

Probabilités et Statistiques élémentaires  
LM 231

Raphaël KRIKORIAN  
Université Pierre et Marie Curie (Paris 6)

Année 2013-2014



# Table des matières

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Rappels de théorie des ensembles</b>  | <b>7</b>  |
| 1.1      | Opérations sur les ensembles . . . . .   | 7         |
| 1.2      | Applications entre ensembles . . . . .   | 8         |
| 1.3      | Dénombrément . . . . .   | 10        |
| 1.4      | Dénombrabilité . . . . .   | 15        |
| <b>2</b> | <b>Espaces Probabilisés</b>  | <b>17</b> |
| 2.1      | Espaces probabilisés . . . . .   | 17        |
| 2.2      | Tribus . . . . .   | 17        |
| 2.3      | Probabilités . . . . .   | 18        |
| 2.3.1    | Définition . . . . .   | 18        |
| 2.3.2    | Exemples généraux . . . . .  | 19        |
| 2.3.3    | Premières propriétés . . . . .   | 19        |
| 2.4      | Probabilités sur un ensemble fini . . . . .  | 22        |
| 2.4.1    | Probabilités uniformes et lien avec la combinatoire . . .                                | 23        |
| 2.4.2    | Jeu de $n$ Pile ou Face : le modèle (fini) de Bernoulli .                                | 25        |
| 2.5      | Le jeu infini de pile ou face : le modèle de Bernoulli . . . . .                         | 27        |
| 2.5.1    | L'espace des états . . . . .   | 27        |
| 2.5.2    | Construction de la tribu des événements . . . . .  | 28        |
| 2.5.3    | Construction de la probabilité . . . . .   | 29        |
| 2.6      | Evénements indépendants . . . . .  | 29        |
| 2.7      | Probabilités conditionnelles . . . . .   | 30        |
| <b>3</b> | <b>Variables aléatoires réelles</b>  | <b>33</b> |
| 3.1      | Variables Aléatoires réelles . . . . .   | 33        |
| 3.1.1    | Définition et premières propriétés . . . . .   | 33        |
| 3.1.2    | Le cas particulier des v.a à valeurs dans un ensemble<br>fini ou dénombrable . . . . .   | 35        |
| 3.2      | Loi d'une variable aléatoire réelle . . . . .  | 36        |
| 3.2.1    | Loi des variables aléatoires à valeurs dans un ensemble<br>fini ou dénombrable . . . . . | 37        |

|          |   |           |
|----------|---|-----------|
| 3.2.2    | Loi de variables aléatoires admettant une densité . . .                     | 39        |
| 3.3      | Espérance d'une v.a. à valeurs dans un ensemble fini (ou dénombrable) . . . | 42        |
| 3.3.1    | Définition . . . . .  | 42        |
| 3.3.2    | Premières propriétés . . . . .  | 43        |
| 3.3.3    | Formule de transfert . . . . .  | 44        |
| 3.4      | Espérance des v.a.r. : cas général . . . . .                                | 45        |
| 3.5      | Espérance des v.a.r. admettant une densité . . . . .                        | 47        |
| 3.5.1    | Résultat fondamental . . . . .  | 47        |
| 3.5.2    | La formule de transfert . . . . .   | 48        |
| 3.5.3    | Application au calcul de densité . . . . .                                  | 48        |
| 3.6      | Variance . . . . .  | 51        |
| 3.6.1    | Variations aléatoires de carré intégrable . . . . .                         | 51        |
| 3.6.2    | Variance . . . . .  | 52        |
| 3.6.3    | Calculs de variance des v.a.r. à valeurs dans $\mathbb{N}$ . . . . .        | 53        |
| 3.6.4    | Cas des v.a admettant une densité . . . . .                                 | 55        |
| 3.7      | Inégalité de Markov et de Bienaymé-Tchebychev . . . . .                     | 57        |
| 3.8      | Vecteurs aléatoires . . . . .   | 58        |
| 3.8.1    | Loi d'un vecteur aléatoire . . . . .  | 58        |
| 3.8.2    | Formules de transfert . . . . .   | 59        |
| 3.8.3    | Loi de d'une somme de v.a.r. . . . .  | 60        |
| 3.9      | Variations aléatoires indépendantes . . . . .                               | 61        |
| 3.9.1    | Définition . . . . .  | 61        |
| 3.9.2    | Cas des v.a à valeurs dans un ensemble discret . . . . .                    | 61        |
| 3.9.3    | Cas des v.a admettant des densités . . . . .                                | 63        |
| 3.9.4    | Loi d'une somme de v.a.r. indépendante . . . . .                            | 63        |
| 3.9.5    | Espérance des produits de v.a indépendantes . . . . .                       | 64        |
| 3.9.6    | Critères d'indépendance . . . . .   | 65        |
| 3.9.7    | Variance d'une somme de v.a.r. indépendantes . . . . .                      | 65        |
| <b>4</b> | <b>Sommes de variables aléatoires indépendantes</b> . . . . .               | <b>67</b> |
| 4.1      | Lois des grands nombres . . . . .   | 68        |
| 4.1.1    | Loi faible des grands nombres dans le cas $\mathcal{L}^2$ . . . . .         | 68        |
| 4.1.2    | Loi forte des grands nombres . . . . .                                      | 69        |
| 4.2      | Théorème de la limite centrale . . . . .                                    | 71        |
| 4.2.1    | Fonctions de répartition . . . . .  | 72        |
| 4.2.2    | Convergence en loi . . . . .  | 72        |
| 4.2.3    | Fonctions caractéristiques . . . . .  | 73        |
| 4.2.4    | Démonstration du théorème de la limite centrale . . . . .                   | 75        |
| 4.3      | Quelques remarques sur les diverses notions de convergence . . . . .        | 77        |

|          |   |           |
|----------|---|-----------|
| <b>5</b> | <b>Rudiments de Statistiques</b>            | <b>79</b> |
| 5.1      | Statistiques . . . . .                      | 79        |
| 5.1.1    | Terminologie . . . . .                      | 79        |
| 5.1.2    | Modélisation . . . . .                      | 79        |
| 5.2      | Estimateurs . . . . .                       | 81        |
| 5.2.1    | Définition . . . . .                        | 81        |
| 5.2.2    | Estimateur sans biais . . . . .             | 81        |
| 5.2.3    | Risque quadratique . . . . .                | 81        |
| 5.2.4    | Moyenne et variance empiriques . . . . .    | 82        |
| 5.2.5    | Maximum de vraisemblance . . . . .          | 82        |
| 5.3      | Intervalles de confiance . . . . .          | 84        |
| 5.4      | Sondages . . . . .                          | 85        |
| 5.5      | Statistiques gaussiennes . . . . .          | 86        |
| 5.5.1    | Sondages gaussiens . . . . .                | 86        |
| 5.5.2    | Loi du chi-deux et loi de Student . . . . . | 86        |
| 5.5.3    | Le théorème fondamental . . . . .           | 87        |
| 5.6      | Test d'hypothèse . . . . .                  | 88        |
| 5.7      | Test du chi-deux . . . . .                  | 89        |



# Chapitre 1

## Rappels de théorie des ensembles

Nous rappelons dans ce chapitre quelques notions élémentaires de théorie des ensembles.

### 1.1 Opérations sur les ensembles

**Ensemble, ensemble fini/infini, cardinal** Un ensemble est intuitivement une collection d'éléments. Etant donné un ensemble  $E$  et un élément  $a$  on écrit  $a \in E$  si  $a$  est un élément de  $E$ . Il existe un unique ensemble ne contenant aucun élément ; on le note  $\emptyset$ . On dit qu'un ensemble est *fini* s'il ne contient qu'un nombre fini d'éléments et *infini* sinon. Si  $A$  est un ensemble fini on appelle *cardinal* de  $A$  le nombre d'éléments de  $A$  et on note ce nombre entier  $\#A$  ou  $\text{card } A$ . Si  $A$  est infini, on pose  $\#A = \infty$ .

**Inclusion, complémentaire** Si  $E$  et  $A$  sont deux ensembles on dit que  $A$  est inclus dans  $E$  ou que  $A$  est un sous-ensemble de  $E$  si tout élément de  $A$  est un élément de  $E$  et on écrit  $A \subset E$ . On peut alors définir le *complémentaire* de  $A$  dans  $E$  qui est l'ensemble des éléments de  $E$  qui n'appartiennent pas à  $A$ . On le notera dans ce cours  $E - A$  ou  $A^c$  ; cette dernière notation cesse d'être ambiguë si l'on suppose  $E$  fixé une fois pour toute, ce que nous ferons.

**Ensemble des parties d'un ensemble** Si  $E$  est un ensemble, l'ensemble constitué des sous-ensembles de  $E$  s'appelle l'ensemble des parties de  $E$  et se note  $\mathcal{P}(E)$ .

**Union, intersection** Si  $(A_i)_{i \in I}$  est une collection d'ensembles inclus dans  $E$ , la réunion des  $A_i$  est l'ensemble  $\bigcup_{i \in I} A_i$  des  $a \in E$  pour lesquels *il existe*  $i \in I$  tel que  $a \in A_i$ . De même l'intersection des  $A_i$  est l'ensemble  $\bigcap_{i \in I} A_i$  des  $a \in E$  pour lesquels  $a \in A_i$  *pour tout*  $i \in I$ . On dit que deux ensembles sont *disjoints* si leur intersection est vide. On dit que les ensembles  $A_i, i \in I$  constituent une *partition* de l'ensemble  $E$  si i) ils sont non vides, ii) leur union sur  $i \in I$  vaut  $E$  iii) ils sont disjoints deux à deux ( $A_i \cap A_j = \emptyset$  si  $i \neq j$ ); on dit aussi que  $E$  est union disjointe des  $A_i, i \in I$ .

On a les formules

$$\left( \bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c, \quad \left( \bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c.$$

et

$$B \cap \left( \bigcup_{i \in I} A_i \right) = \bigcup_{i \in I} (B \cap A_i).$$

**Produits d'ensembles** Si  $A_1, \dots, A_n$  sont des ensembles on peut définir le produit cartésien de ces ensembles comme étant l'ensemble des  $n$ -uplets  $(a_1, \dots, a_n)$  où  $a_1 \in A_1, \dots, a_n \in A_n$ . On note cet ensemble  $A_1 \times \dots \times A_n$ .

Quand les  $A_i$  sont finis son cardinal est le produit des cardinaux des  $A_i$ .

## 1.2 Applications entre ensembles

**Injectivité, surjectivité, bijectivité** Si  $A$  et  $B$  sont deux ensembles, une application associe à tout élément  $a$  de  $A$  un unique élément noté  $f(a)$  de  $B$ . On dit que  $f(a)$  est l'image de  $a$  par  $f$ . Un élément de  $B$  peut n'être l'image d'aucun élément de  $A$  ou au contraire être l'image de plusieurs éléments de  $A$ . On dit qu'une application est *injective* si tout élément de  $B$  est l'image d'au plus un élément de  $A$ , *surjective* si tout élément de  $B$  est l'image d'au moins un élément de  $A$  et *bijective* si elle est injective et surjective.

S'il existe une injection de  $A$  dans  $B$  on a  $\#A \leq \#B$ .

S'il existe une surjection de  $A$  sur  $B$  on a  $\#A \geq \#B$ .

S'il existe une bijection entre  $A$  et  $B$  on a  $\#A = \#B$ .

**Ensemble des applications de  $A$  dans  $B$**  On note  $B^A$  l'ensemble des applications de  $A$  dans  $B$ . Quand  $A$  et  $B$  son finis son cardinal vaut  $(\#B)^{\#A}$ .



**Fonctions indicatrices, codages** Si  $E$  est un ensemble fixé, l'ensemble des parties de  $E$  est en bijection avec l'ensemble des applications de  $E$  dans l'ensemble à deux éléments  $\{0, 1\}$ . Cette bijection est la suivante : à tout ensemble  $A \subset E$  on associe sa *fonction caractéristique* ou *fonction indicatrice*  $\mathbf{1}_A : E \rightarrow \{0, 1\}$  définie par  $\mathbf{1}_A(e) = 1$  si  $e \in A$  et  $\mathbf{1}_A(e) = 0$  sinon. Réciproquement si  $f$  est une application de  $E$  dans  $\{0, 1\}$  l'ensemble  $A$  des  $e \in E$  tels que  $f(e) = 1$  est tel que  $\mathbf{1}_A(\cdot) = f(\cdot)$ . En particulier, cela démontre que quand  $E$  est fini le cardinal de  $\mathcal{P}(E)$  est  $2^{\#E}$

$$\#\mathcal{P}(E) = 2^{\#E}.$$

**Cardinal et fonctions caractéristiques** Si  $A \subset E$  on a

$$\#A = \sum_{x \in E} \mathbf{1}_A(x).$$

Si  $A_1, \dots, A_n$  sont des sous-ensembles de  $E$  on a

$$\mathbf{1}_{A_1 \cap \dots \cap A_n} = \prod_{i=1}^n \mathbf{1}_{A_i}.$$

**Pré-image** Si  $f$  est une application de  $E$  dans  $F$  on définit pour tout  $B \subset F$  l'ensemble  $f^{-1}(B)$  comme étant l'ensemble des  $e \in E$  tels que  $f(e) \in B$ . (Cette définition a un sens même si  $f$  n'est pas inversible.) On dit que  $f^{-1}(B)$  est la pré-image de  $B$  par  $f$ .

On a toujours

$$f^{-1}\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f^{-1}(A_i), \quad f^{-1}\left(\bigcap_{i \in I} A_i\right) = \bigcap_{i \in I} f^{-1}(A_i), \quad f^{-1}(A^c) = \left(f^{-1}(A)\right)^c.$$

Attention le comportement par image directe n'est pas aussi bon.

**Exercice 1.2.1** i) Montrer que si  $A, B$  sont deux sous-ensembles de  $E$  on a

$$1 - \mathbf{1}_{A \cup B} = (1 - \mathbf{1}_A)(1 - \mathbf{1}_B),$$

et en déduire que

$$\#(A \cup B) = \#A + \#B - \#(A \cap B).$$

ii) En généralisant la formule précédente montrer que

$$\#(A_1 \cup \dots \cup A_n) = \sum_{p=1}^n (-1)^{p-1} \sum_{1 \leq i_1 < \dots < i_p \leq n} \#(A_{i_1} \cap \dots \cap A_{i_p}).$$

*Solution.* i) Pour tout ensemble  $F \subset E$

$$\mathbf{1}_{F^c} = 1 - \mathbf{1}_F.$$

Donc

$$\begin{aligned} 1 - \mathbf{1}_{A \cup B} &= \mathbf{1}_{A^c \cap B^c} \\ &= \mathbf{1}_{A^c} \mathbf{1}_{B^c} \\ &= (1 - \mathbf{1}_A)(1 - \mathbf{1}_B). \end{aligned}$$

On a donc

$$\begin{aligned} \mathbf{1}_{A \cup B} &= \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \cdot \mathbf{1}_B \\ &= \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A \cap B}. \end{aligned}$$

Or, pour tout ensemble  $F \subset E$

$$\#F = \sum_{e \in E} \mathbf{1}_F(e).$$

On a donc bien la conclusion.

ii) De façon générale,

$$1 - \mathbf{1}_{A_1 \cup \dots \cup A_n} = \prod_{i=1}^n (1 - \mathbf{1}_{A_i}),$$

et donc

$$\begin{aligned} 1 - \mathbf{1}_{A_1 \cup \dots \cup A_n} &= 1 + \sum_{p=1}^n (-1)^p \sum_{1 \leq i_1 < \dots < i_p \leq n} \mathbf{1}_{A_{i_1}} \cdots \mathbf{1}_{A_{i_p}} \\ &= 1 + \sum_{p=1}^n (-1)^p \sum_{1 \leq i_1 < \dots < i_p \leq n} \mathbf{1}_{A_{i_1} \cap \dots \cap A_{i_p}}, \end{aligned}$$

et en sommant sur  $e \in E$  on obtient bien la formule annoncée.

### 1.3 Dénombrement

**Cardinal d'une union disjointe finie.** Si  $A_1, \dots, A_n$  sont des ensembles finis *disjoints deux à deux* tels que  $A_1 \cup \dots \cup A_n = E$  alors  $E$  est fini et

$$\#E = \sum_{i=1}^n \#A_i.$$

**Cardinal d'un produit.** Si  $A_1, \dots, A_n$  sont des ensembles finis le cardinal du produit  $A_1 \times \dots \times A_n$  est donné par

$$\#(A_1 \times \dots \times A_n) = (\#A_1) \cdots (\#A_n).$$

**Cardinal de l'ensemble des applications de  $A$  dans  $B$ .** Si  $A$  et  $B$  sont des ensembles finis, l'ensemble des applications de  $A$  dans  $B$  est fini et à pour cardinal

$$\#(B^A) = (\#B)^{\#A}.$$

**Nombre d'injections entre deux ensembles finis. Nombre d'arrangements** Si  $A$  et  $B$  sont deux ensembles finis avec  $\#A = p$ ,  $\#B = n$ , l'ensemble des applications injectives de  $A$  vers  $B$  a un cardinal égal à

$$\begin{cases} 0 & \text{si } \#A > \#B \\ n(n-1) \cdots (n-p+1) & \text{si } p \leq n. \end{cases}$$

En effet, supposons  $A = \{a_1, \dots, a_p\}$ ; si  $p > n$ , il ne peut y avoir d'applications injective de  $A$  vers  $B$ , tandis que si  $p \leq n$ , il y a  $n$  choix possibles pour la valeur  $f(a_1)$ ,  $n-1$  choix possibles pour la valeur de  $f(a_2)$  (comme  $f$  est injective  $f(a_2)$  ne peut pas prendre la même valeur que  $f(a_1)$ ) etc.  $n - (p-1) = n - p + 1$  choix possibles pour  $f(a_p)$

C'est aussi le nombre de  $p$ -uplet (ordonnés), on dit aussi d'**arrangements**,  $(e_1, \dots, e_p)$  où  $e_i \in E$ .

**Nombre de bijections de  $A$  vers  $A$ . Nombre de permutations, factorielles** Si  $A$  est un ensemble de cardinal  $n$ , une application de  $A$  vers  $A$  est bijective si et seulement si elle est injective et par conséquent le nombre de bijection de  $A$  vers  $A$  (on dit aussi le nombre de **permutations** de  $A$ ) égale

$$n! = n(n-1) \cdots 2 \cdot 1.$$

Par convention  $0! = 1$ .

**Cardinal de  $\mathcal{P}(E)$ .** Si  $E$  est fini de cardinal  $n$ , le nombre de sous-ensembles de  $E$  est égal au nombre d'applications de  $E$  vers  $\{0, 1\}$  et vaut donc

$$\#\mathcal{P}(E) = 2^n.$$

**Nombre de sous-ensembles de cardinal  $p$  d'un ensemble à  $n$  éléments.**

**Nombre de combinaisons** Si  $E$  est un ensemble fini de cardinal  $n$ , le nombre de sous-ensembles de  $E$  de cardinal exactement  $p$  (on parle aussi de **combinaison** égale

$$\binom{n}{p} = C_n^p = \frac{n(n-1)\cdots(n-p+1)}{p!} = \frac{n!}{p!(n-p)!}.$$

On pose  $\binom{n}{k} = 0$  quand  $k > n$  ou quand  $n < 0$  et  $\binom{n}{0} = 1$ . Remarquer que  $\binom{n}{k} = \binom{n}{n-k}$ .

En effet, un sous-ensemble  $\{a_1, \dots, a_p\}$  de  $E$  peut être vu comme un  $p$ -uplet d'élément de  $E$  où l'on oublie l'ordre des éléments. Or, étant donné  $p$  éléments de  $E$  on peut former  $p!$  (nombre de bijections de  $\{a_1, \dots, a_p\}$  dans lui-même)  $p$ -uplets. Ainsi, le nombre de sous-ensembles de cardinal  $p$  d'un ensemble à  $n$  éléments égale le nombre d'injection de l'ensemble  $\{1, \dots, p\}$  dans  $E$  (i.e le nombre de  $p$ -uplets de  $E$ ) divisé par  $p!$ .

**Binôme de Newton** Considérons l'expression  $(a+b)^n = (a+b)\cdots(a+b)$  (où  $a$  et  $b$  sont des nombres réels ou complexes ou des éléments d'un anneau commutatif). Quand on développe le produit, on obtient une somme de produits de  $a$  et de  $b$  et on voit que le coefficient de  $a^p b^{n-p}$  est égal au nombre de façons de choisir  $p$  éléments (les  $a$ ) parmi  $n$ . On a donc

$$(a+b)^n = \sum_{p=0}^n \binom{n}{p} a^p b^{n-p}.$$

**Triangle de Pascal** En utilisant la formule  $\binom{n}{k} = \frac{n!}{k!(n-k)!}$  on obtient pour tous entiers  $0 \leq k \leq n$  la relation

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}.$$

(Exerc. Le vérifier).

**Obtenir de nouvelles formules sur les coefficients binomiaux** Voici quelques méthodes utiles pour obtenir de nouvelles formules sur les coefficients binomiaux. On considère le polynôme  $(1+X)^n$ . D'après la formule du binôme

$$(1+X)^n = \sum_{k=0}^n \binom{n}{k} X^k. \quad (1.1)$$

En faisant  $X = 1$  dans cette expression on voit que

$$2^n = \sum_{k=0}^n \binom{n}{k}.$$

(**Exo.** Comment démontrer cette formule de façon purement combinatoire?).

De même, si on dérive chacun des membres de la formule (1.1) on obtient

$$n(1+X)^{n-1} = \sum_{k=0}^n \binom{n}{k} k X^{k-1},$$

c'est-à-dire en faisant  $X = 1$

$$n2^{n-1} = \sum_{k=0}^n \binom{n}{k} k.$$

En dérivant plusieurs fois la formule (1.1) on obtiendrait de la même manière d'autres expressions impliquant les coefficients binomiaux.

On peut aussi écrire  $(1+X)^{n+m} = (1+X)^n(1+X)^m$  et utiliser la formule du binôme :

$$\sum_{r=0}^{n+m} \binom{n+m}{r} X^r = \sum_{k=0}^n \binom{n}{k} X^k \sum_{l=0}^m \binom{m}{l} X^l. \quad (1.2)$$

Chacun des membres de l'équation précédente est un polynôme de degré  $n+m$ . Le coefficient du monôme de degré  $r$  de ce polynôme est  $\binom{n+m}{r}$  (quand on regarde le membre de gauche) et également  $\sum_{k+l=r} \binom{n}{k} \binom{m}{l}$ . On a donc

$$\binom{n+m}{r} = \sum_{k=0}^r \binom{n}{k} \binom{m}{r-k}.$$

**Exercice 1.3.1** Une urne contient  $N$  boules noires et  $M$  boules blanches.

i) On effectue  $n$  tirages sans remise. Quel est le nombre total de tels tirages ? Combien de tirages donnent  $x$  ( $x \leq n$ ) boules noires ?

ii) ) On effectue  $n$  tirages avec remise. Quel est le nombre total de tels tirages ? Combien de tirages donnent  $x$  ( $x \leq n$ ) boules noires ?

*Solution.*

On note  $\{1, \dots, N\}$  l'ensemble des boules noires et  $\{N+1, \dots, N+M\}$  l'ensemble des boules blanches.

i) Un tirage sans remise est équivalent à la donnée d'une injection de  $\{1, \dots, n\}$  dans  $\{1, \dots, N + M\}$  (ou à une suite ordonnée, un  $n$ -uplet  $(x_1, \dots, x_n)$ ,  $x_i \in \{1, \dots, N + M\}$ ). Il y a donc  $(N + M) \cdots (N + M - n + 1)$  tirages sans remise.

Un tirage où  $x$  boules noires sont tirées est équivalent à la donnée d'un sous-ensemble  $A$  de  $\{1, \dots, n\}$  à  $x$  éléments (si on pense au tirage comme à une expérience,  $A$  est l'ensemble des temps où le résultat de notre expérience est "boule noire") et de deux injections, une de  $A$  dans l'ensemble des boules noires, une seconde du complémentaire de  $A$  dans  $\{1, \dots, n\}$  dans l'ensemble des boules blanches : on a donc

$$\binom{n}{x} \cdot N(N-1) \cdots (N-x+1) \cdot M(M-1) \cdots (M-(n-x)+1)$$

choix possibles, c'est-à-dire

$$\binom{n}{x} \cdot N(N-1) \cdots (N-x+1) \cdot M(M-1) \cdots (M-n+x+1)$$

choix possibles.

Remarquons que la proportion du nombre de tirages sans remise où  $x$  boules noires sortent dans l'ensemble des tirages sans remise est

$$\begin{aligned} & \frac{\binom{n}{x} \cdot N(N-1) \cdots (N-x+1) \cdot M(M-1) \cdots (M-n+x+1)}{(N+M) \cdots (N+M-n+1)} \\ &= \frac{\binom{n}{x} \cdot \binom{N}{x} x! \cdot \binom{M}{n-x} (n-x)!}{\binom{N+M}{n} n!} \\ &= \frac{\binom{N}{x} \binom{M}{n-x}}{\binom{N+M}{n}} \end{aligned}$$

ii) Un tirage avec remise est équivalent à la donnée d'une application (pas nécessairement injective) de  $\{1, \dots, n\}$  vers  $\{1, \dots, N + M\}$  (ou encore d'un  $n$ -uplet  $(e_1, \dots, e_n)$  de  $\{1, \dots, N + M\}^n$ ) ; il y a donc  $(N + M)^n$  choix possibles.

Un tirage où  $x$  boules noires sont tirées est équivalent à la donnée : d'un sous-ensemble  $A$  de  $\{1, \dots, n\}$  à  $x$  éléments, d'une application (pas nécessairement injective) de  $A$  dans  $\{1, \dots, N\}$  (ou encore d'un  $x$ -uplet de  $\{1, \dots, N\}^x$ ) et d'une application de  $\{1, \dots, n\} - A$  dans  $\{N + 1, \dots, N + M\}$  (ou encore un  $(n - x)$ -uplet de  $\{N + 1, \dots, N + M\}$ ). Il y a donc

$$\binom{n}{x} \cdot N^x \cdot M^{n-x}$$

choix possibles. Remarquons que la proportion du nombre de tirages avec remise où  $x$  boules noires sortent dans l'ensemble des tirages avec remise est

$$\binom{n}{x} \frac{N^x M^{n-x}}{(N+M)^n} = \binom{n}{x} p^x (1-p)^{n-x},$$

où  $p = N/(N+M)$ .

## 1.4 Dénombrabilité

**Définition 1.4.1** *Un ensemble est dit dénombrable s'il est en bijection avec un sous-ensemble de l'ensemble  $\mathbb{N}$  des entiers naturels.*

Nous étendrons cette définition en disant qu'un ensemble est dénombrable s'il est fini ou en bijection avec  $\mathbb{N}$ .

De façon plus concrète, un ensemble est dénombrable si on peut énumérer ses éléments.

