\pi}} \int_{-a}^a e^{-x^2/2} dx$  proche de 1 (souvent 0,9 ou 0,95 ou 0,99 en pratique.) Puis on dit qu'avec proba 90%, ou 95%, ou 99%, suivant la précision choisie, on a  $\frac{S_n - nE(X)}{\sigma\sqrt{n}} \in ]-a, a[$ .

**Démonstration :** Remarquons que puisque  $E(X) = 0$ , on a  $\text{Var}(X) = \sigma^2 = E(X^2)$ .

Comme  $E(X^2) < \infty$ , la fonction caractéristique  $\varphi^X$  des  $X_i$  est deux fois dérivable. De plus,  $(\varphi^X)'(0) = iE(X) = 0$  et  $(\varphi^X)''(0) = i^2E(X^2) = -E(X^2)$ . La formule de Taylor donne  $\varphi^X(t) = 1 - \frac{t^2}{2}E(X^2) + t^2\varepsilon(t^2)$ . On en déduit qu'à  $t$  fixé, quand  $n \rightarrow \infty$ ,  $\varphi^{S_n/\sigma\sqrt{n}}(t) = (\varphi^X(\frac{t}{\sigma\sqrt{n}}))^n = \left(1 - \frac{t^2}{\sigma^2 n}E(X^2) + \frac{t^2}{\sigma^2 n}\varepsilon(\frac{t^2}{\sigma^2 n})\right)^n \rightarrow e^{-t^2/2}$  quand  $n \rightarrow \infty$ . La fonction  $\varphi^{S_n/\sigma\sqrt{n}}$  converge donc vers la fonction caractéristique d'une loi normale  $\mathcal{N}(0, 1)$ . Autrement dit,  $S_n/\sigma\sqrt{n}$  converge en loi vers la loi normale  $\mathcal{N}(0, 1)$ .  $\square$

Voici une bibliographie très incomplète. Allez voir vous même à la B.U.. Gardez en mémoire qu'un bon livre est un livre qui vous donne envie d'apprendre et de travailler son contenu !

# Bibliographie

- [L] Lesigne, Emmanuel *Pile ou Face* ,.
- [B-L] Barbe, Ledoux *Probabilité* ,
- [CGDM] M. Cottrell, V. Genon-Catalot, C Duhamel, T Meyre, *exercices de probabilités, licence, master, ingénieurs*, Cassini, 2005.
- [D] Durrett *Probability, theory and examples*
- [F-F] Foata Fuchs *Probabilités* ,
- [JP] Jean Jacod, Philip Protter, *L'essentiel en théorie des probabilités*, Cassini, 2003.
- [N] Norris, *Markov Chains*
- [R1] Daniel Revuz, *Mesure et intégration*, Hermann, 1994.
- [R2] Daniel Revuz *Probabilités*, Hermann, 1997.
- [Capes] *Probabilités et statistique, Cours et exercices corrigés* Capes de mathématiques.