**Proposition 1.4.1** *Si  $A$  et  $B$  sont deux ensembles.*

- a) *S'il existe une injection de  $A$  dans  $B$  et si  $B$  est dénombrable alors  $A$  est dénombrable*
- b) *S'il existe une surjection de  $A$  dans  $B$  et si  $A$  est dénombrable, alors  $B$  est dénombrable.*

**Théorème 1.4.1** a) *Si  $A_1, \dots, A_n$  sont des ensembles dénombrables, le produit  $A_1 \times \dots \times A_n$  est également dénombrable.*

b) *Si  $(A_i)_{i \in I}$  est une famille dénombrable (c'est-à-dire  $I$  est dénombrable) d'ensembles dénombrables (pour tout  $i \in I$ ,  $A_i$  est dénombrable) alors la réunion  $\bigcup_{i \in I} A_i$  est également dénombrable.*

*Démonstration.* —

a) On peut supposer  $A_1 = \dots = A_n = \mathbb{N}$ . Notons  $p_1, \dots, p_n$  les  $n$  premiers nombres premiers ( $p$  est premier s'il est divisible uniquement par 1 et par  $p$ ) et considérons l'application qui à  $(l_1, \dots, l_n) \in \mathbb{N}^n$  associe le nombre  $2^{l_1} \cdot 3^{l_2} \cdot \dots \cdot p_n^{l_n}$  est une injection de  $\mathbb{N}^n$  dans  $\mathbb{N}$  car la décomposition en facteurs premiers d'un nombre est unique. La proposition 1.4.1 a) permet de conclure.

b) Considérons l'application de  $\mathbb{N} \times \mathbb{N}$  dans  $\bigcup_{i \in I} A_i$  qui au couple  $(n, m)$  associe le  $m$ -ième élément de l'ensemble  $A_{i_n}$  où  $i_n$  est le  $n$ -ième élément de  $I$ . C'est une surjection. La proposition 1.4.1 b) donne la conclusion.  $\square$

**Corollaire 1.4.1** *L'ensemble des entiers relatifs  $\mathbb{Z}$  et l'ensemble des nombres rationnels  $\mathbb{Q}$  sont dénombrables.*

*Démonstration.* —

L'ensemble  $\mathbb{Z}$  est dénombrable car l'application de l'ensemble dénombrable  $\{1, -1\} \times \mathbb{N}$  dans  $\mathbb{Z}$  qui au couple  $(\epsilon, n)$  associe le produit  $\epsilon n$  est une surjection. De même,  $\mathbb{Q}$  est dénombrable car l'application de l'ensemble dénombrable  $\mathbb{Z} \times (\mathbb{N} - \{0\})$  dans  $\mathbb{Q}$  qui au couple  $(p, q)$  associe le rationnel  $p/q$  est une surjection.  $\square$

On peut démontrer que

**Théorème 1.4.2** *L'ensemble des nombres réels  $\mathbb{R}$  n'est pas dénombrable.*

**Corollaire 1.4.2** *L'ensemble des nombres irrationnels n'est pas dénombrable.*

*Démonstration.* —

Car sinon,  $\mathbb{R}$  qui est réunion de  $\mathbb{Q}$  et de l'ensemble des nombres irrationnels serait dénombrable (comme union dénombrable d'ensembles dénombrables).  $\square$



# Chapitre 2

## Espaces Probabilisés

### 2.1 Espaces probabilisés

Un espace probabilisé est la donnée

- d'un espace  $\Omega$  que l'on appelle *l'espace des états*. Quand on modélise une situation concrète  $\Omega$  est l'ensemble des états du système que l'on considère. Bien souvent cet espace est inaccessible à l'expérience ;
- d'un sous-ensemble  $\mathcal{B}$  de  $\mathcal{P}(\Omega)$  qui est *l'ensemble des événements*. Dans une situation concrète c'est l'ensemble de tous les résultats d'expériences que l'on peut effectuer sur le système. En théorie des probabilités (donc quand on fait des mathématiques) cet ensemble  $\mathcal{B}$  sera une *tribu* (on dit aussi une  $\sigma$ -algèbre), cf. définition 2.2.1 ;
- d'une *probabilité*  $\mathbf{P}$  : pour tout événement  $A \in \mathcal{B}$  le réel  $\mathbf{P}(A)$  est le degré de vraisemblance de l'événement  $A$  ; c'est un nombre réel compris entre 0 et 1. Mathématiquement, une probabilité est une application  $\mathbf{P} : \mathcal{B} \rightarrow [0, 1]$  vérifiant les propriétés décrites dans la définition 2.3.1.

Nous précisons dans la suite les deux derniers points.

### 2.2 Tribus

Soit  $\Omega$  un ensemble fixé (l'espace des états).

**Définition 2.2.1** Une tribu ou encore une  $\sigma$ -algèbre de  $\Omega$  est un ensemble de parties de  $\Omega$  (donc un sous-ensemble de  $\mathcal{P}(\Omega)$ , l'ensemble des parties de  $\Omega$ ) qui contient l'ensemble vide, est stable par passage au complémentaire et est stable par union dénombrable :

- $\emptyset \in \mathcal{B}$
- pour tout  $A \in \mathcal{B}$  on a  $A^c \in \mathcal{B}$

– pour toute famille (dénombrable)  $(A_i)_{i \in \mathbb{N}}$  d'éléments de  $\mathcal{B}$  l'union

$$\bigcup_{i \in \mathbb{N}} A_i$$

est également dans  $\mathcal{B}$ .

Il est clair que  $\Omega$  est toujours élément de la tribu (c'est le complémentaire de l'ensemble vide) et qu'une intersection dénombrable d'éléments de la tribu est encore dans la tribu (car  $\bigcap_{i \in \mathbb{N}} A_i = (\bigcup_{i \in \mathbb{N}} A_i^c)^c$ ).

### Exemples

1) Si  $\Omega$  est un ensemble quelconque on peut toujours définir deux tribus :

la tribu triviale qui est  $\mathcal{B} = \{\emptyset, \Omega\}$

la tribu totale qui est  $\mathcal{B} = \mathcal{P}(\Omega)$ .

2) Si  $\Omega = \{1, 2, 3\}$  le sous-ensemble de  $\mathcal{P}(\Omega)$ ,  $\mathcal{B} = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$  est une tribu de  $\Omega$ .

3) (**Exercice :**) Si  $\Omega$  est un ensemble le sous-ensemble de  $\mathcal{P}(\Omega)$  constitué des ensembles qui sont dénombrables ou dont le complémentaire est dénombrable est une tribu.

## 2.3 Probabilités

### 2.3.1 Définition

**Définition 2.3.1** Si  $\Omega$  est un ensemble et  $\mathcal{B}$  est une tribu de  $\Omega$ , une probabilité  $\mathbf{P}$  est une application de  $\mathcal{B}$  dans  $[0, 1]$  telle  $\mathbf{P}(\Omega) = 1$  et telle que pour toute famille dénombrable  $(A_i)_{i \in \mathbb{N}}$  d'événements de  $\mathcal{B}$  **disjoints 2 à 2** on a

$$\mathbf{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i=0}^{\infty} \mathbf{P}(A_i). \quad (2.1)$$

où l'égalité précédente signifie la chose suivante : la probabilité  $\mathbf{P}(\bigcup_{i \in \mathbb{N}} A_i)$  est égale à la limite de la suite croissante de nombres réels  $\sum_{i=0}^N \mathbf{P}(A_i)$  quand  $N$  tend vers l'infini. (Cette limite existe toujours car la suite en question est croissante et bornée.)

**Remarque** L'intérêt d'autoriser la stabilité par unions (intersections) dénombrables dans la définition d'une tribu permet de construire à partir d'événements simples des événements beaucoup plus intéressants que ceux qu'on obtiendrait en ne supposant que la stabilité par unions (intersections) finies. En revanche, si on autorisait la stabilité par unions (intersections) quelconques on ne pourrait pas construire beaucoup de probabilités. La stabilité par unions (intersections) dénombrable est donc le bon compromis.

### 2.3.2 Exemples généraux

#### Mesures de Dirac

Sur tout ensemble  $\Omega$  muni d'une tribu  $\mathcal{B}$  il est possible de construire des mesures de la façon suivante : pour tout  $\alpha \in \Omega$  définissons l'application  $\delta_\alpha : \mathcal{B} \rightarrow [0, 1]$  qui à un ensemble  $A \in \mathcal{B}$  associe le réel 1 si  $\alpha \in A$  et 0 sinon. Cette application  $\delta_\alpha$  est une mesure de probabilité que l'on appelle la mesure de Dirac au point  $\alpha$ . Vérifions rapidement que c'est bien une mesure : déjà  $\delta_\alpha(\Omega) = 1$  puisque  $\alpha \in \Omega$ ; par ailleurs si  $A_i \in \mathcal{B}$ ,  $i \geq 0$  est une famille dénombrable d'ensembles de la tribu disjoints deux à deux on a

$$\delta_\alpha\left(\bigcup_{i \geq 0} A_i\right) = \delta_\alpha(A_i),$$

car :

- soit  $\alpha$  appartient à  $\cup_{i \geq 0} A_i$ ; mais alors il existe un  $i \geq 0$  pour lequel  $\alpha \in A_i$  et cet indice  $i$  est unique car les  $A_i$  sont disjoints deux à deux. L'égalité précédente se réduit à  $1 = 1$ ;
- soit  $\alpha$  n'appartient pas à  $\cup_{i \geq 0} A_i$  et de ce fait n'appartient à aucun des  $A_i$  : l'égalité se réduit à  $0 = 0$ .

#### Sommes pondérées de probabilités

Si  $\mathbb{P}_1, \dots, \mathbb{P}_n$  sont des probabilités sur  $\mathcal{B}$  ( $\mathcal{B}$  étant une tribu) et si  $\lambda_1, \dots, \lambda_n$  sont des réels positifs tels que  $\lambda_1 + \dots + \lambda_n = 1$  alors  $\mathbb{P} := \lambda_1 \mathbb{P}_1 + \dots + \lambda_n \mathbb{P}_n$  est aussi une probabilité.

### 2.3.3 Premières propriétés

Mentionnons tout d'abord deux propriétés immédiates des probabilités :

**Proposition 2.3.1** *Soit  $(\Omega, \mathcal{B}, \mathbb{P})$  un espace probabilisé.*

a) Si  $A \in \mathcal{B}$ ,

$$\mathbf{P}(A^c) = 1 - \mathbf{P}(A).$$

b)  $\mathbf{P}(\emptyset) = 0$

c) (Positivité) Si  $A, B \in \mathcal{B}$  vérifient  $A \subset B$  alors  $\mathbf{P}(A) \leq \mathbf{P}(B)$ .

d) Si  $A, B \in \mathcal{B}$  alors

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B).$$

*Démonstration.* —

a) Il suffit décrire  $\Omega$  comme l'union disjointe finie  $\Omega = A \cup A^c$  : comme  $\mathbb{P}$  est une probabilité  $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$ .

b) suit de la formule précédente et du fait que  $\mathbb{P}(\Omega) = 1$ .

c) On écrit  $B$  comme l'union disjointe  $B = A \cup (B \cap A^c)$  et  $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$ . Comme  $\mathbb{P}(B \cap A^c) \geq 0$  on a bien  $\mathbb{P}(B) \geq \mathbb{P}(A)$ .

d) De l'union disjointe  $A \cup B = A \cup (B \cap A^c)$  on déduit  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$ . Mais de l'union disjointe  $B = (B \cap A^c) \cup (B \cap A)$  on obtient  $\mathbb{P}(B) = \mathbb{P}(B \cap A^c) + \mathbb{P}(B \cap A)$ . De ces deux égalités on déduit la formule d)  $\square$

La preuve des propriétés qui suivent n'est pas difficile mais, à la différence de la démonstration des propriétés précédentes, ne pourrait se faire sans autoriser des unions *dénombrables* :

**Proposition 2.3.2** a) Si  $A_i, i \in \mathbb{N}$  est une famille croissante d'éléments de  $\mathcal{B}$  dont l'union est  $A$  alors  $A \in \mathcal{B}$  et la suite  $\mathbf{P}(A_n)$  (qui est croissante bornée) converge vers  $\mathbf{P}(A)$  :

$$\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \mathbf{P}(A);$$

b) Si  $A_i, i \in \mathbb{N}$  est une famille décroissante d'éléments de  $\mathcal{B}$  dont l'union est  $A$  alors  $A \in \mathcal{B}$  et la suite  $\mathbf{P}(A_n)$  (qui est décroissante positive) converge vers  $\mathbf{P}(A)$  :

$$\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \mathbf{P}(A);$$

c) Si  $A_i, i \in \mathbb{N}$  est une famille dénombrable d'ensembles appartenant à  $\mathcal{B}$  on a toujours (même si les  $A_i$  ne sont pas disjoints deux à deux)

$$\mathbf{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) \leq \sum_{i=0}^{\infty} \mathbf{P}(A_i),$$

(où le membre de droite de l'inégalité précédente qui est la limite de la suite croissante peut éventuellement être infini).

*Démonstration.* —

a) Définissons les ensembles  $B_n$ ,  $n \geq 0$  de la façon suivante :  $B_0 = A_0$ , et pour  $n \geq 1$ ,  $B_n = A_n \cap A_{n-1}^c$ . Les  $B_n$  constituent une famille dénombrable d'ensembles disjoints deux à deux d'éléments de  $\mathcal{B}$  et on peut donc écrire

$$\sum_{k=0}^{\infty} \mathbb{P}(B_k) = \mathbb{P}\left(\bigcup_{k \in \mathbb{N}} B_k\right).$$

c'est-à-dire

$$\lim_{N \rightarrow \infty} \sum_{k=0}^N \mathbb{P}(B_k) = \mathbb{P}\left(\bigcup_{k \in \mathbb{N}} B_k\right),$$

ou encore, puisque les  $B_k$  sont disjoints deux à deux

$$\lim_{N \rightarrow \infty} \mathbf{P}\left(\bigcup_{k=0}^N B_k\right) = \mathbb{P}\left(\bigcup_{k \in \mathbb{N}} B_k\right),$$

Mais

$$\bigcup_{k=0}^N B_k = A_N, \quad \bigcup_{k=0}^{\infty} B_k = A$$

ce qui établit la preuve de a).

b) Il suffit de passer au complémentaire et d'utiliser a)

c) Pour  $\omega \in \Omega$  définissons l'entier  $\nu(\omega)$  comme étant le plus petit entier  $k \geq 0$  pour lequel  $\omega \in A_k$ . L'ensemble  $C_n$  des  $\omega \in \Omega$  pour lesquels  $\nu(\omega) = n$  est l'ensemble

$$C_n = \{\omega \in \Omega, \nu(\omega) = n\} = A_n \cap (A_{n-1} \cup \dots \cup A_0)^c$$

qui est clairement dans  $\mathcal{B}$ . Les ensembles  $C_n$  sont de toute évidence disjoints deux à deux et leur union pour  $n \geq 0$  est  $\bigcup_{n \in \mathbb{N}} A_n$  car pour tout  $\omega$  dans  $\bigcup_{n \in \mathbb{N}} A_n$  il existe un  $n$  tel que  $\nu(\omega) = n$  c'est-à-dire il existe un  $n$  tel que  $\omega \in C_n$ . On a donc

$$\mathbb{P}\left(\bigcup_{n \geq 0} A_n\right) = \mathbb{P}\left(\bigcup_{n \geq 0} C_n\right) = \sum_{n=0}^{\infty} \mathbb{P}(C_n),$$

et comme  $\mathbb{P}(C_n) \leq \mathbb{P}(A_n)$  (puisque  $C_n \subset A_n$ ) on obtient la conclusion du c).  $\square$

Les deux propriétés précédentes a) et b) sont des propriétés de *continuité* (dans un sens à préciser) des probabilités.

## 2.4 Probabilités sur un ensemble fini

Les espaces probabilisés les plus simples sont ceux où l'espace des états  $\Omega$  est fini. On choisit en général comme tribu  $\mathcal{B}$  l'ensemble  $\mathcal{P}(\Omega)$  de toutes les parties de  $\Omega$  (qui est bien une tribu). C'est ce que nous ferons (car le cas où  $\mathcal{B}$  est une tribu plus petite que  $\mathcal{P}(\Omega)$  s'y ramène). Cela étant, il reste à définir la probabilité. Supposons donc que  $\mathbb{P}$  soit une probabilité sur  $\mathcal{B} = \mathcal{P}(\Omega)$ ; tout ensemble  $A \in \mathcal{B} = \mathcal{P}(\Omega)$  est fini (car inclus dans  $\Omega$  qui est fini) et est par conséquent l'union (finie donc dénombrable) des singletons  $\{a\}$  où  $a$  décrit  $A$  :

$$A = \bigcup_{a \in A} \{a\}.$$

Comme cette union est disjointe et finie on a

$$\mathbb{P}(A) = \sum_{a \in A} \mathbb{P}(\{a\}).$$

Supposons que  $\Omega = \{\omega_1, \dots, \omega_n\}$  et notons  $p_i = \mathbb{P}(\{\omega_i\})$ ; on a donc  $p_i \in [0, 1]$ . D'autre part on a

$$\mathbb{P}(A) = \sum_{\{i: \omega_i \in A\}} p_i.$$

Si on choisit  $A = \Omega$  on voit que les  $p_i$ , qui sont dans  $[0, 1]$ , vérifient

$$\sum_{i=1}^n p_i = 1.$$

En conclusion :

**Proposition 2.4.1** *Dans le cas où  $\Omega = \{\omega_1, \dots, \omega_n\}$  est fini, une probabilité  $\mathbb{P}$  sur  $\mathcal{B} = \mathcal{P}(\Omega)$  est déterminée par ses valeurs sur les singletons de  $\Omega$ . Réciproquement si on se donne  $n$  nombres réels positifs  $p_1, \dots, p_n$  dont la somme vaut 1 ( $p_1 + \dots + p_n = 1$ ) alors, l'application  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  qui à  $A \in \mathcal{P}(\Omega)$  associe le réel (dans  $[0, 1]$ )*

$$\mathbb{P}(A) = \sum_{\{i: \omega_i \in A\}} p_i$$

*est une probabilité.*

*Démonstration.* — Nous avons démontré la première partie de la proposition, démontrons la réciproque. Il suffit de vérifier que pour toute famille  $A_k$ ,  $k \in \mathbb{N}$ ,  $A_k \subset \Omega$  où les  $A_k$  sont deux à deux disjoints, la formule (2.1) est

vérifiée. Comme  $\Omega$  est fini, il en est de même de  $\mathcal{B} = \mathcal{P}(\Omega)$  et il suffit donc de considérer le cas où la famille  $(A_k)_{k \in \mathbb{N}}$  est finie et est  $A_0, \dots, A_N$ ,  $A_k \neq A_l$  si  $k \neq l$ . Si on note  $I_k$  l'ensemble des indices  $i \in \{1, \dots, n\}$  pour lesquels  $\omega_i \in A_k$ , les ensembles  $I_0, \dots, I_N$  sont deux à deux disjoints et leur union  $I$  est l'ensemble des indices  $i \in \{1, \dots, n\}$  pour lesquels  $\omega_i \in \bigcup_{k=0}^N A_k$ . On a donc

$$\mathbb{P}\left(\bigcup_{k=0}^N A_k\right) = \sum_{i \in I} p_i = \sum_{k=0}^N \sum_{i \in I_k} p_i = \sum_{k=0}^N \mathbb{P}(A_k).$$

□

### 2.4.1 Probabilités uniformes et lien avec la combinatoire

Un cas important est celui où tous les  $p_i$ ,  $1 \leq i \leq n$  précédents sont égaux. Comme leur somme doit valoir 1 ceci signifie que  $p_1 = \dots = p_n = \frac{1}{n}$ . On dit dans ce cas que la probabilité  $\mathbb{P}$  est *uniforme*. On a alors, pour tout sous-ensemble  $A$  de  $\Omega$

$$\mathbb{P}(A) = \sum_{\{i: \omega_i \in A\}} p_i = \#\{i \in \{1, \dots, n\}, \omega_i \in A\} \cdot \frac{1}{n}$$

soit

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

Ainsi, quand on travaille avec une probabilité uniforme sur un ensemble fini, déterminer la probabilité d'un événement revient à calculer son cardinal : on voit apparaître le lien avec la combinatoire.

#### Lancer de deux dés

On se propose de modéliser le lancer de deux dés (numérotés de 1 à 6). On appellera  $p_i$  la probabilité pour chaque dé, quand on le lance, d'obtenir la face  $i$  et on suppose de façon implicite que le lancer de chacun des deux dés est indépendant de l'autre (dans un sens qui pour l'instant n'est pas bien défini).

La modélisation du problème se fait de la façon suivante.

- *Espace des états* : on pose  $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$  l'ensemble des couples  $(i, j)$  avec  $1 \leq i \leq 6$ ,  $1 \leq j \leq 6$ .
- *Espace des événements* : on posera  $\mathcal{B} = \mathcal{P}(\Omega)$ .

- *Probabilité* : Si on suppose que quand on lance un dé, la probabilité de sortie de la face  $i$  égale  $p_i$  il est naturel de penser que la probabilité  $p_{ij}$  de sortie du couple  $(i, j)$  ( $i$  pour le 1er dé et  $j$  pour le second) est égal au produit  $p_i p_j$  de la probabilité d’obtenir  $i$  sur le premier dé par la probabilité d’obtenir  $j$  sur le second. Naturellement, les  $p_i$  vérifient  $p_i \geq 0$  et  $\sum_{i=1}^6 p_i = 1$ . On décide donc de définir, ayant à l’esprit la Proposition 2.4.1, la probabilité  $\mathbb{P}$  sur  $\mathcal{B}$  par  $\mathbb{P}(\{(i, j)\}) = p_i p_j$ . Pour que cela soit possible il faut vérifier que les  $p_i p_j$  sont positifs, ce qui est évident, et que  $\sum_{(i,j) \in \Omega} p_i p_j = 1$ . Mais cela aussi est clair car

$$\sum_{1 \leq i, j \leq 6} p_i p_j = \left( \sum_{i=1}^6 p_i \right) \left( \sum_{j=1}^6 p_j \right) = 1 \times 1 = 1.$$

Avec cette modélisation, on peut ainsi déterminer la probabilité pour que la somme des deux dés après lancer vaille 7. Cet événement, appelons-le  $A$ , s’écrit de façon mathématique

$$A = \{(i, j) \in \Omega : i + j = 7\}$$

ou de façon explicite

$$A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

La probabilité de  $A$  vaut

$$\mathbb{P}(A) = \sum_{\{1 \leq i, j \leq 6 : i+j=7\}} p_i p_j.$$

Si on suppose que pour tout  $1 \leq i \leq 6$ ,  $p_i = 1/6$  on obtient  $\mathbb{P}(A) = \#A/\#\Omega = 6/36 = 1/6$ . De la même façon l’événement  $B$  : “après le lancer des deux dé la face du premier dé est un nombre pair”, est mathématiquement l’ensemble

$$B = \{(i, j) : i = 2, 4, 6\}.$$

Si on suppose que pour tout  $1 \leq i \leq 6$ ,  $p_i = 1/6$  on obtient  $\mathbb{P}(B) = \#B/\#\Omega = 18/36 = 1/2$ .

La conjonction de ces deux événements, “la somme des deux dés vaut 7” et “le résultat que donne le premier dé est un nombre pair” est mathématiquement l’ensemble  $A \cap B = \{(2, 5), (4, 3), (6, 1)\}$ . Sa probabilité (si on suppose que pour tout  $1 \leq i \leq 6$ ,  $p_i = 1/6$ ) égale  $\#(A \cap B)/\#\Omega = 3/36 = 1/12$ .

**Exercice 2.4.1** *On tire cinq cartes d’un jeu de 32 cartes. Quelle est la probabilité d’obtenir un full c’est-à-dire deux cartes de même valeur et trois autres cartes de même valeur. On supposera chaque tirage équiprobable.*



**Exercice 2.4.2** (cf. exercice 1.3.1) Une urne contient  $n$  boules noires et  $b$  boules blanches.

a) On effectue  $N$  tirages sans remises. Quelle est la probabilité d'obtenir  $x$  boules noires ?

b) Même question si les tirages sont avec remises.

On supposera les tirages équiprobables. On prendra soin de bien définir l'espace probabilisé sur lequel on travaille.

### 2.4.2 Jeu de $n$ Pile ou Face : le modèle (fini) de Bernoulli

On se propose de modéliser un jeu où l'on lance  $n$  fois une pièce (Pile/Face). De façon équivalente un expérimentateur réalise  $n$  expérience le résultat de chaque une d'entre elles pouvant être positif (1) ou négatif (0). Il est naturel de décrire le jeu ou l'expérience précédentes de la façon suivante : on choisit comme espace des états l'ensemble  $\Omega$  de toutes les suites de longueur  $n$  constituées de 0 ou de 1. Une telle suite est donc un  $n$ -uplet  $\omega = (\omega_1, \dots, \omega_n)$  chaque  $\omega_i$ ,  $1 \leq i \leq n$  appartenant à l'ensemble à deux éléments  $\{0, 1\}$ . Ainsi

$$\Omega = \{0, 1\}^n,$$

et a  $2^n$  éléments. Nous choisirons comme tribu  $\mathcal{B}_n = \mathcal{P}(\Omega)$ , l'ensemble des parties de  $\Omega$  (qui a donc  $2^{2^n}$  mais cela n'a pas d'importance). Cette tribu nous permet de décrire des événements. Par exemple l'événement (A) "obtenir  $k$  Pile lors des  $n$  lancers" est décrit par l'ensemble  $A \in \mathcal{B}_n$  (Pile=1, Face=0)

$$A = \{\omega = (\omega_1, \dots, \omega_n), \sum_{i=1}^n \omega_i = k.\}$$

L'événement (B) "on tire au moins un Pile" est décrit par l'ensemble

$$B = \{\omega = (\omega_1, \dots, \omega_n), \exists i \in \{1, \dots, n\} \quad \omega_i = 1\}.$$

L'événement "(A) et (B)" est décrit par l'intersection  $A \cap B$ , l'événement "non A" est décrit par  $A^c$ , l'événement "A ou B" par  $A \cup B$  etc.

Le choix de la probabilité sur notre ensemble est dicté par le jeu ou l'expérience que l'on modélise. Ainsi, on ne modélisera pas de la même façon un jeu ou pile et face ont les mêmes chances de sortir qu'un jeu où pile a deux fois plus de chance de sortir que face. Dans le premier cas, il est naturel<sup>1</sup> de choisir comme probabilité  $\mathbb{P}$  la probabilité uniforme

$$\mathbb{P}(A) = \frac{\#A}{\Omega} = \frac{\#A}{2^n}.$$

---

1. en fait cela sera encore plus naturel quand on aura défini la notion d'indépendance

Noter que la probabilité d'un événement élémentaire "on a tiré la suite  $(\epsilon_1, \dots, \epsilon_n)$ " c'est-à-dire la probabilité du singleton  $\{(\epsilon_1, \dots, \epsilon_n)\}$  vaut  $1/2^n$  (ceci quel que soit  $\epsilon_1, \dots, \epsilon_n$ ). En revanche, dans le second cas, on définira la probabilité d'un événement élémentaire  $\{\epsilon_1, \dots, \epsilon_n\}$  comme étant  $(2/3)^k(1/3)^{n-k}$  où  $k$  est le nombre de 1 dans la suite  $\epsilon_1, \dots, \epsilon_n$ .

Modélisons donc un jeu de  $n$  lancers *indépendants* de 1/0 ou à chaque lancer 1 à une probabilité  $p$  de sortir et 0 une probabilité  $1-p$  d'apparaître.

- *Espace des états* : on pose  $\Omega = \{0, 1\}^n$  l'ensemble des  $n$ -uplets  $(\omega_1, \dots, \omega_n)$   $\omega_i \in \{0, 1\}$ ,  $1 \leq i \leq n$ .
- *Espace des événements* : on posera  $\mathcal{B}_n = \mathcal{P}(\Omega)$ .
- *Probabilité* : Du fait de l'indépendance (notion que l'on appréhende pour l'instant de façon intuitive) il est naturel de penser que la probabilité  $p_\omega$  d'apparition de la suite  $\omega = (\epsilon_1, \dots, \epsilon_n)$ ,  $\epsilon_i \in \{0, 1\}$  est  $p^r(1-p)^{n-r}$  où  $r$  est le nombre de fois où 1 sort (et  $n-r$  est donc le nombre de 0). On remarquera que

$$r = \sum_{i=1}^n \epsilon_i.$$

Il est clair que  $p_\omega \geq 0$ ; vérifions que  $\sum_{\omega \in \Omega} p_\omega = 1$ . On a

$$\begin{aligned} \sum_{\omega \in \Omega} p_\omega &= \sum_{(\omega_1, \dots, \omega_n) \in \{0, 1\}^n} p^{\omega_1 + \dots + \omega_n} (1-p)^{n - (\omega_1 + \dots + \omega_n)} \\ &= \sum_{(\omega_1, \dots, \omega_{n-1}) \in \{0, 1\}^{n-1}} \sum_{\omega_n=0 \text{ ou } 1} p^{\omega_1 + \dots + \omega_{n-1} + \omega_n} (1-p)^{(n-1) - (\omega_1 + \dots + \omega_{n-1}) + (1 - \omega_n)} \\ &= \sum_{(\omega_1, \dots, \omega_{n-1}) \in \{0, 1\}^{n-1}} p^{\omega_1 + \dots + \omega_{n-1}} (1-p)^{(n-1) - (\omega_1 + \dots + \omega_{n-1})} (p + (1-p)) \\ &= \sum_{(\omega_1, \dots, \omega_{n-1}) \in \{0, 1\}^{n-1}} p^{\omega_1 + \dots + \omega_{n-1}} (1-p)^{(n-1) - (\omega_1 + \dots + \omega_{n-1})} \end{aligned}$$

et donc

$$\begin{aligned} \sum_{\omega \in \Omega} p_\omega &= \sum_{(\omega_1, \dots, \omega_{n-1}) \in \{0, 1\}^{n-1}} p^{\omega_1 + \dots + \omega_{n-1}} (1-p)^{(n-1) - (\omega_1 + \dots + \omega_{n-1})} \\ &= \sum_{(\omega_1, \dots, \omega_{n-2}) \in \{0, 1\}^{n-2}} p^{\omega_1 + \dots + \omega_{n-2}} (1-p)^{(n-2) - (\omega_1 + \dots + \omega_{n-2})} \\ &= \dots \\ &= 1 \end{aligned}$$

D'après la Proposition 2.4.1 on peut donc définir une probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{B}_n)$  telle que

$$\mathcal{P}(\{(\omega_1, \dots, \omega_n)\}) = p^{\omega_1 + \dots + \omega_n} (1-p)^{n - (\omega_1 + \dots + \omega_n)}.$$

## 2.5. LE JEU INFINI DE PILE OU FACE : LE MODÈLE DE BERNOULLI 27

On appellera l'espace probabilisé ainsi construit le modèle (fini) de Bernoulli de paramètre  $p$  ( $0 \leq p \leq 1$ ).

Mentionnons une propriété importante du modèle de Bernoulli.

**Proposition 2.4.2** *Dans le modèle de Bernoulli de paramètre  $p$ , l'événement*

$$A = \{(\omega_1, \dots, \omega_n) : \omega_1 + \dots + \omega_n = k\}$$

*admet pour probabilité*

$$\mathbb{P}(A) = \binom{n}{k} p^k (1-p)^{n-k}.$$

*Démonstration.* — On a en effet

$$\begin{aligned} \mathbb{P}(A) &= \sum_{(\omega_1, \dots, \omega_n) \in A} p^{\omega_1 + \dots + \omega_n} (1-p)^{n - (\omega_1 + \dots + \omega_n)} \\ &= p^k (1-p)^k \#A. \end{aligned}$$

Or, le cardinal de  $A$  égale le nombre de façon de choisir  $k$  éléments parmi  $n$ .  
□

**Exercice 2.4.3** *Calculer dans chacun des cas  $p = 1/2$  ou  $p = 2/3$  les probabilités des événements  $A$  et  $B$  définis précédemment.*

## 2.5 Le jeu infini de pile ou face : le modèle de Bernoulli

### 2.5.1 L'espace des états

Nous présentons dans ce paragraphe la modélisation du jeu de pile ou face où l'on joue une infinité de fois. Il est naturel d'introduire comme espace des états l'ensemble  $\Omega$  des suites  $\omega = (\omega_1, \omega_2, \dots)$  où les  $\omega_i$  valent 0 ou 1. On a ainsi  $\Omega = \{0, 1\}^{\mathbb{N} - \{0\}}$ . Un problème plus délicat est de trouver une tribu raisonnable sur cet ensemble. On aimerait par exemple pouvoir décrire un événement du type : “en moyenne pile sort deux fois plus souvent que face” qui de façon ensembliste est l'ensemble des  $\omega = (\omega_1, \dots) \in \Omega$  pour lesquels la limite quand  $n$  tend vers l'infini de la suite

$$\frac{1}{n} \sum_{i=1}^n \omega_i$$

existe et vaut  $2/3$ . Un moment de réflexion montre que cet événement n'appartient à aucune des tribus  $\mathcal{B}_n$  qui modélisent un jeu de  $n$  pile/face<sup>2</sup>.

## 2.5.2 Construction de la tribu des événements

### Trinu engendrée

Sauf dans le cas où l'espace  $\Omega$  est fini, les exemples précédents de tribus sont trop simples pour être utiles. La proposition suivante donne un moyen très commode de construire des tribus non-triviales.

**Proposition 2.5.1** *Soit  $\Omega$  un ensemble et  $\mathcal{S}$  un sous-ensemble de parties de  $\Omega$  (un sous-ensemble de  $\mathcal{P}(\Omega)$ ) sans structure particulière. Il existe une unique tribu  $\mathcal{B}$  qui contient  $\mathcal{S}$  et qui est minimale pour cette propriété c'est-à-dire :*

- ( $\mathcal{S} \subset \mathcal{B}$ ) : pour tout  $A \in \mathcal{S}$  on a  $A \in \mathcal{B}$
- (minimale) : si  $\mathcal{B}'$  est une autre tribu telle que  $\mathcal{S} \subset \mathcal{B}'$  alors  $\mathcal{B} \subset \mathcal{B}'$ .

On appelle  $\mathcal{B}$  la tribu engendrée par la partie  $\mathcal{S}$  et on la note (dans ce cours)  $\mathcal{B} = \mathcal{B}(\mathcal{S})$ .

*Démonstration.* —

Considérons l'ensemble  $\mathbf{E}$  des tribus  $\mathcal{C}$  de  $\Omega$  tel que  $\mathcal{S} \subset \mathcal{C}$ . Cet ensemble  $\mathbf{E}$  est non vide puisqu'il contient la tribu  $\mathcal{P}(\Omega)$  et puisque  $\mathcal{S} \subset \mathcal{P}(\Omega)$ . Notons  $\mathcal{B}$  l'intersection des  $\mathcal{C}$  quand  $\mathcal{C}$  décrit  $\mathbf{E}$ . C'est un sous-ensemble de  $\mathcal{P}(\Omega)$  qui contient  $\mathcal{S}$  mais c'est également une tribu comme il est facile de vérifier (nous recommandons au lecteur de vérifier ce point). Ainsi,  $\mathcal{B}$  est une tribu contenant  $\mathcal{S}$  et appartient donc à  $\mathbf{E}$ . Comme  $\mathcal{B}$  est l'intersection des  $\mathcal{C}$  décrivant  $\mathbf{E}$  on a pour toute tribu  $\mathcal{C}$  contenant  $\mathcal{S}$  l'inclusion  $\mathcal{B} \subset \mathcal{C}$  : cela signifie que  $\mathcal{B}$  est la plus petite tribu contenant  $\mathcal{S}$ .  $\square$

### La tribu du modèle de Bernoulli

Nous définirons la tribu  $\mathcal{B}$  sur  $\Omega$  de la façon suivante : la tribu  $\mathcal{B}$  est la tribu engendrée (au sens de la proposition 2.5.1) par tous les événements  $C_{i,\epsilon}$

$$C_{i,\epsilon} = \{\omega \in \Omega, \omega_i = \epsilon\},$$

où  $i$  décrit  $\mathbb{N} - \{0\}$  et  $\epsilon$  décrit  $\{0, 1\}$ .

---

2. On peut toujours considérer un jeu de  $n$  pile/face comme un cas particulier d'un jeu infini de pile/face : Il suffit d'associer à toute suite  $\omega = (\omega_1, \dots, \omega_n)$  de  $\{0, 1\}^n$  la suite  $\tilde{\omega} \in \{0, 1\}^{\mathbb{N}} - \{0\}$  définie par  $\tilde{\omega}_i = \omega_i$  si  $1 \leq i \leq n$  et  $\tilde{\omega}_i = 0$  si  $i \geq n + 1$

### 2.5.3 Construction de la probabilité

Il reste à présent à construire une probabilité sur  $\mathcal{B}$  ce qui est assez délicat. Si on joue avec une pièce qui donne Pile (resp. Face) avec probabilité  $p$  (resp.  $1 - p$ ) il est naturel d'attribuer à tout événement  $C_{i,\epsilon}$  la probabilité  $p$  si  $\epsilon = 1$  et 0 sinon : il est également naturel de demander que la probabilité d'un événement de la forme<sup>3</sup>

$$\{\omega \in \Omega, \omega_{i_1} = \epsilon_1, \dots, \omega_{i_n} = \epsilon_n\} = C_{i_1, \epsilon_1} \cap \dots \cap C_{i_n, \epsilon_n}$$

soit égale à  $p^r(1-p)^{n-r}$  où  $r$  est le nombre de  $1 \leq k \leq n$  pour lesquels  $\epsilon_k = 1$  ; on remarquera que  $r = \sum_{i=1}^n \epsilon_i$ . Il n'est en revanche pas du tout clair que l'on puisse attribuer à tout événement de la tribu  $\mathcal{B}$  une probabilité qui soit compatible avec ces choix. En fait c'est possible :

**Théorème 2.5.1** *Il existe une unique mesure de probabilité  $\mathbf{P}$  définie sur  $(\Omega, \mathcal{B})$  telle que pour tous  $n \geq 1$ ,  $i_1, \dots, i_n \in \mathbb{N} - \{0\}$ ,  $\epsilon_1, \dots, \epsilon_n \in \{0, 1\}$  on ait*

$$\mathbf{P}(C_{i_1, \epsilon_1} \cap \dots \cap C_{i_n, \epsilon_n}) = p^{\epsilon_1 + \dots + \epsilon_n} (1-p)^{n - (\epsilon_1 + \dots + \epsilon_n)}.$$

## 2.6 Événements indépendants

**Définition 2.6.1** *Une famille d'événements  $(A_i)_{i \in I}$  est dite indépendante (ou encore les événements de la famille  $(A_i)_{i \in I}$  sont dits mutuellement indépendants) si pour toute sous famille finie  $A_{i_1}, \dots, A_{i_n}$  on a*

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_n}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_n}).$$

**Exercice 2.6.1** *Démontrer que si  $A, B$  sont deux événements indépendants les quatre familles  $(A, B)$ ,  $(A, B^c)$ ,  $(A^c, B)$ ,  $(A^c, B^c)$  sont chacune indépendantes.*

*Solution.* a) Montrons par exemple que  $(A^c, B)$  est indépendante :

$$\mathbb{P}(A^c \cap B) = \mathbb{P}((E - A) \cap B) = \mathbf{P}(B - (A \cap B)) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

et comme  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(B)\mathbb{P}(A^c).$$

**Exercice 2.6.2** *Trouver un exemple où chacune des familles  $(A, B)$ ,  $(B, C)$ ,  $(C, A)$  est indépendante mais pas la famille  $(A, B, C)$ .*

---

3. Cet événement décrit l'expérience suivante : au temps  $i_1, \dots, i_n$ , on observe  $\epsilon_1, \dots, \epsilon_n$  et on ne précise pas ce qui se passe aux autres temps

Un résultat important qui permet de construire des familles d'événements indépendants est le suivant :

**Théorème 2.6.1** *Soient  $(A_i)_{i \in I}$  une famille d'événements indépendants et  $I = I_1 \cup \dots \cup I_p$  une partition de  $I$ . Notons pour  $1 \leq k \leq p$ ,  $\mathcal{B}_k$  la tribu engendrée par les événements  $A_i$ ,  $i \in I_k$ . Si  $B_1, \dots, B_p$  sont des événements tels que  $B_k \in \mathcal{B}_k$ , alors la famille d'événements  $(B_k)_{1 \leq k \leq p}$  est indépendante.*

*Démonstration.* — Nous illustrons la preuve dans le cas où  $p = 2$ . Démontrons déjà que pour tout événement  $A_i$ ,  $i \in I_1$  et tout événement  $B_2$  dans  $\mathcal{B}_2$  on a

$$\mathbf{P}(A_i \cap B_2) = \mathbf{P}(A_i)\mathbf{P}(B_2). \quad (2.2)$$

Pour cela, notons  $\mathcal{C}_2$  l'ensemble des événements  $B_2$  pour lesquels cette relation est satisfaite pour tout  $i \in I_1$ . On constate déjà que  $\mathcal{C}_2$  est une tribu (**exercice**). Ensuite, on observe que d'après l'hypothèse d'indépendance des  $A_i$ ,  $i \in I$ , les événements  $A_j$ ,  $j \in I_2$  appartiennent à  $\mathcal{C}_2$ . Par définition de la tribu engendrée et sa minimalité, cela signifie que  $\mathcal{B}_2 \subset \mathcal{C}_2$ . Par conséquent, la relation (2.2) est vraie pour tout  $B_2 \in \mathcal{B}_2$ .

Démontrons à présent que pour tout événement  $B_1 \in \mathcal{B}_1$  et tout événement  $B_2$  dans  $\mathcal{B}_2$  on a

$$\mathbf{P}(B_1 \cap B_2) = \mathbf{P}(B_1)\mathbf{P}(B_2). \quad (2.3)$$

Notons  $\mathcal{C}_1$  l'ensemble des événements  $B_1$  pour lesquels cette relation est satisfaite pour tout  $B_2 \in \mathcal{B}_2$ . On constate comme précédemment que  $\mathcal{C}_1$  est une tribu (**exercice**) et qu'elle contient tous les  $A_i$ ,  $i \in I_1$ . La définition du fait que  $\mathcal{B}_1$  est la tribu engendrée par les  $A_i$ ,  $i \in I_1$  montre que  $\mathcal{B}_1 \subset \mathcal{C}_1$  et donc la relation (2.3) est satisfaite pour tout  $B_1 \in \mathcal{B}_1$  et tout  $B_2 \in \mathcal{B}_2$ . □

## 2.7 Probabilités conditionnelles

**Définition 2.7.1** *Si  $A$  et  $B$  sont deux événements de la tribu  $\mathcal{B}$  et si  $\mathbb{P}(B) > 0$  on définit la probabilité de  $A$  sachant  $B$  comme étant*

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

En fait tout événement  $B \in \mathcal{B}$  définit une nouvelle probabilité sur  $(\Omega, \mathcal{B})$  :

**Proposition 2.7.1** *Si  $B \in \mathcal{B}$  est tel que  $\mathbb{P}(B) > 0$ , l'application  $\mathbb{P}_B : \mathcal{B} \rightarrow [0, 1]$  qui à  $A \in \mathcal{B}$  associe  $\mathbb{P}_B(A) = \mathbb{P}(A|B)$  est une probabilité.*

La preuve de cette proposition est laissée en exercice au lecteur.

Supposons à présent que nous ayons une partition de  $\Omega$  en événements  $B_1, \dots, B_r$  (c'est-à-dire que les événements  $B_1, \dots, B_r$  sont disjoints deux à deux et que leur union soit égale à  $\Omega$ ) et que pour tout  $i$   $\mathbb{P}(B_i) > 0$ . On peut donc définir  $r$  probabilités  $\mathbb{P}(\cdot|B_i)$ .

**Théorème 2.7.1 (Formule des causes)** *Sous les hypothèses précédentes*

$$\mathbb{P}(A) = \sum_{i=1}^r \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

*Démonstration.* — Il suffit de remarquer que  $A$  est l'union disjointe des  $(A \cap B_i)$  et donc

$$\begin{aligned} \mathbb{P}(A) &= \sum_{k=1}^r \mathbb{P}(A \cap B_k) \\ &= \sum_{i=1}^r \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)} \cdot \mathbb{P}(B_i) \end{aligned}$$

ce qui est la formule annoncée.  $\square$

Dans la pratique un problème courant est de calculer  $\mathbb{P}(B_i|A)$  connaissant les  $\mathbb{P}(A|B_j)$ .

**Théorème 2.7.2 (Formule de Bayes)** *Sous les hypothèses précédentes :*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^r \mathbb{P}(A|B_j)\mathbb{P}(B_j)}.$$

*Démonstration.* — Il suffit d'écrire

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(B_i \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)},$$

et d'utiliser la formule des causes.  $\square$





# Chapitre 3

## Variables aléatoires réelles

### 3.1 Variables Aléatoires réelles

#### 3.1.1 Définition et premières propriétés

Dans toute la suite on suppose que  $(\Omega, \mathcal{B}, \mathbb{P})$  est un espace de probabilité.

**Définition 3.1.1** *Une variable aléatoire réelle (en abrégé v.a.r. mais nous écrirons plus simplement v.a.) est une application  $X : \Omega \rightarrow \mathbb{R}$  telle que pour tout intervalle  $I$  de  $\mathbb{R}$  l'ensemble  $X^{-1}(I)$  des  $\omega \in \Omega$  tels que  $X(\omega) \in I$ , appartient à la tribu  $\mathcal{B}$ .*

En fait

**Proposition 3.1.1** *Si  $X : \Omega \rightarrow \mathbb{R}$  est une application alors les trois propositions suivantes sont équivalentes*

1.  $X$  est une variable aléatoire ;
2. pour tout intervalle  $I$  de la forme  $I = ]-\infty, a]$ ,  $X^{-1}(I)$  appartient à  $\mathcal{B}$  ;
3. pour tout intervalle  $I$  de la forme  $]a, b[$ ,  $X^{-1}(I)$  appartient à  $\mathcal{B}$ .

*Démonstration.* — Que 1 implique 2 est évident. Démontrons donc que 2 implique 3. On constate que tout intervalle  $]a, b[$  peut s'écrire sous la forme  $] - \infty, a]^c \cap ] - \infty, b]$  et que  $]a, b[ = \bigcup_{n \geq 1} ]a, b - \frac{1}{n}]$ . On a donc (en utilisant le fait que les opérations d'intersection, union, passage au complémentaire se comportent bien par images inverses)

$$X^{-1}(I) = \bigcup_{n \geq 1} (X^{-1}(] - \infty, a]))^c \cap X^{-1}(-\infty, b - \frac{1}{n}]).$$

Comme les ensembles apparaissant dans le membre de droite de l'égalité sont des éléments de  $\mathcal{B}$  et que  $\mathcal{B}$  est une tribu, on en déduit que le membre de gauche est également dans  $\mathcal{B}$ .

La démonstration du fait que 3 implique 1 se démontre de manière analogue. Démontrons par exemple que pour  $a \leq b$ ,  $X^{-1}([a, b])$  est dans  $\mathcal{B}$ . On observe que  $[a, b] = \bigcap_{n \geq 1} ]a - \frac{1}{n}, b + \frac{1}{n}[$  et donc

$$X^{-1}([a, b]) = \bigcap_{n \geq 1} X^{-1}(]a - \frac{1}{n}, b + \frac{1}{n}[)$$

est dans  $\mathcal{B}$ . Les autres cas se traitent de façon analogue.  $\square$

**Notation** Dans la suite du cours, quand  $X$  est une v.a et  $A \subset \mathbb{R}$  nous noterons  $\{X \in A\}$  ou  $[X \in A]$  ou  $(X \in A)$  l'ensemble  $\{\omega \in \Omega : X(\omega) \in A\}$ .

La proposition qui suit permet de construire de v.a.

**Proposition 3.1.2** a) Si  $X_n$ ,  $n \geq 1$  est une famille de v.a alors  $Z = \sup_{n \geq 1} X_n$  (resp.  $Z = \inf_{n \geq 1} X_n$ ) est une v.a

b) Si  $X_n$ ,  $n \geq 1$  est une famille de v.a alors  $Z = \limsup_{n \geq 1} X_n$  (resp.  $Z = \liminf_{n \geq 1} X_n$ ). En particulier si pour tout  $\omega \in \Omega$ ,  $X_n(\omega)$  converge vers  $X(\omega)$  alors  $X$  est une v.a.

c) Si  $X_1, \dots, X_n$  sont des v.a et  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  est une application continue alors  $Z = f(X_1, \dots, X_n)$  est une v.a. En particulier  $X_1 + X_2$ ,  $\lambda X$ , ( $\lambda \in \mathbb{R}$ ),  $X_1 \cdot X_2$  sont des v.a.

*Démonstration.* —

a) Soit  $\omega$  tel que  $\sup_{n \geq 1} X_n(\omega) > a$ . Alors par définition du sup, il existe un  $n$  pour lequel  $X_n(\omega) > a$  et  $\omega$  est donc dans l'union  $\bigcup_{n \geq 1} \{X_n > a\}$ . Réciproquement si  $\omega \in \bigcup_{n \geq 1} \{X_n > a\}$  alors il existe  $n$  tel que  $X_n(\omega) > a$  et *a fortiori*  $\sup_n X_n(\omega) > a$ . Nous avons donc démontré que les deux ensembles  $\{Z > a\}$  et  $\bigcup_{n \geq 1} \{X_n > a\}$  sont égaux. Mais ce dernier ensemble est une union dénombrable d'éléments de la tribu  $\mathcal{B}$  (car chaque  $X_i$  est une v.a). Ainsi pour tout  $a$  l'événement  $\{\sup_{n \geq 1} X_n(\omega) > a\}$  est dans  $\mathcal{B}$  et il en est de même de son complémentaire  $\{\sup_{n \geq 1} X_n(\omega) \leq a\}$ . La proposition 3.1.1 permet de conclure.

b) Par définition  $\limsup_{n \rightarrow \infty} X_n(\omega) = \inf_{p \in \mathbb{N}} \sup_{k \geq p} X_k(\omega)$  (resp.  $\liminf_{n \rightarrow \infty} X_n(\omega) = \sup_{p \in \mathbb{N}} \inf_{k \geq p} X_k(\omega)$ ). Il suffit donc d'appliquer deux fois a). La dernière assertion résulte du fait que  $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$  si et seulement si  $\liminf X_n(\omega) = \limsup X_n(\omega) = X(\omega)$ .

c) Si  $I$  est un intervalle ouvert de  $\mathbb{R}$ ,  $Z^{-1}(I)$  est l'ensemble des  $\omega \in \Omega$  tels que  $(X_1(\omega), \dots, X_n(\omega)) \in f^{-1}(I)$ . Comme  $f$  est continue,  $f^{-1}(I)$  est un ensemble *ouvert* de  $\mathbb{R}^n$  et, par conséquent, est une union dénombrable de pavés ouverts c'est-à-dire d'ensembles  $P$  de la forme  $]a_1, b_1[ \times \dots \times ]a_n, b_n[$  (nous admettrons ce fait qui n'est par ailleurs pas très difficile à démontrer). Par conséquent l'ensemble des  $\omega \in \Omega$  tels que  $(X_1(\omega), \dots, X_n(\omega)) \in f^{-1}(I)$  est une union dénombrable d'ensembles de la forme  $\{\omega \in \Omega, (X_1(\omega), \dots, X_n(\omega)) \in ]a_1, b_1[ \times \dots \times ]a_n, b_n[\}$  c'est-à-dire d'ensembles de la forme  $\{\omega \in \Omega, X_1(\omega) \in ]a_1, b_1[, \dots, X_n(\omega) \in ]a_n, b_n[\} = X_1^{-1}(]a_1, b_1[) \cap \dots \cap X_n^{-1}(]a_n, b_n[)$  qui sont clairement dans  $\mathcal{B}$ .

□

En particulier

**Proposition 3.1.3** *a) Si  $X : \Omega \rightarrow \mathbb{R}$  est une v.a. et  $f : \mathbb{R} \rightarrow \mathbb{R}$  est une application continue, alors la fonction  $Y : \Omega \rightarrow \mathbb{R}$  définie par  $Y = f \circ X$  (c'est-à-dire  $Y(\omega) = f(X(\omega))$  pour tout  $\omega \in \Omega$ ) est encore une v.a. On la note  $Y = f(X)$ .*

*b) Si  $X, Y$  sont deux v.a l'application  $Z = \max(X, Y)$  est une v.a*

*c) Si  $X$  et  $Y$  sont deux v.a,  $aX + bY$  est également une v.a.*

### 3.1.2 Le cas particulier des v.a à valeurs dans un ensemble fini ou dénombrable

Il s'agit du cas où  $X(\Omega)$  l'ensemble des valeurs prises par  $X$  est un ensemble fini ou dénombrable de  $\mathbb{R}$ . Dans ce cas la caractérisation des variables aléatoires est plus simple :

**Proposition 3.1.4** *Si  $X : \Omega \rightarrow \mathbb{R}$  est à valeurs dans un ensemble fini ou dénombrable  $E$  alors  $X$  est une variable aléatoire si et seulement si pour tout  $e \in E$ ,  $X^{-1}(\{e\}) \in \mathcal{B}$ .*

*Démonstration.* — Supposons que pour tout  $e \in E$ ,  $X^{-1}(\{e\}) \in \mathcal{B}$  et soit  $I$  un intervalle de  $\mathbb{R}$ . Comme  $X$  est à valeurs dans  $E$ , on a

$$X^{-1}(I) = \bigcup_{e \in I \cap E} X^{-1}(\{e\})$$

et comme  $E \cap I \subset E$  est dénombrable on déduit que  $X^{-1}(I)$  est dans  $\mathcal{B}$ .

Réciproquement, si  $X^{-1}(I)$  est dans  $\mathcal{B}$  pour tout intervalle, on a en prenant  $I = \{e\}$ ,  $e \in E$  que  $X^{-1}(\{e\}) \in \mathcal{B}$ . □

**Exemple** Revisitons l'exemple du jeu infini de Pile/Face :  $\Omega = \{0, 1\}^{\mathbf{N}}$  et  $\mathcal{B}$  est la tribu engendrée par les ensembles  $C_{i,\epsilon_i} = \{\omega = (\omega_0, \dots) \in \Omega, \omega_i = \epsilon_i\}$ . Pour  $n \in \mathbf{N}$  l'application  $X_n : \{0, 1\}^{\mathbf{N}} \rightarrow \{0, 1\}$  qui à  $\omega = (\omega_0, \omega_1, \dots)$  associe  $\omega_n$  est une variable aléatoire. Il suffit en effet de vérifier que pour  $\epsilon = 0$  ou  $\epsilon = 1$  l'ensemble des  $\omega$  pour lesquels  $\omega_n = \epsilon$  appartient à  $\mathcal{B}$ . Or, cet ensemble est le cylindre  $C_{n,\epsilon}$  qui par définition est dans  $\mathcal{B}$ . En fait, la tribu  $\mathcal{B}$  a été construite de façon que toutes les applications  $X_n : \Omega \rightarrow \{0, 1\}$  ( $n \geq 0$ ) soient des variables aléatoires (c'est d'ailleurs la plus petite tribu ayant cette propriété).

**Exercice** On considère le jeu infini de Pile/Face  $(\Omega, \mathcal{B}, \mathbb{P})$  et on garde les notations de l'exemple précédent. Définissons pour tout  $\omega \in \Omega$ , l'entier  $\nu(\omega)$  comme étant le plus petit entier  $k$  pour lequel  $X_k(\omega) = 1$  (en d'autres termes  $\nu(\omega)$  est le premier temps où on tire Pile). Démontrer que  $\nu$  est une variable aléatoire.

*Solution* : Si  $\{\nu = 0\} = \{X_0 = 1\}$  est dans  $\mathcal{B}$  car  $X_0$  est une v.a et pour tout  $n \in \mathbf{N}$ , ( $n \geq 1$ )

$$\{\nu = n\} = \{X_0 = 0\} \cap \dots \cap \{X_{n-1} = 0\} \cap \{X_n = 1\};$$

c'est une intersection finie d'éléments de  $\mathcal{B}$  (car, puisque chaque  $X_i$  est une v.a, les ensembles  $\{X_i = 0\}$ ,  $1 \leq i \leq n-1$  et  $\{X_n = 1\}^c$  sont dans  $\mathcal{B}$ )

## 3.2 Loi d'une variable aléatoire réelle

Soit  $X : \Omega \rightarrow \mathbb{R}$  une v.a. On sait que pour tout intervalle  $I$  de  $\mathbb{R}$  l'ensemble  $X^{-1}(I)$  est un événement (appartient à  $\mathcal{B}$ ). Il est donc possible de parler de la probabilité  $\mathbb{P}(X \in I)$  de l'événement  $\{X \in I\}$ .

**Définition 3.2.1** Soit  $X$  une v.a.r. L'application qui à tout intervalle  $I$  de  $\mathbb{R}$  associe  $\mathbb{P}(X \in I)$  s'appelle la loi de  $X$ .

**Exemple 3.2.1** Si  $A \in \mathcal{B}$ , sa fonction indicatrice  $\mathbf{1}_A : \Omega \rightarrow \mathbb{R}$  (qui prend la valeur 1 si  $\omega \in A$  et 0 sinon) est une v.a.r. Réciproquement, si une v.a. ne prend que les valeurs 0 et 1 elle est la fonction indicatrice de  $A = \mathbf{1}_A^{-1}(\{1\}) \in \mathcal{B}$ .

### Compléments

En fait la loi d'une v.a.r. a une signification plus profonde. Appelons  $Bor(\mathbb{R})$  le sous-ensemble de  $\mathcal{P}(\mathbb{R})$  défini comme étant la tribu engendrée par les intervalles de  $\mathbb{R}$ . Cette tribu s'appelle la *tribu borélienne* (ou de Borel); les éléments de la tribu borélienne s'appellent les boréliens de  $\mathbb{R}$ . On a alors le résultat suivant :

**Proposition 3.2.1** *Si  $X : \Omega \rightarrow \mathbb{R}$  est une v.a.r, alors pour tout borélien  $A \in \text{Bor}(\mathbb{R})$ ,  $X^{-1}(A) \in \mathcal{B}$ .*

*Démonstration.* — Si on note  $\mathcal{C}$  l'ensemble des  $A \subset \mathbb{R}$  tels que  $X^{-1}(A) \in \mathcal{B}$ , il est facile de voir que  $\mathcal{C}$  est une tribu ; par ailleurs, il est clair que  $\mathcal{C}$  contient les intervalles. Comme la tribu borélienne est la tribu engendrée par les intervalles, c'est-à-dire la plus petite tribu de  $\mathbb{R}$  contenant les intervalles, on a  $\text{Bor}(\mathbb{R}) \subset \mathcal{C}$ . Mais alors, pour tout borélien  $A$  de  $\mathbb{R}$ ,  $X^{-1}(A) \in \mathcal{B}$ .  $\square$

La proposition suivante montre que toute v.a.r. définit naturellement une probabilité sur  $\mathbb{R}$  muni de sa tribu borélienne.

**Proposition 3.2.2** *L'application  $\mu_X : \text{Bor}(\mathbb{R}) \rightarrow [0, 1]$  qui à tout borélien de  $\mathbb{R}$  associe le réel  $P(X \in A)$  de  $[0, 1]$  est une probabilité sur  $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ . On appelle cette probabilité la loi de la v.a  $X$ .*

*Démonstration.* — Il suffit de démontrer que si  $(A_i)_{i \in \mathbb{N}}$  est une famille dénombrable de boréliens de  $\mathbb{R}$  disjoints deux à deux alors

$$\mathbb{P}(X \in \bigcup_{i \in \mathbb{N}} A_i) = \sum_{i=0}^{\infty} P(X \in A_i),$$

ce qui est clair car l'événement  $\{X \in \cup_{i \in \mathbb{N}} A_i\}$  est l'union dénombrable *disjointe* des événements  $\{X \in A_i\}$ . Enfin la condition  $P(X \in \mathbb{R}) = 1$  achève la preuve.  $\square$

Il faut retenir que la loi d'une v.a est une *probabilité* sur  $\mathbb{R}$  (muni de sa tribu borélienne). Ceci illustre le fait qu'il est possible de construire de nombreuses mesures de probabilités sur  $\mathbb{R}$  muni de sa tribu borélienne.

### 3.2.1 Loi des variables aléatoires à valeurs dans un ensemble fini ou dénombrable

Si  $X : \Omega \rightarrow \mathbb{R}$  prend ses valeurs dans un ensemble  $E = \{e_0, e_1, \dots\}$  qui est fini ou dénombrable on a pour tout intervalle  $I \subset \mathbb{R}$

$$\mathbb{P}(X \in I) = \mathbb{P}(X \in I \cap E) = \sum_{e \in I \cap E} \mathbb{P}(X = e).$$

La loi de  $X$  est donc parfaitement déterminée par les réels  $p_X(e) = \mathbb{P}(X = e)$ , ( $e \in E$ ) et dans la pratique quand on demande de déterminer la loi de  $X$  on demande de calculer les réels  $p_X(e) = \mathbb{P}(X = e)$ .

La loi  $\mu_X$  de  $X$  est la probabilité sur  $(\mathbb{R}, \text{Bor}(\mathbb{R}))$  définie par

$$\mu_X = \sum_{e \in E} \mathbb{P}(X = e) \delta_e$$

où  $\delta_e$  est la mesure de Dirac (pour  $A \in \text{Bor}(\mathbb{R})$ ,  $\delta_e(A) = 1$  si  $e \in A$  et 0 sinon). En effet pour tout borélien (ou tout intervalle)  $A$

$$\mathbb{P}(X \in A) = \mathbb{P}(X \in A \cap E) = \sum_{e \in A \cap E} \mathbb{P}(X = e),$$

et cette somme n'est rien d'autre que

$$\sum_{e \in E} \mathbb{P}(X = e) \delta_{\mathbb{E}}(A).$$

### Quelques lois classiques de variables aléatoires à valeurs dans un ensemble fini ou dénombrable

**Loi de Bernoulli** C'est la loi d'une v.a.r prenant pour valeur 0 ou 1 et  $\mathbb{P}(X = 1) = p$ ,  $\mathbb{P}(X = 0) = 1 - p$ . On dit que  $p \in [0, 1]$  est le paramètre de la loi.

**Loi géométrique** On dit qu'une v.a  $X : \Omega \rightarrow \mathbb{N}$  (à valeurs dans  $\mathbb{N}$ ) suit une loi géométrique de paramètre  $a$  ( $0 < a < 1$ ) si

$$\mathbb{P}(X = n) = (1 - a)a^n.$$

On remarquera que l'on a bien  $\sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$  ( $\sum_{k \geq 0} a^k = 1/(1 - a)$ ).

La v.a  $\nu$  de l'exercice de la section 3.1.2 suit une loi géométrique de paramètre  $1/2$ . En effet

$$\begin{aligned} \nu^{-1}(n) &= \{\omega = (\omega_0, \omega_1, \dots) \in \Omega, \omega_0 = 0, \dots, \omega_{n-1} = 0, \omega_n = 1\} \\ &= C_{0,0} \cap \dots \cap C_{n-1,0} \cap C_{n,1} \end{aligned}$$

et d'après le théorème 2.5.1

$$\mathbb{P}(\nu = n) = (1/2)^n \cdot (1/2).$$

**Loi binomiale** On dit qu'une variable aléatoire  $Z$  à valeurs dans  $\{0, \dots, n\}$  suit une loi binomiale  $(n, p)$  si

$$\mathbb{P}(Z = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

où  $\binom{n}{k}$  est le coefficient binomial

$$\binom{n}{k} = C_n^k = \frac{n!}{(n-k)!k!} = \frac{n(n-1) \cdots (n-k+1)}{k!}.$$

On a bien (formule du binôme de Newton)  $\sum_{k=0}^n \mathbb{P}(Z = k) = (p+1-p)^n = 1$ .

**Exemple** Jouons  $n$  fois au jeu de pile/face où pile sort avec probabilité  $p$  et face avec probabilité  $1-p$  et notons  $Z$  la variable aléatoire :  $Z$  est le nombre de

pile qui sortent (après avoir joué  $n$  fois). Si on note  $X_i$  les variables aléatoires  $X_i(\omega) = \omega_i$  ( $\omega = (\omega_1, \dots, \omega_n)$ ) on a

$$Z = X_1 + \dots + X_n.$$

C'est bien une variable aléatoire  $Z : \mathcal{B} \rightarrow \mathbb{N}$  à valeurs dans l'ensemble fini  $\{0, \dots, n\}$  ( $\mathcal{B} = \mathbb{P}(\Omega)$ ) et

$$\mathbb{P}(Z = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

**Loi de Poisson** Une variable aléatoire  $Z : \Omega \rightarrow \mathbb{N}$  suit une loi de Poisson de paramètre  $\lambda > 0$  si

$$\mathbb{P}(Z = n) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

On vérifie encore que  $\sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$  (cf. le développement en série de  $e^\lambda$ ).

**Exercice** Soit  $X_n$  une v.a suivant une loi binomiale  $(n, p_n)$ . Montrer que si  $\lim_{n \rightarrow \infty} np_n = \lambda$  on a pour tout  $k \in \mathbb{N}$

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

(On dit que  $X_n$  converge en loi vers une loi de Poisson de paramètre  $\lambda$ )

### 3.2.2 Loi de variables aléatoires admettant une densité

**Définition 3.2.2** On dit que la variable aléatoire  $X : \Omega \rightarrow \mathbb{R}$  admet une densité continue (resp. continue par morceaux etc.) s'il existe une fonction positive continue (resp. continue par morceaux etc.)  $\rho_X : \mathbb{R} \rightarrow [0, \infty[$  telle que  $\int_{-\infty}^{\infty} \rho_X(t) dt = 1$  et telle que pour tout intervalle  $I$

$$\mathbb{P}(X \in I) = \int_I \rho_X(t) dt.$$

#### Fonction de répartition

**Définition 3.2.3** Si  $X$  est une v.a.r. on définit sa fonction de répartition  $F_X : \mathbb{R} \rightarrow [0, 1]$  par

$$F_X(x) = \mathbb{P}(X \leq x).$$

Voici quelques propriétés des fonctions de densité.

**Proposition 3.2.3** Si  $F_X : \mathbb{R} \rightarrow [0, 1]$  est la fonction de répartition d'une v.a.r.  $X$

- la fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$  est croissante : si  $x_1 \leq x_2$  alors  $F_X(x_1) \leq F_X(x_2)$ .
- on a  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  et  $\lim_{x \rightarrow \infty} F_X(x) = 1$ .
- $F_X$  est continue à droite en tout point.

**Exemple** Si  $X$  est une v.a.r. prenant seulement deux valeurs 0 ou 1 et  $\mathbb{P}(X = 1) = p$ ,  $\mathbb{P}(X = 0) = 1 - p$  ( $X$  est une v.a. de Bernoulli de paramètre  $p$ ) alors sa fonction de répartition  $F_X(t) = \mathbb{P}(X \leq t)$  vaut 0 si  $x < 0$ ,  $1 - p$  si  $0 \leq x < 1$  et 1 si  $x > 1$ .

### Fonction de répartition d'une v.a.r. admettant une densité

Faisons une remarque importante : si une v.a.  $X$  admet une densité  $\rho_X$  alors sa fonction de répartition

$$F_X(x) = \mu_X([-\infty, x]) = \int_{-\infty}^x \rho_X(t) dt$$

est continue.

De la définition précédente on déduit facilement la proposition suivante.

**Proposition 3.2.4** Une v.a.r.  $X$  admet une densité  $\rho$  continue par morceaux si et seulement si sa fonction de répartition  $F_X$  est continue et dérivable par morceaux. On a alors en tout point où  $F_X$  est dérivable  $F'_X(x) = \rho_X(x)$ .

Si on note  $\mu_X$  la loi de  $X$  ( $\mu_X(A) = \mathbb{P}(X \in A)$  pour  $A \in \text{Bor}(\mathbb{R})$ ) on a donc  $\mu_X(I) = \int_I \rho_X(t) dt$  pour tout intervalle  $I$  et  $F_X(x) = \mu_X([-\infty, x]) = \int_{-\infty}^x \rho_X(t) dt$ . Il existe donc des variables aléatoires n'admettant pas de densité : par exemple une v.a  $X$  à valeurs dans  $\mathbb{R}$  ne prenant que deux valeurs 0 ou 1 et telle que  $\mathbb{P}(X = 0) = p$  avec  $p = 1/2$  ne peut posséder de densité car sa fonction de répartition  $F_X(x)$  vaut 0 si  $x < 0$ ,  $1/2$  si  $0 \leq x < 1$  et 1 si  $1 \leq x$  : elle est discontinue en 0 et en 1 (mais bien continue à droite).

### Quelques exemples de loi admettant une densité

**Loi uniforme** La variable aléatoire  $X : \Omega \rightarrow \mathbb{R}$  suit une loi uniforme sur l'intervalle  $[a, b]$  si sa densité est donnée par

$$\rho_X(x) = \frac{1}{b-a} \cdot \mathbf{1}_{[a,b]}.$$

On a bien  $\rho_X(t) \geq 0$  pour tout  $t$  et  $\int_{\mathbb{R}} \rho_X(t) dt = 1$ . Cette loi est caractérisée par

$$\mathbb{P}(X \in [c, d]) = \frac{1}{b-a} \text{longueur}([a, b] \cap [c, d]).$$



(En effet,

$$\begin{aligned}\mathbb{P}(X \in [c, d]) &= \int_{[c, d]} \frac{1}{b-a} \cdot \mathbf{1}_{[a, b]}(x) dx \\ &= \frac{1}{b-a} \int_{\mathbb{R}} \mathbf{1}_{[c, d]}(x) \cdot \mathbf{1}_{[a, b]}(x) dx \\ &= \frac{1}{b-a} \int_{\mathbb{R}} \mathbf{1}_{[c, d] \cap [a, b]}(x) dx.\end{aligned}$$

**Loi exponentielle de paramètre  $\theta$**  La v.a admet une densité  $\rho_X$  exponentielle de paramètre  $\theta$  si

$$\rho_X(x) = \theta e^{-\theta x} \mathbf{1}_{[0, \infty[}(x).$$

La fonction de répartition est

$$F_X(x) = \int_{-\infty}^x \rho_X(t) dt = (1 - e^{-\theta x}) \mathbf{1}_{[0, \infty[}(x),$$

et converge bien vers 1 en  $\infty$ . Intuitivement, la loi uniforme sur l'intervalle  $[a, b]$  modélise une expérience où la probabilité d'un point de tomber dans un intervalle de taille  $2\Delta x$ ,  $]x - \Delta x, x + \Delta x[ \subset [a, b]$  ne dépend pas de  $x$  (et est linéaire en  $\Delta x$ ).

**Loi normale  $N(\mu, \sigma^2)$**  C'est la loi de densité

$$\rho_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}.$$

Il n'est pas complètement évident que  $\int_{\mathbb{R}} \rho_X(x) dx = 1$  (ce qui est indispensable pour que  $\rho_X$  soit une densité). Ceci résulte, après le changement de variable  $u = (x - \mu)/\sigma$  de l'égalité (voir plus loin pour une preuve)

$$\int_{-\infty}^{\infty} e^{-u^2/2} du = \sqrt{2\pi}.$$

La loi normale  $N(0, 1)$ , donc de densité,

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

est dite loi normale centrée réduite.

### 3.3 Espérance d'une v.a. à valeurs dans un ensemble fini (ou dénombrable)

#### 3.3.1 Définition

Soit  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire ne prenant qu'un nombre fini de valeurs  $x_1, \dots, x_r$ . On définit l'espérance de  $X$  comme étant le nombre réel

$$\mathbb{E}(X) = \sum_{i=1}^r x_i \cdot \mathbb{P}(X = x_i).$$

Remarquons que si  $X$  prend ses valeurs dans un ensemble infini dénombrable la quantité

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i \cdot \mathbb{P}(X = x_i),$$

qui semble être un bon candidat pour la définition de l'espérance peut ne pas exister car la série peut ne pas converger. Pour garantir cette convergence il suffit de demander que la série précédente soit absolument convergente.

On adoptera donc la définition suivante :

**Définition 3.3.1** *Si  $X$  est une v.a.r. prenant un nombre fini ou dénombrable de valeurs dans  $E \subset \mathbb{R}$ , telle que*

$$\sum_{x \in E} |x| \mathbb{P}(X = x) < \infty \tag{3.1}$$

*on définit l'espérance de  $X$  comme étant le nombre réel*

$$\mathbb{E}(X) = \sum_{x \in E} x \mathbb{P}(X = x).$$

*Quand la condition (3.1) est vérifiée on dit que  $X$  est dans  $\mathcal{L}^1(\Omega, \mathcal{B}, \mathbb{P})$ .*

#### Remarque

- Quand  $X$  est positive, et dans ce cas seulement, si la série  $\sum_{x \in E} |x| \mathbb{P}(X = x) = \infty$  diverge on pose  $\mathbb{E}(X) = \infty$ .
- La variance représente intuitivement la *valeur moyenne* prise par la variable aléatoire  $X$ .

### 3.3. ESPÉRANCE D'UNE V.A. À VALEURS DANS UN ENSEMBLE FINI (OU DÉNOMBRABLE)

#### 3.3.2 Premières propriétés

**Théorème 3.3.1** *L'espérance des v.a.r. prenant un nombre fini ou dénombrable de valeurs vérifie les conditions suivantes :*

a) si  $A \in \mathcal{B}$  on a  $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$ .

b) Si  $X, Y$  sont des v.a.r. positives prenant un nombre fini ou dénombrable de valeurs telles que  $X \leq Y$  (c'est-à-dire pour tout  $\omega \in \Omega$   $X(\omega) \leq Y(\omega)$ ) alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .

c) Si  $X, Y$  sont des v.a dans  $\mathcal{L}^1(\Omega, \mathcal{B}, \mathbb{P})$  prenant un nombre fini ou dénombrable de valeurs et  $a, b \in \mathbb{R}$  on a (linéarité de l'espérance)

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

d) Si  $X$  est une v.a positive prenant un nombre fini ou dénombrable de valeurs telle que  $\mathbb{E}(X) = 0$  alors  $X$  est nulle  $\mathbb{P}$ -presque sûrement c'est-à-dire que l'ensemble des  $\omega \in \Omega$  pour lesquels  $X(\omega) > 0$  a une probabilité nulle.

*Démonstration.* — Le point a) est immédiat. Démontrons le point b). Notons  $(x_i)_i$  l'ensemble des valeurs prises par  $X$  et  $(y_j)_j$  celles prises par  $Y$ . Définissons  $A_i = X^{-1}(x_i)$ ,  $B_j = Y^{-1}(y_j)$  qui sont dans la tribu des événements  $\mathcal{B}$ . Les  $(A_i)_i$  constituent une partition de  $\Omega$  tout comme les  $(B_j)_j$ . Par conséquent les  $(A_i \cap B_j)_{i,j}$  constituent également une partition de  $\Omega$ . Si  $A_i \cap B_j \neq \emptyset$  alors  $x_i \leq y_j$  car  $X(\omega) \leq Y(\omega)$  pour  $\omega \in A_i \cap B_j$ ; sinon  $\mathbb{P}(A_i \cap B_j) = 0$ . Donc dans tous les cas  $\mathbb{P}(A_i \cap B_j)x_i \leq \mathbb{P}(A_i \cap B_j)y_j$ . A présent écrivons

$$\begin{aligned} \mathbb{E}(X) &= \sum_i \mathbb{P}(A_i)x_i \\ &= \sum_{i,j} \mathbb{P}(A_i \cap B_j)x_i \\ &\leq \sum_{i,j} \mathbb{P}(A_i \cap B_j)y_j \\ &\leq \sum_j \mathbb{P}(B_j)y_j = \mathbb{E}(Y). \end{aligned}$$

Démontrons maintenant le point c). En utilisant la décomposition précédente

on voit que

$$\begin{aligned}
 \mathbb{E}(aX + bY) &= \sum_{i,j} \mathbb{P}(A_i \cap B_j)(ax_i + by_j) \\
 &= a \sum_{i,j} \mathbb{P}(A_i \cap B_j)x_i + b \sum_{i,j} \mathbb{P}(A_i \cap B_j)y_j \\
 &= a \sum_i \mathbb{P}(A_i)x_i + b \sum_{i,j} \mathbb{P}(B_j)y_j \\
 &= a\mathbb{E}(X) + b\mathbb{E}(Y).
 \end{aligned}$$

Enfin, pour le point d) nous remarquons (avec les notations précédentes) que si  $\mathbb{E}(X) = 0$  on a  $\sum_i \mathbb{P}(A_i)x_i = 0$ ; mais comme  $X$  est positive, chacun des termes de la somme précédente est positif ou nulle et donc pour tout  $i$ ,  $\mathbb{P}(A_i)x_i = 0$ . Cela démontre que pour tout  $i$  pour lequel  $x_i > 0$  on doit avoir  $\mathbb{P}(A_i) = 0$ . Ainsi  $\mathbb{P}(\bigcup_{i,x_i \neq 0} A_i) = 0$ . Cela démontre le point d.  $\square$

### 3.3.3 Formule de transfert

Il est important dans la pratique de savoir calculer des espérances de v.a aléatoires de la forme  $Y = f(X)$  où  $X : \Omega \rightarrow \mathbb{R}$  est une v.a et  $f : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction (disons continue).

**Proposition 3.3.1** *Si  $X$  prend un nombre fini ou dénombrable de valeurs dans  $E \subset \mathbb{R}$  et si*

$$\sum_{e \in E} |f(e)|\mathbb{P}(X = e) < \infty$$

*alors l'espérance de la variable aléatoire  $Y = f(X)$  est donnée par*

$$\mathbb{E}(f(X)) = \sum_{e \in E} f(e)\mathbb{P}(X = e).$$

*Démonstration.* — Supposons que  $X : \Omega \rightarrow \mathbb{R}$  prenne ses valeurs dans un ensemble fini  $E \subset \mathbb{R}$  et soit  $f : E \rightarrow \mathbb{R}$ . Comme  $Y = f(X)$ , la v.a  $Y$  ne prend qu'un nombre fini de valeurs  $e'$  qui sont dans  $E' = f(E)$ . Par définition de l'espérance

$$\mathbb{E}(Y) = \sum_{e' \in E'} e'\mathbb{P}(Y = e').$$

Pour chaque  $e' \in E'$ , notons  $A_{e'}$  l'ensemble des  $e \in E$  tels que  $f(e) = e'$  ( $A_{e'} = f^{-1}(\{e'\})$ ) et constatons que  $E'$  est l'union disjointe des  $A_{e'}$ ,  $e' \in E'$ .

On a donc,

$$\begin{aligned}
 \mathbb{E}(Y) &= \sum_{e' \in E'} e' \mathbb{P}(Y = e') \\
 &= \sum_{e' \in E'} e' \mathbb{P}(f(X) = e') \\
 &= \sum_{e' \in E'} e' \mathbb{P}(X \in f^{-1}(e')) \\
 &= \sum_{e' \in E'} e' \sum_{e \in f^{-1}(e')} \mathbb{P}(X = e) \\
 &= \sum_{e' \in E'} \sum_{e \in f^{-1}(e')} f(e) \mathbb{P}(X = e) \\
 &= \sum_{e \in E} f(e) \mathbb{P}(X = e) \\
 &= \mathbb{E}(X).
 \end{aligned}$$

Un examen de la preuve que nous venons de présenter montre qu'elle s'étend au cas où  $X$  prend un nombre infini dénombrable de valeurs à condition que l'hypothèse de la proposition soit vérifiée.

□

### 3.4 Espérance des v.a.r. : cas général

En s'inspirant de la définition de l'espérance que nous avons exposée dans la section précédente pour les v.a ne prenant qu'un nombre fini ou dénombrable de valeurs, il est possible de définir, sous certaines conditions, l'espérance de v.a.r. en général. La construction de l'espérance dans le cas général est cependant un peu fastidieuse et nous la laisserons donc de côté.

**Théorème 3.4.1 (Espérance : cas positif)** *A toute variable aléatoire réelle positive  $X$  il est possible d'associer un élément de  $[0, \infty[ \cup \{\infty\}$  que l'on appelle l'espérance de  $X$  et que l'on note  $\mathbb{E}(X)$  et qui vérifie les propriétés suivantes :*

- a) si  $A \in \mathcal{B}$  on a  $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$ .
- b) Si  $X, Y$  sont des v.a positives telle que  $X \leq Y$  (c'est-à-dire pour tout  $\omega \in \Omega$   $X(\omega) \leq Y(\omega)$ ) alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .
- c) Si  $X, Y$  sont des v.a positives et  $a, b \in \mathbb{R}$  on a (linéarité de l'espérance)

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

d) Si  $X$  est une v.a positive telle que  $\mathbb{E}(X) = 0$  alors  $X$  est nulle  $\mathbb{P}$ -presque sûrement c'est-à-dire que l'ensemble des  $\omega \in \Omega$  pour lesquels  $X(\omega) > 0$  a une probabilité nulle.

### Remarque

- Notons que  $\mathbb{E}(a) = a$  si  $a$  est une constante.
- Il est facile de voir que les conditions précédentes imposent que l'espérance coïncide avec celle que nous avons définie pour les v.a. prenant un nombre fini ou dénombrable de valeurs.

**Remarque** On dit qu'une propriété  $P_\omega$  qui dépend de  $\omega \in \Omega$  est vraie  $\mathbb{P}$ -presque sûrement si l'ensemble des  $\omega \in \Omega$  pour lesquels  $P_\omega$  est fausse est de  $\mathbb{P}$ -probabilité nulle ( $\mathbb{P}(\{\omega : P_\omega \text{ fausse}\}) = 0$ ).

**Définition 3.4.1** On dit qu'une v.a.r. est dans  $\mathcal{L}^1(\Omega, \mathcal{B}, \mathbb{P})$  (on note aussi  $\mathcal{L}^1(\Omega, \mathbb{P})$  ou  $\mathcal{L}^1(\mathbb{P})$  ou simplement  $\mathcal{L}^1$ ) si l'espérance de la v.a.r. **positive**  $|X|$  est finie :  $\mathbb{E}(|X|) < \infty$ . Dans ce cas on dit que la v.a.r.  $X$  est ( $\mathbb{P}$ -)intégrable.

Si on pose  $X_+ = \max(0, X)$  et  $X_- = -\min(0, X)$ , on a  $X_+ \geq 0$ ,  $X_- \geq 0$

$$X = X_+ - X_-$$

et

$$|X| = X_+ + X_-,$$

et en particulier

$$0 \leq X_+ \leq |X|, \quad 0 \leq X_- \leq |X|.$$

Donc si  $X \in \mathcal{L}^1(\Omega, \mathcal{B}, \mathbb{P})$  on a aussi  $X_+, X_- \in \mathcal{L}^1(\Omega, \mathcal{B}, \mathbb{P})$ .

**Définition 3.4.2** Si  $X \in \mathcal{L}^1(\Omega, \mathcal{B}, \mathbb{P})$  on a pose

$$\mathbb{E}(X) = \mathbb{E}(X_+) - \mathbb{E}(X_-).$$

**Théorème 3.4.2 (Espérance : cas  $\mathcal{L}^1$ )** a) si  $A \in \mathcal{B}$  on a  $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$ .

b) Si  $X, Y$  sont des v.a dans  $\mathcal{L}^1(\Omega, \mathbb{P})$  telles que  $X \leq Y$  (c'est-à-dire pour tout  $\omega \in \Omega$   $X(\omega) \leq Y(\omega)$ ) alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .

c) Si  $X, Y$  sont des v.a dans  $\mathcal{L}^1(\Omega, \mathbb{P})$  et  $a, b \in \mathbb{R}$  on a (linéarité de l'espérance)

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

d) Si  $X \in \mathcal{L}^1(\Omega, \mathcal{B})$  on a toujours  $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$  et on a égalité si et seulement si  $\mathbb{P}$ -ps  $X \geq 0$  ou  $\mathbb{P}$ -ps  $X \leq 0$ .

## 3.5 Espérance des v.a.r. admettant une densité

### 3.5.1 Résultat fondamental

Dans la section précédente 3.4 nous avons vu comment étendre la notion d'espérance définie pour des v.a.r prenant un nombre fini ou dénombrable de valeurs à des v.a.r. positives ou  $\mathcal{L}^1$ . Dans cette section nous donnons une formule qui permet d'exprimer l'espérance d'une v.a.r. admettant une densité en fonction de la densité. Nous *admettrons* ainsi le théorème important suivant :

**Théorème 3.5.1** *Soit  $X$  une v.a. r.  $X$  admet une densité  $\rho_X$ .*

– a) *Si  $X$  est positive on peut toujours écrire*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \rho_X(x) dx.$$

– b) *Si  $X$  est de signe quelconque, on a  $X \in \mathcal{L}^1(\Omega, \mathcal{B}, \mathbb{P})$  si et seulement si*

$$\int_{-\infty}^{\infty} |x| \rho_X(x) dx < \infty$$

*et dans ce cas*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \rho_X(x) dx.$$

**Exemple** Soit  $X$  une v.a suivant une loi normale centrée réduite : Une telle loi admet une densité

$$\rho_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Comme

$$\mathbb{E}(|X|) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |x| e^{-x^2/2} dx$$

est finie  $X$  est dans  $\mathcal{L}^1(\Omega, \mathcal{B}, \mathbb{P})$ . L'espérance de  $X$  est donc définie et vaut

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Il s'agit de l'intégrale d'une fonction intégrable, impaire sur un intervalle symétrique : cette intégrale est donc nulle.

### 3.5.2 La formule de transfert

Étant donnée une v.a.r.  $X$  admettant une densité  $\rho_X$  et une fonction continue (ou continue par morceaux) on se propose de déterminer l'espérance de la v.a.r.  $Y := f(X)$ . Nous *admettrons* le résultat suivant :

**Théorème 3.5.2** *Si  $X : \Omega \rightarrow \mathbb{R}$  est une v.a admettant une densité  $\rho_X$  et  $f : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction continue (ou continue par morceaux) alors la v.a  $Y = f(X)$  est  $\mathbb{P}$ -intégrable si et seulement si l'intégrale  $\int_{-\infty}^{\infty} |f(x)|\rho_X(x)dx$  est finie et dans ce cas l'espérance de  $Y = f(X)$  est donnée par*

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(x)\rho_X(x)dx.$$

**Exercice 3.5.1** *Calculer  $\mathbb{E}(Y)$  où  $Y = X^2$  est le carré d'une v.a suivant une loi normale centrée réduite.*

**Solution** On a

$$\begin{aligned} \mathbb{E}(X^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(xe^{-x^2/2}) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \frac{d}{dx} (-e^{-x^2/2}) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \\ &= 1 \end{aligned}$$

(on a effectué une intégration par parties). □

**Remarque :**

Dans le cas général des v.a de la forme  $Y = f(X)$  où  $X : \Omega \rightarrow \mathbb{R}$  est une v.a et  $f : \mathbb{R} \rightarrow \mathbb{R}$  est continue, la forme générale de la formule de transfert fait intervenir la *loi* de  $X$  :

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

toutes les fois où l'intégrale converge.

### 3.5.3 Application au calcul de densité

Le problème qui nous intéresse dans cette section est le suivant : Étant donnée une v.a  $X$  dont on connaît la densité  $\rho_X$ , déterminer la densité, si elle existe de la v.a  $Y = f(X)$ , où  $f$  est une fonction continue de  $\mathbb{R}$  dans  $\mathbb{R}$ .



Le résultat qui va nous permettre d'aborder cette question est le théorème suivant que nous admettrons :

**Théorème 3.5.3** *Si  $X$  est une v.a telle que pour toute fonction continue bornée  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  on a*

$$\mathbb{E}(\phi(X)) = \int_{-\infty}^{\infty} \phi(x)\rho_X(x)dx,$$

alors  $X$  admet  $\rho_X$  pour densité.

### La méthode pour calculer la densité de $Y = f(X)$

Supposons que  $Y$  admette une densité  $\rho_Y$ . On doit alors avoir pour toute fonction continue bornée  $\phi : \mathbb{R} \rightarrow \mathbb{R}$

$$\mathbb{E}(\phi(Y)) = \int_{-\infty}^{\infty} \phi(y)\rho_Y(y)dy.$$

Mais  $\phi(Y) = \phi(f(X)) = \phi \circ f(X)$  et on a donc,

$$\mathbb{E}(\phi(Y)) = \mathbb{E}(\phi \circ f(X)) = \int_{-\infty}^{\infty} \phi \circ f(x)\rho_X(x)dx.$$

Supposons que  $\phi$  soit une bijection dérivable de  $\mathbb{R}$  dans  $\mathbb{R}$  envoyant  $\mathbb{R}$  sur  $\mathbb{R}$ . La formule classique de changement de variable montre que ( $x = f^{-1}(y)$ ,  $dx = 1/f'(f^{-1}(y))dy$ )

$$\int_{-\infty}^{\infty} \phi \circ f(x)\rho_X(x)dx = \int_{-\infty}^{\infty} \phi(y)\frac{1}{|f'(f^{-1}(y))|}\rho_X(f^{-1}(y))dy.$$

En conclusion, pour toute fonction  $\phi$  continue de  $\mathbb{R} \rightarrow \mathbb{R}$

$$\int_{-\infty}^{\infty} \phi(y)\rho_Y(y)dy = \int_{-\infty}^{\infty} \phi(y)\frac{1}{|f'(f^{-1}(y))|}\rho_X(f^{-1}(y))dy$$

et il est naturel de penser que

$$\rho_Y(y) = \frac{1}{|f'(f^{-1}(y))|}\rho_X(f^{-1}(y)). \quad (3.2)$$

Justifions le fait que c'est effectivement le cas. Le calcul précédent montre que pour toute fonction continue bornée  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  on a

$$\mathbb{E}(\phi(Y)) = \int_{-\infty}^{\infty} \phi(y)\frac{1}{|f'(f^{-1}(y))|}\rho_X(f^{-1}(y))dy.$$

Mais d'après le théorème 3.5.3 cela implique que  $Y$  admet une densité et que celle-ci est donnée par la formule (3.2).

On pourrait démontrer de la même manière :

**Théorème 3.5.4** Soient  $X$  une v.a de densité  $\rho_X$  prenant ses valeurs dans un intervalle  $I$  (fini ou infini) et  $f : I \rightarrow J$  est une application de classe  $C^1$  pas nécessairement bijective mais telle que tout point  $y \in J$  ait un nombre fini d'antécédents. Alors, la v.a  $Y = f(X)$  admet une densité  $\rho_Y$  dont l'expression est donnée par

$$\rho_Y(y) = \sum_{x \in f^{-1}(y)} \frac{\rho_X(x)}{|f'(x)|} \cdot \mathbf{1}_J.$$

L'expression précédente peut prendre la valeur  $\infty$  mais la fonction positive  $\rho_Y$  restera d'intégrale 1.

Appliquons ce qui précède à un exemple.

**Exercice** Supposons que  $X$  admette une densité  $\rho_X$ . Déterminer la densité, si elle existe de  $Y = X^2$ . Application au cas où  $X$  suit une loi normale  $N(0, 1)$ .

**Solution** On a  $Y = f(X)$  où  $f(x) = x^2$  est une bijection de  $I_- = ]-\infty, 0[$  sur  $]0, \infty[$  et de  $I_+ = ]0, \infty[$  sur  $]0, \infty[$  ( $f$  est une fonction continue strictement décroissante sur  $I_- = ]-\infty, 0[$  et strictement croissante sur  $I_+ = ]0, \infty[$ .) Pour toute fonction  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  continue et bornée

$$\mathbb{E}(\phi(Y)) = \mathbb{E}(\phi \circ f(X)) = \int_{-\infty}^{\infty} \phi(f(x))\rho_X(x)dx.$$

Ecrivons,

$$\int_{-\infty}^{\infty} \phi(f(x))\rho_X(x)dx = \int_{-\infty}^0 \phi(x^2)\rho_X(x)dx + \int_0^{\infty} \phi(x^2)\rho_X(x)dx$$

et effectuons dans chacune des intégrales du membre de droite le changement de variable  $y = x^2$  :

$$\int_{-\infty}^0 \phi(x^2)\rho_X(x)dx = \int_0^{\infty} \phi(y)\rho(-\sqrt{y})\frac{dy}{2\sqrt{y}},$$

$$\int_0^{\infty} \phi(x^2)\rho_X(x)dx = \int_0^{\infty} \phi(y)\rho(\sqrt{y})\frac{dy}{2\sqrt{y}},$$

si bien que

$$\begin{aligned} \mathbb{E}(\phi(Y)) &= \int_0^{\infty} \phi(y) \left( \rho(-\sqrt{y}) + \rho(\sqrt{y}) \right) \frac{1}{2\sqrt{y}} dy \\ &= \int_{\mathbb{R}} \phi(y) \left( \rho(-\sqrt{y}) + \rho(\sqrt{y}) \right) \frac{1}{2\sqrt{y}} \cdot \mathbf{1}_{]0, \infty[}(y) dy \end{aligned}$$

et comme cette formule est vraie pour toute fonction  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  continue bornée, on peut conclure que  $Y$  admet une densité  $\rho_Y$  égale à

$$\rho_Y(y) = \left( \rho(-\sqrt{y}) + \rho(\sqrt{y}) \right) \frac{1}{2\sqrt{y}} \mathbf{1}_{]0, \infty[}(y).$$

(Ne pas oublier le terme  $\mathbf{1}_{]0, \infty[}(y)$ .)

Si  $X$  suit une loi normale  $N(0, 1)$  sa densité est  $\rho_X(x) = (1/\sqrt{2\pi})e^{-(x^2/2)}$  et la densité de  $Y = X^2$  vaut

$$\rho_Y(y) = \frac{1}{\sqrt{2\pi}} \frac{e^{-y}}{\sqrt{y}} \mathbf{1}_{]0, \infty[}(y).$$

**Exercice** Si  $X$  suit une loi normale centrée réduite, déterminer la loi de  $Y = \sigma X + \mu$ .

*Solution* On a  $Y = f(X)$  où  $f(x) = ax + b$ . C'est une bijection  $C^1$  de  $\mathbb{R}$  sur  $\mathbb{R}$ . La formule du théorème 3.5.4 montre donc que la densité de  $Y$  est

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2},$$

c'est-à-dire est une loi  $N(0, \sigma^2)$ .

## 3.6 Variance

### 3.6.1 Variables aléatoires de carré intégrable

Commençons par une définition.

**Définition 3.6.1** On dit qu'une v.a  $X$  est dans  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  si son carré est  $\mathbb{P}$ -intégrable,

$$\mathbb{E}(|X|^2) < \infty.$$

Une des propriétés importantes de ces espaces  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  est qu'ils sont stables par combinaisons linéaires :

**Théorème 3.6.1** L'espace  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  est un  $\mathbb{R}$ -espace vectoriel c'est-à-dire que si  $a, b \in \mathbb{R}$  et  $X, Y \in \mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  on a aussi  $aX + bY \in \mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$ . En outre, la propriété suivante est toujours vérifiée (inégalité de Minkowski) :

$$\mathbb{E}(|X + Y|^2)^{1/2} \leq \mathbb{E}(|X|^2)^{1/2} + \mathbb{E}(|Y|^2)^{1/2}.$$

Une propriété très utile des espaces  $\mathcal{L}^2$  est la propriété de Cauchy-Schwarz (qui permet entre autres choses de démontrer l'inégalité de Minkowski).

**Théorème 3.6.2 (Cauchy-Schwarz)** *Si  $X$  et  $Y$  sont des v.a.r dans  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  alors le produit  $XY$  est dans  $\mathcal{L}^1(\Omega, \mathcal{B}, \mathbb{P})$  et on a*

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|X|^2)^{1/2} \mathbb{E}(|Y|^2)^{1/2}$$

avec égalité si et seulement si  $X$  et  $Y$  sont colinéaires.

### 3.6.2 Variance

Nous pouvons à présent définir la variance d'une v.a.r. de carré intégrable.

Si  $X$  est dans  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  la v.a  $X - \mathbb{E}(X)$  est également dans  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  puisque c'est une somme de deux v.a de  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  (une v.a constante est toujours dans  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$ ).

**Définition 3.6.2** *La variance d'une v.a.r. dans  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  est la quantité*

$$\mathbf{Var}X := \mathbb{E}(|X - \mathbb{E}(X)|^2).$$

La racine carrée  $\sigma$  de ce nombre s'appelle l'écart-type de  $X$ .

La variance est donc la moyenne (l'espérance) des carrés des écarts de  $X$  par rapport à  $\mathbb{E}(X)$ . Elle mesure en quelque sorte le caractère plus ou moins diffus de la variable aléatoire  $X$ .

Le calcul suivant

$$\begin{aligned} \mathbb{E}(|X - \mathbb{E}(X)|^2) &= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \end{aligned}$$

démontre

**Proposition 3.6.1** *Si  $X \in \mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$ , on a  $\mathbf{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ .*

**Remarque** De façon plus générale on peut définir pour tout  $p \geq 1$  l'espace  $\mathcal{L}^p(\Omega, \mathcal{B}, \mathbb{P})$  des v.a  $X$  dont le moment d'ordre  $p$  est fini c'est-à-dire telles que  $\mathbb{E}(|X|^p) < \infty$ . Ce sont encore des espaces vectoriels et la quantité  $\mathbb{E}(|X|^p)^{1/p}$  définit une (semi-)norme sur  $\mathcal{L}^p(\Omega, \mathcal{B}, \mathbb{P})$  qui en fait un espace de Banach (un espace vectoriel normé complet).

### 3.6.3 Calculs de variance des v.a.r. à valeurs dans $\mathbb{N}$

#### Fonctions génératrices

Un outil très utile pour calculer les moments d'ordre  $p$  d'une v.a à valeurs dans  $\mathbb{N}$  est d'introduire la *fonction génératrice* de  $X$ .

**Définition 3.6.3** *La fonction génératrice d'une v.a  $X$  à valeurs dans  $\mathbb{N}$  est la fonction définie par*

$$\begin{aligned}\phi_X(t) &= \mathbb{E}(t^X) \\ &= \sum_{k \in \mathbb{N}} t^k \mathbb{P}(X = k).\end{aligned}$$

L'intérêt de  $\phi_X$  réside dans la proposition suivante :

**Proposition 3.6.2** *On a toujours*

$$\lim_{t \rightarrow 1, t < 1} \frac{d}{dt} \phi_X(t) = \mathbb{E}(X)$$

et de façon plus générale

$$\lim_{t \rightarrow 1, t < 1} \frac{d^p}{dt^p} \phi_X(t) = \mathbb{E}(X(X-1) \cdots (X-p+1)).$$

*Démonstration.* —

Afin de simplifier la démonstration nous supposons que  $X$  ne prend qu'un nombre fini de valeurs dans  $\{0, 1, \dots, N\}$ . Il suffit de calculer

$$\frac{d^p}{dt^p} \left( \sum_{k=0}^N t^k \mathbb{P}(X = k) \right) = \sum_{k=0}^N k(k-1) \cdots (k-p+1) t^{k-p} \mathbb{P}(X = k)$$

En faisant  $t = 1$  on obtient le résultat d'après la formule de transfert.

Dans le cas général où  $X$  prend ses valeurs dans  $\mathbb{N}$  on peut procéder de la façon suivante : pour  $0 \leq t < 1$ ,

$$\frac{d^p}{dt^p} \left( \sum_{k=0}^{\infty} t^k \mathbb{P}(X = k) \right) = \sum_{k=0}^{\infty} k(k-1) \cdots (k-p+1) t^{k-p} \mathbb{P}(X = k),$$

car les séries sont uniformément convergentes. La série du membre de droite converge quand  $t \rightarrow 1^-$  vers  $\sum_{k=0}^{\infty} k(k-1) \cdots (k-p+1) \mathbb{P}(X = k)$  d'après le théorème de convergence monotone et cette quantité égale  $\mathbb{E}(X \cdots (X-p+1))$  d'après le théorème de transfert.  $\square$

Le calcul de la variance est alors clair puisque

$$\begin{aligned}\mathbf{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2 \\ &= \phi_X''(1) + \phi_X'(1) - (\phi_X'(1))^2.\end{aligned}$$

### Exemples

**Loi géométrique** Dans ce cas  $X$  prend ses valeurs dans  $\mathbb{N}$  et  $P(X = k) = (1-a)a^k$  ( $0 \leq a < 1$ ). Ainsi

$$\phi_X(t) = (1-a) \sum_{k=0}^{\infty} t^k a^k = (1-a) \frac{1}{1-ta},$$

pourvu que  $0 \leq t < a^{-1}$ . On a

$$\phi_X'(t) = a \frac{1-a}{(1-ta)^2}, \quad \phi_X''(t) = 2a^2 \frac{1-a}{(1-ta)^3},$$

et donc

$$\mathbb{E}(X) = \frac{a}{1-a}, \quad \mathbf{Var}(X) = \frac{a}{(1-a)^2}.$$

**Loi binomiale** Si la v.a  $X$  suit une loi  $(p, n)$  elle prend ses valeurs dans  $\{0, 1, \dots, n\}$  et  $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ . On a

$$\phi_X(t) = \sum_{k=0}^n \binom{n}{k} t^k p^k (1-p)^{n-k} = (tp + 1 - p)^n,$$

si bien que

$$\phi_X'(t) = pn(tp + 1 - p)^{n-1}, \quad \phi_X''(t) = p^2 n(n-1)(tp + 1 - p)^{n-2},$$

et donc

$$\mathbb{E}(X) = np, \quad \mathbf{Var}(X) = np(1-p).$$

**Loi de Poisson** Si la v.a  $X$  suit une loi de Poisson de paramètre  $\lambda$ , elle prend ses valeurs dans  $\mathbb{N}$  et  $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ . Il vient

$$\phi_X(t) = e^{-\lambda} \sum_{k=0}^{\infty} t^k \lambda^k \frac{1}{k!} = e^{-\lambda} e^{t\lambda} = e^{(t-1)\lambda}.$$

Ainsi,

$$\phi'_X(t) = \lambda e^{(t-1)\lambda}, \quad \phi''_X(t) = \lambda^2 e^{(t-1)\lambda},$$

et donc

$$\mathbb{E}(X) = \lambda, \quad \mathbf{Var}(X) = \lambda.$$

Mentionnons enfin une propriété très utile des fonctions génératrices : une fonction génératrice caractérise de façon unique la loi de la variable aléatoire qui la définit.

**Proposition 3.6.3** *Soient  $X$  et  $Y$  deux v.a.r. à valeurs dans  $\mathbb{N}$  et supposons que pour tout  $0 \leq t < 1$  (ou même pour tout  $t \in ]a, b[ \subset [0, 1[$ ) on ait  $g_X(t) = g_Y(t)$ . Alors,  $X$  et  $Y$  ont même loi : pour tout  $k \in \mathbb{N}$ ,  $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$ .*

### 3.6.4 Cas des v.a admettant une densité

Si  $X$  est une v.a admettant une densité  $\rho_X$ , alors d'après la formule de transfert,  $X$  est dans  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  si et seulement si

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 \rho_X(x) dx,$$

est finie.

**Loi uniforme** La variable aléatoire  $X : \Omega \rightarrow \mathbb{R}$  suit une loi uniforme sur l'intervalle  $[a, b]$  si sa densité est donnée par

$$\rho_X(x) = \frac{1}{b-a} \cdot \mathbf{1}_{[a,b]}(x).$$

On a

$$\begin{aligned} \mathbb{E}(X) &= \int_{\mathbb{R}} x \cdot \frac{1}{b-a} \cdot \mathbf{1}_{[a,b]}(x) dx \\ &= \frac{1}{b-a} \int_a^b x dx \\ &= \frac{1}{b-a} \left[ \frac{x^2}{2} \right]_a^b \\ &= \frac{a+b}{2}, \end{aligned}$$

ce qui est conforme à l'intuition : la position en moyenne d'un point jeté au hasard sur l'intervalle  $(a, b)$  sera située au milieu de l'intervalle  $(a, b)$ .

Calculons la variance

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{\mathbb{R}} x^2 \cdot \frac{1}{b-a} \cdot \mathbf{1}_{[a,b]}(x) dx \\ &= \frac{1}{b-a} \int_a^b x^2 dx \\ &= \frac{1}{b-a} \frac{b^3 - a^3}{b-a} \\ &= \frac{a^2 + b^2 + ab}{3},\end{aligned}$$

( $b^3 - a^3 = (b-a)(b^2 + ab + a^2)$ ) et donc

$$\begin{aligned}\mathbf{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X) &= \frac{a^2 + b^2 + ab}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{(b-a)^2}{12}.\end{aligned}$$

### Exemples

**Loi exponentielle de paramètre  $\theta$**  La v.a admet une densité

$$\rho_X(x) = \theta e^{-\theta x} \mathbf{1}_{[0,\infty[}(x).$$

On a

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} x \theta e^{-\theta x} \mathbf{1}_{[0,\infty[}(x) dx \\ &= \int_0^{\infty} x \theta e^{-\theta x} dx \\ &= \frac{1}{\theta} \int_0^{\infty} u e^{-u} du \\ &= \frac{1}{\theta},\end{aligned}$$

(après changement de variable et intégration par parties). Le moment d'ordre 2 s'obtient de façon analogue,

$$\begin{aligned}\mathbb{E}(X^2) &= \int_0^{\infty} x^2 \theta e^{-\theta x} dx \\ &= \frac{1}{\theta} \int_0^{\infty} u e^{-u} du \\ &= \frac{2}{\theta^2},\end{aligned}$$



(après changement de variable et deux intégrations par parties). On a donc

$$\mathbf{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{\theta^2}.$$

**Loi normale**  $N(\mu, \sigma^2)$  On sait (cf. l'exercice de la section 3.5.3) que si  $Y$  suit une loi  $N(\mu, \sigma^2)$  alors elle est de la forme  $\sigma X + \mu$  où  $X$  suit une loi normale  $N(0, 1)$  de densité

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

On sait que pour une telle loi,  $\mathbb{E}(X) = 0$  et  $\mathbf{Var}(X) = 1$ . Par conséquent,  $\mathbb{E}(Y) = \sigma\mathbb{E}(X) + \mu = \mu$  et  $\mathbf{Var}(Y) = \sigma^2\mathbf{Var}(X) = \sigma^2$ .

### 3.7 Inégalité de Markov et de Bienaymé-Tchebychev

L'intérêt de considérer les moments d'une v.a  $X$  réside dans les propositions suivantes.

**Proposition 3.7.1** *Si  $X$  est une v.a dans  $\mathcal{L}^1(\Omega, \mathbb{P})$  on a pour tout  $\lambda > 0$*

$$\mathbb{P}(|X| > \lambda) \leq \frac{\mathbb{E}(|X|)}{\lambda}.$$

*Démonstration.* — La v.a  $|X|$  peut s'écrire

$$|X| = |X| \cdot \mathbf{1}_{\{|X| > \lambda\}} + |X| \cdot \mathbf{1}_{\{|X| \leq \lambda\}},$$

et par additivité et positivité de l'espérance on a

$$\mathbb{E}(|X|) \geq \mathbb{E}(|X| \cdot \mathbf{1}_{\{|X| > \lambda\}}).$$

Or,

$$|X| \cdot \mathbf{1}_{\{|X| > \lambda\}} \geq \lambda \cdot \mathbf{1}_{\{|X| > \lambda\}},$$

et par conséquent

$$\mathbb{E}(|X|) \geq \lambda \mathbb{E}(\mathbf{1}_{\{|X| > \lambda\}}),$$

c'est-à-dire

$$\mathbb{E}(|X|) \geq \lambda \mathbb{P}(\{|X| > \lambda\}),$$

ce qui est la conclusion de la proposition.  $\square$

La proposition précédente est une version quantitative du fait que la probabilité que  $X$  prenne de grandes valeurs a tendance à être petite.

Si on a des informations sur les moments d'ordre supérieurs l'estimation précédente est meilleure :

**Proposition 3.7.2** *Si  $X$  est une v.a dans  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  on a pour tout  $\lambda > 0$*

$$\mathbb{P}(|X| > \lambda) \leq \frac{\mathbb{E}(|X|^2)}{\lambda^2}.$$

*Démonstration.* —

Il suffit de remarquer que  $\{X > \lambda\} = \{X^2 > \lambda^2\}$  et d'appliquer la proposition précédente à la v.a  $Y = X^2$ .  $\square$

Appliquée à la v.a  $Y = X - \mathbb{E}(X)$  la proposition précédente donne le théorème de Bienaymé-Tchebychev :

**Théorème 3.7.1** *Si  $X$  est une v.a dans  $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$  on a pour tout  $\lambda > 0$*

$$\mathbb{P}(|X - \mathbb{E}(X)| > \lambda) \leq \frac{\mathbf{Var}(|X|)}{\lambda^2}.$$

*Si on note  $\sigma = \sqrt{\mathbf{Var}(X)}$  l'écart type on a donc*

$$\mathbb{P}(|X - \mathbb{E}(X)| > \lambda\sigma) < \frac{1}{\lambda^2},$$

*(d'où le nom d'écart type donné à  $\sigma$ )*

Le théorème de Bienaymé-Tchebychev permet d'obtenir les probabilités des déviations importantes de la v.a  $X$  par rapport à sa moyenne.

## 3.8 Vecteurs aléatoires

Un vecteur aléatoire est un  $n$ -uplet de variables aléatoires réelles  $X = (X_1, \dots, X_n)$ .

### 3.8.1 Loi d'un vecteur aléatoire

**Définition 3.8.1** *Si  $X_1, \dots, X_n$  sont des v.a, on appelle loi du vecteur aléatoire  $X = (X_1, \dots, X_n)$  la donnée de tous les quantités*

$$\mathbb{P}(X_1 \in I_1, \dots, X_n \in I_n)$$

*pour tous intervalles  $I_1, \dots, I_n$  de  $\mathbb{R}$ .*

Dans la suite nous distinguerons deux cas, suivant que les  $X_i$  prennent des valeurs dans un ensemble fini ou dénombrable ou sont des v.a.r. ayant une densité.

### Cas des v.a. prenant un nombre fini ou dénombrable de valeurs

Si chacune des v.a.  $X_i$ , prend ses valeurs dans un ensemble  $E_i$  fini ou dénombrable ( $i = 1, \dots, n$ ), la loi du vecteur  $X = (X_1, \dots, X_n)$  est déterminée par les nombres

$$\mathbb{P}(X_1 = e_1, \dots, X_n = e_n),$$

où  $(e_1, \dots, e_n) \in E_1 \times \dots \times E_n$ .

### Cas des v.a. admettant une densité

Si chacune des v.a.  $X_i$  admet une densité, il est possible de démontrer que le vecteur  $X = (X_1, \dots, X_n)$  admet une densité, c'est-à-dire qu'il existe une fonction  $\rho(x_1, \dots, x_n)$  telle que pour tous intervalles  $I_1, \dots, I_n$

$$\mathbb{P}(X_1 \in I_1, \dots, X_n \in I_n) = \int_{I_1 \times \dots \times I_n} \rho(x_1, \dots, x_n) dx_1 \cdots dx_n. \quad (3.3)$$

Réciproquement, si le vecteur aléatoire  $X = (X_1, \dots, X_n)$  admet une densité  $\rho$  au sens de (3.3) alors chacune de v.a.r.  $X_i$  admet une densité  $\rho_i$  telle que pour tout intervalle  $I$  de  $\mathbb{R}$  on a

$$\mathbb{P}(X_i \in I) = \int_{\mathbb{R} \times \dots \times \mathbb{R}} \rho(x_1, \dots, x_i, \dots, x_n) dx_1 \cdots \widehat{dx_i} \cdots dx_n \quad (3.4)$$

où le chapeau signifie que l'on n'intègre pas sur la variable  $x_i$ .

### 3.8.2 Formules de transfert

Etant donné un vecteur aléatoire  $X = (X_1, \dots, X_m)$  et

$$\begin{aligned} f : \mathbf{R}^m &\rightarrow \mathbf{R}^p \\ (x_1, \dots, x_m) &\mapsto (f_1(x_1, \dots, x_m), \dots, f_p(x_1, \dots, x_m)) \end{aligned}$$

une application continue, on cherche à calculer l'espérance de  $Y = f(X)$  c'est-à-dire que l'on veut calculer le vecteur  $E(Y) = (E(Y_1), \dots, E(Y_p))$  où  $Y_j = f_j(X_1, \dots, X_m)$ . Le problème se ramène donc au suivant : étant donné  $X = (X_1, \dots, X_m)$  un vecteur aléatoire calculer  $E(\phi(X_1, \dots, X_m))$  où  $\phi$  est une application de  $\mathbf{R}^m$  dans  $\mathbf{R}$ . Ce problème se résout de la même façon que dans le cas  $m = 1$  (variables aléatoires). Nous donnons les résultats sans démonstration (les preuves sont identiques à celles du cas  $m = 1$ ).

**Cas de Vecteurs aléatoires à valeurs dans ensembles finis ou dénombrables**

Si  $X = (X_1, \dots, X_m)$  prend ses valeurs dans  $E_1 \times \dots \times E_m$  on a

$$\mathbb{E}(\phi(X_1, \dots, X_m)) = \sum_{e_1 \in E_1, \dots, e_m \in E_m} \phi(e_1, \dots, e_m) \mathbb{P}(X_1 = e_1, \dots, X_m = e_m).$$

**Cas de Vecteurs aléatoires admettant une densité** Si le vecteur aléatoire  $X = (X_1, \dots, X_m)$  admet une densité  $\rho(x_1, \dots, x_m)$  on a

$$\mathbb{E}(\phi(X_1, \dots, X_m)) = \int_{\mathbf{R}} \dots \int_{\mathbf{R}} \phi(x_1, \dots, x_m) \rho_X(x_1, \dots, x_m) dx_1 \dots dx_m.$$

**3.8.3 Loi de d'une somme de v.a.r.**

**Théorème 3.8.1** Soient  $X$  et  $Y$  deux v.a.r.

- Si les v.a.r.  $X, Y$  prennent leurs valeurs dans des ensembles finis ou dénombrables et si  $p$  est la loi du vecteur aléatoire  $(X, Y)$ , alors la loi  $p_{X+Y}$  de  $X + Y$  vérifie pour tout  $e$

$$p_{X+Y}(e) = \sum_{e_1 + e_2 = e} p(e_1, e_2).$$

- Si le vecteur aléatoire  $(X, Y)$  admet pour densité  $\rho$  alors la v.a.r.  $X + Y$  admet pour densité  $\rho_{X+Y}$  :

$$\rho_{X+Y}(x) = \int_{\mathbf{R}} \rho_X(x - y, y) dy.$$

*Démonstration.* — Faisons la preuve dans le second cas. i) Introduisons le vecteur aléatoire  $Z = (X + Y, Y)$  et calculons sa loi. Pour toute fonction  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$  continue bornée,

$$\mathbb{E}(\phi(Z)) = \mathbb{E}(\phi(X + Y, Y)) = \mathbb{E}(\psi(X, Y)),$$

où  $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$  égale à  $\psi(x, y) = \phi(x + y, y)$ . Comme  $(X, Y)$  admet une densité  $\rho$  on a d'après la formule de transfert

$$\begin{aligned} \mathbb{E}(\psi((X, Y))) &= \int_{\mathbb{R}^2} \psi(x, y) \rho(x, y) dx dy \\ &= \int_{\mathbb{R}^2} \phi(x + y, y) \rho(x, y) dx dy \\ &= \int_{\mathbf{R}} \left( \int_{\mathbf{R}} \phi(x + y, y) \rho(x, y) dx \right) dy. \end{aligned}$$

Effectuons le changement de variables ( $y$  étant fixé)  $u = x + y$ ,  $x = u - y$  dans l'intégrale du milieu :

$$\mathbb{E}(\psi((X, Y))) = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \phi(u, y) \rho(u - y, y) du \right) dy$$

et donc,

$$\mathbb{E}(\phi(Z)) = \int_{\mathbb{R}^2} \phi(u, v) \rho(u - v, v) dudv.$$

Cela étant vrai pour toute fonction  $\phi$  continue bornée on peut dire que  $Z = (U, V) = (X + Y, Y)$  admet une densité donnée par  $\rho_Z(u, v) = \rho(u - v, v)$ .

ii) La formule (3.4) montre que la densité de  $X + Y$  est  $\rho_X$  :

$$\rho_X(x) = \int_{-\infty}^{\infty} \rho(x - v, v) dv.$$

□

## 3.9 Variables aléatoires indépendantes

### 3.9.1 Définition

La notion d'indépendance de  $n$  v.a.r. est la suivante :

**Définition 3.9.1** Une suite  $X_1, \dots, X_n$  de variables aléatoires est dite indépendante si pour tous intervalles  $I_1, \dots, I_n$  de  $\mathbf{R}$ ,

$$\mathbb{P}(X_1 \in I_1, \dots, X_n \in I_n) = \mathbb{P}(X_1 \in I_1) \cdots \mathbb{P}(X_n \in I_n).$$

**Définition 3.9.2** Une famille  $(X_i)_{i \in A}$  de v.a est dite indépendante si toute sous famille finie  $(X_i)_{i \in J}$ ,  $J \subset A$  fini est indépendante.

### 3.9.2 Cas des v.a à valeurs dans un ensemble discret

Si l'on suppose que les  $X_i$  sont à valeurs dans  $E_i$  ensembles fini ou dénombrables la définition précédente se simplifie car on a la proposition suivante dont la preuve est facile :

**Proposition 3.9.1** La famille  $(X_1, \dots, X_n)$  est indépendante si et seulement si pour tout  $(e_1, \dots, e_n) \in E_1 \times \dots \times E_n$  on a

$$\mathbb{P}(X_1 = e_1, \dots, X_n = e_n) = \mathbb{P}(X_1 = e_1) \cdots \mathbb{P}(X_n = e_n).$$

**Définition 3.9.3** Si on note  $p_{X_i}(e) := \mathbb{P}(X_i = e)$  et  $p(e_1, \dots, e_n) = \mathbb{P}(X_1 = e_1, \dots, X_n = e_n)$  on dit que  $p$  est la loi du vecteur aléatoire  $(X_1, \dots, X_n)$ . Ainsi, si  $X_1, \dots, X_n$  sont indépendantes on a

$$p(e_1, \dots, e_n) = p_{X_1}(e_1) \cdots p_{X_n}(e_n).$$

Donnons un exemple important.

**Le jeu de Pile/Face infini** Ici,  $\Omega = \{0, 1\}^{\mathbb{N}}$  muni de la tribu  $\mathcal{B}$  engendrée par les événements élémentaires (les cylindres)  $C_{i,e} = X_i^{-1}(e)$ ,  $e \in \{0, 1\}$ ,  $i \in \mathbb{N}$  (où les  $X_i$  sont définies par  $X_i(\omega) = \omega_i$ ) et de la probabilité  $\mathcal{P}$  décrite en 2.5.3. Pour toute sous famille  $X_{i_1}, \dots, X_{i_n}$  de  $X_1, X_2, \dots$ , on a comme précédemment

$$\begin{aligned} \mathbb{P}(X_{i_1} = \epsilon_1, \dots, X_{i_n} = \epsilon_n) &= \mathbb{P}(\{\omega : \omega_1 = \epsilon_1, \dots, \omega_n = \epsilon_n\}) \\ &= \mathbb{P}(\{\omega : \omega \in C_{i_1, \epsilon_1} \cap \cdots \cap C_{i_n, \epsilon_n}\}) \\ &= p^{\sum_{l=1}^n \epsilon_l} (1-p)^{n - \sum_{l=1}^n \epsilon_l} \end{aligned}$$

par définition de la probabilité  $\mathbb{P}$ . Par ailleurs,

$$\begin{aligned} \mathbb{P}(X_{i_1} = \epsilon_1) \cdots \mathbb{P}(X_{i_n} = \epsilon_n) &= p^{\epsilon_1} (1-p)^{1-\epsilon_1} \cdots p^{\epsilon_n} (1-p)^{1-\epsilon_n} \\ &= p^{\sum_{l=1}^n \epsilon_l} (1-p)^{n - \sum_{l=1}^n \epsilon_l} \\ &= \mathbb{P}(X_{i_1} = \epsilon_1, \dots, X_{i_n} = \epsilon_n). \end{aligned}$$

On a donc bien pour tous  $\epsilon_1, \dots, \epsilon_n$  l'égalité  $\mathbb{P}(X_{i_1} = \epsilon_1, \dots, X_{i_n} = \epsilon_n) = \mathbb{P}(X_{i_1} = \epsilon_1) \cdots \mathbb{P}(X_{i_n} = \epsilon_n)$  ce qui démontre l'indépendance de toute sous-famille  $X_{i_1}, \dots, X_{i_n}$ .

Nous avons donc démontré que la famille  $X_1, X_2, \dots$  est indépendante.

**Remarque** Le résultat précédent est beaucoup plus profond qu'il n'apparaît. La vérification de l'indépendance des v.a  $X_i$  est facile mais la construction de la mesure  $\mathbb{P}$ , qui est justement celle qui rend la famille  $(X_i)_i$  indépendante, est loin d'être triviale.

Il existe en fait un résultat plus général : Etant donnée une famille  $(X_\alpha)_{\alpha \in A}$  de v.a à valeurs dans  $\mathbb{R}$  définies sur un espace probabilisé  $(\Omega', \mathcal{B}', \mathbb{P}')$ , il existe un espace probabilisé  $(\Omega, \mathcal{B}, \mathbb{P})$  et des v.a  $Y_\alpha$ ,  $\alpha \in A$  telles que pour tout  $\alpha \in A$ ,  $Y_\alpha$  a même loi que  $X_\alpha$  et telles que la famille  $(Y_\alpha)_{\alpha \in A}$  soit indépendante.

Ainsi, la construction de v.a de lois données qui sont indépendantes est toujours possible.

### 3.9.3 Cas des v.a admettant des densités

On a la proposition suivante

**Proposition 3.9.2** *Si les  $X_1, \dots, X_n$  sont des v.a indépendantes admettant des densités  $\rho_{X_1}, \dots, \rho_{X_n}$  alors pour tous intervalles  $I_1, \dots, I_n$  de  $\mathbb{R}$  on a*

$$\mathbb{P}(X_1 \in I_1, \dots, X_n \in I_n) = \int_{I_1 \times \dots \times I_n} \rho(x_1, \dots, x_n) dx_1 \cdots dx_n$$

où

$$\rho(x_1, \dots, x_n) = \rho_{X_1}(x_1) \cdots \rho_{X_n}(x_n).$$

On dit que  $\rho$  est la densité du vecteur aléatoire  $X = (X_1, \dots, X_n)$ .

*Démonstration.* —

En effet pour tous intervalles  $I_1, \dots, I_n$ ,

$$\begin{aligned} \mathbb{P}(X \in I_1 \times \dots \times I_n) &= \mathbb{P}(X_1 \in I_1) \cdots \mathbb{P}(X_n \in I_n) \\ &= \left( \int_{I_1} \rho_{X_1}(x_1) dx_1 \right) \cdots \left( \int_{I_n} \rho_{X_n}(x_n) dx_n \right) \\ &= \int_{I_1 \times \dots \times I_n} \rho_{X_1}(x_1) \cdots \rho_{X_n}(x_n) dx_1 \cdots dx_n. \end{aligned}$$

□

### 3.9.4 Loi d'une somme de v.a.r. indépendante

Nous pouvons à présent calculer la densité d'une suite de v.a.r. indépendante en utilisant les résultats des sous-sections précédentes et le théorème 3.8.1

**Théorème 3.9.1** *Soient  $X$  et  $Y$  deux v.a.r. indépendantes*

- *Si  $X$  et  $Y$  sont des v.a.r. prenant leurs valeurs dans des ensembles finis ou dénombrables et admettent respectivement pour lois  $p_X$  et  $p_Y$  alors la loi  $p_{X+Y}$  de  $X + Y$  vérifie pour tout  $e$*

$$p_{X+Y}(e) = \sum_{e_1+e_2=e} p_X(e_1)p_Y(e_2).$$

- *Si  $X$  et  $Y$  sont des v.a.r. admettant respectivement pour densités  $\rho_X$  et  $\rho_Y$  alors la v.a.r.  $X + Y$  admet pour densité  $\rho_{X+Y}$  :*

$$\rho_{X+Y}(x) = \int_{\mathbb{R}} \rho_X(x-y)\rho_Y(y)dy.$$

### 3.9.5 Espérance des produits de v.a indépendantes

Le théorème fondamental de cette section est le suivant :

**Théorème 3.9.2** *Si  $X_1, \dots, X_n$  est une famille de v.a indépendantes et dans  $\mathcal{L}^1(\Omega, \mathcal{B}, \mathbb{P})$  le produit  $Y = X_1 \cdots X_n$  est également une v.a dans  $\mathcal{L}^1(\Omega, \mathcal{B}, \mathbb{P})$  et son espérance égale le produit des espérance des  $X_i$  :*

$$\mathbb{E}(X_1 \cdots X_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n).$$

*Démonstration.* — Donnons une preuve de ce résultat quand les  $X_i$  sont à valeurs dans un ensemble fini ou dénombrable.

Traitons le cas  $n = 2$  le cas général se faisant de la même façon

Supposons que  $X_1, X_2$  sont à valeurs dans un ensemble fini  $E$ . Notons  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  l'application définie par  $f(x_1, x_2) = x_1 \cdot x_2$ . On a d'après la formule de transfert

$$\begin{aligned} \mathbb{E}(f(X_1 \cdot X_2)) &= \sum_{(e_1, e_2) \in E \times E} f(e_1, e_2) \mathbb{P}(X_1 = e_1, X_2 = e_2) \\ &= \sum_{(e_1, e_2) \in E \times E} e_1 e_2 \mathbb{P}(X_1 = e_1) \mathbb{P}(X_2 = e_2) \\ &= \left( \sum_{e_1 \in E} e_1 \mathbb{P}(X_1 = e_1) \right) \left( \sum_{e_2 \in E} e_2 \mathbb{P}(X_2 = e_2) \right) \\ &= \mathbb{E}(X_1) \mathbb{E}(X_2). \end{aligned}$$

□

**Remarque** Dans le cas où les v.a  $X_i$  admettent des densités (continues)  $\rho_{X_i}$  on peut donner la preuve suivante. La formule de transfert appliquée à  $Y = f(X_1, X_2) = X_1 \cdot X_2$  donne

$$\mathbb{E}(X_1 \cdot X_2) = \int_{\mathbb{R} \times \mathbb{R}} f(x_1, x_2) \rho_X(x_1, x_2) dx_1 dx_2.$$

Or on sait que  $\rho_X(x_1, x_2) = \rho_{X_1}(x_1) \rho_{X_2}(x_2)$  si bien que

$$\begin{aligned} \mathbb{E}(X_1 \cdot X_2) &= \int_{\mathbb{R} \times \mathbb{R}} x_1 x_2 \rho_{X_1}(x_1) \rho_{X_2}(x_2) dx_1 dx_2, \\ &= \int_{\mathbb{R}} x_1 \rho_{X_1}(x_1) \int_{\mathbb{R}} x_2 \rho_{X_2}(x_2) \\ &= \mathbb{E}(X_1) \mathbb{E}(X_2). \end{aligned}$$

**Remarque** Attention, la réciproque du résultat précédent est fausse : si deux v.a  $X, Y$  sont telles que  $\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y)$  on ne peut pas conclure que  $X, Y$  sont indépendantes. En revanche, le résultat suivant est vrai :



**Théorème 3.9.3** *Si  $X_1, \dots, X_n$  sont des v.a telles que pour toutes fonctions continues bornées  $\phi_1, \dots, \phi_n$  de  $\mathbb{R} \rightarrow \mathbb{R}$  on a*

$$\mathbb{E}(\phi_1(X_1) \cdots \phi_n(X_n)) = \mathbb{E}(\phi_1(X_1)) \cdots \mathbb{E}(\phi_n(X_n)),$$

*alors la famille de v.a  $X_1, \dots, X_n$  est indépendante.*

### 3.9.6 Critères d'indépendance

Dans la pratique il est important de déterminer si une famille de v.a est indépendante. Un cas qui se présente fréquemment est le suivant : on suppose donnée une famille de v.a indépendantes  $X_1, X_2, \dots$  et on construit, à partir des  $X_i$ , de nouvelles v.a  $Y_1, Y_2, \dots$ . Par exemple, on peut définir  $Y_1 = X_1$ ,  $Y_2 = X_1 + X_2$ ,  $Y_n = X_1 + \dots + X_n$ ...mais on pourrait définir les  $Y_i$  par  $Y_1 = X_1 + X_2$ ,  $Y_2 = X_2 + X_3$ ,  $Y_n = X_n + X_{n+1}$  etc. ou faire des choses plus compliquées. On se propose alors de savoir si la famille ainsi construite est indépendante. Le théorème qui suit permet dans certains cas de répondre à cette question.

**Théorème 3.9.4** *Soient  $(X_i)_{i \in \mathbb{N}}$  une famille de v.a indépendantes et  $(J_i)$  ( $i = 1, 2, \dots$ ) des sous ensembles finis de  $\mathbb{N}$  qui forment une partition de  $\mathbb{N}$  (i.e les  $J_i$  sont non vides,  $\mathbb{N} = \cup_{i \geq 1} J_i$  et  $J_i \cap J_j = \emptyset$  si  $i \neq j$ ). Supposons données des applications (continues, continues par morceaux,...)  $f_i$  de  $\mathbb{R}^{\#J_i} \rightarrow \mathbb{R}$  et posons  $Y_i = f_i(X_{i_1}, \dots, X_{i_{\#J_i}})$  (où  $i_1 < \dots < i_{\#J_i}$  sont les éléments de  $J_i$ ). Alors, la famille de v.a  $Y_1, Y_2, \dots$  est indépendante.*

Ainsi, si la famille de v.a.r.  $X_1, X_2, \dots$  est indépendante, il en est de même de  $X_1 + X_2, X_3 + X_4, \dots, X_{2n-1} + X_{2n}, \dots$ . En revanche, la suite de v.a.r.  $X_1 + X_2, X_2 + X_3, X_3 + X_4, \dots$  ne sera en général pas indépendante.

### 3.9.7 Variance d'une somme de v.a.r. indépendantes

Si  $(X_i)_{i \in \mathbb{N}}$  est une famille de v.a.r. on a toujours

$$\mathbb{E}(X_1 + \dots + X_n) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n).$$

Si en outre on suppose la famille indépendante on peut calculer facilement la variance de  $X_1 + \dots + X_n$ .

**Théorème 3.9.5** *Soit  $(X_i)_{i \in \mathbb{N}}$  une famille de v.a.r. indépendantes. Alors,*

$$\mathbf{Var}(X_1 + \dots + X_n) = \mathbf{Var}(X_1) + \dots + \mathbf{Var}(X_n).$$

*Démonstration.* — Soit  $S := X_1 + \cdots + X_n$ ; il suffit de calculer  $\mathbf{Var}(S) = \mathbb{E}(S^2) - (\mathbb{E}(S))^2$ . On a

$$\begin{aligned}\mathbb{E}(S^2) &= \mathbb{E}\left(\sum_{1 \leq i, j \leq n} X_i \cdot X_j\right) \\ &= \sum_{1 \leq i, j \leq n} \mathbb{E}(X_i X_j).\end{aligned}$$

Or quand  $i \neq j$  on a d'après l'hypothèse d'indépendance  $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j)$ . On a donc en décomposant la somme précédente en  $i = j$  et  $i \neq j$

$$\mathbb{E}(S^2) = \sum_{i=1}^n \mathbb{E}(X_i^2) + \sum_{1 \leq i \neq j \leq n} \mathbb{E}(X_i)\mathbb{E}(X_j).$$

D'autre part,

$$(\mathbb{E}(S))^2 = \sum_{1 \leq i, j \leq n} \mathbb{E}(X_i)\mathbb{E}(X_j).$$

On a donc

$$\begin{aligned}\mathbf{Var}(S) &= \mathbb{E}(S^2) - (\mathbb{E}(S))^2 \\ &= \sum_{i=1}^n \mathbb{E}(X_i^2) - \sum_{i=1}^n \mathbb{E}(X_i)^2 \\ &= \sum_{i=1}^n \mathbf{Var}(X_i).\end{aligned}$$

□

**Exercice 3.9.1** a) Calculer l'espérance et la variance d'une v.a.r. suivant une loi de Bernoulli  $\mathbb{P}(X = 1) = p$ ,  $\mathbb{P}(X = 0) = 1 - p$ .

b) On suppose que les v.a.r.  $(X_i)_{i \in \mathbb{N}}$  sont indépendantes et suivent chacune une loi de Bernoulli de paramètre  $p$ . Calculer l'espérance et la variance de  $X_1 + \cdots + X_n$ .

c) Comparer avec l'espérance et la variance d'une loi binomiale  $(n, p)$ .

# Chapitre 4

## Sommes de variables aléatoires indépendantes

Soient  $X_1, \dots, X_n, \dots$  une famille de v.a indépendantes et introduisons les v.a  $S_n = X_1 + \dots + X_n$ . Si par exemple les v.a  $X_i$  ont même loi, l'intuition que nous avons des probabilités nous incite à penser que les moyennes

$$\frac{1}{n}S_n = \frac{1}{n}(X_1 + \dots + X_n)$$

convergent quand  $n$  tend vers l'infini vers l'espérance  $\mathbb{E}(X_1)$  de  $X_1$ . En effet, si par exemple les  $X_i$  suivent une loi de Bernoulli  $(1/2, 1/2)$  et forment une famille indépendante de v.a– les  $X_i$  modélisent donc un jeu infini de Pile/Face où les tirages sont indépendants– l'expérience ou l'intuition indique que

$$\frac{1}{n}(X_1 + \dots + X_n)$$

convergent vers le nombre (ou la v.a constante)  $1/2$  qui n'est rien d'autre que l'espérance de  $X_1$ . La première difficulté que nous rencontrons est de donner un sens à la convergence précédente. Rappelons la définition suivante qui introduit la notion de convergence presque-sûre, pertinente pour notre propos :

**Définition 4.0.1 (Convergence presque sûre)** *On dit que la suite de v.a  $Y_n$  converge  $\mathbb{P}$ -presque sûrement vers la v.a  $Y$  si l'ensemble (il s'agit en fait d'un événement) des  $\omega \in \Omega$  pour lesquels la suite  $(Y_n(\omega))_n$  converge vers  $Y(\omega)$  est de probabilité 1.*

Le théorème fondamental de ce chapitre et qui est à la base de la théorie mathématique des probabilités est *la loi forte des grands nombres* qui est

**Théorème 4.0.6** Soit  $X_1, \dots, X_n, \dots$  une famille de v.a indépendante où les  $X_i$  ont même loi et sont dans  $\mathcal{L}^1(\Omega, \mathbb{P})$ . Alors, la suite de v.a

$$\frac{1}{n}(X_1 + \dots + X_n)$$

converge  $\mathbb{P}$ -p.s vers la v.a constante  $\mathbb{E}(X_1)$ .

La démonstration de ce théorème fondamental dans cette généralité sort du cadre de ce cours, mais nous allons en donner une preuve sous des hypothèses plus faibles.

Avant de continuer introduisons la définition suivante :

**Définition 4.0.2** On dit qu'une suite de v.a.r.  $(X_n)_{n \in \mathbb{N}}$  est i.i.d (pour indépendante identiquement distribuée) si elle est indépendante et si les  $X_i$  suivent la même loi.

## 4.1 Lois des grands nombres

### 4.1.1 Loi faible des grands nombres dans le cas $\mathcal{L}^2$

Supposons que le v.a.r.  $X_i$  soient i.i.d. et de carré intégrable.

Nous savons déjà d'après la linéarité de l'espérance que

$$\mathbb{E}(S_n) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)$$

et d'après l'indépendance

$$\text{Var}(S_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n),$$

et comme les  $X_i$  ont même loi

$$\mathbb{E}(S_n) = n\mathbb{E}(X_1), \quad \text{Var}(S_n) = n\text{Var}(X_1).$$

Le fait que la variance de la somme des  $n$  v.a  $X_1, \dots, X_n$  se comporte comme  $n$  et non pas comme  $n^2$  (c'est ici où intervient l'hypothèse d'indépendance) est l'observation fondamentale. Appliquons en effet l'inégalité de Bienaymé-Tchebychev :

$$\mathbb{P}(|S_n - n\mathbb{E}(X_1)| > \lambda) \leq \frac{n\text{Var}(X_1)}{\lambda^2},$$

ce qui peut s'écrire

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| > \frac{\lambda}{n}\right) \leq \frac{n\text{Var}(X_1)}{\lambda^2}.$$

Posons à présent  $\lambda = n\epsilon$  où  $\epsilon$  est un réel positif :

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| > \epsilon\right) \leq \frac{\text{Var}(X_1)}{n\epsilon^2}.$$

Nous voyons donc que pour tout  $\epsilon > 0$  on a

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| > \epsilon\right) = 0.$$

Introduisons la définition suivante :

**Définition 4.1.1 (Convergence en probabilité)** *On dit que la suite de v.a  $(Y_n)_n$  converge en probabilité vers la v.a  $Y$  si pour tout  $\epsilon > 0$  on a*

$$\lim_{n \rightarrow \text{infty}} \mathbb{P}(|Y_n - Y| > \epsilon) = 0.$$

Nous avons donc démontré la *loi faible des grands nombres*

**Théorème 4.1.1** *Si  $(X_i)_{i \geq 1}$  est une famille de v.a indépendante et si les  $X_i$  ont même loi et sont dans  $\mathcal{L}^2$  alors la suite  $S_n/n$  converge en probabilité vers  $\mathbb{E}(X_1)$ .*

### 4.1.2 Loi forte des grands nombres

Rappelons l'énoncé de la loi forte des grands nombres :

**Théorème 4.1.2** *Soit  $X_1, \dots, X_n, \dots$  une famille de v.a.r. indépendante où les  $X_i$  ont même loi et sont dans  $\mathcal{L}^1(\Omega, \mathbb{P})$ . Alors, la suite de v.a*

$$\frac{1}{n}(X_1 + \dots + X_n)$$

*converge  $\mathbb{P}$ -p.s vers la v.a constante  $\mathbb{E}(X_1)$ .*

Nous nous proposons dans ce qui suit d'en donner une preuve dans le cas où les v.a sont  $\mathcal{L}^4$ , c'est-à-dire quand  $\mathbb{E}(|X_1|^4) < \infty$ .

*Démonstration.* — *Dans le cas  $\mathcal{L}^4$ .*

i) Remarquons que l'hypothèse  $\mathbb{E}(X_i^4) < \infty$  entraîne que

$$\forall k = 0, 1, 2, 3, 4, \mathbb{E}(|X_i|^k) < \infty. \quad (4.1)$$

Pour  $k = 0$  c'est évident. On a d'après l'inégalité de Cauchy-Schwarz (cf. 3.6.2)

$$\mathbb{E}(X_i^2 \cdot \mathbf{1}) \leq \mathbb{E}(X_i^4)^{1/2} \mathbb{E}(\mathbf{1}^2)^{1/2} = \mathbb{E}(X_i^4)^{1/2} < \infty.$$

On a donc (4.1) pour  $k = 4$  et  $k = 2$ . En appliquant à nouveau Cauchy-Schwarz on a  $\mathbb{E}(|X_i \cdot \mathbf{1}|) \leq \mathbb{E}(X_i^2)^{1/2} \mathbb{E}(\mathbf{1}^2)^{1/2} < \infty$ ; ainsi (4.1) est vraie pour  $k = 1$ . Enfin,  $\mathbb{E}(|X_i|^3) = \mathbb{E}(X_i^2 |X_i|) \leq \mathbb{E}(X_i^4)^{1/2} \mathbb{E}(X_i^2)^{1/2} < \infty$ . On a donc bien établi (4.1) pour  $k = 0, 1, 2, 3, 4$ .

ii) Posons  $\bar{X}_k = X_k - \mathbb{E}(X_k)$  et  $\bar{S}_n = \bar{X}_1 + \dots + \bar{X}_k$ . La famille de v.a  $(\bar{X}_i)_i$  est indépendante et les  $\bar{X}_i$  sont de même loi et d'espérance nulle :  $\mathbb{E}(\bar{X}_i) = 0$ . Vérifions que

$$\forall k = 0, 1, 2, 3, 4, \mathbb{E}(|\bar{X}_i|^k) < \infty. \quad (4.2)$$

Pour cela, il suffit de constater que  $\mathbb{E}(|\bar{X}|^k) \leq \mathbb{E}\left((|X| + \mathbb{E}(|X|))^k\right)$  et que cette dernière quantité est une combinaison linéaire de terme de la forme  $\mathbb{E}(|X_i|^l) \mathbb{E}(|X|)^m$  pour  $l + m = k$ ,  $0 \leq l, m \leq k$ . On utilise alors pour conclure (4.1).

iii) Posons  $\bar{S}_n = \sum_{i=1}^n \bar{X}_i$  et remarquons que

$$\bar{S}_n^4 = \left(\sum_{i=1}^n \bar{X}_i\right)^4 = \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} \bar{X}_{i_1} \bar{X}_{i_2} \bar{X}_{i_3} \bar{X}_{i_4}$$

et donc

$$\mathbb{E}(\bar{S}_n^4) = \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} \mathbb{E}(\bar{X}_{i_1} \bar{X}_{i_2} \bar{X}_{i_3} \bar{X}_{i_4}). \quad (4.3)$$

On constate à présent que si les indices  $i_r$ ,  $r = 1, 2, 3, 4$  sont distincts deux à deux on a d'après l'indépendance des  $\bar{X}_i$  et le fait que  $\mathbb{E}(\bar{X}_i) = 0$

$$\mathbb{E}(\bar{X}_{i_1} \bar{X}_{i_2} \bar{X}_{i_3} \bar{X}_{i_4}) = \mathbb{E}(\bar{X}_{i_1}) \mathbb{E}(\bar{X}_{i_2}) \mathbb{E}(\bar{X}_{i_3}) \mathbb{E}(\bar{X}_{i_4}) = 0$$

De la même manière si un indice est différent des trois autres,  $\mathbb{E}(\bar{X}_{i_1} \bar{X}_{i_2} \bar{X}_{i_3} \bar{X}_{i_4}) = 0$ ; en effet si par exemple cet indice est  $i_1$ ,  $\bar{X}_{i_1}$  est indépendant de  $\bar{X}_{i_2} \bar{X}_{i_3} \bar{X}_{i_4}$  et donc

$$\mathbb{E}(\bar{X}_{i_1} \bar{X}_{i_2} \bar{X}_{i_3} \bar{X}_{i_4}) = \mathbb{E}(\bar{X}_{i_1}) \mathbb{E}(\bar{X}_{i_2} \bar{X}_{i_3} \bar{X}_{i_4}) = 0 \times \mathbb{E}(\bar{X}_{i_2} \bar{X}_{i_3} \bar{X}_{i_4}) = 0.$$

Cela montre que les seuls termes qui contribuent à la somme (4.3) sont les indices tels que  $\#\{i_1, i_2, i_3, i_4\}$  égale 1 ou 2. Il est alors facile de voir que

$$\mathbb{E}(\bar{S}_n^4) = \binom{4}{2} \sum_{1 \leq i < j \leq n} \mathbb{E}(\bar{X}_i^2 \bar{X}_j^2) + \sum_{1 \leq i \leq n} \mathbb{E}(\bar{X}_i^4)$$

et donc en utilisant le fait qu'il y a  $n(n-1)/2$  termes dans la première somme du membre de droite de cette inégalité et que  $\mathbb{E}(\bar{X}_i^2 \bar{X}_j^2) = \mathbb{E}(\bar{X}_i^2) \mathbb{E}(\bar{X}_j^2) = \mathbb{E}(\bar{X}_1^2)^2$  (les v.a.r.  $\bar{X}_i^2$  et  $\bar{X}_j^2$  sont indépendantes si  $i \neq j$ ) on a

$$\mathbb{E}(\bar{S}_n^4) \leq 3n^2 \mathbb{E}(\bar{X}_1^2)^2 + n \mathbb{E}(\bar{X}_1^4).$$

Par conséquent,

$$\mathbb{E}\left(\left(\frac{\bar{S}_n}{n}\right)^4\right) = \frac{\mathbb{E}(\bar{S}_n^4)}{n^4} \leq \frac{3\mathbb{E}(\bar{X}_1^2)}{n^2} + \frac{\mathbb{E}(\bar{X}_1^4)}{n^3}.$$

iv) La dernière inégalité montre que

$$\sum_{n=1}^{\infty} \mathbb{E}\left(\left(\frac{\bar{S}_n}{n}\right)^4\right) < \infty.$$

Mais on peut démontrer (c'est le théorème de convergence monotone) que

$$\mathbb{E}\left(\sum_{n=1}^{\infty} \left(\frac{\bar{S}_n}{n}\right)^4\right) = \sum_{n=1}^{\infty} \mathbb{E}\left(\left(\frac{\bar{S}_n}{n}\right)^4\right) < \infty.$$

On a donc

$$\mathbb{E}\left(\sum_{n=1}^{\infty} \left(\frac{\bar{S}_n}{n}\right)^4\right) < \infty.$$

Mais si l'espérance d'une v.a.r. positive est finie, cette v.a.r. est finie  $\mathbb{P}$ -presque sûrement. Par conséquent,  $\mathbb{P}$ -p.s. la somme

$$\sum_{n=1}^{\infty} \left(\frac{\bar{S}_n}{n}\right)^4$$

converge et en particulier

$$\frac{\bar{S}_n}{n}$$

converge  $\mathbb{P}$ -p.s. vers 0. Mais, si on se rappelle la définition de  $\bar{S}_n$ , cela est équivalent au fait que  $\mathbb{P}$ -p.s.

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mathbb{E}(X_1) \quad \mathbb{P} - p.s.$$

□

## 4.2 Théorème de la limite centrale

Nous donnons une première version de ce théorème :

**Théorème 4.2.1** Soient  $X_1, \dots, X_n, \dots$  une famille de v.a.r. indépendantes et supposons que les  $X_i$  soient de même loi et dans  $\mathcal{L}^2(\Omega, \mathbb{P})$  (c'est-à-dire de carré intégrable,  $\mathbb{E}(X_i^2) < \infty$ ). Notons  $\mu = \mathbb{E}(X_1)$  et  $\sigma^2 = \text{Var}(X_1)$  (comme les  $X_i$  sont de même loi  $\mathbb{E}(X_i) = \mu$  et  $\text{Var}(X_i) = \sigma$  pour tout  $i$ ). Alors, pour tout intervalle  $I$  de  $\mathbf{R}$

$$\lim_{n \rightarrow \infty} \mathbf{P} \left( \frac{S_n - n\mu}{\sigma\sqrt{n}} \in I \right) = \int_I \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Une autre façon d'énoncer ce théorème est de dire que

$$\lim_{n \rightarrow \infty} \mathbf{P} \left( \frac{\sqrt{n}}{\sigma} \left( \frac{S_n}{n} - \mu \right) \in I \right) = \int_I \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

En d'autres termes,  $S_n/n - \mu$  converge d'après la loi forte des grands nombres vers 0, la déviation des moyennes par rapport à l'espérance, "renormalisée" par le facteur  $\sqrt{n}/\sigma$ , converge dans un certain sens vers une loi gaussienne normalisée (espérance nulle, variance égale à 1).

Avant de passer à la preuve de ce théorème, nous devons faire quelques rappels et introduire quelques notions utiles.

### 4.2.1 Fonctions de répartition

Se reporter à la section 3.2.2

### 4.2.2 Convergence en loi

**Définition 4.2.1** On dit qu'une suite de v.a  $(Y_n)_{n \in \mathbf{N}}$  converge en loi vers une v.a  $Y$  si et seulement si pour toute fonction continue bornée  $f : \mathbf{R} \rightarrow \mathbf{R}$

$$\lim_{n \rightarrow \infty} \mathbb{E}(f(Y_n)) = \mathbb{E}(f(Y)).$$

L'intérêt de cette notion réside dans la proposition suivante (admise) :

**Théorème 4.2.2** Soient  $(Y_n)_{n \in \mathbf{N}}$  une suite de v.a et  $Y$  une v.a. et notons  $F_{Y_n} : \mathbf{R} \rightarrow [0, 1]$  et  $F_Y : \mathbf{R} \rightarrow [0, 1]$  les fonctions de répartition des v.a  $Y_n$ ,  $Y$  :

$$F_{Y_n}(t) = \mathbb{P}(Y_n \leq t), \quad F_Y(t) = \mathbb{P}(Y \leq t).$$

La suite  $(Y_n)_{n \in \mathbf{N}}$  converge en loi vers  $Y$  si et seulement si en tout point  $t_0$  où  $F_Y$  est continue à gauche

$$\lim_{n \rightarrow \infty} F_{Y_n}(t_0) = F_Y(t_0).$$



Un corollaire utile du théorème précédent est le suivant

**Corollaire 4.2.1** *Si les  $Y_n$  convergent en loi vers une v.a  $Y$  qui admet une densité  $\rho_Y$  alors pour tout intervalle  $I$  de  $\mathbf{R}$*

$$\lim_{n \rightarrow \infty} \mathbb{P}(Y_n \in I) = \int_I \rho_Y(y) dy$$

*Démonstration.* — En effet, dans ce cas  $F_Y$  est continue en tout point.  $\square$

### 4.2.3 Fonctions caractéristiques

**Définition 4.2.2** *Si  $Y$  est une v.a.r, la fonction caractéristique de  $Y$  est la fonction continue  $\phi_Y : \mathbf{R} \rightarrow \mathbf{R}$  définie par*

$$\phi_Y(t) = \mathbb{E}(e^{itY}),$$

(où  $i = \sqrt{-1}$ ).

**Remarque i)** Pour  $t$  fixé la v.a  $e^{itY}$  est bornée par 1 (puisque  $Y$  est à valeurs réelles) et est donc intégrable.

ii) La fonction caractéristique d'une v.a ne dépend que de la loi de cette v.a.

On peut préciser le résultat de continuité précédent (admis) :

**Proposition 4.2.1** *Si  $Y$  est une v.a.r intégrable, alors la fonction caractéristique de  $Y$  est de classe  $C^1$  (dérivable et de dérivée continue) et*

$$\phi'_Y(t) = \mathbb{E}\left((iY)e^{itY}\right).$$

*De même, si  $Y$  est dans  $\mathcal{L}^p(\Omega, \mathbb{P})$  la fonction caractéristique de  $Y$  est de classe  $C^p$  et on a*

$$\frac{d^p}{dt^p} \phi_Y(t) = \mathbb{E}\left((iY)^p e^{itY}\right).$$

**Exercice :** Montrer que si  $Z = aY + b$

$$\phi_Z(t) = e^{itb} \phi_Y(ta).$$

Calculons à présent les fonctions caractéristiques de certaines lois classiques.

**v.a discrètes** Si  $Y$  prend un nombre fini de valeurs  $y_1, \dots, y_r$  et si on note  $p_r = \mathbb{P}(Y = y_r)$  on a

$$\begin{aligned}\phi_Y(t) &= \mathbb{E}(e^{itY}) \\ &= \sum_{k=1}^r e^{ity_k} \mathbb{P}(Y = y_k) \\ &= \sum_{k=1}^r (e^{it})^{y_k} \mathbb{P}(Y = y_k)\end{aligned}$$

et on reconnaît (si  $Y$  est à valeurs entières) la fonction génératrice de  $Y$  au point  $e^{it}$ . Le calcul des fonctions caractéristiques de v.a discrètes est exactement le même que celui que nous avons effectué dans un chapitre précédent.

**v.a admettant une densité  $\rho_Y$**  Dans ce cas

$$\begin{aligned}\phi_Y(t) &= \mathbb{E}(e^{itY}) \\ &= \int_{-\infty}^{\infty} e^{ity} \rho_Y(y) dy.\end{aligned}$$

On habituellement  $\hat{\rho}_Y(t) = \int_{-\infty}^{\infty} e^{ity} \rho_Y(y) dy$  et on dit que la fonction  $\hat{\rho}_Y$  est la *transformée de Fourier* de la fonction  $\rho_Y$

**Exemple :** *Fonctions caractéristique d'une gaussienne* Rappelons que si  $Z$  est une v.a suivant une loi gaussienne  $\mathcal{N}(\mu, \sigma)$  on peut l'écrire sous la forme  $Z = \sigma Y + \mu$  où  $Y$  suit une loi gaussienne normalisée  $\mathcal{N}(0, 1)$  de densité

$$\rho(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

On a donc

$$\phi_Y(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ity} e^{-y^2/2} dy.$$

On a (cf. exercice) :

$$\phi_Y(t) = e^{-t^2/2}.$$

Ainsi, la fonction d'une caractéristique d'une v.a suivant une loi gaussienne  $\mathcal{N}(\mu, \sigma)$  est

$$\phi_Z(t) = e^{it\mu - \sigma^2(t^2/2)}.$$

### Liens avec la convergence en loi

Les fonctions caractéristiques jouent un rôle important dans les problèmes où interviennent des convergences en loi. Les deux théorèmes qui suivent illustrent ce fait.

**Théorème 4.2.3** *La loi d'une v.a.r (resp. d'un vecteur aléatoire) est déterminée par sa fonction caractéristique : si  $Y$  et  $Z$  sont deux v.a.r telles que pour tout  $t \in \mathbb{R}$*

$$\phi_Y(t) = \phi_Z(t),$$

*alors la loi de  $Y$  et la loi de  $Z$  sont les mêmes : pour tout intervalle  $I$  de  $\mathbb{R}$*

$$\mathbb{P}(Y \in I) = \mathbb{P}(Z \in I).$$

*En particulier, elles ont la même fonction de répartition.*

La notion de fonction caractéristique est très utile pour donner un critère utile de convergence en loi. On a ainsi le théorème important suivant (admis) :

**Théorème 4.2.4** *La suite de v.a.r (resp. de vecteurs aléatoires)  $(Y_n)_{n \in \mathbb{N}}$  converge en loi vers  $Y$  si et seulement si pour tout  $t \in \mathbb{R}$*

$$\lim_{n \rightarrow \infty} \phi_{Y_n}(t) = \phi_Y(t).$$

### Fonction caractéristique d'une somme de v.a.r. indépendantes

**Théorème 4.2.5** *Si les v.a.r  $X_1, \dots, X_n$  forment une famille indépendante pour tout  $t \in \mathbb{R}$*

$$\phi_{X_1 + \dots + X_n}(t) = \phi_{X_1}(t) \cdots \phi_{X_n}(t).$$

*Démonstration.* — Il suffit de constater que

$$\phi_{X_1 + \dots + X_n}(t) = \phi_{(X_1, \dots, X_n)}(t, \dots, t),$$

et d'appliquer le théorème précédent. □

### 4.2.4 Démonstration du théorème de la limite centrale

Passons à présent à la démonstration du théorème de la limite centrale (ou "Central Limit Theorem")

Soient donc  $X_1, \dots, X_n, \dots$  une famille indépendante de v.a qui sont de même loi et de carré intégrable. Nous posons  $\mu = \mathbb{E}(X_1)$  et  $\sigma = \text{Var}(X_1)$ . Notons  $S_n = X_1 + \dots + X_n$ ,  $\Sigma_n = (X_1 - \mu) + \dots + (X_n - \mu)$  et posons

$$Z_n = \frac{S_n - n\mathbb{E}(X_1)}{\sigma\sqrt{n}} = \frac{\Sigma_n}{\sigma\sqrt{n}}.$$

La formule de l'exercice du début de la section 4.2.3 montre que

$$\phi_{Z_n}(t) = \phi_{\Sigma_n}\left(\frac{t}{\sigma\sqrt{n}}\right),$$

et comme  $\Sigma_n$  est la somme des v.a indépendantes  $X_i - \mu$ ,  $1 \leq i \leq n$  et que ces v.a ont même loi :

$$\phi_{\Sigma_n}(t) = \phi_{X_1 - \mu}(t)^n.$$

Si on pose  $\phi(t) = \phi_{X_1 - \mu}(t)$  on a donc

$$\phi_{Z_n}(t) = \phi\left(\frac{t}{\sigma\sqrt{n}}\right)^n.$$

Puisque la v.a est de carré intégrable la fonction  $\phi$  est de classe  $C^2$ . Par ailleurs,

$$\phi(0) = \mathbb{E}(1), \quad \phi'(0) = i\mathbb{E}(X_1 - \mu), \quad \phi''(0) = -\mathbb{E}((X_1 - \mu)^2),$$

c'est-à-dire

$$\phi(0) = 1, \quad \phi'(0) = 0, \quad \phi''(0) = -\sigma^2.$$

D'après la formule de Taylor

$$\phi(t) = 1 - \frac{\sigma^2}{2}t^2 + o(t^2),$$

et donc pour  $t$  fixé

$$\phi_{Z_n}(t) = \left(1 - \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)\right)^n,$$

ou encore

$$\phi_{Z_n}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n.$$

Pour  $n$  suffisamment grand,  $1 - (t^2)/(2n) + o(1/n)$  est dans la boule de centre 1 et de rayon 1/2 et on peut écrire

$$\phi_{Z_n}(t) = \exp\left(n \log\left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)\right),$$

où  $\log$  est la détermination principale du logarithme dans le plan complexe (qui admet le même développement en série que le logarithme réel). On a donc

$$\phi_{Z_n}(t) = \exp\left(-\frac{t^2}{2} + o(1)\right) = e^{-(t^2/2)} + o(1),$$

et partant,

$$\lim_{n \rightarrow \infty} \phi_{Z_n}(t) = e^{-t^2/2}.$$

On reconnaît dans le membre de droite la fonction caractéristique d'une v.a. gaussienne normalisée et le théorème 4.2.4 montre que  $Z_n$  converge en loi vers une loi gaussienne normalisée.

### 4.3 Quelques remarques sur les diverses notions de convergence

Rappelons les diverses notions de convergence que nous avons rencontrées. Soit  $(\Omega, \mathcal{B}, \mathbb{P})$  un espace probabilisé.

**Définition 4.3.1 (Convergence presque-sure)** *On dit qu'une suite de v.a.r.  $(X_n)_{n \in \mathbb{N}}$  converge  $\mathbb{P}$ -presque sûrement (et on écrit  $\mathbb{P}$ -p.s.) vers une variable aléatoire  $X$  si l'ensemble des  $\omega \in \Omega$  pour lesquels  $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ , qui est un événement, est de probabilité 1.*

**Définition 4.3.2 (Convergence en moyenne)** *On dit qu'une suite de v.a.r.  $(X_n)_{n \in \mathbb{N}}$  converge en moyenne (ou dans  $\mathcal{L}^1$ ) vers une variable aléatoire  $X$  si*

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|) = 0.$$

**Définition 4.3.3 (Convergence en probabilité)** *On dit qu'une suite de v.a.r.  $(X_n)_{n \in \mathbb{N}}$  converge en probabilité vers une variable aléatoire  $X$  si pour tout  $\epsilon > 0$  on a*

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$

**Définition 4.3.4 (Convergence en probabilité)** *On dit qu'une suite de v.a.r.  $(X_n)_{n \in \mathbb{N}}$  converge en probabilité vers une variable aléatoire  $X$  si pour tout  $\epsilon > 0$  on a*

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$

Voici quelques liens entre ces diverses notions de convergence.

**Proposition 4.3.1** *a) La convergence p.s et la convergence en moyenne impliquent la convergence en probabilité; la convergence en probabilité (et donc la convergence p.s et la convergence en moyenne) implique la convergence en loi.*

*b) S'il existe une constante  $A$  telle que  $|X_n| \leq A$  la convergence en probabilité de  $X_n$  est équivalente à sa convergence en moyenne.*

78 CHAPITRE 4. SOMMES DE VARIABLES ALÉATOIRES INDÉPENDANTES

*c) Si une suite de v.a.r. converge en loi vers une constante, alors elle converge en probabilité vers cette constante.*

# Chapitre 5

## Rudiments de Statistiques

### 5.1 Statistiques

#### 5.1.1 Terminologie

On appelle *population* un ensemble sur lequel une étude statistique est menée et on nomme *individus* les éléments de cet ensemble. Un *échantillon* est un sous-ensemble de la population. L'*effectif* d'un échantillon est son cardinal. Un problème important en statistique est de déterminer la *fréquence* de la valeur d'un *caractère* que l'on trouve parmi les individus d'une population ou d'un échantillon : c'est le quotient du cardinal de l'ensemble constitué des éléments de la population ou de l'échantillon pour lesquels le caractère prend cette valeur par le cardinal de la population ou de l'échantillon.

#### 5.1.2 Modélisation

La modélisation d'un problème statistique se fait de la manière suivante. On suppose données  $n$  variables aléatoires  $X_1, \dots, X_n$  définies sur un espace probabilisé  $(\Omega, \mathcal{B}, \mathbb{P})$  (que l'on ne connaît pas). Pour simplifier nous supposons que les v.a.  $X_i$  sont à valeurs dans  $\mathbb{R}$  ou dans  $\mathbb{N}$ . On supposera que :

- les v.a.  $(X_i)_{1 \leq i \leq n}$  sont indépendantes ;
- elles sont identiquement distribuées, c'est-à-dire suivent la même loi  $\mu$  : pour tout  $1 \leq i \leq n$  et tout intervalle  $I \subset \mathbb{R}$ ,  $\mathbb{P}(X_i \in I) = \mu(I)$  ;
- la loi  $\mu$  commune aux v.a.  $(X_i)_{1 \leq i \leq n}$  est dans un certain ensemble de lois  $\Theta$ .

L'ensemble  $\Theta$  des lois peut-être l'ensemble de toutes les lois ou l'ensemble de toutes les lois normales ou l'ensemble de toutes les lois de Poisson *etc.* Souvent les lois de l'ensemble  $\Theta$  sont définies par un paramètre  $\theta \in \mathbb{R}^d$  : par exemple si  $\Theta$  est l'ensemble de toutes les lois normales  $\mathcal{N}(\mu, \sigma^2)$ , les éléments

de  $\Theta$  sont paramétrés par  $\theta = (\mu, \sigma^2) \in \mathbb{R}^2$ ; si  $\Theta$  est l'ensemble de toutes les lois de Poisson  $\mathcal{P}(\lambda)$  le paramètre  $\theta$  sera  $\theta = \lambda \in \mathbb{R}$ . Dans ce cas on peut supposer que  $\Theta$  est (en bijection avec) un sous-ensemble  $T$  de  $\mathbb{R}^d$ .

**Définition 5.1.1** *On dira que le modèle statistique  $\Theta$  est paramétré par  $\theta \in T \subset \mathbb{R}^d$  s'il existe une bijection du sous-ensemble  $T \subset \mathbb{R}^d$  sur  $\Theta$ .*

Une façon équivalente de poser le problème et qui a l'avantage de ne pas faire appel à l'espace  $(\Omega, \mathcal{B}, \mathbb{P})$  est de considérer que les v.a.  $(X_i)_{1 \leq i \leq n}$  sont des v.a. définies sur l'espace de probabilité  $(\mathbb{R}, \text{Bor}(\mathbb{R}), \mu)$  où  $\text{Bor}(\mathbb{R})$  est la tribu borélienne de  $\mathbb{R}$  (la plus petite tribu contenant les intervalles de  $\mathbb{R}$ ) et  $\mu$  est la loi commune des  $X_i$ ; en effet la proposition 3.2.2 montre que la donnée de tous les  $\mathbb{P}(X_i \in I)$ ,  $I$  intervalle de  $\mathbb{R}$ , permet de définir (de façon unique) une mesure de probabilité sur  $\mathbb{R}$  par  $\mu(A) = \mathbb{P}(X \in A)$ . De façon plus générale, si  $\text{Bor}(\mathbb{R}^n)$  est la tribu borélienne de  $\mathbb{R}^n$ , c'est-à-dire celle engendrée par tous les pavés de  $\mathbb{R}^n$  (un pavé de  $\mathbb{R}^n$  est un produit de  $n$  intervalles de  $\mathbb{R}$ ), on peut définir sur  $(\mathbb{R}^n, \text{Bor}(\mathbb{R}^n))$  une mesure  $\mu^{\otimes n}$  par  $\mu^{\otimes n}(I_1 \times \dots \times I_n) = \mu(I_1) \dots \mu(I_n)$ . On peut démontrer que l'étude des  $X_i$  se ramène à celle des v.a.  $\tilde{X}_i$  définies sur  $(\mathbb{R}^n, \text{Bor}(\mathbb{R}^n), \mu^{\otimes n})$ ,  $\tilde{X}_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\tilde{X}_i(x_1, \dots, x_n) = x_i$ . Les v.a.  $\tilde{X}_i$  ont même loi  $\mu$  et sont indépendantes. On peut ainsi voir un  $n$ -échantillon satisfaisant une statistique  $\Theta$  comme le tirage indépendant de  $n$  valeurs  $x_1, \dots, x_n \in \mathbb{R}$  suivant la loi  $\mu$ .

Dans ce contexte

**Définition 5.1.2** *Un  $n$ -échantillon du modèle statistique  $\Theta$  est la donnée de  $n$  variables aléatoires indépendantes de même loi  $\mu \in \Theta$ .*

**Exemple** Imaginons par exemple que l'on veuille contrôler le diamètre des boulons fabriqués dans une certaine usine. Pour cela, on prélève un échantillon de  $n$  boulons et on mesure leurs diamètres  $x_1, \dots, x_n$ . On peut répéter cette expérience tous les jours et on obtiendra alors de nouvelles valeurs pour  $x_1, \dots, x_n$ . Nous modéliserons ce problème en supposant que les valeurs  $x_1, \dots, x_n$  obtenues sont les valeurs de  $n$  variables aléatoires  $X_1, \dots, X_n$  évaluées en un point  $\omega \in \Omega$  d'un certain espace probabilisé  $(\Omega, \mathcal{B}, \mathbb{P})$  que l'on ne connaît pas :  $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$  où  $\omega \in \Omega$ <sup>1</sup>. On fait alors l'hypothèse que ces v.a.r.  $(X_i)$  sont indépendantes et suivent une même loi. L'expérience montre qu'il est raisonnable de supposer que leur loi commune est une loi normale  $\mathcal{N}(\mu, \sigma^2)$ . Le problème est alors d'essayer d'estimer les paramètres  $\mu$  et  $\sigma$  de cette loi normale à partir de l'observation des  $x_1, \dots, x_n$ .

---

1.  $\omega$  dépend du jour où l'on réalise l'expérience; on peut aussi supposer que les variables aléatoires dépendent du jour où l'on fait le prélèvement, ce qui permet de vérifier la qualité de la production de boulons au cours du temps.



## 5.2 Estimateurs

### 5.2.1 Définition

**Définition 5.2.1** Soit  $(X_1, \dots, X_n)$  un  $n$ -échantillon du modèle  $\Theta$ . Un estimateur est une v.a. de la forme  $\varphi(X_1, \dots, X_n)$  où  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ .

On dira souvent par abus de langage que  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  elle même est un estimateur.

### 5.2.2 Estimateur sans biais

**Définition 5.2.2** Soit  $f : \Theta \rightarrow \mathbb{R}$  une fonction. On dit qu'un estimateur  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  est un estimateur sans biais de  $f$  si pour tout  $\mu \in \Theta$  et toutes v.a.r., i.i.d de loi  $\mu$  on a

$$\mathbb{E}(\varphi(X_1, \dots, X_n)) = f(\mu).$$

L'intérêt de la notion précédente réside dans le fait que *quelle que soit* la loi  $\mu \in \Theta$  des  $X_i$  l'espérance de l'estimateur (sa moyenne) est égale à  $f(\mu)$ .

On choisira souvent pour  $f : \Theta \rightarrow \mathbb{R}$  l'application qui à une loi  $\mu$  de  $\Theta$  associe son espérance (ou de façon équivalente sa moyenne)  $m(\mu)$  ou sa variance  $\sigma^2(\mu)$  qui sont définies par  $m(\mu) := \mathbb{E}(X)$  et  $\sigma^2(\mu) := \text{Var}(X)$ ,  $X$  étant une v.a. de loi  $\mu$  ( $\mathbb{E}(X)$  et  $\text{Var}(X)$  sont clairement indépendantes du choix de la v.a.  $X$  pourvu qu'on la choisisse de loi  $\mu$ ).

### 5.2.3 Risque quadratique

Il est également intéressant d'introduire la notion suivante.

**Définition 5.2.3** On appelle *risque quadratique* d'un estimateur  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  le réel

$$R_\mu(\varphi) = \mathbb{E}\left(\left(\varphi(X_1, \dots, X_n) - f(\mu)\right)^2\right)$$

où les  $X_i$  sont i.i.d. de loi  $\mu$  (la quantité précédente est indépendante du choix des v.a.  $X_i$  pourvu qu'elles soient i.i.d. de loi  $\mu$ ).

Notons que quand  $\varphi$  est un estimateur sans biais la quantité précédente n'est autre que la variance de  $\varphi(X_1, \dots, X_n)$  :

**Proposition 5.2.1** Si  $\varphi$  est un estimateur sans biais de  $f$  on a pour toutes v.a. i.i.d.  $(X_i)_{1 \leq i \leq n}$  de loi  $\mu \in \Theta$

$$R_\mu(\varphi) = \text{Var}(\varphi(X_1, \dots, X_n)).$$

Etant donnée  $f : \Theta \rightarrow \mathbb{R}$ , il est alors naturel de considérer les estimateurs sans biais  $\varphi$  qui minimisent  $R_\mu(\varphi)$ .

**Définition 5.2.4** *Un estimateur  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  de  $f : \Theta \rightarrow \mathbb{R}$  est dit sans biais de variance minimum s'il est sans biais et si pour tout estimateur sans biais  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$*

$$R_\mu(\varphi) \leq R_\mu(\psi).$$

### 5.2.4 Moyenne et variance empiriques

**Définition 5.2.5** *Soient  $X_1, \dots, X_n$  un  $n$ -échantillon de loi  $\mu$ .*

1. *La moyenne empirique des  $(X_i)_{1 \leq i \leq n}$  est la v.a.*

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}.$$

2. *La variance empirique des  $(X_i)_{1 \leq i \leq n}$  est la v.a.*

$$\Sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

(noter le dénominateur  $n-1$  au lieu de  $n$ ).

On a alors

**Proposition 5.2.2** *Les estimateurs  $\bar{X}$  et  $\Sigma^2$  sont des estimateurs sans biais respectivement de  $m(\mu)$  et de  $\sigma^2(\mu)$ .*

*Démonstration.* — cf. preuve donnée en amphithéâtre. □

### 5.2.5 Maximum de vraisemblance

#### Cas discret

Soit  $\Theta$  une statistique (un ensemble de lois) dont on suppose toutes les lois discrètes (à valeurs dans  $\mathbb{N}$ ) et paramétrées par un paramètre  $\theta \in T \subset \mathbb{R}^d$  (penser par exemple à des lois de Poisson). On notera  $\mu_\theta$  ou encore par abus de notation  $\theta$  la loi de paramètre  $\theta$ . Supposons que  $X_1, \dots, X_n$  soit un  $n$ -échantillon de loi  $\mu_\theta \in \Theta$ . La loi du vecteur aléatoire  $(X_1, \dots, X_n)$  est la donnée de toutes les valeurs

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

pour  $(x_1, \dots, x_n) \in \mathbb{N}^n$ . Comme les v.a.  $X_i$  sont indépendantes, on a

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n)$$

et si les  $X_i$  sont de loi  $\mu_\theta$

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mu_\theta(x_1) \cdots \mu_\theta(x_n).$$

On notera que comme on est dans le cas discret on peut assimiler  $\mu_\theta$  à la donnée des  $\mu_\theta(x) = \mathbb{P}(X = x)$  (où  $X$  est de loi  $\mu_\theta$ ), c'est-à-dire à une application de  $\mathbb{N} \rightarrow [0, 1]$ ,  $\mu_\theta(x) = \mathbb{P}(X = x)$ .

**Définition 5.2.6** *On appelle fonction de vraisemblance du modèle l'application qui à  $(\theta, x_1, \dots, x_n) \in T \times \mathbb{N}^n$  associe le produit*

$$L(\theta, x_1, \dots, x_n) = \mu_\theta(x_1) \cdots \mu_\theta(x_n).$$

Introduisons alors la définition suivante :

**Définition 5.2.7** *Un estimateur du maximum de vraisemblance de  $\theta \in \mathbb{R}^d$  est une application  $\hat{\theta} : \mathbb{N}^n \rightarrow T \subset \mathbb{R}^d$  telle que pour tout  $(x_1, \dots, x_n) \in \mathbb{N}^n$*

$$L(\hat{\theta}(x_1, \dots, x_n), x_1, \dots, x_n) = \sup_{\theta \in T} L(\theta, x_1, \dots, x_n).$$

*Si  $(X_1, \dots, X_n)$  est un  $n$ -échantillon de statistique  $\Theta$ , on dit alors également que la v.a.  $\hat{\theta}(X_1, \dots, X_n)$  est un estimateur du maximum de vraisemblance de  $\theta$ .*

Ainsi, étant donné un  $n$ -échantillon  $(x_1, \dots, x_n)$  de statistique dans l'ensemble  $\Theta = \{\mu_\theta : \theta \in T\}$ , sans information supplémentaire, le choix le plus raisonnable pour la loi de l'échantillon (le plus vraisemblable) est celui qui maximise la probabilité d'apparition de cet échantillon.

### Cas des lois admettant une densité

On peut généraliser les notions de la sous-section précédente au cas où  $\Theta$  est paramétré par  $\theta \in T \subset \mathbb{R}^d$  et est un ensemble de lois  $\mu_\theta \in T$  admettant chacune une densité  $\rho_\theta$ .

**Définition 5.2.8** *On appelle fonction de vraisemblance du modèle l'application qui à  $(\theta, x_1, \dots, x_n) \in T \times \mathbb{R}^n$  associe le produit*

$$L(\theta, x_1, \dots, x_n) = \rho_\theta(x_1) \cdots \rho_\theta(x_n).$$

**Définition 5.2.9** *Un estimateur du maximum de vraisemblance de  $\theta \in T \subset \mathbb{R}^d$  est une application  $\hat{\theta} : \mathbb{R}^n \rightarrow T \subset \mathbb{R}^d$  telle que pour tout  $(x_1, \dots, x_n) \in \mathbb{R}^n$*

$$L(\hat{\theta}(x_1, \dots, x_n), x_1, \dots, x_n) = \sup_{\theta \in T} L(\theta, x_1, \dots, x_n).$$

*Si  $(X_1, \dots, X_n)$  est un  $n$ -échantillon de statistique  $\Theta$ , on dit alors également que la v.a  $\hat{\theta}(X_1, \dots, X_n)$  est un estimateur du maximum de vraisemblance de  $\theta$ .*

Dans la pratique, déterminer un estimateur du maximum de vraisemblance revient à un problème d'optimisation. Illustrons cela dans le cas où  $d = 1$ . Si  $T \subset \mathbb{R}$  est un ouvert et si la fonction  $L(\theta, x_1, \dots, x_n)$  est disons de classe  $C^2$ , alors

- Si  $\hat{\theta}(x_1, \dots, x_n) \in T$  existe on a

$$\frac{\partial L}{\partial \theta}(\hat{\theta}(x_1, \dots, x_n), x_1, \dots, x_n) = 0. \quad (5.1)$$

- Réciproquement, si  $\hat{\theta}(x_1, \dots, x_n)$  est un point où (5.1) a lieu et si pour tout  $\theta \in T$ ,  $x_1, \dots, x_n$

$$\frac{\partial^2 L}{\partial \theta^2}(\theta, x_1, \dots, x_n) < 0 \quad (5.2)$$

alors  $\hat{\theta}(x_1, \dots, x_n)$  réalise le maximum de  $L(\theta, x_1, \dots, x_n)$ .

En fait, d'un point de vue pratique et théorique, il est beaucoup plus utile d'appliquer les critères donnés par (5.1) et (5.2) au logarithme  $\ln L$  plutôt qu'à  $L$ .

### 5.3 Intervalles de confiance

On suppose toujours que  $\Theta$  l'ensemble des lois possibles modélisant notre modèle est paramétré par  $\theta \in T \subset \mathbb{R}^d$ . Etant donné un  $n$ -échantillon de notre expérience, il est en général impossible de déterminer exactement le paramètre  $\theta \in \mathbb{R}^d$  de notre statistique (ou une fonction  $f(\theta)$  de  $\theta$ ). En revanche, il est possible de donner un intervalle où se situe  $\theta$  (ou  $f(\theta)$ ) avec une grande probabilité.

**Définition 5.3.1** *Soit  $0 < \alpha < 1$ . Un intervalle de confiance de niveau  $1 - \alpha$  pour  $f(\theta)$  est la donnée de deux variables aléatoires réelles  $A, B$  telles que, pour tout  $\theta \in T \subset \mathbb{R}^d$ ,*

$$\mathbb{P}(f(\theta) \in ]A, B]) \geq 1 - \alpha.$$

Les sections suivantes illustrent cette notion dans plusieurs cas.

## 5.4 Sondages

Un sondeur veut déterminer dans une population de  $N$  individus (par exemple  $N = 6.10^7$ ) le nombre de personnes appartenant à une catégorie  $A$  (donc connaître le cardinal de l'ensemble  $A$ ). Pour cela il effectue un sondage sur un échantillon de  $n$  individus (p.ex  $n = 10^3$ ) tirés au hasard. Pour simplifier on supposera que le sondeur procède de la façon suivante. Il réalise  $n$  tirages  $X_1, \dots, X_n$  qui sont autant de variables aléatoires  $X_i : \Omega \rightarrow \{1, \dots, N\}$  suivant une loi uniforme sur  $\{1, \dots, N\}$ . On pose alors  $Y_i = \mathbf{1}_A \circ X_i$  et on définit  $Z_n = Y_1 + \dots + Y_n$  le nombre de personnes interrogées qui appartiennent à la catégorie  $A$ . Il est facile de voir que les  $Y_i$  suivent une même loi de Bernoulli de paramètre  $p$  (où  $p$  est la proportion de personnes dans la population ayant l'opinion  $A$ ) et sont indépendantes pourvu que les  $X_i$  le soient (**Exercice** : Prouver ces faits). On est ainsi dans la situation où l'on dispose d'un  $n$ -échantillon  $(Y_i)_{1 \leq i \leq n}$  où les  $Y_i$  suivent une loi de Bernoulli de paramètre  $p \in [0, 1]$  que l'on veut déterminer (au moins avec grande probabilité). Pour cela on constate que comme les  $Y_i$  sont i.i.d. et suivent une loi de Bernoulli de paramètre  $p$ , leur somme  $Z_n$  suit une loi de Binomiale  $(n, p)$ . Dans la pratique, on utilisera l'approximation normale que donne le Théorème Central Limit : la suite de v.a  $(\sqrt{n}/\sigma)(Z_n - np)$ ,  $\sigma^2 = p(1 - p)$ , converge en loi vers une loi normale centrée réduite. On peut donc écrire

$$\mathbb{P}(\{\omega \in \Omega : p \in [\frac{Z_n(\omega)}{n} - \frac{c\sigma}{\sqrt{n}}, \frac{Z_n(\omega)}{n} + \frac{c\sigma}{\sqrt{n}}]\}) \approx \int_{-c}^c \frac{e^{-t^2/2}}{\sqrt{2\pi}}.$$

[On rappelle que quand  $c \approx 2$  cette intégrale est à peu près égale à 0.95 et quand  $c \approx 2.6$  elle vaut à peu près 0.99.] Dans la pratique cette approximation est bonne quand par exemple  $np \approx 30$ . Pour simplifier nous supposons que nous sommes dans cette hypothèse asymptotique. Ainsi, si on connaît l'écart type  $\sigma$  on dira que l'intervalle  $I_n = [\frac{Z_n}{n} - \frac{2\sigma}{\sqrt{n}}, \frac{Z_n}{n} + \frac{2\sigma}{\sqrt{n}}]$  (resp.  $[\frac{Z_n}{n} - \frac{2.6\sigma}{\sqrt{n}}, \frac{Z_n}{n} + \frac{2.6\sigma}{\sqrt{n}}]$ ) est un *intervalle de confiance* avec une fiabilité de 95% (resp. 99%) pour la valeur de  $p$  : la probabilité pour que  $p$  se trouve dans l'intervalle ainsi déterminé<sup>2</sup> est de 0.95 (resp. 0.99). Malheureusement, on ne connaît pas la valeur de  $\sigma$  (car sinon on connaîtrait déjà celle de  $p$  puisque  $\sigma^2 = p(1 - p)$ ). En revanche, on a toujours l'inégalité  $\sigma = \sqrt{p(1 - p)} \leq 1/2$  si bien que l'on peut écrire dès que  $n$  est assez grand

$$\mathbb{P}(p \in [\frac{Z_n}{n} - \frac{c}{2\sqrt{n}}, \frac{Z_n}{n} + \frac{c}{2\sqrt{n}}]) \geq \sim \int_{-c}^c \frac{e^{-t^2/2}}{\sqrt{2\pi}}.$$

---

2. Attention, cela signifie que  $p = \#A/N$  étant fixé (mais inconnu) la probabilité qu'un sondage portant sur  $n$  personnes fournisse un intervalle  $I_n$  ne contenant pas  $p$  est inférieure à 0.05

Un intervalle de confiance à 0.95 (resp. 0.99) est donc par exemple  $[\frac{Z_n}{n} - \frac{1}{\sqrt{n}}, \frac{Z_n}{n} + \frac{1}{\sqrt{n}}]$  (resp.  $[\frac{Z_n}{n} - \frac{1.3}{\sqrt{n}}, \frac{Z_n}{n} + \frac{1.3}{\sqrt{n}}]$ )

## 5.5 Statistiques gaussiennes

### 5.5.1 Sondages gaussiens

#### Exemple

Dans l'exemple précédent, l'estimation *a priori* sur la variance des  $Y_i$ , rendue possible par le fait que les  $Y_i$  suivent une loi de Bernoulli, est l'élément clé pour obtenir un intervalle de confiance. Revenons à présent au problème que nous avons décrit dans l'exemple de la section 5.1.2. On sait que le diamètre d'un boulon fabriqué dans une certaine usine suit une loi normale de moyenne  $\mu$  et de variance  $\sigma$  inconnues<sup>3</sup>. On veut avoir une estimation de  $\mu$  avec une certitude raisonnable (intervalle de confiance) en effectuant un nombre  $n$  de tirages (indépendants) ce que l'on modélise par  $n$  v.a  $Y_k$ ,  $1 \leq k \leq n$ , i.i.d suivant une même loi normale  $\mathcal{N}(\mu, \sigma^2)$ . Puisque les  $Y_i$  sont indépendantes et gaussiennes leur somme  $Z_n = Y_1 + \dots + Y_n$  est également gaussienne (**Exercice**<sup>4</sup>) et on a donc l'égalité<sup>5</sup>

$$\mathbf{P}(\mu \in [\frac{Z_n}{n} - \frac{c\sigma}{\sqrt{n}}, \frac{Z_n}{n} + \frac{c\sigma}{\sqrt{n}}]) = \int_{-c}^c \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt.$$

Malheureusement on ne connaît pas  $\sigma$  et on ne peut donc pas préciser d'intervalle de confiance. Il faut procéder autrement.

Avant d'énoncer le théorème qui permet de résoudre ce problème introduisons quelques définitions.

### 5.5.2 Loi du chi-deux et loi de Student

**Définition 5.5.1 (Loi du chi-deux  $\chi^2$ )** Une loi du chi-deux de paramètre  $n$  ( $n \geq 2$ ) est la loi d'une somme de carrés de  $n$  v.a indépendantes suivant une même loi normale centrée réduite  $\mathcal{N}(0, 1)$  : si  $X_1, \dots, X_n$  sont des v.a. i.i.d. de loi  $\mathcal{N}(0, 1)$ , la loi de  $X_1^2 + \dots + X_n^2$  est par définition un chi-deux à  $n$  degrés de liberté. On note cette loi  $\chi^2(n)$ . C'est une loi (dite loi gamma  $\gamma_{1/2, n/2}$ ) de densité

$$\mathbf{1}_{]0, \infty[}(x) \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^x t^{(n/2)-1} e^{-t/2} dt.$$

3. cette loi dépend donc de deux paramètres

4. On peut utiliser les fonctions caractéristiques

5. et pas seulement une identité asymptotique

**Définition 5.5.2 (Loi de Student)** La loi de Student de paramètre  $n - 1$ , notée  $\mathcal{T}(n - 1)$  ( $n \geq 2$ ) est la loi de densité

$$c_{n-1} \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}}, \quad c_{n-1} = \frac{\Gamma(\frac{n}{2})}{\sqrt{(n-1)\pi}\Gamma(\frac{n-1}{2})}.$$

Cette densité tend vers la densité gaussienne réduite  $(1/\sqrt{2\pi})e^{-t^2/2}$  quand  $n$  tend vers l'infini ; on a les approximations suivantes

|                   |                 |                 |
|-------------------|-----------------|-----------------|
| $P( T_n  \leq a)$ | = 0.95          | 0.99            |
| si $n = 10$       | pour $a = 2.26$ | pour $a = 3.35$ |
| 20                | 2.09            | 2.86            |
| 30                | 2.04            | 2.76            |
| $\infty$          | 1.96            | 2.58            |

Ces lois sont tabulées.

### 5.5.3 Le théorème fondamental

Un candidat naturel pour avoir une approximation de  $\mu$  est la *moyenne empirique*

$$\bar{Y}_n = \frac{Y_1 + \dots + Y_n}{n}$$

et pour obtenir une approximation de  $\sigma^2$  la variance empirique

$$\Sigma_n^2 = \frac{(Y_1 - \bar{Y}_n)^2 + \dots + (Y_n - \bar{Y}_n)^2}{n - 1}.$$

Le théorème fondamental est alors :

**Théorème 5.5.1** Si les v.a  $Y_1, \dots, Y_n, \dots$  sont i.i.d. de même loi normale  $\mathcal{N}(\mu, \sigma^2)$  alors

(a) la v.a

$$\frac{\sqrt{n}}{\sigma}(\bar{Y}_n - \mu)$$

suit une loi normale centrée réduite  $\mathcal{N}(0, 1)$  (c'est le théorème "Central Limit").

(b) Les v.a.  $\bar{Y}_n$  et  $V_n$  sont indépendantes et la v.a.

$$Z_n = \sqrt{n} \frac{\bar{Y}_n - \mu}{\Sigma_n}$$

suit une loi de Student  $\mathcal{T}(n-1)$  à  $n-1$  degrés de liberté ;

(c) la v.a

$$\frac{(n-1)\Sigma_n^2}{\sigma^2}$$

suit la loi du chi-deux à  $n-1$  degrés de liberté  $\chi^2(n-1)$ .

Dans l'exemple qui nous intéresse on peut donc,

(i) sans connaître la valeur de  $\sigma$ , obtenir un intervalle de confiance pour la moyenne  $\mu$  des diamètres des boulons : la probabilité pour que l'intervalle (aléatoire)

$$\left[ \bar{Y}_n - a \frac{\Sigma_n}{\sqrt{n}}, \bar{Y}_n + a \frac{\Sigma_n}{\sqrt{n}} \right]$$

contienne le réel  $\mu$  (inconnu) est égale à  $\mathbb{P}(|Z_n| \leq a)$  (on a utilisé le (b) du théorème précédent) ; cette dernière quantité s'estime en utilisant une table où est tabulée la loi de Student.

(ii) comparer l'écart-type empirique  $\Sigma_n$  et l'écart-type réel  $\sigma$ . Par exemple, si  $n = 10$ , la probabilité que  $9\Sigma_{10}^2 > 16.9\sigma^2$  est à peu près inférieure à 0.05 car si  $F_{\chi^2(9)}(\cdot)$  est la fonction de répartition d'un chi-deux à 9 degrés de liberté on a  $1 - F_{\chi^2(9)}(16.9) \approx 0.05$ . On a utilisé le (c) du théorème précédent.

## 5.6 Test d'hypothèse

Il est commode à ce stade d'introduire la notion de *test d'hypothèse*. Le type de problème que l'on se propose d'étudier est le suivant : à partir d'un  $n$ -échantillon de même loi  $\mu \in \Theta$  décider si une hypothèse  $\mathcal{H}_0$  portant sur la loi  $\mu$  de ces v.a. est raisonnable ou pas. Si elle ne l'est pas on la rejette, sinon on l'accepte, ce qui alors signifie seulement qu'elle est plausible. Dans ce type de discussion on introduit un *seuil*  $\alpha$  qui est un réel compris entre 0 et 1 et quantifie le degré de certitude que l'on a lors de la procédure de rejet ou non de l'hypothèse  $\mathcal{H}_0$ . La procédure est la suivante : on considère une *statistique* c'est-à-dire une v.a  $Z_{n,\mathcal{H}_0}$  qui est une fonction des v.a  $X_1, \dots, X_n$  et dont on connaît la loi (et donc la fonction de répartition) *pourvu que* l'hypothèse  $\mathcal{H}_0$  soit vérifiée. On décide de rejeter ou d'accepter l'hypothèse au seuil  $\alpha \in (0, 1)$  si par exemple, les données expérimentales donnent  $|Z_{n,\mathcal{H}_0}| \geq t_\alpha$  où  $t_\alpha$  est définie par  $\mathbb{P}_{\mathcal{H}_0}(|Z_{n,\mathcal{H}_0}| \geq t_\alpha) = \alpha$ . En d'autres termes on rejette l'hypothèse  $\mathcal{H}_0$  si la réalisation  $\omega \in \Omega$  qui donne  $|Z_{n,\mathcal{H}_0}(\omega)| \geq t_\alpha$  fait partie d'un événement très peu probable.

Dans l'exemple (ii) de la section 5.5.3 précédent l'échantillon est une suite de v.a de loi  $\mathcal{N}(\mu, \sigma^2)$  qui dépend des deux paramètres  $\mu$ , et  $\sigma$  *a priori*



inconnus et on veut tester l'hypothèse suivante portant sur  $\sigma$  :  $\mathcal{H}_0 : 9\Sigma_{10}^2 \geq 20\sigma^2$ . Pour cela on construit une statistique, en l'occurrence la v.a  $Z_{10} = \frac{9\Sigma_{10}^2}{\sigma^2}$  dont on sait qu'elle suit une loi du chi-deux à 9 degrés de liberté et pour laquelle la fonction de répartition est tabulée. Au seuil  $\alpha = 0.05$  on rejette l'hypothèse car

$$\mathbb{P}(Z_{10} \geq 20) \leq \mathbb{P}(Z_{10} \geq t_\alpha) = \mathbb{P}(\chi_9^2 \geq t_\alpha)$$

où  $t_\alpha = 16.9$  et  $\mathbb{P}(\chi_9^2 \geq t_\alpha) = 1 - F_{\chi^2(9)}(16.9) \approx 0.05$  est plus petite que le seuil  $\alpha = 0.05$  que l'on s'était fixé.

## 5.7 Test du chi-deux

**Déterminer si un dé est pipé ou non** : On veut savoir si un dé à 6 faces présente chacune de ses faces de façon équiprobable. Pour cela on jette le dé  $n$  fois ( $n$  grand). Le résultat du  $k$ -ème lancer est modélisé par une v.a  $X_k$  à valeurs dans  $E = \{1, \dots, 6\}$  et on suppose que les v.a  $X_k$  sont indépendantes et identiquement distribuées. Dire que le dé est pipé c'est dire que  $(p_1, p_2, p_3, p_4, p_5, p_6) \neq (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$  (on a noté  $p_i = \mathbb{P}(X_k = i)$ ). On veut donc tester l'hypothèse  $\mathcal{H}_0$  : "la loi  $(p_i)_{i \in E}$  est uniforme". Pour fixer les idées supposons qu'on réalise 600 lancers et qu'on obtienne 60 fois le 1, 108 fois le 2, 108 fois le 3, 102 fois le 4, 108 fois le 5 et 114 fois le 6. Le dé est-il pipé ou non? Pour répondre à cette question, on utilise le théorème suivant :

**Théorème 5.7.1** *Soit  $X_1, \dots, X_n, \dots$  une suite de v.a. i.i.d. à valeurs dans un ensemble fini  $E = \{1, \dots, r\}$ . Nous noterons  $(p_i)$  leur loi commune avec  $p_i = \mathbb{P}(X_n = i)$ . Posons*

$$N_i^{(n)} = \sum_{k=1}^n \mathbf{1}_i \circ X_k,$$

la fréquence empirique de sortie de  $i$ . Alors,

(a) le vecteur aléatoire

$$Z_n = \left( \frac{N_1^{(n)} - np_1}{\sqrt{np_1}}, \dots, \frac{N_r^{(n)} - np_r}{\sqrt{np_r}} \right)$$

converge en loi vers un vecteur gaussien  $Z$  c'est-à-dire un vecteur aléatoire dont toute combinaison linéaire des composantes suit une loi gaussienne;

(b) la suite de v.a

$$T_n = \frac{(N_1^{(n)} - np_1)^2}{np_1} + \dots + \frac{(N_r^{(n)} - np_r)^2}{np_r},$$

converge en loi quand  $n \rightarrow \infty$  vers un  $\chi^2$  à  $r - 1$  degrés de liberté.

Ainsi, si l'hypothèse  $\mathcal{H}_0$  est vérifiée la v.a

$$T_{600} = \frac{(N_1^{600} - 100)^2}{100} + \dots + \frac{(N_6^{600} - 100)^2}{100},$$

doit suivre (approximativement) une loi du chi-deux à 5 ( $= 6 - 1$ ) degrés de liberté. Or, pour une telle loi  $\mathbb{P}(\mathcal{T}_5 \geq 0.412) \leq 0.005$ . Dans notre expérience on a obtenu

$$T_{600}(\omega) = \frac{(60 - 100)^2}{100} + \dots + \frac{(114 - 100)^2}{100} = 3.92$$

Comme l'événement  $3.92 > 0.412$  on décide au seuil  $\alpha = 0.005$  de rejeter l'hypothèse  $\mathcal{H}_0$  (le fait d'observer  $\mathcal{T}_5 \geq 0.412$  est un événement rare).